Machine learning for the prediction of intrinsic magnetic properties and microstructure characterization of sintered FeNdB type permanent magnets

Zur Erlangung des akademischen Grades eines

DOKTORS DER INGENIEURWISSENSCHAFTEN (Dr.-Ing.)

von der KIT-Fakultät für Maschinenbau des

Karlsruher Instituts für Technologie (KIT)

angenommene

DISSERTATION

von

M.Sc. Choudhary, Amit Kumar

Tag der mündlichen Prüfung: 17.07.2025

Hauptreferent: Prof. Dr. rer. nat. Gerhard Schneider

Korreferenten: apl. Prof. Dr. Ralf Mikut

Abstract

The properties of material, which are the measurable responses of a material under various conditions, are determined by the properties of its phases and microstructure. Microstructure is the arrangement of phases and defects within a material characterized by various attributes, including type, size, orientation, shape, and amount. While intrinsic properties are studied through experiments and simulations, microstructure characterization typically relies on the microscopy image analysis. These factors collectively influence the overall properties and behavior. The extrinsic properties of a material arise from the intrinsic properties of its individual phases and the microstructure of the material. However, traditional methods for determining these properties can be tedious, expensive, subjective, and error prone due to manual efforts. Therefore, to address these challenges, a data-driven approach has been developed to obtain a better trade-off between accuracy and time-efficiency.

In this thesis, the generalization capabilities of machine learning approaches for intrinsic property prediction tasks and quantitative microstructure characterization using microscopy image data are demonstrated. More specifically, the use of Newtonian space of chemical composition of the 14:2:1 hard magnetic phase ($TM_{14}RE_2B$; TM = transition metal, RE = rare earth, B = boron) to predict the intrinsic magnetic properties at room temperature, while Kerr microscopy (KM) is used for microstructure characterization of the sintered FeNdB type permanent magnet. The 14:2:1 phase enables permanent magnets with excellent magnetic properties, and from an application perspective, Curie temperature (T_c), saturation magnetization (M_s), and anisotropy constant (K_1) are essential parameters. The regression models to predict the mass density in [g/cm^3], T_c in [K], and M_s in [μ_B/f.u.] of the 14:2:1 phase at room temperature are developed using a supervised learning approach and a dataset that includes the experimentally reported values across various literature sources. These models show strong predictive performance and are generalized to be valid across ternary to senary alloys, with an average deviation of 1.5% for mass density, a mean absolute error of less than 16 K which accounts for an average deviation of less than 3% for T_c, and an average deviation of less than 3% for M_s when compared against reference values.

For microstructure characterization, KM holds vital information for analyzing the distribution of grains, crystal orientation, and magnetic domain structures. It is less expensive and requires less time for sample preparation, image acquisition, and material analysis compared to the electron backscatter diffraction (EBSD) approach, especially for a large cross-section of samples. Within this thesis, deep learning models were developed using supervised learning approaches to analyze grain size and crystallographic grain orientation in FeNdB magnets from KM images. These models achieved a better trade-off between accuracy and time-efficiency, being 10x faster with a prediction error of less than 4% for grain size and less than 5º deviation for orientation analysis when compared to the EBSD approach, with the comparison including image acquisition, processing, and material analysis.

Finally, the developed data-driven approaches could serve as the basis for developing novel magnetic 14:2:1 phases from chemical composition and emerge as a cost effective tool for rapidly analyzing large cross-sectional areas of the samples to quantify heterogeneities in grain size and crystallographic grain orientation.

Acknowledgments

I express my heartfelt gratitude to the many people and research groups who supported me throughout my Ph.D journey. Their contributions have been invaluable and played a significant role in completing this thesis.

First, I am deeply thankful to my Ph.D advisor, Prof. Dr. Gerhard Schneider, for his unwavering guidance, patience, mentorship, and insight. I would also like to thank Dr. Timo Bernthaler for his invaluable guidance and for allowing me to conduct productive research at the Institut für Materialforschung Aalen (IMFAA). I extend my gratitude to Prof. Dr. Dagmar Goll and her team for providing resources, evaluating research results, and offering constant guidance. Their expertise and dedication have been instrumental in shaping my research.

Financial support for this research was provided by the German Federal Ministry of Education and Research (BMBF) within the "FHProfUnt" program (Project KLEVER, Grant No. 13FH255PA6), the Carl Zeiss Foundation and the German Research Foundation (DFG) (Subproject MEMORI, Grant No. P2018-03-002 and subproject MaKerr, Grant No. 413993866). I am indebted to them for their support and belief in my work.

I am fortunate to have had the opportunity to collaborate with many outstanding colleagues and researchers. In particular, I am grateful to my team members: Andreas Jansche, Tvrtko Grubesa, Florian Trier, Patrick Krawczyk, Anoop Kini, Dominic Hohs, and Olatomiwa Badmos. Their inputs, discussions, and shared experiences have enriched my work and made the research journey enjoyable.

I extend my gratitude to all the members of IMFAA for their inclusiveness and support throughout my Ph.D. Thanks to the machine learning research group for their invaluable discussions and ideas. Their involvement was critical to the success of this research.

Lastly, I want to express my gratitude to my parents, siblings, my wife, and my friend Khaja Saif Ahmed. Their unwavering belief in me and their support have been immeasurable.

List of abbreviations

AB	**A**daptive **B**oosting
AB-DT	**A**daptive **B**oosted **D**ecision **T**rees
AM	**A**tomic **M**ass
ANNs	**A**rtificial **N**eural **N**etworks
AP	**A**xial **P**ress
ASTM	**A**merican **S**ociety for **T**esting and **M**aterials
BCE	**B**inary **C**ross-entropy
BSDS	**B**erkeley **S**egmentation **D**ataset
CC	**C**hemical **C**omponents
C	**C**olor
CNNs	**C**onvolutional **N**eural **N**etworks
COCO	**C**ommon **O**bjects in **C**ontext
DFT	**D**ensity **F**unctional **T**heory
DL	**D**eep **L**earning
DNNs	**D**eep **N**eural **N**etworks
DoPACOM	**D**omain **P**attern **A**nalysis and **C**rystallographic **O**rientation **M**easurement
DRM	**D**irectional **R**eflectance **M**icroscopy
DRP	**D**irectional **R**eflectance **P**rofile
DT	**D**ecision **T**rees
DTL	**D**istance **T**ransformation **L**oss
EBSD	**E**lectron **B**ackscattered **D**iffraction
ECD	**E**quivalent **C**ircle **D**iameter
EDX	**E**nergy **D**ispersive **X**-ray **S**pectroscopy
ERMA	**E**uropean **R**aw **M**aterials **A**lliance
ET	**E**xtra **T**ree **R**egressor
EV	**E**lectric **V**ehicles
FC	**F**ully **C**onnected
FeNdB	**I**ron **N**eodymium **B**oron
FFT	**F**ast **F**ourier **T**ransformation
FL	**F**ocal **L**oss
GB	**G**radient **B**oosting
GD	**G**radient **D**escent
GLCM	**G**ray **L**evel **Co-occurence M**atrix
Grad-CAM	**G**radient-weighted **C**lass **A**ctivation **M**apping
HED	**H**olistically **N**ested **E**dge **D**etection
HP	**H**aralick **P**arameters
HR	**H**uber **R**egressor
HSV	**H**ue **S**aturation **V**ariance
IoU	**I**ntersection over **U**nion
IP	**I**sostatic **P**ress
IPF	**I**nverse **P**ole **F**igure
IQR	**I**nter **Q**uartile **R**ange
KM	**K**err **M**icroscopy
LM	**L**ight **M**icroscopy
LOOCV	**L**eave-**O**ne-**O**ut-**C**ross-**V**alidation
LR	**L**inear **R**egressor

M	'Fe' **S**ubstituting **E**lement in 2:14:1 phase
MAE	**M**ean **A**bsolute **E**rror
MAPE	**M**ean **A**bsolute **P**ercentage **E**rror
MDI	**M**ean **D**ecrease in **I**mpurity
ML	**M**achine **L**earning
MLP	**M**ulti-**L**ayer **P**erceptron
MOKE	**M**agneto-**O**ptic **K**err **E**ffect
MSW	**M**ean **S**quared **W**eight
N-UW	**N**on-**U**niform **W**eightage
ODS	**O**ptimal **D**ataset **S**cale
OIM	**O**rientation **I**maging **M**icroscopy
PSD	**P**ole **S**trength **D**ifference
PSPP	**P**rocessing, **S**tructure, **P**roperties, and **P**erformance
QMA	**Q**uantitative **M**icrostructure **A**nalysis
RE	**R**are **E**arth
REE	**R**are **E**arth **E**lements
ReLU	**R**ectified **L**inear **U**nits
RF	**R**andom **F**orest
RGB	**R**ed, **G**reen and **B**lue
RMSE	**R**oot **M**ean **S**quared **E**rror
ROI	**R**egion **O**f **I**nterest
SAM	**S**egment **A**nything **M**odel
SC	**S**trip **C**asting
SD	**S**tandard **D**eviation
SEM	**S**canning **E**lectron **M**icroscopy
SHAP	**SH**apley **A**dditive ex**P**lanations
SLIC	**S**imple **L**inear **I**terative **C**lustering
SVM	**S**upport **V**ector **M**achines
SVR	**S**upport **V**ector **R**egression
TEM	**T**ransmission **E**lectron **M**icroscope
TM	**T**ransition **M**etal
TP	**T**ransversal **P**ress
UW	**U**niform **W**eightage
VR	**V**oting **R**egressor
WCE	**W**eighted **C**ross-**E**ntropy
XGB	**E**xtreme **G**radient **B**oosting
XRD	**X**-**r**ay **D**iffraction

Notations

$(BH)_{max}$	Maximum Energy Product
B_r	Magnetic Remanence
H_{ext}	External Magnetic Field
K_1	Anisotropy Constant
J_s	Saturation Polarization
T_c	Curie Temperature
M	Saturation Magnetization in emg/g
μ_0	Vaccum Permeability

M_s	Saturation Magnetization
ρ_m	Mass Density
α_Z	Magnetic Dipole Momnet Misalignment Angle
μ_B	Bohr Magneton
H	Externally Applied Magnetizing Force
B	Magnetic Flux Density
H_a	Anisotropy Field
$f.u.$	Formula Unit
T	Tesla
K	Kelvin
Φ	Equivalent Circle Diameter
H_{ci}	Coercivity
φ_1	First Parameter of Euler Angle
ϕ	Second Parameter of Euler Angle
φ_2	Third Parameter of Euler Angle
θ	Theta Orientation
ρ	Rho Orientation
η	Distance Field Parameter
Λ	Weight for Distance Transform Loss
R^2	Coefficient of Determination
R_{ccc}	Concordance Correlation Coefficient
β	Temperature Coefficient of Saturated Magnetization
α	Volume Fraction of Hard Magnetic Phase
γ	Performance Ratio to Optimize the Weightage to MSE and MSW

List of figures

List of tables

Table of contents

1. Introduction

1.1. Motivation

The design, discovery and sustainability of advanced materials are conceptualized by the materials tetrahedron that highlights the importance and interdependence of the processing, structure, properties and performance (PSPP) of materials [1]. Over the years, experiments or simulation-based models have been developed to understand the physics of materials, which can later be used for better a interpretation of their characteristics.

The shift in the paradigm for material development starts from pure empirical methods followed by a model or theoretical-based approach and a computational science approach that involves simulations such as density functional theory (DFT) [2]. These paradigms have led to vast advances in data and computer science, leading to the emergence of data-driven approaches that can generate both forward and inverse models to find a better relationship in PSPP. Forward models have been used to study the causes and effects of processing, composition, and structure on material properties. On the other hand, inverse models have been used to design and optimize materials with application goals [2], [3], [4]. The fourth paradigm of materials development has been intensively supported by rapid progress in the field of machine learning (ML) [5] and deep learning (DL) [6]. Their involvement created a positive impact on the various tasks involved in different domains that cover health care [7], the pharma industry [8], mineralogy [9], materials informatics [4], robotics [7], automotive [10], etc. They have often proved very effective in solving tasks that require distinguished human expertise or tasks in which human intelligence finds its limitations.

In materials science, these ML and DL techniques are used in materials discovery, design, and process optimization, as well as in the material properties. Computer vision tasks and predictive analysis are two important areas that find application in materials sciences. Different algorithms and learning methods are adopted depending on the size and type of dataset, the task's complexity, and hardware resources. ML algorithms often require less dataset and hardware resources to converge, but are less robust than the DL. Computer vision tasks such as image segmentation have been used to identify different phases in a material's microstructure. Detection of defects or regions of interest is possible with object detection methods; classification of the regions or materials with similar features, such as texture, color, and shape, etc., is possible with image classification methods. Furthermore, the information in the microstructure can be extracted and linked to the properties or processing conditions of the materials using regression models. A supervised learning approach involves training models with input and expected output(s) called ground truth or labels. For a model to be trained successfully, labels have to be unbiased and should be prepared by subject experts. If there is not sufficient number of labels available for large input images, then an unsupervised learning approach can be adopted.

These advances in ML and DL are particularly relevant when addressing global challenges related to the rare earth (RE) permanent magnets. RE permanent magnets are essential for clean energy and an economic framework that aims to increase economic prosperity with consideration that they are environmentally and socially compatible. Many countries aim to become climate neutral by 2050 or reduce their dependence on non-renewable energy resources. An effective and successful transition is highly dependent on the development and

application of cleaner energy in an environmentally sustainable and economical way. This would require massive energy-efficient materials with many high-tech applications of strategic importance, covering renewable energy, robotics, electronics, communication devices, the automotive industry, and defense and aerospace applications. According to the European raw materials alliance's (ERMA) report on RE magnets and motors in 2021 [11], 95% of electric vehicles (EV) use traction motors with RE permanent magnets in them, and its demand is expected to increase to 14 times by 2030. Other applications such as wind turbines, robotics, sensors, and machine tools consume more than 1 million tons of rare earth magnets annually [11]. The RE permanent magnets constitute elements that are scarcely available and have adverse effects on the environment due to mining and processing. Due to geo-political reasons, this has led to limited and controlled supply of rare earth elements (REE) into the markets, causing a considerable gap between demand and supply, leading to a an increase in cost [12].

Sintered iron-neodymium-boron (FeNdB) magnets are currently the widely used permanent magnets due to their excellent extrinsic properties such as high remanence (B_r) and high coercivity (H_{ci}) [13]. Moreover, in the last few decades, the energy density ($(BH)_{max}$) of FeNdB has improved, making it an ideal candidate for applications requiring high energy to volume ratio, such as EVs and wind turbines [14]. Such applications also require high coercive FeNdB magnets to maintain sufficient performance even at higher operating temperatures. For high operating temperatures a high Curie temperature (T_c) is favorable. Manufacturing high coercive FeNdB magnetic alloys requires suitable alloying elements and fine-tuning of the microstructure [15]. The primary phase in sintered FeNdB magnets is $Fe_{14}Nd_2B$, the hard magnetic phase. The grain size and volume fraction of the hard magnetic phase directly correlate to B_r and $(BH)_{max}$. The Nd-rich phase affects the magnets' hardening and occupies grain boundaries of hard magnetic phases highly influenced by processing conditions [16]. To improve the coercivity and produce the thermally stable FeNdB magnets, the microstructure should be such that the hard magnetic phase (grains) is fine and the Nd-rich phase distribution is optimized. In addition to that, careful addition of other REEs can enhance the grain's magneto crystalline anisotropy field, which positively correlates with coercivity [15]. Small amounts of heavy REEs, such as dysprosium (Dy) and terbium (Tb), are added to replace the neodymium (Nd) in the FeNdB sintered magnets to increase coercivity. The rare-earth FeNdB permanent magnet has nearly 30% REEs by weight, with Nd being a major part and Dy or Tb in less percentage. These REEs account for nearly 50% of the magnet cost [16].

Therefore, it is essential to determine the intrinsic properties and microstructural features of FeNdB magnets, such as phase distribution, grain morphology, grain orientation, and study the influence of alloying REEs and processing conditions to characterize them efficiently [13]. The intrinsic properties of magnets are strongly influenced by their chemical composition, which in turn affects the magnetic phases. These phases are crucial in determining the physical properties that govern the magnets' overall magnetic behavior. However, the high computational costs associated with these analysis methods pose significant challenges to their widespread practical implementation. To address these challenges, a deep understanding of the underlying chemical and physical phenomena in materials must be built on reliable experimental data. Over time, the chemical composition of FeNdB magnets has been carefully adjusted and optimized to enhance their performance. This optimization process involves fine-tuning the concentrations of various alloying elements within the material. As a result, scientists have significantly improved the magnet's capacity to generate and sustain strong magnetic fields. However, rare

earth FeNdB magnets possess a complex microstructure, which requires detailed characterization across different scales using advanced microscopy techniques such as optical and scanning electron microscopes. Despite this progress, current methods such as SEM electron backscattered diffraction (EBSD) come with limitations such as high costs and time-consuming workflows. To overcome these issues, there is a growing focus on integrating Kerr microscopy (KM) to overcome the challenges of analyzing large, inhomogeneous samples while reducing both capital investment and time-intensive analysis.

This thesis demonstrates the ways to explore, develop, and validate the effectiveness of ML and DL-based approaches using supervised learning techniques to predict the intrinsic properties from the chemical composition of $Fe_{14}Nd_2B$ phases (14:2:1) followed by the characterization of the sintered FeNdB permanent magnets using automated quantitative microstructure analysis (QMA) involving optical and electron microscopy data.

1.2. Objective of the research

The research aims to develop an automated framework based on data-driven models to predict intrinsic magnetic properties of 14:2:1 phases based on their chemical composition and efficiently characterizing the sintered FeNdB permanent magnets using microstructural information obtained using the Kerr microscopy. For the intrinsic property prediction task, the specific objective is to develop and identify potential applications of data-driven models that can assist in developing novel magnetic phases in less time. On the other hand, for the microstructure characterization task for FeNdB type magnet, the primary objective is to use Kerr microscopy to extract grain size distribution as well as to obtain orientation of grains from domain pattern information. The models and approaches developed are evaluated against existing state-of-the-art approaches to measure performance and degree of effectiveness.

Within the scope of this research, different ML and DL algorithms have been explored to determine the relationship between the processing conditions, structure, properties, and performance of the sintered FeNdB magnets. Different scales of information, such as chemical composition and its effect on intrinsic magnetic properties, microscopy data from optical light microscopy (LM) [17], KM [18], and SEM [17] have been utilized for microstructure characterization.

1.3. Approach and methodologies

Using a supervised learning approach, classical machine learning regression models are developed to analyze the influence of chemical composition of 14:2:1 hard magnetic phases on intrinsic properties such as mass density, Curie temperature, and magnetic saturation. The approach is limited to the intrinsic properties at ambient temperature. The ground truth for regression models are values reported in the various literatures. Moreover, for some phases, the property values are reported by different literature and values differ by large, and such cases are called repeated phases.

For characterizing FeNdB magnets and identifying PSPP relationships at different scales [19] there are different techniques. Specifically, microstructural characterization involves data acquisition using microscopes ranging from optical light microscopes to electron microscopes [17], [20]. Due to rapid advancement in the field of microscopy, especially in automation of

microscopes, a large amount of data is being generated and the need for automated tools to analyze large data with high accuracy and robustness in less time is on the rise.

Each microscope provides a different scale of information and changes the field of view accordingly. Phase fractions and distribution and their morphologies can be quantified mainly by optical light microscopy under bright field light. Kerr microscopy effectively captures information on the grain size distribution and the micromagnetic domain patterns in each grain [18], [21] that carry information about the crystallographic orientation of each individual grain. The processing conditions affects the alignment of the micro-magnetic domain patterns. Scanning electron microscopes are well known for capturing the information on crystals or grains and providing the three-dimensional (3D) grain orientation [21]. The existing approaches to extract information from different microscopy images of magnets are either manual or semi-automatic. They are highly dependent on the user's subjective knowledge and have risks of inducing fatigue errors. The above challenges often restrict the microstructure analysis of large-scale microstructural samples. To overcome the above challenges, feature-based classical machine learning and advanced deep learning classification models are developed to detect and quantify the pores and oxide distribution, and grain size analysis from the Kerr microscopic images and finally, using the correlative Kerr and SEM-EBSD method for the crystallographic orientation analysis.

Since models are trained using the supervised learning approach, ground truth significantly influences model performance and robustness. Therefore, to remove bias and inaccuracies, the subject experts validate the ground truth based on the information from different sources and results from different techniques. These include values reported from various sources, including established physical models and DFT calculations for intrinsic properties, manually creating labels for the detection of pores and oxides and grain size, which are subsequently validated using SEM-EBSD and energy dispersive X-ray spectroscopy (EDX) [17]. For grain orientation, SEM-EBSD is considered as the ground truth.

1.4. Outline of the thesis

The objectives of this thesis mentioned in Section 1.2 are structured into several chapters.

Chapter 2 discusses the fundamentals of sintered FeNdB permanent magnets starting with intrinsic magnetic properties followed by extrinsic magnetic properties, including microscopy techniques for magnet characterization, domain pattern analysis from Kerr micrographs correlative microscopy approaches, and an exploration of the anisotropy of FeNdB-type permanent magnets. The chapter also explores the basics of data-driven models that include machine learning and deep learning techniques. It also provides a review of the literature on the application of machine learning in materials science, focusing on the prediction of intrinsic properties of 14:2:1 phases and quantitative microstructure analysis.

Chapter 3 covers the experiments conducted, describing materials and sample preparation, creating datasets for model development, ground truth or reference data used for performance assessments, and specifications for software and hardware used. This chapter also details models for intrinsic magnetic property prediction from chemical composition, grain boundary detection, and crystallographic grain orientation prediction from Kerr micrographs.

In Chapter 4, the results of the experiments are presented, including evaluations of the models and their applications.

Chapter 5 offers a comprehensive discussion of the results, evaluating the models' performance, their application to various samples, and their implications. It also provides insight into the strengths and limitations of the developed models.

Finally, Chapter 6 provides conclusions summarizing the findings, highlighting the contributions of the developed approaches, and outlining potential areas for future research and development.

2. Fundamentals

2.1. Fundamentals of the sintered FeNdB permanent magnets

Sintered rare earth FeNdB magnets are widely used permanent magnets for converting electrical energy to mechanical energy and vice versa. They have applications ranging from everyday use devices to high-end technologically advanced equipment to build a more sustainable future [22], [23], [24], [25]. The $Fe_{14}Nd_2B$ alloy crystalline structure has excellent magnetic properties, such as high energy density and coercivity, that support miniaturization and are resistant to demagnetization, respectively, compared to magnets made from materials such as ferrite and AlNiCo. In these magnets, the chemical composition of the hard magnetic 14:2:1 phase ($TM_{14}RE_2B$; TM = transition metal, RE = rare earth, B = boron) significantly dictates their exceptional magnetic properties alongside microstructure. As a result of these improvements, the 14:2:1 magnets have achieved an impressive high energy product of around 451 kJ/m^3 [26], [27], [28]. The 14:2:1 phase's exceptional intrinsic magnetic characteristic, such as saturation polarization (J_s > 1.2 T) and magnetocrystalline anisotropy (K_1 > 106 J/m^3), are essential prerequisites for achieving significant remanence and coercivity levels, respectively.

Compared to other magnets like ferrite, AlNiCo, and Samarium Cobalt, FeNdB permanent magnets have the highest magnetic field strength and resistance to demagnetization, but rather low operating temperatures of 80 °C to 100 °C. Above 100 °C, the irreversible loss of coercivity occurs. By adding heavy rare earth metals such as Tb or Dy to the alloy, the maximum operating temperature can be increased up to 180 °C to 200 °C with some reduction in remanence [24]. However, these heavy RE metals are costly and there are limited reserves around the globe. The magnetic properties of the FeNdB magnets show a strong temperature dependence, and based on the application, different elements replace Nd or Fe in FeNdB. If the application requires a magnet with high coercivity and operating temperature, then REE such as Dy is added to the FeNdB. If a higher magnetic strength is needed at high operating temperatures, then Co is added to FeNdB [29]. Tb positively influences coercivity at higher temperatures more than Dy. However, these REEs are scarce resources that make them expensive, and for geopolitical reasons, the supply of these elements is highly strategic. Most of the REE reserves are in China (~ 75%), followed by the USA (~ 13%), India (~ 4%) and Australia (~ 1.5%). The mining and processing of the REEs is challenging and has adverse environmental effects therefore, there is a need for FeNdB magnets with either a lesser REE content or a magnet free of REE that has properties close to the rare earth magnets.

In FeNdB permanent magnets, the relationship between intrinsic and extrinsic magnetic properties is crucial for understanding their performance. Saturation magnetization (M_s), representing the maximum magnetic moment per unit volume, is typically high in $Fe_{14}Nd_2B$ (32.5 μ_B/f.u.) phase due to strong ferromagnetic coupling between Nd and Fe atoms. A high M_s is also needed to achieve a high energy product. Similarly, Curie temperature, the temperature at which the material transitions from ferromagnetic to paramagnetic state, is determined by the magnetic interactions between atoms. These intrinsic properties depend on the chemical composition of the phase [24]. Extrinsic properties, on the other hand, significantly impact the magnetic properties of the material. The grain orientation of the 14:2:1 grains in the magnet, for instance, plays an important role in determining magnetic properties. While intrinsic properties are determined by chemical composition and crystal structure of the phase, the

extrinsic properties influence manufacturing processes, which, for example, can help optimizing the grain alignment. Furthermore, grain size and texture are extrinsic features that can be engineered to enhance magnet performance. Smaller grain sizes and well-aligned textures are desirable for maximizing magnetic properties, such as coercivity and remanence. Processes such as powder metallurgy and annealing allow control over grain size and texture or phase composition, enabling tailored magnet properties for specific applications. Therefore, understanding and optimizing both intrinsic and extrinsic properties are essential for designing FeNdB magnets with desired magnetic properties for diverse industrial applications [30], [31], [32].

Composition of the sintered FeNdB magnets

The ternary phase diagram of FeNdB suggests that six phases are relevant for sintered FeNdB magnets. $Fe_{14}Nd_2B$ is the main phase and the only hard magnetic component. Other phases are the Nd-rich phase, Fe, $Fe_{17}Nd_2$, $Fe_{17}Nd_5$, Fe_2B, and $Fe_4Nd_{1.1}B_4$ [33]. The hard magnetic phase is a tetragonal compound with the easy direction of magnetization along the c-axis and is referred to as a grains. In a ternary FeNdB sintered magnet, the main phase with > 80 vol% is 14:2:1 [34]. During liquid phase sintering this phase is surrounded by a Nd rich liquid. During cooling from the sintering temperature, several additional phases can crystalline, first the 4:1:4 phase. The residual liquid solidifies eutectically. During this solidification, metastable softmagnetic phases have been reported to be formed [35]. The Fe_4NdB_4 phase does not significantly affect the coercivity, whereas the soft magnetic phases negatively affect it. Therefore, the composition of the phases after sintering is considered such that soft magnetic phases like $Fe_{17}Nd_2$ or Fe are avoided by increasing the Nd and B contents [33], [36]. RE permanent magnets are primarily based on the $Fe_{14}Nd_2B$ phase, which features a tetragonal lattice structure with a space group of $P4_2/mnm$.

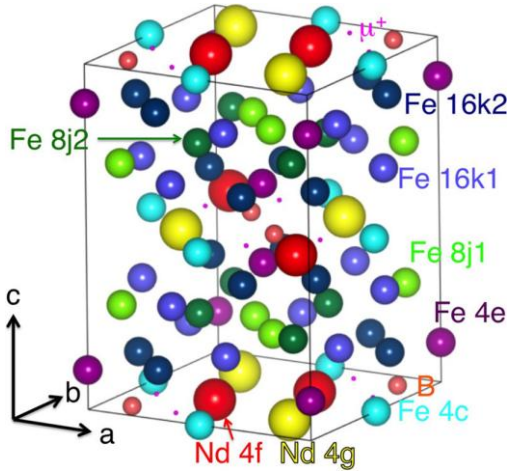


Figure 1: The crystal structure of $Fe_{14}Nd_2B$ in tetragonal symmetry with space group $P4_2/mnm$. Tiny spheres in pink are the muon site (μ^+) predicted by first-principles calculation [37].

Figure 1 shows the unit cell of $Fe_{14}Nd_2B$. Each unit cell comprises 4 formula units, amounting to 68 atoms, with two distinct RE positions (4*f* and 4*g*), six unique iron sites (16k_1, 16k_2, 8j_1, 8j_2, 4*e*, and 4*c*), and a single boron site (4*f*) [37]. Finally, different rare earth elements are added to FeNdB magnets based on the application and desired magnetic properties such as to increase coercivity and temperature stability. The process and alloying content of REE influence the microstructure and intrinsic behavior of the main or matrix phase [36]. There is also a possibility

that, due to the formation of oxides and pores, the homogeneous growth of the $Fe_{14}Nd_2B$ phase is affected.

Magnetic properties sintered FeNdB magnets

The magnetic properties of sintered FeNdB magnets are given by the hard magnetic 14:2:1 phase and its intrinsic properties and the microstructure of the magnet. The extrinsic properties are obtained by combining the intrinsic properties of the phases in the magnet and microstructural parameters such as grain size, grain orientation, phase distribution etc. Intrinsic properties originate on the atomic scale and involves fundamental quantum phenomena and include T_c, M_s, and K_1 and they get influenced by the chemical composition and crystal structure [38]. Moreover, in a magnet there can be more than one magnetic phase with each having distinguished characteristics, and therefore, it is possible that one magnet has different intrinsic properties.

Conversely, extrinsic properties reflect the microstructure of the magnet and combine the interplay between different phases. These include the size and shape of the $Fe_{14}Nd_2B$ hard magnetic phase (referred as grains) and their crystallographic orientation and distribution of other phases in the magnet. Examples of such properties are B_r, H_{ci} and $(BH)_{max}$.

The hysteresis curve shows the response behavior under the magnetic field. Equation (1) shows the relationship between the externally applied magnetizing force (H) and the induced magnetic flux density (B). Therefore, it is also referred as the B-H curve.

$$B = \mu_0(H + M_s) \tag{1}$$

B-H or J-H curve can be used to characterize the magnets by determining the number of different magnetic properties such as saturation magnetization (M_s), intrinsic coercivity (H_{ci}), remanence, and maximum energy product.

Saturation magnetization is the maximum flux density or magnetization that a material can have when it is fully magnetized by applying a very high external magnetic field.

Remanence: Once the external magnetic field is removed, the amount of magnetization left in the material is referred to as B_r.

Intrinsic coercivity is the reverse magnetic field required to demagnetize a fully magnetized material to zero.

Maximum energy product: It measures the maximum magnetic energy stored in a material and represents the magnet strength. It is the largest rectangular area that can be drawn in the B-H curve in the 2nd quadrant starting from the origin..

Manufacturing process for the sintered FeNdB magnets

The production of the sintered FeNdB magnets is done through liquid phase sintering process which starts with the production of the alloy by melting from the raw materials. Using strip-casting (SC) a very fine grained microstructure can be produced. The SC alloy is converted to fine powder using hydrogen decrepitation followed by jet-milling. The next step involves filling the powder into the desired molds and pressing it under an external magnetic field (up to 500 MPa). The pressing of the molds under the influence of an external magnetic field (1 Tesla) changes the grain orientation's alignment, which influences the magnets' remanence. It is then sintered in a vacuum furnace to about 1000 °C - 1100 °C in an inert argon gas chamber and

heat treatment (600 °C to 800 °C) is carried out to induce full coercivity [33]. However, the sintering temperature is decided based on the size of the powder particles, oxygen content and composition. Finally, the magnet is machined to the desired shape or dimensions and coated with corrosion resistant coatings.

2.1.1. Intrinsic magnetic properties

In the context of FeREB based magnets, the $Fe_{14}RE_2B$ phases serve as the fundamental components. These phases are crucial to understanding the physical properties that primarily determine the magnetic characteristics of these magnets. Notably, $Co_{14}RE_2B$ phase is not utilized for manufacturing permanent magnets, while cobalt addition to iron-based compounds is a common practice aimed at producing permanent magnets. In the context of permanent magnets, magnetic properties at room temperature are of paramount interest. Magnetic studies conducted on single crystals at room temperature have provided valuable insights into the behavior of these phases. For instance, in the case of $Fe_{14}Pr_2B$, it has been observed that the magnetization of the hard direction reaches saturation in external fields of approximately 10 T. However, this value is smaller than that observed in the easy direction, indicating an anisotropy in saturation magnetization [39]. Similar observations regarding pronounced anisotropy in saturation magnetization have been reported for $Fe_{14}Er_2B$ and $Fe_{14}Tm_2B$ phases [40]. The T_c of $Fe_{14}RE_2B$ compounds is noted to be relatively low, which can adversely affect the magnet properties associated with these materials. T_c values primarily depend on the Fe substituting element (M) sublattice, prompting extensive studies to understand how various substitutions at the iron sites impact their magnetic properties. The composition dependencies of T_c in $Fe_{14-x}M_xNd_2B$ compounds have been extensively investigated. Studies have shown that substituting iron with elements such as Co, Ni, Si, Cu, or Ga leads to an increase in T_c, while replacement with V, Cr, Mn, Ru, Al, Ge, or Be results in a decrease. This complex relationship arises from various physical processes, including changes in exchange interactions and mean iron moments [41], [42]. *Figure 2* shows the composition dependence of T_c for various substitution elements [41], [43], [44], [45], [46], [47], [48]. Notably, substitution of RE sites has also been explored to obtain a variety of solid solutions. Isomorphous substitution is particularly interesting due to its potential to influence anisotropy, shift spin-reorientation temperature, and modify the temperature dependence of anisotropy fields, thereby improving the performance of alloys as permanent magnets.

Further, FeNdB based materials exhibit excellent magnetic properties at room temperature. However, their behavior above room temperature is often unsatisfactory due to the lower T_c. Efforts have been made to enhance magnetic parameters at elevated temperatures with a focus on improving coercivity. Partial substitution of Nd with heavy rare earths such as Dy or Tb has shown promise in increasing the anisotropy field and coercivity [40]. However, this comes with a compromise, since an increase in the Dy content leads to a decrease in the saturation magnetization due to the antiparallel coupling of the Dy and Fe moments. Consequently, practical considerations often restrict the Dy content to small atomic percentages in the manufacture of permanent magnet materials [41], [49].

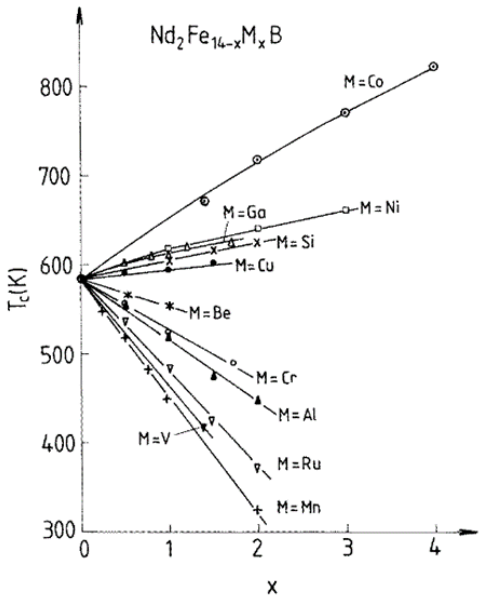


Figure 2: The composition dependence of T_c [K] for various Fe substitution elements (M) for $Fe_{14-x}M_xNd_2B$ at room temperature. Depending on the substitution element and the solubility limit, substitution amount (x) varies between 1 to 4 [41]. (The plot has been obtained after permission from [41], Reports on Progress in Physics, 1998)

In addition to T_c, M_s, or J_s is a crucial property for various applications involving magnetic materials. M_s quantifies the maximum magnetic moment per unit volume that a material can achieve when subjected to an external magnetic field. Saturation magnetization is commonly reported in the literature using the unit [μ_B/f.u.] (where μ_B denotes the Bohr magneton and f.u. represents the formula unit). This unit can be converted to [emu/g] (cgs unit), or Tesla [T] (SI unit) depending on the engineering requirements. The conversion to Tesla requires knowledge of the mass density of the material. This calculation relies on the established physical formula shown in the Equation (2) and Equation (3) detailed in the literature [24], [41].

$$J_s = \frac{m * \rho_m * 4\pi}{10000} \tag{2}$$

$$m = \frac{M_s * 9.274 * 10^{-24}}{(\sum AM) * 1.6605 * 10^{-27}} \tag{3}$$

Here m refers to the saturation magnetization in emu/g, ρ_m is mass density in [g/cm^3], M_s is saturation magnetization in [μ_B/f.u.], and AM is the formula mass of the 14:2:1 phase.

The magnetic saturation polarization J_s can also be calculated from M_s using another physical formula shown in Equation (4) from the literature [24], [41] provided M_s has been measured or reported in Ampere per meter [A/m].

$$J_s = \mu_0 * M_s \tag{4}$$

where M_s is saturation magnetization measured in [A/m] and μ_0 is vacuum permeability in [Vs/Am].

Hund's rule dictates the magnetic behavior of TM and RE elements within the 14:2:1 phase [50]. For compounds with light rare earth elements (Pr, Nd, Pm, Sm, Eu, and Gd), there is a negative coupling of spins between TM and RE moments, leading to parallel alignment and ferromagnetic behavior, resulting in larger J_s values. Conversely, compounds with heavy rare earth elements (Tb, Dy, Ho, Er, Tm, and Yb) exhibit antiparallel alignment and ferrimagnetic behavior, leading to smaller J_s values. Different contributions to M_S may arise from various elements within the compound. For instance, in $Fe_{14}RE_2B$ compounds, the mean magnetic moments of Fe are around 2.2 μ_B/atom, while in $Co_{14}RE_2B$ compounds, the mean Co moments are approximately 1.4 μ_B/atom [24], [41]. *Table 1* shows the calculated M_s [μ_B/f.u.], T_c [K], and H_a [kOe] for $Fe_{14}RE_2B$ compounds at room temperature.

Table 1: The calculated saturation magnetization (M_s), Curie temperature (T_c) and anisotropy field (H_a) of the $Fe_{14}RE_2B$ compounds at room temperature [24] [41].

Compound ($Fe_{14}RE_2B$)	**M_s [μ_B/f.u.]**	**T_c [K]**	***H_a* [kOe]**
$Fe_{14}Nd_2B$	32.5	585	73
$Fe_{14}Pr_2B$	31.9	565	75
$Fe_{14}La_2B$	28.4	530	20
$Fe_{14}Ce_2B$	23.9	424	26
$Fe_{14}Sm_2B$	30.2	616	>150
$Fe_{14}Gd_2B$	17.5	661	24
$Fe_{14}Tb_2B$	14.0	620	~ 220
$Fe_{14}Dy_2B$	14.0	598	~ 150
$Fe_{14}Ho_2B$	15.9	573	75
$Fe_{14}Er_2B$	17.7	554	8
$Fe_{14}Tm_2B$	22.6	541	8
$Fe_{14}Lu_2B$	22.5	535	26
$Fe_{14}Y_2B$	27.8	565	26
$Fe_{14}Th_2B$	14.1	481	26

The limits within the compositional space, which allow the thermodynamic stability of the 14:2:1 phase, are dictated by the solubility limit of elements in a 14:2:1 phase. Beyond the solubility limit, 14:2:1 phase is not a single crystal anymore. Understanding solubility limits can offer further avenues for tailoring compositions of a 14:2:1 phase to meet application-specific magnetic properties. These findings underscore the importance of analyzing the magnetic properties of FeREB compounds, particularly their behavior at room temperature, for the development and optimization of permanent magnets. Understanding the anisotropy of saturation magnetization in different crystallographic directions is crucial to tailor the magnetic properties of these compounds to meet specific application requirements.

2.1.2. Extrinsic magnetic properties

Grain size and texture are extrinsic properties that can be optimized to enhance magnet performance. An ideal FeNdB sintered magnet would have grains of $Fe_{14}Nd_2B$ phase that are small, and in uniform geometry [16], [51] with an easy magnetization axis being well aligned [52] and grain boundaries filled with the thin layers of non-magnetic Nd-rich phase to avoid the easy spread of the demagnetization from one grain to the other [36]. The adequate characterization of $Fe_{14}Nd_2B$ grains and their orientation [53] during different manufacturing steps is important to understand magnetic properties such as coercivity, energy product, and

saturation polarization. Along with morphology, the crystallographic orientation of the $Fe_{14}Nd_2B$ phase, which is the only hard magnetic phase in the microstructure, influences different properties. Almost every $Fe_{14}Nd_2B$ grain in the sintered magnet has a magnetization in a uniform direction. The manufacturing process leads to the formation of the magnetic decoupling via a non-ferromagnetic grain boundary.

The texture of a polycrystalline material, i.e., the total of the orientations of the individual crystallites, is sensitive to details of the preparation of the considered material. In many cases, the texture can be well described statistically by pole figures or an orientation distribution function, which have to be determined from experimental data obtained by neutron or X-ray diffraction. However, this procedure is difficult and expensive [8]. A further simplification is to characterize the texture by simple quantities called texture parameters or degree of texture. In particular, such a description is sensible if the texture has a high degree of symmetry for example, being spatially homogeneous and rotationally symmetric. In $Fe_{14}Nd_2B$, the c-axis of the tetragonal unit cell is the desired direction of the magnetization. The degree of texture is often deduced from measurements of the magnetic properties [10,11,12].

Another possibility to determine the crystallographic texture is EBSD. This method gives complete information regarding the local texture of very small areas (under ideal conditions now down to several tens of nanometers). EBSD is a technique used in SEM to analyze microstructures and obtaining crystallographic insights. The diffraction patterns obtained from EBSD are pivotal to obtain the exact orientation of particular grains as well as information about grain boundary phases [54], [55]. The reliability of these patterns, crucial for accurate indexing, is highly dependent on meticulous sample preparation, demanding perfectly flat and undamaged surfaces to produce high-quality Kikuchi patterns [56]. Achieving this level of preparation for EBSD is important and time-consuming. The resolution of EBSD is dependent on several factors, including sample preparation, the SEM equipment utilized, and the nature of the material under examination. For instance, spatial resolution typically ranges up to approximately 20 nm for dense materials such as magnets and extends to about 50 nm for lighter materials such as aluminum. Furthermore, EBSD samples are often tilted 70° toward the detector, resulting in a resolution deterioration that is approximately three times worse along the tilted surface [57], [58].

In FeNdB magnets, EBSD is capable of producing orientation maps that provide information on the distribution of crystallographic orientations. This data holds significant importance in fine-tuning magnet processing techniques and enhancing performance [54], [59], [60]. Additionally, permanent magnet materials often exhibit multiple phases, comprising the primary magnetic phase (e.g., $Fe_{14}Nd_2B$ in FeNdB magnets) and secondary phases (e.g., Nd-rich or grain boundary phases). EBSD can identify these phases based on their crystallographic characteristics, facilitating the refinement of material composition and the processing methods to optimize magnetic properties and adverse phase alterations. The preferred orientation or texture of grains within permanent magnet materials is pivotal in obtaining magnetic anisotropy and magnetization behavior, thus necessitating this information for the optimization of magnet processing to achieve desired magnetic properties [61], [62], [63].

An alternative method for determining the crystallographic texture involves use of the Kerr effect that involves the configuration of magnetic domains that correlates with the 'easy' magnetization direction, which can be observed optically under polarized light. However, the

determination of texture on a mesoscopic (~1 μm) scale as determined by EBSD is related to magneto-optical Kerr microscopy. Thus, a quantitative description of local texture could be effectively correlated with the domain patterns of individual grains and with backscattered electron images. Khlopkov et al. [64] demonstrated the efficacy of EBSD in microtexture analysis, revealing intricate orientation maps and domain patterns crucial for understanding the magnetic properties of materials, demonstrated by the highest-energy density magnet with $(BH)_{max}$ = 451 kJ/m³. Moreover, the insights provided by magneto-optical Kerr microscopy further corroborated these findings, particularly in understanding the domain structures across grains. Wang et al. [65] extended these insights by showcasing a novel pathway for redesigning FeNdB magnets, using dual-step heat treatments to optimize texture and near-surface microstructure. The domain structures for the samples by acquiring KM images at the same location where EBSD maps were acquired. Through microstructural investigations, they established correlations between microstructures and magnetic properties (B_r and H_{ci}), offering a promising avenue for enhancing magnet performance. This work not only redefined conventional notions of bulk magnet design, but also highlighted the potential applicability of such techniques beyond specific magnet types. Opelt et al. [66] researched comprehensive microstructural analyzes of FeNdB magnets, using the combination of techniques including SEM, EDX, and KM. Their findings emphasized the intricate relationship between microstructure and composition, shedding light on the effect of processing parameters on the magnetic properties such as remanence and coercivity. Moreover, the delineation of boundaries and domains through advanced microscopy techniques paved the way for a deeper understanding of magnet behavior at the mesoscopic level. Yuan et al. [53] work provided crucial insights into boundary plane misorientations in sintered FeNdB magnets, showcasing the impact of grain boundaries on the remanence, coercivity of magnetic properties. Through detailed characterization, they highlighted the significance of boundary configurations in influencing coercivity, lattice distortion, and domain formation, underscoring the need for a nuanced understanding of microstructural features.

X-ray diffraction is another approach for measuring the crystallographic texture of FeNdB alloys. The results are typically represented as arbitrary values of a larger specimen area. Due to the intricate crystal structure of FeNdB alloys, complete textures by X-ray diffraction are sporadic and are often depicted using standard techniques such as pole figures and orientation distribution functions [67].

The above methods provide insights into crystallographic texture that are either qualitative or indicative of the overall grain orientation of the sample. However, employing EBSD enables the precise determination of individual grain orientations, not just the c-axis direction, and provides details regarding grain boundary characteristics and misorientation. These measured orientations directly correlate with the microstructure, highlighting deviations from the typical behavior of materials depending on the application. Consequently, EBSD gives more comprehensive insights into the microstructural evolution during FeNdB magnet processing.

2.1.3. Microscopy techniques for magnet characterization

The quantitative analysis of microstructure in sintered permanent magnets includes a range of microscopy techniques, such as LM, polarized KM, SEM, transmission electron microscope (TEM), magnetic force microscopy, X-ray diffraction and EBSD, based on the required scale of information and characterization. LM and KM offer faster processing and a broader field of

view compared to diffraction-based microscopy methods [68]. Consequently, for QMA of sintered permanent magnets, the data obtained from LM and KM are used to characterize phases, crystallographic orientation, and intrinsic properties [68], [69], [70]. Furthermore, it is essential to recognize that the performance and capabilities of EBSD as a technique are determined not only by the physics of EBSD pattern formation but also by the specific analytical approaches used to capture and analyze EBSD patterns. However, the grain size or the grain orientation analysis using the EBSD over a larger area is very time-consuming.

Quantitative microstructure analysis involves quantifying material structures and microstructures based on their appearance on micrographs [69], [71], [72], [73], [74]. Over time, the accuracy of such analysis has heavily relied on the subjective expertise of the examiner. Microstructural information such as phase distribution, phase/object morphology, grain size distribution, crystallographic grain orientation, and defect detection constitute some of the extensively measured microstructural features. Moreover, microscopy has progressed greatly, resulting in high-resolution, automated focus adjustment, and improved scanning stage leading to automation of microscopy, leading to the wide-field microstructural data and enhancing material characterization. Therefore, an abundance of data is generated for comprehensive material evaluation. Several digital image analysis methods have been developed to analyze such tasks quantitatively. However, these methods are mostly manual or semi-automated, requiring supervision from experienced experts. For instance, Kerr microscopy can provide information about grain distribution in sintered permanent magnets, offering a faster and broader field of view than the more expensive state-of-the-art SEM-EBSD approach. However, grain analysis through Kerr microscopy is predominantly carried out manually by subject experts [69].

Similarly, determining phase distribution using LM in sintered permanent magnets poses challenges due to the complex microstructure comprising multiple phases. Analyzing such data involves a semi-automated process. This type of analysis can be tedious, leading to potential fatigue-induced errors or biases and resulting in inconsistent output based on the subjective knowledge of experts. Overcoming these obstacles requires an automated framework that ensures impartiality and delivers highly accurate quantitative outcomes. In this thesis, the application of various data-driven approaches for QMA of sintered permanent magnets has been discussed, leveraging LM, polarized KM, and SEM data, and their associations with magnetic properties have been explored.

2.1.4. Domain pattern analysis in permanent magnets using Kerr microscopy

According to the domain theory of magnetism [75], ferromagnetic materials have regions where the magnetization is in a uniform direction. These regions referred as domains have atoms with magnetic dipole moments aligned to each other and oriented in the same direction. Each domain has magnetization oriented in a uniform direction, but different domains have different magnetization. The formation of domains in a ferromagnetic material is primarily due to the spin-spin exchange of electrostatic coupling between electron orbits, magnetocrystalline (spin-orbit coupling), and magnetostatic energy.

The energy required to align or rotate the magnetization out of the preferred one is called magnetocrystalline anisotropy energy. The presence of the large single domain will create a large amount of the magnetic field that extends outside the domain and requires a large amount of magnetostatic energy stored in the magnetic field. Therefore, the domain is split into multiple small domains such that the magnetic field passes through each domain as a loop but in opposite directions, reducing the loss of magnetic field outside the domain [76].

When the ferromagnetic materials are cooled below Curie temperature, large domains lose thermal stability and are split into small domains to reduce internal energy. A small domain is stable and will not split further easily. Every time a domain is split, magnetic dipoles with magnetization pointing in different directions are adjacent near the domain wall. This exchange interaction creates a magnetization force that supports aligning nearby dipoles to a point in the same direction. In order for adjacent dipoles to point in different directions, energy is proportional to the area of the domain wall [76]. Therefore, the net amount of energy reduced during domain split equals the difference between the saved magnetic field and the energy required for the creation of the domain wall. Further, magnetization field energy is proportional to the cube of domain size and domain wall energy is the square of domain size. This means that the domain size decreases, the net amount of energy reduced also decreases. The domain split continues until the net energy reduced equals the field energy saved. Different microscopy methods can visualize domains and patterns within the domains (grains). Some microscopy methods include Kerr, Lorentz, and magnetic force. The range of domain observation techniques spans from optical and electron microscopy to magnetic force microscopy. Among these, the magneto-optic Kerr effect (MOKE) can visualize domains with a spatial resolution of 0.1 μm, recording time of several hundred milliseconds to several seconds, and an information depth of up to 20 nm using an optical microscope [77]. The Kerr effect describes the interaction between a beam of linearly polarized light and a ferromagnetic sample. Kerr microscopy works on the principle of MOKE, enabling the visualization of magnet domains.

By utilizing the Kerr effect, changes in the polarization of light reflected from a magnetic surface can be captured. This effect has proven effective in visualizing and characterizing magnets by quantifying vital magnet domain information [23], [68], [41]–[43]. In short, the reflected polarized light undergoes a rotation known as the Kerr rotation angle because of the magnetized state. The rotation of light caused by the magnetic surface is detected by using a filter called a polarized analyzer. This filter is set at a specific angle, and as the light's polarization changes, the filter lets through different amounts of light. The changes in light intensity help measure the rotation caused by the magnetization. Finally, a KM image of the sample is generated [78]. Note that the Kerr microscopy setup used here is aligned with the magnetization direction of the sintered permanent magnet samples, which lies parallel to the plane of light incidence. This aligns with the polar MOKE configuration. Kerr microscopy emerges as a cost-effective and non-invasive tool, effectively contributing to domain characterization within magnets. Another facet of domain characterization involves crystallographic orientation analysis, a task often performed by SEM-EBSD techniques that is a state of the art technique due to their high precision. Utilizing this technique, one can evaluate the crystallographic nature of the sample, along with other features such as grain size, local grain orientation, texture analysis and phase distribution [21]–[23], [68], [32], [34], [44].

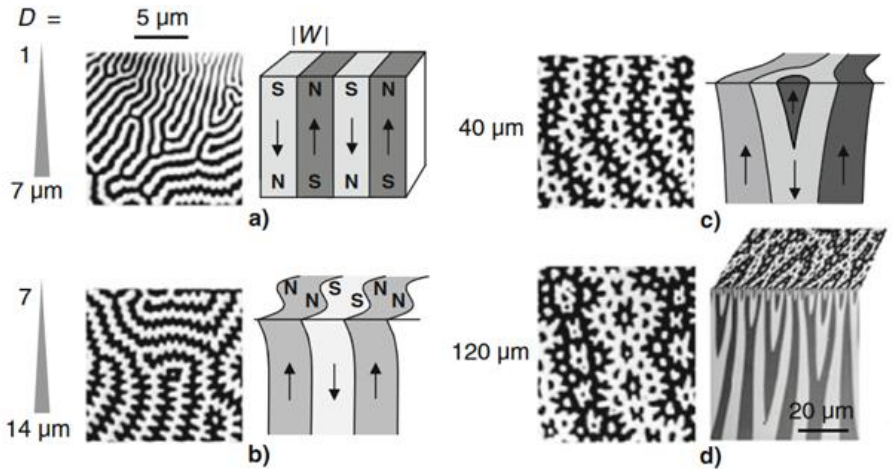


Figure 3: The schematic breakdown and KM images of the Dy-modified FeNdB crystal for samples with varying thickness: (a) 1 μm, (b) 14 μm, (c) 40 μm and (d) 120 μm. W represents the domain width, and D is the sample width [81]. (The figures has been obtained after permission from [81], Springer Nature, 2021)

*Figure 3 s*hows the images of magnetic domains of the Dy-modified FeNdB crystals with different thickness (D) observed under the Kerr effect. The images are acquired on top surface of the sample, while the sketch shows the possible side views. *Figure 3(a)* is for the thin film sample (D value between 1 μm and 7 μm) while *Figure 3(b, c, d)* is for samples with D value more than 7 μm. The domain width (W) increases with an increase in thickness of the samples according to $W \sim \sqrt{D}$.

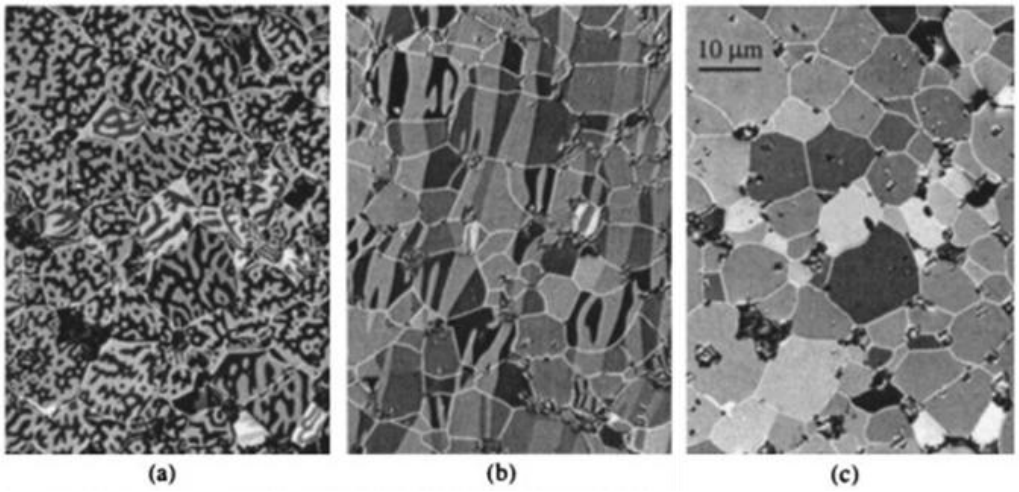


Figure 4: Kerr microscopic images of domains in sintered FeNdB magnet in a thermally demagnetized state with axis of anisotropy perpendicular to the imaging plane (a) and parallel to the imaging plane (b). For the same sample local as in (a), the image without any domain structures can be seen after application of the saturation pulse (c) [82]. (The figures has been obtained after permission from [82], Oxford, 2001)

The magnetization state and the grain's orientation concerning the incident polarized light plane influence the rotation. In instances like thermally demagnetized FeNdB magnets, the grains exhibit no net magnetization. Nevertheless, to minimize internal stray fields at the micro-magnetic scale, magnetic domains with opposing magnetization directions form within individual grains, which can be seen via Kerr microscopy. Notably, in materials characterized by high uniaxial crystalline anisotropy, domain magnetization aligns with the crystalline anisotropy axis, which coincides with the preferential axis of magnetization, referred to as the magnetic easy axis [82]. The *Figure 4(a & b)* shows the KM image of the sample when the axis

of anisotropy is perpendicular and parallel to the imaging plane, respectively. In *Figure 4(c)*, domain patterns are not visible as the sample is in remanent state after the application of saturation pulse. Note that sample location here is the same as shown in *Figure 4(a)*.

2.1.5. Correlative microscopy approach for magnet characterization

The correlative microscopy involving KM and SEM offers an advanced approach to getting insight into the microstructure of magnets. This approach delivers multi-scale insights, enabling high-resolution data analysis while including the broader perspective provided by low-magnification overviews of magnets [83].

Both KM and SEM-EBSD offer valuable insights into magnet microstructures, each with distinct advantages and drawbacks. EBSD, while offering high accuracy in crystallographic characterization, presents limitations in terms of a smaller field of view, higher cost, longer acquisition time, and often requiring specialized sample preparation. On the other hand, KM, being an optical microscopy technique, can analyze magnetic domains in materials over a broader field of view. It generates red, blue, and green (RGB) images, requires less acquisition time, and is comparatively more cost-effective than SEM-EBSD. Due to these characteristics, KM and EBSD complement each other. To take advantage of the strengths of both techniques, they can be combined through correlative microscopy, resulting in a comprehensive approach for efficient materials characterization.

The microstructure of a sintered FeNdB-type magnet is shown in *Figure 5* using correlative microscopy, both an axis of anisotropy perpendicular to the image plane *(a, b, c)* and one within the image plane *(d, e, f)*. The magnet sample shown in *Figure 5* is a FeNdB sample sectioned at different angles to the axis of anisotropy. It is an axially pressed sample and anisotropic in nature having a mean equivalent circle diameter (ECD) of 10 µm, B_r of 1.11 [T] and $(BH)_{max}$ of 235 kJ/m^3. The bright field microscopy images in *Figure 5(a & b)* effectively display distinct phases, such as the hard magnetic phase or grains, pores, oxides, and the phases constituting grain boundaries. In contrast, Kerr microscopy unveils the domain structures within the hard magnetic phases, with their visibility influenced by the grain orientation. EBSD mapping, as shown in *Figure 5(c & f)*, uses an inverse pole figure (IPF) color coding to represent the orientation of the hard magnetic phases.

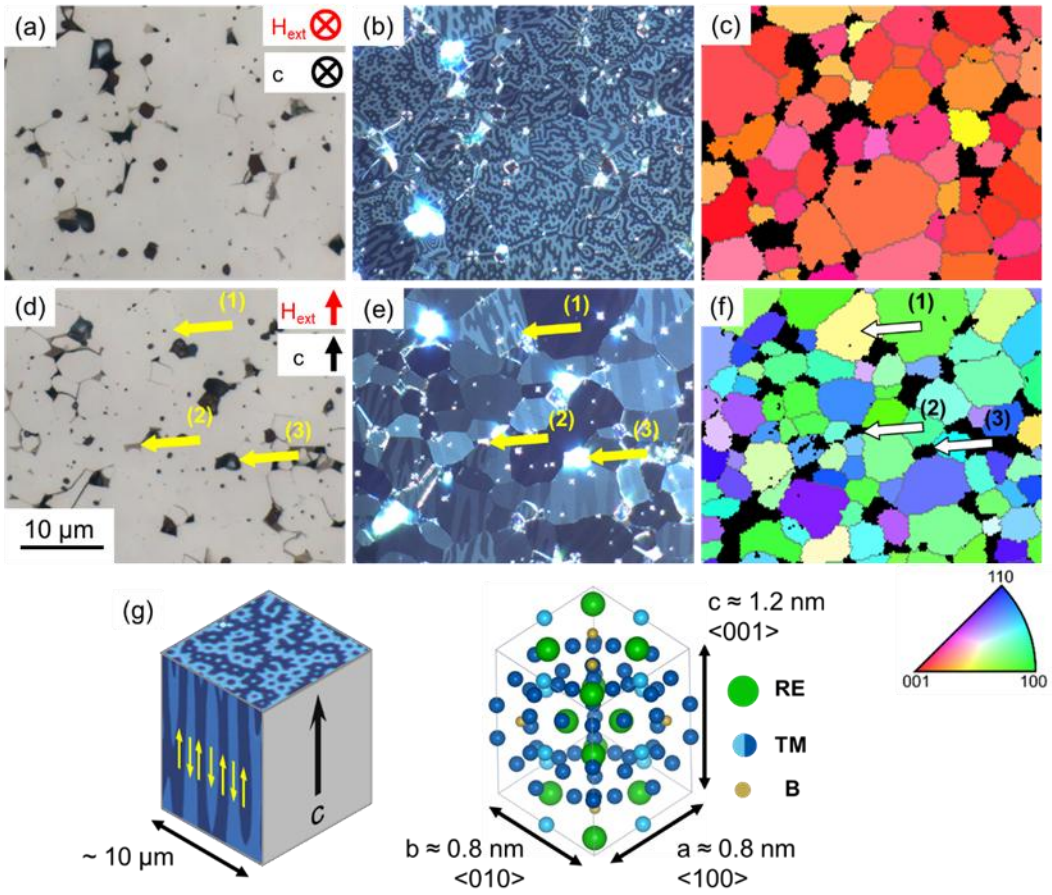


Figure 5: Correlative microscopy analysis of a sintered FeNdB-type magnet, capturing its microstructure in two orientations: one with the axis of anisotropy perpendicular to the image plane (a, b, c) and another within the image plane (d, e, f). In the bright field images (a, d), we observe three distinct features: (1) the hard magnetic phase, (2) the grain boundary phase, and (3) RE-oxides and pores. KM images (b, e) provide insight into the domain structures and the underlying grain orientation of the hard magnetic phase. The crystallographic orientation of the grains is further correlated with a color-coded SEM-EBSD IPF map (c, f). Importantly, the appearance of the domain pattern depends on the crystallographic axis of the FeNdB crystal structure is observable in the plane of observation (g) [68].

2.1.6. Anisotropy of FeNdB-type permanent magnets

The manufacturing process of sintered FeNdB permanent magnets leads to a complex microstructure characterized by a highly anisotropic alignment of $Fe_{14}Nd_2B$ grains, aiming to achieve magnetic decoupling through non-ferromagnetic grain boundary phases [53]. The extent of alignment varies based on the chosen green body powder compaction method [84]. Industrial processes use axial press (AP), transversal press (TP), and isostatic press (IP) compaction methods determined by the geometric configuration of the pressing tool and the orientation of the external magnetic field (H_{ext}) (shown schematically in *Figure 6)*.

The interaction between the H_{ext} and the pressing direction impacts the texture of the magnet. Mechanical factors such as compaction forces, friction, and magneto-static interactions among the particles (typically 1-5 µm) affect the alignment along magnetic field lines. As the easy axis aligns parallel to the pressing direction, friction between the particles or between particles and the die during compaction disrupts the grain alignment. Consequently, AP compaction yields

higher grain misorientation compared to samples produced through TP or IP compaction [52], [84], [85].

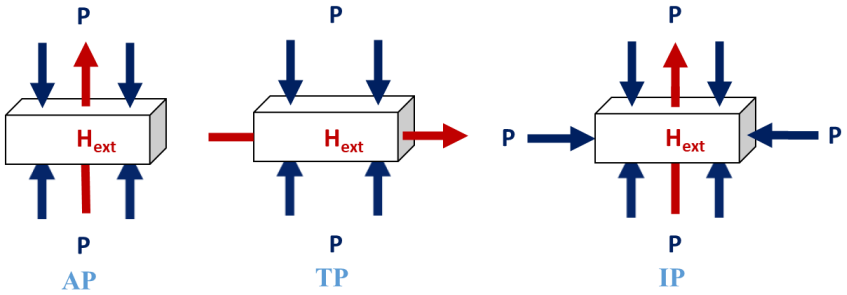


Figure 6: The compaction process of green body powder using different geometric configurations: AP, TP, and IP compaction. Mechanical pressure is applied to align particles parallel to an external magnetic field (H_{ext}). The goal is to create a uniaxial grain orientation distribution (texture) along the specified direction of H_{ext}.

The IP compaction typically results in a well-defined uniaxial texture characterized by uniformity throughout the magnet volume; TP and AP compaction samples can display variations in their uniaxial texture due to process-related geometric constraints. As a result, it becomes crucial to assess grain alignment or orientation quantitatively across various length scales. This assessment ensures quality assurance by examining the crystallographic texture through diverse techniques. Further, it is also possible to establish a correlation between the grain orientations within the sample and the magnetic properties, such as magnetic remanence. Sawatzki et al. [63] reported that by calculating the remanence values of sintered FeNdB magnets using the grain orientation histogram derived from EBSD, it is feasible to effectively demonstrate the alignment of the grain's magnetic measurements. As described in the Equation (5), the remanence and degree of grain alignment in the magnet along with other non-varying parameters can be correlated [68], [86].

$$B_r \propto (J_s * \beta) * \left\{\left(\frac{\rho}{\rho_0}\right) * (1 - \alpha)\right\} * f \tag{5}$$

Here, J_s represents the saturation magnetization of the grains, β is the temperature coefficient of saturated magnetization, $\frac{\rho}{\rho_0}$ denotes the ratio of the actual to theoretical density of the magnet, $1 - \alpha$ indicates the volume fraction of the hard magnetic $Fe_{14}Nd_2B$ phase, and f variable parameter signifies the fraction of grains aligned with the easy magnetization direction and thus serves as a potential indicator for assessing the degree of grain alignment. Other parameters remain constant during the compaction stage of sample manufacturing.

During the manufacturing process, uniform magnetization is always desired and is aimed to have a uniform alignment of the grains so that the inhomogeneity of flux density between the north and south poles can be minimized [84]. This difference is referred as pole strength difference (PSD) as shown in *Figure 7.*

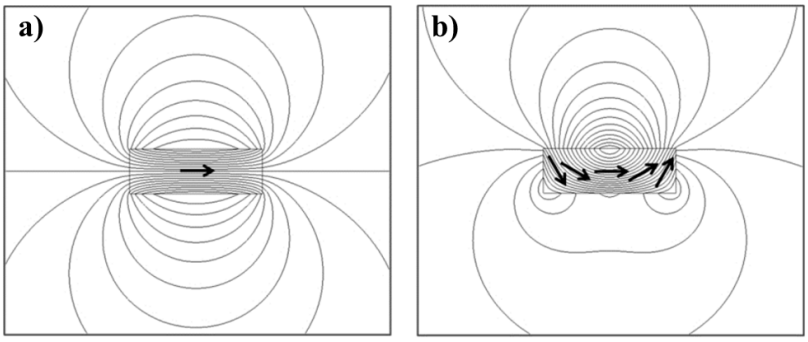


Figure 7: (a) Uniform and (b) non-uniform magnetization in the magnet sample. Pole strength difference seen in non-uniform magnetized sample is caused due to non-symmetric alignment of magnet [87]. (The figures are takenwith permission from [87], Springer, 1998)

In real-world applications, depending on the manufacturing process, PSD < 2% is possible. However, there are limited applications where non-uniform anisotropy is desired such as segments with radial magnetization or shape field magnets where PSD more than 2% (possibly upto 20%) is required. To achieve such non-uniform anisotropy, an optimized non-uniform magnetic field is applied during the compaction step of manufacturing process. Magnets with higher PSD values exhibit a non-uniform distribution of grain orientations or broader global misalignment, serving specific application requirements [87].

2.2. Fundamentals of the data-driven models

Machine learning and deep learning are the two widely used data-driven techniques that rely primarily on statistical learning methods. Data-driven approaches to materials science integrating high-throughput experiments and computing for materials data would prove beneficial compared to traditional experiment and computational simulation [88]. Data, learning approach, and algorithms are three main components of data-driven technique. Materials science data can be structured, unstructured, or both. More than 80% of the data generated from different sources around the globe is unstructured data that require users to extract meaningful parameters, and its growth is exponential [89]. Automated processes to extract meaningful information from such data require standardized approaches and different algorithms.

Feature engineering

Unlike traditional physics-based or rule-based models that require the function to process the input data to output data, the data-driven approach is fed with relevant input samples and the expected output data, which then comes up with the rules or the function by finding the statistical structures. If the input and expected data are available in sufficient numbers to train the model, then it is termed supervised learning. This would mean that the model should be able to transform the input data meaningfully so that the input's valuable features can be correlated to the expected output. This process of extracting the representable features from the input data is called feature engineering. The need for feature engineering is higher in the case of unstructured data such as images or videos than for structured data. The ML approach requires the feature engineering step to be carried out manually from the input data. For example, in the case of the optical microscopy image of the sintered FeNdB permanent magnet shown in *Figure 14*, features such as color, edge, and texture can be extracted. *Table 2* gives an overview of the different features and types that can be extracted from the image depending on the task and

region of interest. In case of non-availability of the expected labels for the input data, then an unsupervised learning approach is adopted, which forms clusters of data with similar features and provides useful insight about the data. However, the challenge lies when feature engineering has to be implemented for large amounts of data, or the extracted features do not represent meaningful input information to the model. In such cases, advanced data-driven approaches such as DL techniques are used to automatically learn the hidden features in the input data and with higher accuracy than the ML techniques, provided the data are extensive. Further, depending on the type of task that the ML or DL models have to solve, it can be divided into classification and regression tasks. If the task involves discrete output data, it is a classification task; if output data are continuous values, it is a regression task. For example, detecting the different phases in the microstructure of a material is a classification task, whereas predicting the material's properties is a regression task.

The success of machine learning models in addressing the challenges of developing complex multivariate functions by integrating human expert insights has been critical to their effectiveness. The approach follows a traditional ML framework that includes feature extraction and the application of DL models. Depending on the pre-set target, image features like color, edge, texture, shape, etc. can differ. *Table 2* shows a selection of commonly used image features and their categories. The table categorizes various features used in image analysis based on their types and classes.

Table 2: Overview of the commonly used features that can be extracted from an image alongside the associated image information. These features and their corresponding data are essential inputs for training machine learning models.

Features	Feature type	Feature class
RGB channels	Color	Local
HSV channels	Color	Local
Gray value	Color	Local
GLCM	Texture	Global
Structure tensor	Texture	Local/Global
Histogram of gradients	Texture	Local/Global
Entropy	Texture	Global
Haralick parameters	Texture	Global
Bilateral filter	Edge	Local
Sobel filter	Edge	Local
Difference of Gaussians	Edge	Local
Hessian Matrix	Edge	Local/Global
Area	Shape	Local/Global
Eccentricity	Shape	Local/Global
Feret min or max	Shape	Local/Global

Feature Types: The feature types in image analysis can be categorized into three main groups: Color, texture, and edge [90]. Color features encompass aspects of an image's chromatic characteristics, including RGB channels, hue, saturation, variance (HSV) channels, and gray value. These features are instrumental in tasks such as object recognition based on color patterns or tracking colored objects. Texture features explore repetitive patterns and variations within the image and include the gray-level co-occurrence matrix (GLCM), structure tensor, histogram

of gradients, entropy, and Haralick parameters [91]. Texture analysis is crucial for tasks such as material identification, surface characterization, and anomaly detection. Shape features includes morphological parameters of regions of interest in image such as area, eccentricity and ferret min or max.

Texture depends on the pixel values and its coordinates, which greatly depends on the image scale and resolution as well. Using GLCM, which provides information about different combinations of gray values occurring in an image, one can extract second order statistical texture features referred as Haralick parameters. Miyamoto et al. [92] describes computation of a number of statistical properties of the GLCM first, followed by the computation of the Haralick texture features [91]. Let us consider an image matrix *(x,y)* where for each matrix element, a pixel value is an integer over the range, 0,1,...N−1. *P(i,j)* is the GLCM and *p(i,j)* is the normalized GLCM where *i* and *j* indicates row and column indices of matrix respectively.

Then the 14 Haralick parameters *(f_1 to f_{14})* are calculated using Equations (6) to (31) as follows:

Angular Second Moment (*f_1*): The angular second moment feature is a measure of the homogeneity of the image. In a homogeneous image there are very few dominant gray-tone transitions. Angular Second Moment is high when the image has very good homogeneity or when the pixels are very similar. Noisy images have many small entries.

$$f_1 = \sum_i \sum_j p(i,j)^2 \tag{6}$$

Contrast (*f_2*): The contrast feature is a difference moment of the GLCM and is a measure of the contrast or the amount of local variations present in the image.

$$f_2 = \sum_{n=0}^{N-1} n^2 \left[\sum_{\substack{i=1 \\ |i-j|=n}}^{N} \sum_{j=1}^{N} p(i,j) \right] \tag{7}$$

Correlation (*f_3*): The correlation measures the linear dependence of gray levels of neighboring pixels. It measures the correlation between pairs of pixel values.

$$f_3 = \frac{\sum_i \sum_j (ij) * p(i,j) - \mu_x * \mu_y}{\sigma_x * \sigma_y} \tag{8}$$

Where, μ_x, μ_y, σ_x, σ_y are the means and standard deviations of p_x and p_y

Sum of Squares Variance (*f_4*): The sum of squares variance measures the spread or variance of the gray values in the image around the mean gray values. Increasing weight given to greater gray value differences.

$$f_4 = \sum_i \sum_j (i-\mu)^2 * p(i,j) \tag{9}$$

Where, μ is the mean gray value of the normalized GLCM.

Inverse Difference Moment (*f_5*): The inverse difference moment is the local homogeneity. It is high when the local gray level is uniform and the inverse GLCM is high. Maximum when neighboring pixels have same value.

$$f_5 = \sum_i \sum_j \frac{1}{1+(i-j)^2} * p(i,j) \tag{10}$$

Sum Average (f_6): Average sum of gray levels.

$$f_6 = \sum_{i=2}^{2N} i * \mathrm{p}_{x+y}(i) \tag{11}$$

Sum Variance (f_7): Variance of the sum of gray values.

$$f_7 = \sum_{i=2}^{2N} (i - f_8)^2 * \ p_{x+y}(i) \tag{12}$$

Sum Entropy (f_8): The uniform or flat distribution of sum of gray levels has a maximum entropy.

$$f_8 = -\sum_{i=2}^{2N} p_{x+y}(i) * \log\left(p_{x+y}(i)\right) \tag{13}$$

Entropy (f_9): The uniform (flat) joint distribution of gray levels has maximum joint entropy.

$$f_9 = -\sum_i \sum_j p(i,j) * \log\big(p(i,j)\big) \tag{14}$$

Difference Variance (f_{10}): Variance of the difference of gray levels.

$$f_{10} = \text{variance of } p_{x-y} \tag{15}$$

Difference Entropy (f_{11}): The uniform (flat) distribution of the difference of the gray levels has a maximum entropy.

$$f_{11} = -\sum_{i=0}^{N-1} p_{x-y}(\mathrm{i}) * \log\big[p_{x-y}(\mathrm{i})\big] \tag{16}$$

Information Measures of Correlation-1 (f_{12}): Normalized mutual information.

$$f_{12} = \frac{f_9 - HXY1}{\max(HX, HY)} \tag{17}$$

Information Measures of Correlation-2 (f_{13}): Difference between joint entropy and joint entropy assuming independence.

$$f_{13} = \sqrt{(1 - exp[-2 * (HXY2 - f_9)]} \tag{18}$$

Maximal Correlation Coefficient (f_{14}): Relates to how fast the Markov chain converges.

$$f_{14} = \sqrt{(\text{second largest eigen value of } Q)} \tag{19}$$

Where,

$$\mathrm{R} = \sum_{i=1}\sum_{j=1} p(i,j) \tag{20}$$

$$p(i,j) = \frac{P(i,j)}{R} \tag{21}$$

$$p(i) = \sum_{j=1}^{N} p(i,j) \tag{22}$$

$$p(j) = \sum_{i=1}^{N} p(i,j) \tag{23}$$

$$p_{x+y}(k) = \sum_{i=1}^{N}\sum_{j=1}^{N} p(i,j), k = 2,3 \ldots 2N, i + j = k \tag{24}$$

$$p_{x-y}(k) = \sum_{i=1}^{N}\sum_{j=1}^{N} p(i,j), k = 0,1, \ldots N-1, \ \mid i-j \mid = k \tag{25}$$

$$p_{x-y}(k) = \sum_{i=1}^{N}\sum_{j=1}^{N} p(i,j), k = 0,1, \ldots N-1, \ \mid i-j \mid = k \tag{26}$$

$$HXY1 = \ -\sum_{i}\sum_{j} p(i,j) * \ log\left(p_x(i) * p_y(j)\right) \tag{27}$$

$$HXY2 = \ -\sum_{i}\sum_{j} p_x(i) * p_y(j) * \ log\left(p_x(i) * p_y(j)\right) \tag{28}$$

$$Q = Q(i,j) = \sum_{k} \frac{p(i,k) * p(j,k)}{p_x(i) * p_y(k)} \tag{29}$$

$$\mathrm{HX} = \text{entropy of } P_x \tag{30}$$

$$\mathrm{HY} = \text{entropy of } P_y \tag{31}$$

N is the number of distinct gray levels in the quantified image. *R* is the normalizing constant, which is the number of neighboring pixel pairs used in computing a particular gray-level spatial dependence matrix. $p_{x+y}(k)$ and $p_{x-y}(k)$ represents the probability distribution of the sum and difference of the two gray levels respectively. *Q* is a transition matrix for a Markov chain of neighboring pixel gray levels. *Table 3* provides a classification of Haralick parameters against the type of information extracted from the image.

Lastly, edge features concentrate on identifying abrupt transitions within an image, often indicating object boundaries or structural changes. This category comprises the Bilateral filter, Sobel filter, Difference of Gaussians, and Hessian matrix and is indispensable for tasks involving object segmentation and understanding image structure [93], [94], [95].

Table 3: The texture information extracted from an KM image as the Haralick parameters are listed and categorized.

Haralick parameters	Information extracted from image
f_1, f_2, f_3	Measures of gray value variation
f_4, f_5	Measure of difference moments
f_6, f_7, f_{10}	Sum and difference of neighboring pixels
f_8, f_9, f_{11}	Measure of entropy (uncertainty)
f_{12}, f_{13}	Information theoretic measures

Feature Classes: Characteristic classes are classified as local, global, or a combination of both (local / global), addressing the spatial scope of their analysis [96]. Local features focus on specific regions within an image, capturing detailed information about localized patterns or attributes. Examples include RGB channels, HSV channels, Gray value, Structure tensor, Bilateral filter, Sobel filter, Difference of Gaussians, and Hessian matrix, making them particularly useful for detailed analyses and fine-grained information extraction. In contrast, global features provide insights that encompass the entire image, such as GLCM, Histogram of gradients, entropy, and Haralick parameters, offering a holistic view of image characteristics and aiding in understanding overall patterns and distribution. Additionally, some features, such as the structure tensor and Histogram of gradients, offer a combination of local and global insights, allowing them to analyze localized details and general trends, striking a balance between fine-grained and comprehensive analysis [96], [97]. In essence, the categorization of features by type and class demonstrates the breadth of information that can be extracted from images. This structured approach allows one to select features tailored to specific analysis objectives. Moreover, this categorization aligns with the requirements of various applications, enabling one to make informed decisions on feature selection for image analysis tasks across different domains.

Moreover, the process of extracting image features is dependent on the desired model output and the complexity in the microstructure of the sample. The features obtained from the input have a direct influence on the performance of the model. Therefore, users need to be well versed in diverse image processing techniques to formulate an appropriate feature vector containing vital information, thereby enabling the model to establish the relationship between input and expected output.

Increasing the number of features in an ML model can lead to improved performance, provided they are a meaningful representation of the data [69], [98]. Adding additional features to a dataset has profound implications for the performance of ML models. Firstly, it enhances their discriminative power, allowing them to discern finer details and complexities within the data. Providing a model with a more comprehensive set of information makes it better equipped to distinguish between different classes or patterns, resulting in more accurate classifications or regressions [60]. Moreover, the inclusion of these additional features can counter the underfitting model. When a model is initially simple and lacks the necessary complexity to grasp the intricate relationships in the data, it often performs poorly. Here, introducing more features proves invaluable, empowering the model to capture the subtler patterns and align closely with the training data [5], [99]. Furthermore, a richer feature set also contributes to improved generalization. It assists the model with a broader range of representations, enabling it to better extrapolate unseen data points. This means that the model can recognize patterns

across a wider spectrum of scenarios, enhancing performance when tested on new, unfamiliar data. To address this complexity effectively, models require the ability to capture these non-linear patterns. More features come to the rescue by providing the model with a means to represent these intricate relationships between variables. This greatly enhances the model's predictive capabilities.

Feature engineering requires meaningful, data-specific features that captures the dataset's distinct characteristics [68], [100]. However, it is important to note that incorporating more features is not always beneficial and can have downsides. First, the "curse of dimensionality" is a significant concern [101]. This phenomenon arises when too many irrelevant or redundant features are introduced, causing the model to be managed with a high-dimensional feature vector. In such cases, the model may struggle to find meaningful patterns amidst the vast dimensions, ultimately leading to overfitting. Furthermore, the quality of the data used to create these additional features becomes paramount. Low-quality or noisy features can be counterproductive, deteriorating the model's overall performance rather than enhancing it [101]. This underscores the importance of ensuring that the supplementary features are reliable and meaningful. While expanding the feature set offers substantial benefits, it must be approached carefully.

Using the advanced ML methods, including DL techniques, which does the automatic extraction of high-level features from input data. This overcomes the dependence on domain expertise for manual feature extraction steps [102]. Such models have shown improved accuracy and robustness compared to conventional feature-based ML models [99]. However, it is worth noting that DL models need a comparatively large number of datasets and also incur more computational costs during the training process.

2.2.1. Machine learning techniques

Machine learning models aim to extract the visual information from the hand-crafted or manually extracted multi-dimensional or higher dimensional feature set of input and correlate it with the expected output labels. They are also referred to as classical, traditional, or shallow ML and depending on application, different supervised and unsupervised ML methods are used. For the supervised learning approach, linear models (linear, Huber and logistic regression), tree-based approaches such as decision trees (DT) and random forest (RF), kernel-based support vector machines (SVM), and simple artificial neural networks (ANNs) such as multi-layer perceptron (MLP) are widely used. Clustering based approaches such as K-means is one of the popular unsupervised learning methods.

Tree-based machine learning models

Decision trees are non-parametric models that split the dataset into small homogeneous subsets iteratively. At each node in the tree, a rule-based decision divides the dataset so that inhomogeneity or impurity is reduced and results in the formation of the leaf node every time. The splitting of the training dataset is continued until no further splitting is possible or when the inhomogeneity or impurity is minimum or zero. Some of the impurity metrics are information entropy and Gini impurity. Because of the non-parametric nature of the model, they do not require data preprocessing such as scaling or normalization and work well for both classification and regression tasks. Additionally, they are tolerant to outliers in the dataset.

However, such models are prone to overfitting because it creates a large tree structure when working with a large feature set and is also affected by the variance of the dataset, specifically when the dataset is of small size [103].

Nevertheless, there are different tuning parameters that can be used to prevent overfitting such that model is not complex and too specific and reduce the variance in predictions. Early stopping of tree growth is achieved by optimizing the maximum depth of the tree, the number of features to be considered for the splitting, and the minimum number of samples required to split a node. Since all the features do not contribute equally to the model's learning process, the features with low importance can also be removed so that the model is not complex and thus avoid overfitting.

Figure 8 shows a schematic illustration of the bagging and boosting ensemble approaches adopted by tree-based models to avoid overfitting. Bootstrap aggregation or bagging is an ensemble method that involves training of the multiple small decision trees on the different sub-samples from the same dataset. This is to reduce the variance. Random forest model [104] builds multiple DTs from subsets of features such that the correlation between each features are removed. Later, the outputs from each of the multiple DTs are averaged to get the final output. As multiple DTs are trained, RF can handle higher dimensional feature sets well.

Boosting is another type of ensemble learning where a number of weak learners or simple DTs are trained on the same dataset sequentially [105]. Each simple DT generalizes the dataset poorly, and its weights are adjusted to improve the learning criteria by following simple DT. This is performed in a way such that data points with low prediction error are given less importance or weightage when compared to data points with high prediction error. This is continued until the model gives low prediction error to all data points. In this way, each of the simple DT, contributes to the learning of the model.

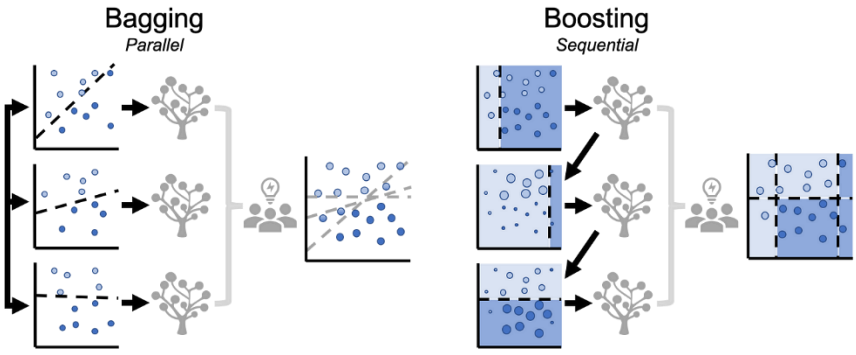


Figure 8: Schematic illustration of the bagging and boosting ensemble approach for tree-based models [106].

There are different boosting algorithms such as adaptive boosting (AB) [107], gradient boosting (GB) [105], and extreme gradient boosting (XGB) [108] etc. AB assigns low weightage to the incorrectly predicted data points from weak learners or simple DTs. However, it gets negatively affected by the presence of noisy data points and is comparatively slower than the other boosting methods. GB and XGB are more advanced boosting methods which work well on the noisy data, complex and non-linear datasets [108], [109]. *Figure 9* shows the evolution of the tree-based algorithms that includes both bagging and boosting approaches to overcome the challenges of model generalization.

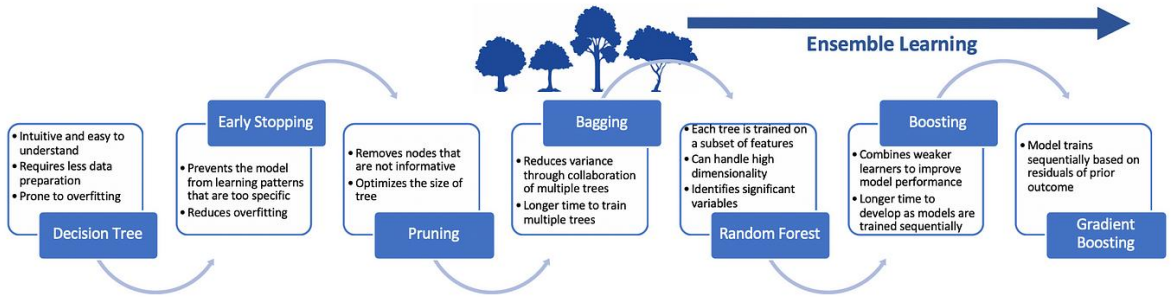


Figure 9: Schematic illustration of the evolution of tree-based algorithms from simple decision trees to extreme gradient boosting [106].

Multi-layer perceptron model

Multi-layer perceptron [6] is a fully connected feed forward ANN. It consists of a perceptron which mimics the behavior of biological neurons that receive input, process the information, and pass it on to the output layers. MLP has input layer to receive the input features, hidden layers with many neurons or nodes, and output layers to predict the outputs. Each node has a weight assigned to it depending on its relative importance to other nodes and an activation function that processes the weighted sum of all the inputs. The activation function forms the non-linear relationship between the input features and outputs. Depending on the complexity of the task, different activation functions are used. The data flow is from input to output layer, that is, feed forward, and the learning process is via the backpropagation method. Once the data pass through input to output layer with initial random weights assigned to each node, the error in the prediction when compared to the labeled data is calculated, and accordingly using the backpropagation method, the weights of nodes are adjusted. This process is repeated until the optimum results are achieved. However, advanced or more complex ANNs involve convolutional neural networks (CNNs).

Importance of tree-based models over ANNs on structured datasets

Tree-based models often outperform ANNs on structured datasets such as tabular data due to several key advantages. One primary reason is their suitability for structured datasets, which require less hyperparameter tuning and consume fewer computational resources while maintaining comparable or even superior generalization abilities compared to ANNs. This advantage stems from the inductive biases inherent in different ML algorithms. Inductive bias represents the assumptions and tendencies incorporated into a model when learning from data and making predictions [110]. It guides the learning process and influences how a model generalizes from training data to unseen examples. DTs possess specific inductive biases that align well with tabular data, distinguishing them from ANNs [110], [111], [112]. However, it is to be noted that in case of correlated features, ANNs such as MLP outperform tree-based models. For several reasons, ANNs excel with homogeneous features such as images, texts, or audio, but struggle with tabular data. Firstly, tabular data often have non-homogeneous feature importance, with certain attributes strongly correlated with the target variable. Secondly, ANNs are prone to overfitting when dealing with small tabular datasets and tend to produce overly smooth solutions [113]. Additionally, non-informative features have a more detrimental impact on ANNs compared to tree-based methods [110].

Studies, such as those conducted by Ullah et al. in [111], have shown that tree-based models can occasionally outperform DL models on medium-sized datasets. Moreover, tree-based models offer a significant advantage in terms of computational efficiency, with up to 100 times less computational cost while maintaining nearly identical accuracy. This reduced need for tuning and processing effort further enhances their appeal [111]. Shwartz-Zvi et al. in [112] have identified several challenges that ANNs, particularly deep neural networks (DNNs), have when applied to tabular data. These challenges include a lack of locality, data sparsity (often resulting from missing values), mixed feature types (numerical, ordinal and categorical), and a lack of prior knowledge about dataset structure. Furthermore, DNNs are often perceived as „black box" models, lacking transparency and interpretability in how they transform input data into model outputs [114].

Despite the strengths of DL methods, the 'no free lunch" principle still applies, meaning no single model works best for all problems. However, tree-ensemble algorithms, such as XGBs are frequently recommended for real-world tabular data problems. The results from various studies indicate that XGBs consistently outperforms most DL methods across different datasets [105], [115]. While no single DL method excels on all datasets, XGBs consistently delivers strong performance [112]. In summary, the inherent inductive biases, efficiency, and performance of tree-based models make them a preferred choice for tabular data tasks in many practical scenarios.

2.2.2. Deep learning techniques

Deep learning models aim to overcome the challenges of traditional ML approaches, primarily when dealing with unstructured data. DL models further leverage the advantage of data augmentation techniques that traditional ML models typically lack. This augmentation strategy helps the DL models with exposure to a comprehensive spectrum of data variations, enhancing their adaptability across diverse situations. A pivotal aspect of DL models, especially U-net, is their inbuilt ability to exploit the spatial hierarchies intrinsic to images. These models are adept at detecting features at varying levels of abstraction, effectively capturing both local and global patterns. This distinctive characteristic helps DL models discern relevant features within complex images, making them more robust when faced with variations and noise. Another distinct advantage of the DL model is its usage for transfer learning. DL models gain a substantial advantage in learning meaningful features from images using pre-trained models, often developed on vast and diverse datasets such as ImageNet. These pre-trained models can be fine-tuned for specific tasks, a capability typically lacking in traditional ML models.

Lastly, the scalability factor emerges prominently with the traditional ML models and may encounter difficulties when confronted with the escalating complexity of data. On the contrary, DL architectures, characterized by multiple layers and many parameters, prove highly capable in managing different datasets. This capacity enables them to capture fine or subtle patterns effectively, which may not be possible with traditional ML models.

DL techniques have superiority in solving problems concerning unstructured data because they automate the most crucial step of extracting the features from the input data. Its performance and robustness highly depend on the dataset size and quality. One of the pre-requisites for training a DL model is the large dataset. This is needed for the model architecture to extract

different low-level and high-level features from the input data and subsequently perform feature distillation to learn the most important features affecting the model's learning process. Creating an abundance of a labeled dataset is always desirable, but it is practically challenging because of the lack of capacity and computational efforts. Therefore, using transfer learning technique pre-trained networks can be used to develop domain-specific models with lesser amount of data. Additionally, the data augmentation step is adopted to populate the existing dataset by manipulating the existing dataset based on geometrical and image processing functions [116].

The four main components of the DL models are layers, network, loss function, and optimizer.

Layers

Layers are the building blocks of the DL that receives the input in various forms or processes and converts it using different sets of linear or non-linear functions and forwards them to output layer. The first layer is the input layer, the last layer is the output layer, and the remaining layers in between are hidden layers. Depending on the type of transformations required for the input data and tasks, different hidden layers are adopted such as normalization layer, convolutional layer, pooling layer, activation layer, dense layer, recurrent layer, etc.

Convolutional layers

Convolution is a mathematical operation that defines the relationship between spatial and time domains using convolution theorems. It is also interpreted as a filter (kernel filters) that extracts the information such as color, edge or texture, etc., from the input tensor or image. This is similar to the operations seen in signal processing and Fourier transformations. It converts all the pixels in the receptive field into a single value and therefore the size of the image decreases as one proceed deeper into the architecture and finally has a feature vector as output from all the convolutional layers. Further, the layers close to the input image learn to extract the simple filters or learn fewer kernel filters compared to the layers that are deeper in the network architecture [6].

Conv2D is a 2D convolutional layer that creates a kernel filter that slides over the 2D input tensors and is defined in Equation (32) as the multiplication of the two functions, X which is input and K is the kernel to get output O.

$$O(i,j) = \sum_{p}\sum_{q} K(p,q)X(i-p, j-q) \tag{32}$$

Since kernel filters operate on small regions in the image, they are defined by a specific size denoted by (p,q) and because the kernel is flipped in both horizontal and vertical directions, the values p and q are negative as seen in Equation (32). However, this does not have considerable effect on the learning process of the network, and using the cross-correlation operation, this can be set to positive values. This can be expressed as Equation (33):

$$O(i,j) = \sum_{p}\sum_{q} K(p,q)X(i+p, j+q) \tag{33}$$

In case of the 3D input such as the 3D tensor X, the convolution operates by sliding kernel K over the three directions of the input and computes the dot product between kernel and the input patch. As input is a 3D tensor, it has height, weight, and number of channels that can be

represented as [*l, w, c*] and the kernel is also a 3D filter of size [*p, q, r*] where *p* is height, *q* is width, and *r* is depth.

This can be expressed as Equation (34):

$$O(i,j) = \sum_{p=1}^{l} \sum_{q=1}^{w} \sum_{r=1}^{c} K(p,q,r)X(i+p, j+q, r) \tag{34}$$

Further, the convolution can be performed by padding the input with zeros to control the output dimension and prevent the loss of the information on the borders. If *P* is the padding width and *F* is the size of the filter, the two are correlated with the Equation (35):

$$P = \frac{F-1}{2} \tag{35}$$

The kernel shifts over the input in the number of pixels steps, which is defined as a stride (*S*). A higher stride will reduce the dimensions of the activation map and vice versa. Considering the padding and stripe in to the account, the output from the convolutional layer for 3D input tensor of size (*H, W, C*) can be given as in Equation (36):

$$O = \left[\frac{H+2P-F}{S} + 1, \frac{W+2P-F}{S} + 1, K\right] \tag{36}$$

Additionally, a diluted convolution can be performed with a wider kernel such that pixel spaces are inserted between the elements of the kernel to reduce the number of output parameters and memory required for processing the convolution operations [117]. Similarly, there are partial convolution and shuffled convolution that have application depending on the use-cases. The output from the convolution is also referred to as the activation map or feature map, which can be used to visualize the type of information the convolutional layer has learned or extracted.

Pooling layers

The feature map as a result of the convolutional layer has high dimension and contains few irrelevant details. Therefore, to extract useful information and also reduce computational cost, pooling layers are added to the network. It has two basic purposes, first, it reduces the number of parameters or weights, and second, it helps in reducing the likelihood of overfit in network. It has to be noted that the pooling layer does not add any learnable parameters, instead it removes some of the learned parameters from the previous convolutional layer. The spatial filter (F_L) and strides (S_L) in the pooling layer (L) operate by defining the window of size F_L x F_L and reducing the data inside window to a single value, and this window is moved in steps of strides after each operation. The size of the window can be either square or rectangle. These steps also lead to translational invariance and capture the input data's important features in smaller space or dimension.

For the input volume of size M_{L-1} x N_{L-1} x O_{L-1} in the pooling layer (L), the output from the pooling layer can be given as in Equations (37) to (39):

$$M_L = M_{L-1} \tag{37}$$

$$N_L = \frac{(N_{L-1} - F_L)}{S_L} + 1 \tag{38}$$

$$O_L = \frac{(O_{L-1} - F_L)}{S_L} + 1 \tag{39}$$

Max pooling and average pooling are two of the most common ones. Max pooling finds the maximum value within the window and skips rest of the values, whereas average pooling calculates mean values inside the window. In comparison to other pooling approaches, max pooling converges faster and performs better as it avoids negative elements in window from being canceled and thus prevents formation of the blurred activation maps.

Fully connected (FC) layers

The activation maps of the convolutional layers (and pooling layers) are flattened and passed through the FC layers to make predictions. FC layers are the last layers of a CNN and are feed-forward neural network that captures the global feature or information in the data. While the convolutional layer and the pooling layers extract features from the data, FC layers are the prediction layers where classification or regression tasks are performed. They consist of neurons and hidden layers that are interconnected to each other such that non-linear relationship between the input features and ground truth can be formed effectively. Due to the dense structure of the FC, having a few layers of FC can result in large network weights and could also induce overfitting depending on the size of the feature vector and dataset.

Residual Connections

The CNNs have quite a large number of layers, and stacking more layers could help in better generalization of the dataset and subsequently learn complex functions. However, more depth in the network and the presence of the non-linear activation functions cause the gradients to vanish or explode when information is passed through the layers. During the back propagation, the feedback from the last layer starts to vanish or in some cases explode as it passes through the layers of the network. This vanishing effect of the gradients makes the network untrainable. Therefore, residual connections are included to overcome this by creating another path for information to reach later layers by skipping some layers. For example, in the case of the traditional feed forward connection, the data flows from layer (i) to layer ($i+n$), and function (F) represents these layers. The input (x) would pass through the layers one by one and produce the output $F(x)$. In the case of residual connections, the identity map is first applied to input (x) followed by the element wise addition $F(x)+x$ so that the information flow is easy and also the features are preserved. Such connections replicate the ensemble of small shallow neural networks and avoid gradient vanishing or exploding rather than solving them.

Networks

Convolutional neural network architectures or convnets have achieved or exceeded the results that were considered entirely dependent on the human realm. The basic building blocks of the CNNs are convolutional layers, pooling layers, and fully connected layers that correspond to feature extraction layers, dimensionality reduction and classification or regression layers respectively. Depending upon the application and type of the dataset, different architectures have been designed. For the higher dimensional input such as images, CNNs require a large amount of data so that it can automatically extract useful information from the training set and avoid manual feature extraction. The network size and depth affect the amount of data required for training the model. The network topology also defines the inductive bias that constrains the hypothesis space.

Further, training a CNN from scratch requires lots of data (millions) and high requirements computation hardware resources. However, such networks learn local and translation-invariant global features, making them effective in perceptual or vision problems. They are also repurposable and transfer of the feature extraction kernels learned from one domain can be transferred to the other domain using transfer learning. The transfer learning approach is highly effective as it reduces the size of the training set, and few hundreds or thousands of data are good enough to train the model with minor modifications. In the transfer learning approach, the fully connected layers that perform the classification or regression task are often replaced by the custom layers that keep the architecture's feature extraction layers undisturbed. However, it has to be noted that features extracted by the lower or initial layers have information about the input data. On the contrary, higher or deeper layers provide less information on the input data and more information about the target. This makes CNNs act as information distillation pipeline, and during the transfer learning, it is recommended to freeze all feature extraction layers or lower layers as they do not have information on target or use-case it has been trained on. In the computer vision domain, many pre-trained models such as VGG16, InceptionV3, EfficientNets, etc. are trained on large benchmark datasets such as ImageNet and MS COCO.

Loss functions

Loss function gives the quantitative measure of model's performance during the training process. It is also referred as objective function. This function has to be minimized during training process as it measures the difference between the model's prediction and ground truth. Depending on the type of task, data and hypothesis space, different loss functions can be considered. One can also use more than one loss function and assign weights to them accordingly. Some of the widely use loss functions for classification and regression tasks are L1 loss or mean absolute error, L2 loss or mean squared error, Huber loss, Cross-entropy loss, dice loss, and focal loss etc.

For semantic segmentation tasks, the loss functions can be categorized into four groups depending on its function. They are distribution-based loss, region-based loss, boundary-based loss, and compound loss. Distribution-based loss considers the dissimilarity between two distributions to calculate the loss value. Region-based loss aims to increase the overlap between predicted and ground truth results. Compound loss combines distribution and region-based loss. The boundary-based loss is a sub-category of region based loss function that is specifically used to compute the loss when the task is to learn the edges or boundary conditions (edge detection tasks) [118].

Optimizers

The optimizers update the network parameters based on the loss function values. This is a repetitive process that takes place after every training epoch and is implemented using variants of gradient descent (GD). GD takes a single scalar loss value as input and therefore, in case there are multiple loss function, the losses are combines to give a single scalar value [119]. There are different variants of the GD such as stochastic gradient descent, batch gradient descent, mini-batch gradient descent, adaptive gradients, Adam, RMSProp, etc. The choice of the optimizer is dependent on network size and other hyper-parameters such as learning rate, activation function primarily. A faster optimizer with a good learning rate is always desirable while training the large network.

For permanent magnet samples, the visual representation of grains in the KM images reveals subtle color variations between individual grains, as illustrated in *Figure 14*. Therefore, the classification of each grain demands a feature vector of higher dimensions. Moreover, extracting features from such complex microstructures presents a laborious challenge. Furthermore, the process of generating labels for KM images, particularly grain boundaries, is a tedious task that relies on manual input from domain experts [69]. To address these complexities, a deep learning model like U-net becomes imperative. This model is good at handling small datasets and overcomes the need for labor-intensive manual feature extraction steps [120].

The U-net architecture is a CNN specifically crafted for semantic segmentation tasks, offering high accuracy even with a relatively modest dataset for training. As illustrated in *Figure 10*, the network exhibits a distinctive contracting and expanding structure that aligns with the core principles of a conventional convolutional network.

The contracting path encompasses multiple blocks, each featuring a sequence of 3 × 3 convolutions followed by rectified linear units (ReLU) for non-linearity, and a 2 × 2 max-pooling operation with a stride of 2. This arrangement effectively achieves down-sampling, facilitating the capture of essential features and patterns. The number of kernels (also known as feature maps) doubles with each subsequent block, empowering the network to discern intricate structures and representations within the data. Through its contracting path, U-net captures features at various scales and resolutions using a series of convolutional layers and pooling operations. This multi-resolution context capturing is essential for recognizing both global and local patterns in an image.

At the bottom layer, a pivotal connection emerges between the contraction and expansion layers. Here, two successive 3 × 3 CNN layers are coupled with a 2 × 2 up-convolutional layer. However, the strength of the U-net lies in its expansion layer, which houses multiple expansion blocks featuring 3 × 3 CNN layers, each accompanied by a 2 × 2 up sampling layers. Throughout this process, the feature map undergoes progressive halving to ensure symmetry. The use of transposed convolutional layers for up sampling is a defining characteristic of U-net's expanding path. These layers enable the model to recover spatial information lost during down-sampling, leading to finer and more detailed segmentations.

A significant innovation is the integration of features from the corresponding contraction layers during expansion. After each block, the feature map is concatenated with the feature map of the respective contraction layer. This strategy ensures that the acquired features are optimally harnessed for the reconstruction process, reinforcing the network's ability to produce accurate segmentation. U-net employs short connections, also known as skip connections, that directly link layers with similar spatial resolutions in the contracting and expanding paths. This facilitates the transfer of information between different scales, enabling the model to capture multi-resolution contextual information effectively.

The ultimate layer incorporates a 1 × 1 convolutional layer, merging the individual feature vectors derived from the 64 components. These feature vectors are then coupled with the desired labels, culminating in a unified representation that captures the intricate relationships between the features and their corresponding labels [120]. In conclusion, the U-net architecture's symmetry, feature concatenation, specialized focus on semantic segmentation,

and its ability to handle small datasets effectively make it a unique and influential model in the realm of deep learning and image analysis.

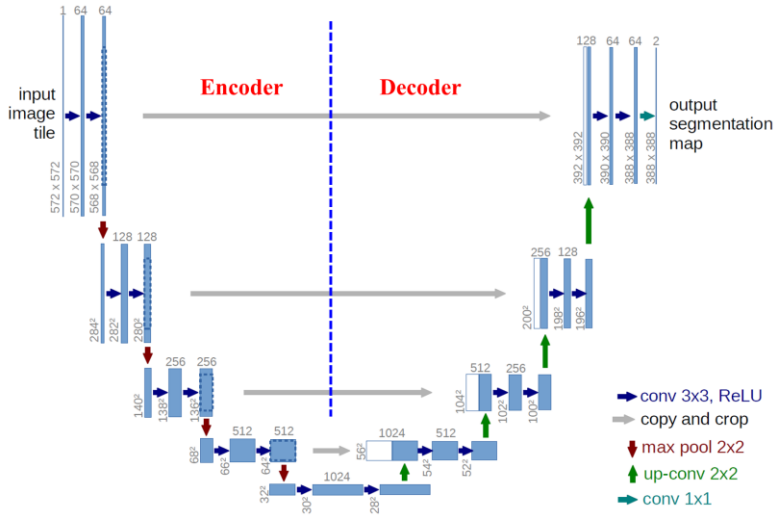


Figure 10: Architecture of the U-net algorithm, showing a contracting path on the left side and an expansion path on the right side. The blue boxes depict multi-channel feature maps, while the white boxes represent replicated feature maps. The various colored arrows symbolize distinct operations, as detailed in the legend at the bottom right [120].

2.3. Literature Review - ML for magnet research

To characterize a material, a good understanding of the relationship between the process, structure, and properties is important. This understanding enables the application space to expand and overcome the limitations of existing applications. *Figure 11* shows the materials science tetrahedron for characterizing the material [121].

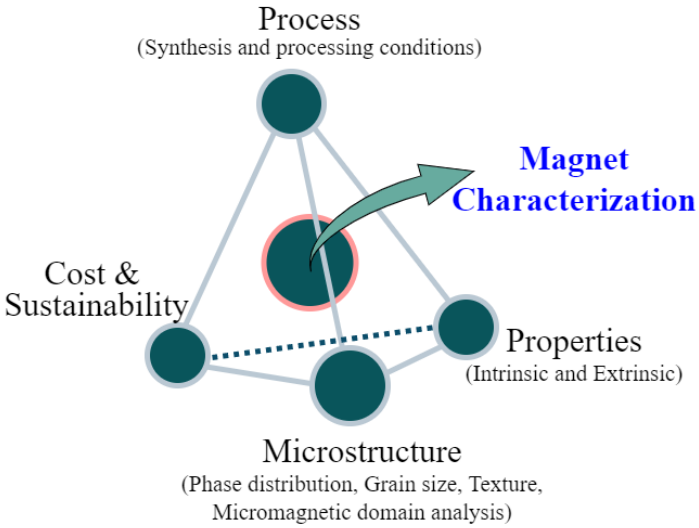


Figure 11: Materials science tetrahedron showing the correlation between process, performance, microstructure, and material's properties for characterizing the material [121]. Note that performance is replaced by cost and sustainability for magnet characterization.

Many specific methods have also been developed over the years to characterize rare earth permanent magnets. In recent years, the application of ML techniques in materials science has emerged as a powerful tool for understanding and optimizing the properties of advanced materials. This includes, at the macroscopic scale, changes in the intrinsic magnetic properties for different chemical compositions of the magnets within the solubility limits that could be useful for optimizing new desirable materials. Additionally, automated quantitative microstructure analysis has gained significant importance as a desired technique for ensuring quality control in materials, for tasks like phase quantification, grain size, defect detection, and texture analysis.

2.3.1. Property prediction for magnets

The acceleration of material property prediction has been predominantly driven by computational science, incorporating simulation-based methods such as DFT and molecular dynamics [122]. However, the high computational costs of these methods impose limitations on their practical implementation. Understanding the chemical and physical phenomena underlying materials is fundamentally linked to the experimental data. ML enables the extraction of such insights from experimental datasets, allowing the learning of laws governing these phenomena [123].

The composition of the magnets is a critical factor in determining their magnetic performance. The chemical makeup of these magnets has been adjusted and fine-tuned to optimize their magnetic properties. This optimization involves considering various elements and their concentrations within the material. By carefully manipulating these compositional elements, scientists have enhanced the magnet's ability to generate and maintain a strong magnetic field. Given solubility allowances, elemental substitutions in the 14:2:1 lattice, replacing Fe, Nd, or B at regular TM, RE, and B sites, are possible [24], [41], [48], [124], [125], [126]. However, there are limitations to the way the composition can be modified. These limitations are influenced by the principles of thermodynamics and the behavior of the 14:2:1 phase. In particular, the thermodynamic stability of the 14:2:1 phase is directly influenced by the solubility limit of elements within it. The solubility limit refers to the maximum concentration of a particular element that can be dissolved in a given material while maintaining its stability. Beyond a specific range of compositions, the material is anticipated to consist of a mixture of multiple phases [124], [125], [126], [127]. For example, in the $Fe_{14-x}TM_xNd_2B$ system [128], the presence of multiple phases becomes evident when exceeding the solubility limit denoted by 'x'. This limit could differ by values such as x is 2 for chromium (Cr), 1.5 for copper (Cu), or 3 for aluminum (Al). This solubility limit varies when considering a system with rare earth substitutions, as seen in the $Fe_{14-x}TM_xPr_2B$ system, where the limit changes to x as 3 for chromium (Cr), 2 for silicon (Si), or 1 for gallium (Ga) [126]. Therefore, the crystal structure of the material is influenced by the introduction of substitutions at either the TM or RE sites within the 14:2:1 unit cell [24], [41], [128].

The potential occupation of interstitial sites remains a consideration, though at present, hydrogen stands out as the most recognized element to occupy these positions [41], [129]. A thorough understanding of interstitial and substitutional site behaviors can present possibilities for designing compositions within the 14:2:1 framework. Studying $Fe_{14}Nd_2B$'s crystal structure

through diffraction techniques reveals how elements occupy specific lattice sites. X-ray diffraction (XRD) on single crystals precisely locates atomic positions, which are crucial for determining site occupancy. Neutron diffraction on powder samples is key to identifying light elements like Boron (B) in the presence of heavier elements. It also elucidates magnetic structures by interacting with magnetic moments. XRD on powder samples provides an averaged view of site occupancy and crystal structure. Key findings include site-specific occupancies for Nd, Fe, and B atoms, insights into magnetic properties influenced by Fe arrangements, and data on unit cell parameters and stability variations. Nd atoms occupy sites 4*f* and 4*g*, Fe atoms are distributed in distinct positions (16k_1, 16k_2, 8j_1, 8j_2, 4*e*, and 4*c*), and B atoms are at 4*f* site as shown in *Figure 1* [130].

The reported site occupancy of elements replacing Fe in $Fe_{14}Nd_2B$ varies between different studies. Substituting Al, Mn, and Si for Fe in the $Fe_{14}Nd_2B$ crystal structure has distinct impacts on both the structural and magnetic properties of the material. Al preferentially occupies the k_2 and j_1 sites, leading to an increase in the axial ratio (*c/a*) and altering magnetic interactions due to the weakening contributions from these Fe sites [124]. Mn follows a similar substitution pattern, preferring the k_2 and j_1 sites, but significantly impacts the magnetic properties by introducing antiferromagnetic interactions, resulting in a dramatic reduction in T_c. This is attributed to Mn's tendency for antiferromagnetic interactions at close distances and its large atomic size that affects the local environment [124]. However, Marasinghe et al. [131] have reported that it occupies j_2. Si shows a more complex site occupancy behavior. Initial studies indicated a preference for the 4*c* site and a moderate preference for the 8j_1 site, while avoiding the 16k_1, 8j_2 and 4*e* sites. However, Grössinger et al. [132] suggest that Si can occupy a broader range of sites, including the 16k_1 and 16k_2 sites, and some studies indicating preferences on the order of 16k_2, then 16k_1, and finally 4*e*. Pringle et al. [133] suggests that Si can occupy the 16k_1, 16k_2, 8j_1, 4*c*, and 4*e* sites. This variability in site occupancy reflects Si's flexibility in integrating into the crystal lattice, influenced by coordination with rare earth atoms such as Nd or Y, which stabilizes the crystal structure and modifies magnetic properties [41]. For Cr replacing Fe in $Fe_{14-x}Cr_xY_2B$, Bara et al. [134] indicates Cr occupies site 8j_2, while Moze et al. [135] reported that it occupies site 4*e*. These differences can arise from variations in crystal structure conventions, experimental conditions, measurement techniques, and sample preparation methods used in each study. Similarly, the site occupancy by Ga varies among different reported studies particularly at higher concentrations. Grössinger et al. [132] proposed that Ga prefers the 16k_2 site, whereas Wiessinger et al. [136] suggested that it occupies the 8j_1 sites. Therefore, consensus on the precise location of Ga within the pure $Fe_{14}Nd_2B$ structure, especially at higher Ga concentrations, has yet to be established. For $Fe_{13}GaNd_2B$ and $Fe_{12.2}Ga_{1.8}Nd_2B$, specific site occupancies have been reported in the literature: Ga atoms have been found to occupy the 4*c*, 8j_1 and 16k_2 sites within these alloys [137]. This indicates that Ga can be substituted into multiple positions within the $Fe_{14}Nd_2B$ crystal structure depending on the exact composition and synthesis conditions of the alloy. And for Cu replacing Fe in $Fe_{14-x}Cu_xNd_2B$, j_1 and k_2 are preferred sites [41]. For $Fe_{14-x}Nb_xPr_2B$, Jurczyk et al. [138] reported that Nb occupies j_2 and k_1.

In the study of phases analogous to those based on boron but with carbon substitutions, similar trends in the preferential occupation of iron sites have been observed. Mn in $Fe_{14-x}Mn_xRE_2C_1$ phases has been noted to favor the 8j_2 sites. For phases where nickel (Ni) substitutes iron, a strong preference for the 16k_2 sites, with some occupancy also observed in the adjacent 16k_1 sites. Silicon, another substituting element, exhibits a distribution across the 4*c*, 8j_1, and 16k_2

sites within these structures. Moreover, Cu and vanadium (V) tend to predominantly occupy the $16k_2$ sites, although they also exhibit some presence at the $16k_1$ sites. These findings underscore the complexity and specificity of site preferences for different substituent elements within the crystal lattice, reflecting their atomic sizes and likely interactions with neighboring atoms in these materials [41]. By exploring the boundaries of what elements can be incorporated while maintaining stability, researchers can tailor the composition of the 14:2:1 magnet to optimize their performance in various technological contexts.

Traditionally, the prediction of properties in materials has relied on employing physics-based models such as DFT, atomistic spin models, and first principle approaches. Often, it is not feasible to accurately model all underlying physics due to complexity or cost constraints, leading to necessary simplifications and approximations that introduce limitations. Unlike conventional prediction approaches, another strategy for developing prediction methods involves finding the underlying patterns from experimental data. Experimental data directly reflect the chemical and physical complexities of the material. ML can identify the underlying phenomenon hidden within a dataset. The field of materials has seen a surge of innovative ML applications that have significantly improved the discovery of new materials [2], [4], [139], [140]. In recent years, a wide array of alloy properties can be forecasted using ML techniques. Aspects like heat treatment performance, mechanical characteristics, plasticity, superplastic behaviors, and various other material properties can be effectively predicted through ML methods [141], [142], [143], [144].

Over the last decade, ML has also been applied to predict various properties in magnets. Park et al. [145] developed an ML-driven property prediction model focused on FeNdB magnets and $Fe_{14}Nd_2B$ phase, specifically targeting coercivity ($\mu_0 H_{ci}$) and $(BH)_{max}$. Their approach involved ANN and SVR models, achieving correlation coefficient (R^2) scores of 0.89 and 0.87, respectively. The input data for their model was derived from micromagnetic simulations. On the other hand, Möller et al. [146] used ML approach to predict K_1 for permanent magnets; however, their study was confined to permanent magnets based on $REA_{12}X$ phase, where RE is Ce or Nd, X is element B, C or N and A comprises Mn, Fe, Co, Ni, Ti, V, Cr, Cu, Zn, Al, Si and P. Moreover, when it comes to linearity of trained ML models, linear models have proven suitable for characterizing physical properties such as lattice parameters within unit cells. Examples include perovskite systems such as ABX_3 [147], [148], or $A_2B_2X_6$ [149], [150], where A represents a large cation with varying valence, B denotes a transition metal, and X stands for oxides and halides. A linear model for the lattice parameters has also been applied to the $NiFe_2O_4$ system [151]. However, this category of multi-linear models [152] for spinel structure has recently been used as a benchmark for evaluating Gaussian process regression models in Zhang et al.'s work [153]. It has to be noted that the materials dataset for developing a property prediction model from the structured datasets such as tabular data is associated with some notable challenges such as lack of large generalized dataset or pre-trained model that can be used for fine tuning on limited dataset, which is quite common in computer vision tasks. The presence of outliers and bias in the experimental and simulation results reported across the literatures is unknown most of the times.

It is well known that tabular materials datasets differ from those used for common ML applications, such as computer vision [154]. Additionally, the perception that a well generalized ML model requires massive dataset has to be carefully reviewed as some of the benchmark

datasets such as the Iris dataset which is widely used for classification and clustering tasks has 50 samples from each of the three classes (species) [155]. Therefore, the major challenge for ML application in the materials science field is strategically identifying the tasks to be solved, input data and features effecting the target parameter. When dealing with a small amount of data, such as less than 1000, the quality of the data and the number of useful features tend to be more significance than relying on extensive computational resources with large amounts of data and weakly correlated features [156]. An optimized way involves having a trade-off between bias and variance by considering optimal feature selection techniques or creating a model structure that aligns better with the underlying physics. Another approach could be to narrow the scope of the machine learning task by confining the material design possibilities or using machine learning to address a specific subset of the broader problem. This targeted approach acknowledges the data availability challenges and still provides an effective predicting model.

Moreover, the choice of regression model for the property prediction task is highly dependent on the size of the training set, the features extracted, the distribution hypothesis and inductive bias of the model. The trade-off between bias and variance, based on the interaction between the data and the model, should always be considered. The ability of the learning model to generalize the information from the training set to find the optimal relationship between input and the expected target value is determined by bias and the model's sensitivity to perturbations in the training set, which is the variance [157]. The choice of a single model for a specific regression task is highly subjective and can be either linear or non-linear. It could also be either parametric or non-parametric. Moreover, it is well known that in the structured dataset involving tabular data, the tree-based models often outperform the advanced DL models, and one major reason for its superiority is that there are ensembled models [110].

For the property prediction task, approach involving ensembling of traditional ML models has been adopted as the material's database is small and tabular, which could potentially result in overfitting. Ensembling the model involves combining multiple models for the learning task using different methods such as bagging, boosting, and stacking. Stacking involves specifically heterogeneous weak or base learners trained independently on the same dataset. The output of each base learner is combined with either a uniform or non-uniform weightage [158]. Nevertheless, constructing a stacked model is scientific and requires the right choice of base learners and distribution of the optimized weight. Further, it has been proven that the optimal weight distribution among base learners leads to better bias and variance trade-off than individual base learners [159], [160]. The stacked model approach is also termed the voting approach. For the voting approach in regression task, either uniform weightage (UW) or non-uniform weights (N-UW) can be assigned to all base learners. N-UW involves assigning floating values between 0 and 1.

Constructing an effective voting regressor (VR) model involves three steps, choice of base learners, parameter optimization of base learners, and weight assigned to each base learner. Note that the optimized hyperparameters should be considered for each chosen base learner. There are different approaches for calculating the weightage of base learners in VR models, and it is mainly split into two categories: performance-based [161], [162], [163] and rank-based weights [164], [165]. The performance-based approach involves using the cross-validation metric value evaluated on the same validation set as the weights. Each base learner's hyperparameters are fine-tuned on the training set and evaluated for performance on the

validation set using metrics such as mean absolute error (MAE), root mean squared error (RMSE), R^2 score etc. The rank-based approach assigns rank to each base learner based on its performance on the validation set. This decides the number of votes each learner has to calculate the final prediction through the N-UW VR model. The rank is a non-zero and a positive integer. Equation (40) shows the working of the VR with four base regressors for the prediction of target (Y) from the feature vector (f). For a given input feature vector (f), each base regressor $F_i(f)$ predicts the output that is multiplied by the assigned weight (w_i). VR calculates the average weighted prediction from each base regressor to obtain the final output.

$$\text{Prediction } (Y) = \frac{F_1(f) * w_1 + F_2(f) * w_2 + F_3(f) * w_3 + F_4(f) * w_4}{(w_1 + w_2 + w_3 + w_4)} \tag{40}$$

$$\text{if } w_1 = w_2 = w_3 = w_4 \text{ (uniform VR)}$$

For example, if the performance of each of the trained four base regressors in the validation set when measured using the metric R^2 score are 0.97, 0.91, 0.89, and 0.94, respectively, then these metric values are considered as weights in the case of the performance-based weight approach. In the rank-based approach, the base regressors are assigned a rank based on validation error such that the regressor with lowest error is assigned the highest rank and the least rank is assigned to the regressor with the highest error. This would mean that the base regressor with the highest rank has more weight during the final prediction calculation. Based on the above R^2 scores, the ranking would be 4, 2, 1 and 3, respectively; hence, these are the assigned weights. Equations (41) and (42) show the performance-based and rank-based approach for 4 base learners.

$$\text{Prediction } (Y_P) = \frac{F_1(f) * 0.97 + F_2(f) * 0.91 + F_3(f) * 0.89 + F_4(f) * 0.94}{(0.97 + 0.91 + 0.89 + 0.94)} \tag{41}$$

$$\text{Prediction } (Y_R) = \frac{F_1(f) * 4 + F_2(f) * 2 + F_3(f) * 1 + F_4(f) * 3}{(4 + 2 + 1 + 3)} \tag{42}$$

Density of the 14:2:1 magnetic phase

From an application perspective, magnetization of the 14:2:1 phases holds paramount significance [26], [27], [28]. While the magnetic moment of phases is reported in two different units such as [μ_B/f.u.] ([μ_B]: Bohr magneton, [f.u.]: formula unit) and [emu/g] (emu: electromagnetic unit) [166], [167], [168], deriving magnetization from this magnetic moment needs their specific mass density (ρ_m) values. However, there are cases where the ρ_m information for that specific 14:2:1 phase is not available.

The mass density of phases can significantly vary depending on their composition, particularly due to the inclusion of alloying elements. In 14:2:1 phases, the ρ_m is influenced by the characteristics of its constituent elements. Replacing Fe and Nd with different rare earth or transition metal elements in 14:2:1 can significantly impact the resulting alloys' density. For example, in the FePrB phase, praseodymium (Pr) replaces neodymium, potentially resulting in a less denser phase due to its slightly lower atomic mass. On the contrary, in the FeCoB 14:2:1 phase, Co substitutes Fe, offering a higher atomic mass and potentially increasing the alloy

density. Furthermore, using heavier rare earth elements such as holmium (Ho), Tb, or Dy in alloys like FeHoB, FeTbB, and FeDyB, respectively, could result in higher densities due to their higher atomic masses compared to neodymium.

Table 4: The lattice constants (a and c) and density (ρ_m) of the tetragonal $Fe_{14}RE_2B$ phases at room temperature [24].

Compound ($Fe_{14}RE_2B$)	**a [Å]**	**c [Å]**	**ρ_m [g/cm³]**
$Fe_{14}Nd_2B$	8.80	12.20	7.60
$Fe_{14}Pr_2B$	8.80	12.23	7.54
$Fe_{14}La_2B$	8.82	12.34	7.40
$Fe_{14}Ce_2B$	8.76	12.11	7.67
$Fe_{14}Sm_2B$	8.80	12.15	7.72
$Fe_{14}Gd_2B$	8.79	12.09	7.87
$Fe_{14}Tb_2B$	8.77	12.05	7.96
$Fe_{14}Dy_2B$	8.76	12.01	8.05
$Fe_{14}Ho_2B$	8.75	11.99	8.12
$Fe_{14}Er_2B$	8.73	11.95	8.22
$Fe_{14}Tm_2B$	8.73	11.93	8.26
$Fe_{14}Lu_2B$	8.70	11.85	8.47
$Fe_{14}Y_2B$	8.76	12.00	7.00
$Fe_{14}Th_2B$	8.80	12.17	8.86

Table 4 shows the ρ_m for $Fe_{14}RE_2B$ phases at room temperature. In each scenario, while efforts are made to maintain the specified 14:2:1 ratio, variations in density occur due to differences in atomic masses and concentrations of substituting elements.

When using the Archimedes approach for experimental density measurement of a 14:2:1 phase, the material must be free of any internal porosity or the presence of other phases. Otherwise, including such internal porosity or additional phases in volume measurement can result in an inaccurate determination of the ρ_m for the 14:2:1 phase. This method is also labor-intensive and expensive, requiring a pure single-phase material.

In the work of Herbst et al. [169], a calculation of X-ray-based mass density for 14:2:1 phases was conducted. In addition to the chemical composition, this calculation critically relies on the availability of measured lattice parameters. The authors highlighted that within each unit cell, there are precisely 4 formula units or 68 atoms [169], comprising sites for TM, RE, and B. In contrast to the density determination of 14:2:1 phases, the theoretical density of sintered magnets with multiple phases is frequently reported [41], [170]. These sintered magnets have not only the 14:2:1 phase but also include other constituents such as grain boundary phases, oxide phases, porosity, and various intermetallic phases. Empirical models estimate the density of the 14:2:1 phase in such cases. However, applying these empirical models relies on accurate knowledge of the volume fractions and types of all phases present in the microstructure.

Therefore, there remains a gap in the literature concerning ML-based empirical models tailored specifically for predicting the ρ_m of the 14:2:1 phase. In the work [98], a data-driven approach has been reported to fill the existing gap by predicting the ρ_m from the chemical composition of the 14:2:1 magnetic phase.

Curie temperature of the 14:2:1 magnetic phase

An essential requirement for any potential 14:2:1 phase candidate is its ability to maintain sufficient magnetic properties (magnetization and coercivity) under operational temperatures, a characteristic determined by its Curie temperature. These operational temperatures typically span from 220 K to 450 K, necessitating that the T_c of the phase exceeds approximately 600 K [171], [172]. Irrespective of the temperature range, prediction of the T_c holds particular significance due to its influence on various temperature-dependent intrinsic features, such as the magnetocrystalline anisotropy constant and saturation magnetization, as well as extrinsic attributes such as remanence or coercivity.

Determining the T_c for the potential novel 14:2:1 phase through the experimental process can be time-consuming. However, using DFT presents scope to predict T_c by assessing the spin magnetic moment at each atomic site and the magnetic interactions between these sites [173]. This methodology is particularly suited for evaluating the behavior of 3*d* valence electrons. However, it has limitations when dealing with 4*f* electrons, as seen in rare earth elements. The mathematical description of the spin-orbit interaction or the robust correlation effects characteristic of 4*f* electrons is challenging [174]. To overcome this concern, a partial solution involving the integration of DFT modeling with Dynamical Mean Field Theory [128] has been proposed. Although this hybrid approach is promising, its full potential is still being established [175]. Even in relatively straightforward systems like the body-centered cubic iron system, this method's predictions fail to provide precise T_c values. Instead, the predicted T_c values span a considerable range, extending from 800 K to 1800 K, demonstrating the complexity of the task. Adding to this computational process, Drebov et al. [176] introduced an alternative approach. They showcased the utility of combining DFT calculations with Monte Carlo simulations for high-throughput screening. This computational strategy allows the efficient evaluation of potential candidates with high T_c. By harnessing the power of simulation, this method offers a streamlined process for identifying promising 14:2:1 phase candidates based on their T_c values.

Recently, ML-based approaches have gained popularity for their ability to predict T_c across various categories of permanent magnets. As the complexity of the data increases, characterized by an increase in the number of chemical constituents, there has been a significant increase in the associated prediction errors [173], [172], [177]. For example, in binary magnets, which involve systems composed of two components, specifically those consisting of TMs from the 3*d* series and REEs from the 4*f* series, Nguyen et al. [173] demonstrated that the T_c prediction error, quantified in terms of the MAE, could be minimized to a greater extent, reaching as low as 32 K. Similarly for the ternary magnets, with a notable example being the Fe-Co-Al system, Nelson et al. [172] displayed the trajectory of the error, indicating that upon tuning a random forest regression model, the error could be reduced to around 57 K in terms of MAE. Recent work by Long et al. [177] illustrated that employing random forest regression yielded a predictive MAE of approximately 69 K in the cases that involve higher complexity, including ferromagnetic and antiferromagnetic quaternary and quinary magnets. However, achieving such a high level of predictive accuracy necessitated the use of a substantial array of features, amounting to 139, which includes both the chemical composition and the crystallographic attributes of the materials under consideration.

An average absolute error of 69 K might prove inadequate for multi-component 14:2:1 magnet, which typically includes 4 or 5 distinct elements, primarily when dealing with T_c values approaching the vicinity of 600 K. This situation implies that a novel magnet composition could yield a predicted T_c that falls within the operating temperature range, potentially resulting in magnets prone to unreliable demagnetization under working conditions. In the work [178], using the chemical composition as input features, a VR model has been trained to predict T_c at room temperature with an MAE of less than 20 K in unseen 14:2:1 magnetic phases.

Saturation magnetization of the 14:2:1 magnetic phase

In intermetallic compounds involving RE and TM such as Fe and Co, the high J_s and T_c of the transition metals are combined with the elevated K_1 of the RE metals, such as Nd, Pr, and Sm. The M_s or J_s of the 14:2:1 phase is crucial for various applications. In the literature reported values, magnetization is typically expressed using the [μ_B/f.u.] and can be conveniently converted to [emu/g]. The Tesla [T] unit (SI) is commonly used in practical engineering devices. Knowledge of the ρ_m is required to convert from [emu/g] to [T].

Hund's rule governs the interaction between the 3*d* spins of TM and the 4*f* spins of REEs, resulting in an antiparallel coupling between them. In compounds featuring light rare earth elements such as Pr, Nd, Pm, Sm, Eu, and Gd, this negative spin coupling leads to a parallel alignment of RE and TM moments, inducing a ferromagnetic temperature behavior characterized by more significant J_s values. Conversely, compounds containing heavy rare earth elements like Tb, Dy, Ho, Er, Tm, and Yb exhibit an antiparallel alignment of RE and TM moments due to the same negative spin coupling, resulting in a ferrimagnetic temperature behavior with smaller J_s values. In cases where rare earth elements lack a 4*f* moment, such as La, Ce, Lu, Y, and Th, the Fe sublattices display a ferromagnetic coupling. Furthermore, M_s can lead to distinct magnetic contributions [24], [41]. In $Fe_{14}RE_2B$ compounds, the average Fe moments are approximately 2.2 μ_B per atom, while in $Co_{14}RE_2B$ compounds, the average Co moments are about 1.4 μ_B per atom. These variations in magnetic behavior and M_s contribute to the diverse range of magnetic properties observed in materials containing the 14:2:1 phase [24], [41].

The evolution of data-driven methodologies has led to a paradigm shift in material design, prominently driven by materials informatics. Over recent decades, the use of ML techniques has increased, aiming to predict the magnetic properties of materials and expedite the search for new compositions [100], [146], [178], [179], [180]. Notably, the prediction of M_s has received attention through ML-based regression models, particularly in the domain of hard magnetic phases. For instance, Möller et al. in [146] developed models for rare earth intermetallic $REA_{12}X$ compounds (RE as Ce or Nd, A representing elements like Mn, Fe, Co, Ni, Ti, V, Cr, Cu, Zn, Al, Si, or P, and X as B, C, or N). In another approach, Giaremis et al. [181] used ANN to establish the correlation between the structure of Sm-Co permanent magnets and their saturation magnetization. At the same time, Wang et al. [182] introduced an ML-based strategy to predict saturation magnetization in Fe-based soft magnets. Although data-driven methods for predicting M_s have been explored for various phases beyond the 14:2:1 magnetic phases [183], [184], [185], research efforts have been limited when it comes to using ML for predicting intrinsic properties like density and T_c using elemental composition as features for 14:2:1 magnetic phases [98], [178].

2.3.2. Quantitative microstructure analysis of rare earth permanent magnets

Microscopy techniques are used to observe the samples at different magnifications. Depending on the type of microscopy technique used, different essential information is obtained from the images. Optical light microscopy, Kerr microscopy, scanning electron microscopy, and transmission electron microscopy are some examples that provide information at different scales. *Table 5* gives an overview of the information that can be obtained for rare earth magnets using LM, KM, and SEM techniques. However, effective and accurate quantitative analysis of the information in the microstructure of the samples acquired using microscopy techniques requires various analytical tools. As discussed in section 2.1, optimal control of the magnet's microstructure is important to ensure optimal performance of the permanent magnets.

Table 5: The different information that can be obtained from the magnet using optical light, Kerr, and scanning electron microscopy techniques. Different microscopy techniques are considered depending on the information needed for magnet characterization.

Microscopy techniques	Sample information
Optical light microscopy	Various occurring phases, volume fraction of phases, grain size, grain shape, distribution of phases
Kerr microscopy	Magnetic domain structures, grain size and grain orientation
Scanning electron microscopy	Various occurring phases, volume fraction of phases, grain size, grain shape, distribution of phases, crystal orientation texture, elemental distribution

Grain size and domain structure analysis from microscopy data

In recent years, substantial research has been carried out to apply ML strategies to detect grain boundaries in material microstructures from LM, SEM and KM images.

Pusch et al. [186] developed a semi-automated technique for quantifying domain structure through correlative imaging. They employed a sample of FeNdB sinter magnet to acquire correlative images using a polarized light microscope (polished surface) and an optical microscope (etched surface) to identify grain boundaries. The alignment of these two images facilitated the detection of grain boundaries. By applying the maximum and second maximum thresholds derived from the histogram of each grain, the method extracted domain patterns, as shown in *Figure 12*. Later, these extracted domains were categorized into six groups to facilitate analysis. This approach necessitates meticulous sample preparation to ensure the visibility of grain boundaries in light microscopy images. Moreover, precise acquisition of correlative images is crucial, with human intervention required in cases where the grain boundaries in the two images do not align accurately.

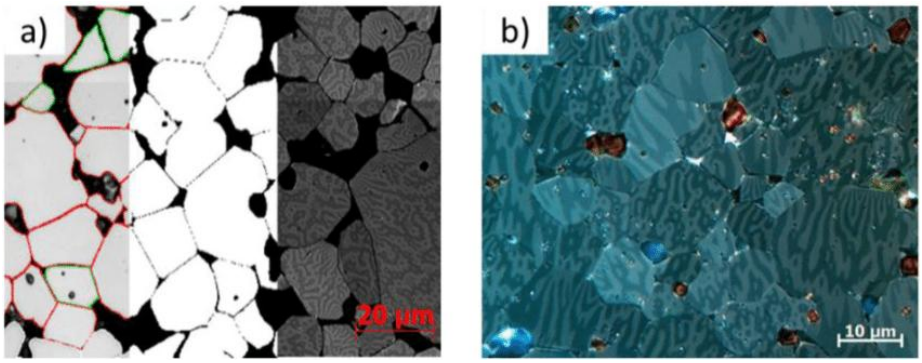


Figure 12: (a) A light microscopic image of the FeNdB magnet sample, revealing the grain boundaries through the threshold-based image segmentation and their correlation with a Kerr micrograph image for grain extraction (b) a Kerr micrograph image of the sample [186].

Exl et al. [187] investigated the effect of grain size and its position on magnetization reversal in the synthetic dataset that was ideally structured with large-size grained FeNdB permanent magnets using machine learning approaches. Furat et al. [71], in his study, used the U-net architecture to detect grain boundaries in time-resolved computer tomography data of an AlCu sample. They evaluated various U-net configurations, including 2D U-net, multichannel 2D U-net, and 3D U-net, to process images and compare segmentation results. Remarkably, the 3D U-net outperformed manually labeled ground truth data, indicating its accuracy in identifying grain boundaries. The study emphasizes the role of precise dataset labeling for enhancing predictive performance. Overall, Furat et al.'s [71] work showcases U-net's efficacy for grain boundary detection while underscoring the importance of meticulous data labeling.

Neuman et al. [188] introduced a hybrid algorithm that combines a random forest classifier and a watershed algorithm to trinarize the image data of the Ibuprofen tablets. This combined method aimed to detect pores, active pharmaceutical ingredients, and microcrystalline cellulose. The random forest classifier was trained on a higher-dimensional feature set to categorize the three distinct regions of interest. The study compared the performance of image trinarization using the watershed algorithm, the random forest classifier, and the hybrid model. The hybrid algorithm resulted in superior overall performance, optimizing the accuracy of the task. However, this approach has limitations. The watershed algorithm's stability is not promising, often necessitating the user to pre-define markers for satisfactory results. In a separate instance, Jiang et al. [189] also adopted a feature-based machine learning approach in their work. Their approach uses superpixel extraction, followed by creating a novel higher-dimensional feature set that includes color, texture, and edge information from each superpixel. This methodology was applied for segmenting grains in thin-section images of rocks under a polarized light microscope. Multiple images were captured at varying angles, and the simple linear iterative clustering (SLIC) technique was utilized to oversample each image based on pixel intensity criteria. This method achieved an F1-score of 0.87 after validation against the maximum intensity image.

In the context of grain boundary detection in material samples, the task is parallel to edge detection in natural or biomedical images. Prominent DL models for edge detection include DeepContour [190], DeepEdge [191], Sketch Tokens [192], and Holistically Nested Edge Detection (HED) [99]. These models were assessed using the Berkeley segmentation dataset (BSDS) and benchmark [193], featuring 200 training, 100 validation, and 200 testing images.

The HED model framework yielded superior outcomes to those of recent CNN-based edge detectors. It achieved a fixed contour threshold optimal dataset scale (ODS) of 0.78, closely approaching the human ODS of 0.80. While these models excelled in natural images, they failed in detecting grain boundaries or essential edges in images from of materials such as Kerr micrographs of magnets.

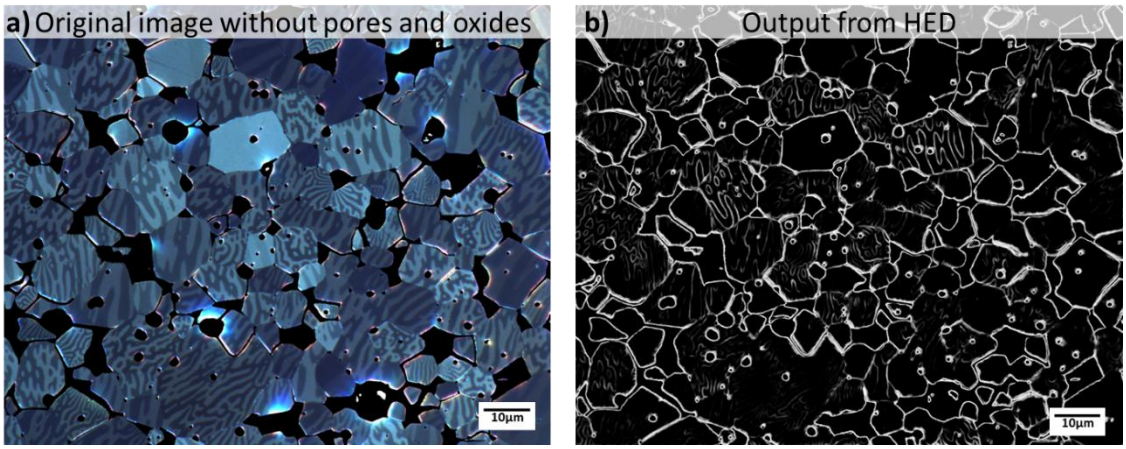


Figure 13: (a) Kerr microscopy image of FeNdB sintered magnet sample acquired at 1000x magnification after pre-processing step to remove the pores and oxides (b) predicted grain boundaries from the pre-trained HED model. The HED model has been trained on the BSDS dataset [99].

Figure 13 demonstrates the HED model's performance on a sample magnet image obtained through Kerr microscopy. While the HED model effectively detected grain boundaries in specific image regions, it struggled to differentiate grain boundaries from domain patterns. This contrast highlights the necessity for a customized edge detection model, tailored to magnet samples analyzed using Kerr microscopy.

The alternative conventional method for grain analysis is the EBSD technique, as detailed by Hohs et al. [59]. EBSD has proven to be an exceptionally effective and valuable tool for analyzing grain structures in diverse materials [194]. This technique offers numerous benefits, including the assessment of grain size distribution, texture, and orientation [194], [195], [196]. Nevertheless, when analyzing grain size or orientation across larger areas, such as obtaining size or orientation distributions for entire components, EBSD becomes impractical due to its time-consuming nature.

To address this limitation, an attempt has been made to combine ML techniques with computer vision methods [69]. This approach aims to directly determine the grain size distribution from light microscopic images of magnetic materials. This approach offers sufficient accuracy and reduced processing time than EBSD and other manual methods.

Methods for quantifying anisotropy in magnets

Magnetic materials are generally analyzed through magnetometric techniques. When evaluating manufacturing irregularities, a broad-scale analysis is often necessary, typically achieved by measuring the magnetic dipole moment of the magnet concerning its geometric axes. This approach offers insights on a comprehensive level and necessitates the complete

magnetization of the sample. Additionally, the measurement of remanence and coercivity, which are essential parameters for magnet performance, is influenced by anisotropy.

Diffraction-based methods: For the quantitative analysis of microstructures such as grain orientation, size, and material textures, can be carried out through a less conventional strategy for acquiring locally resolved texture information on substantial length scales involves macroscopic diffraction techniques such as XRD, as proposed by Bunge [67]. While this method gives valuable insights into crystallographic anisotropy and homogeneity, it does not offer microstructural-level details. However, the SEM-EBSD technique has emerged as a state-of-the-art tool. This diffraction-based method has high precision, but is best suited for the analysis of small areas within samples [197]. Therefore, to obtain higher resolution at the microstructural level, microscopy techniques are imperative.

Optical microscopy methods: Kerr microscopy, applied to both permanent and soft magnetic materials, has become a well-established technique for visualizing magnetic domain structures [77], [78]. Additionally, polarized light optical microscopy has proven to be useful in determining the grain orientation of materials with hexagonal structures and has found applications in geological studies covering extensive specimen areas [198], [199], [200], [201]. Jin et al. [202] determined the orientation of α-titanium using polarized light optical microscopy and later compared it to the EBSD approach. This method uses a physics-based forward model capable of simulating the interaction between polarized light and the material, thereby predicting grain orientation. The difference between the predicted values and the EBSD measurements ranged from 2.5° to 20°. Gaskey et al. [203] conducted a study that involved the characterization of crystal orientation in pure nickel coins and polycrystalline silicon solar cells using directional reflectance microscopy (DRM). DRM is an optical microscopy technique that captures crystal lattice information through DRM maps generated when the sample is observed under varying illumination angles. Through mathematical analysis, a directional reflectance profile (DRP) and the Funk-Radon transform (FRT) are used to achieve 3D orientation indexing. This method resulted in orientation differences of 4° to 10° compared to EBSD measurements. Similar efforts to extract partial grain orientation information from DRM have been performed by Seita et al. [204] and Hara et al. [205]. Other methods to determine grain orientation from microscopy data include spatially resolved acoustic microscopy [206], Raman microscopy [207], and ultrasound methods [208].

Data-driven methods: Advancements in data-driven approaches and analyses have prompted the use of machine learning and deep learning techniques to accelerate quantitative microstructure analysis, overcome the limitations of conventional methods, and enhance the characterization of materials by facilitating crystallographic grain orientation analysis. Deep learning models have been developed to forecast the orientation of the EBSD pattern indices, presenting a solution to the challenges associated with conventional EBSD indexing [21], [209], [210], [211], [212]. Moreover, there has been an active application of correlative microscopy data with machine learning techniques to develop transfer functions for knowledge integration from diverse microscopy methods [70], [213], [214]. In a study by Fitzek et al. [213], Raman-SEM-EDX data were used to classify phases within WO_3-WS_2 powder and volcanic rock samples using a random forest classifier. Wittwer et al. [214] developed the EulerNet model, which takes the correlative microscopy data from DRM and EBSD to predict the crystal orientation of Inconel 718. DRM captures the local surface reflection intensity across various

illumination angles as a continuous 2D array of 6x72 values for each grain. The EulerNet model takes DRM input signals as 2D arrays for each grain and maps them to Euler angles derived from EBSD. The trained EulerNet achieved a disorientation angle of 6.7° in the validation set. In [68], we explored the possibilities of combining ML approaches with the correlative microscopy approach that combines Kerr microscopy information with SEM-EBSD to train a model to extract the crystallographic grain orientation as reduced Euler angles. Once trained, the model can predict the grain orientation from the Kerr microscopy images only.

3. Experiments

3.1. Materials and sample preparation

For the experiments on the magnet characterization using the ML approaches, non-commercially available sintered FeNdB type permanent magnets were used. All the samples were in a demagnetized state, consisting of both anisotropic and isotropic types.

Table 6 lists all the samples that were used for the experiments. The characterization of the magnet here involves grain size analysis and texture analysis from the sectioned samples using the correlative microscopy method. They were prepared for microscopy using water-free metallographic polishing to avoid oxidation of sensitive phases. The final preparation step was made with 0.25 µm diamond suspension polish. The magnet samples have ECD values ranging between 3 µm to 14 µm. ECD refers to the diameter of a hypothetical circle that has the same area as the grain being measured.

Table 6: List of all the FeNdB sintered magnet samples used for the experiments. The sample characteristics of each cross-section label are shown in the table. The classification of the samples into train and test set for grain boundary detection and orientation prediction task is provided.

Sample code	**Measured ECD Φ**	**Remannence B_r**	**Coercivity H_{ci}**	**Maximum energy product $(BH)_{max}$**	**Alignment type**	**Sample type**	**Dataset type**	
	[µm]	**[T]**	**(kA/m)**	**kJ/m³**			**Grain boundary**	**Grain orientation**
FeNdB-A	7	1.30	1355	325	AP	Anisotropic	Train	Not conisdered
FeNdB-B	10	1.08	2625	225	AP	Anisotropic	Test	Test
FeNdB-E	10	1.11	2625	235	AP	Anisotropic	Train	Train
FeNdB-AP	10	1.34	1115	340	AP	Anisotropic	Test	Test
FeNdB-TP	10	1.41	1115	385	TP	Anisotropic	Test	Test
FeNdB-IP	10	1.44	1115	400	IP	Anisotropic	Test	Test
FeNdB-PSD-01	9	1.11	2625	235	AP	Anisotropic	Test	Test
FeNdB-PSD-05	9	1.11	2625	235	AP	Anisotropic	Test	Test
FeNdB-PSD-11	9	1.11	2625	235	AP	Anisotropic	Test	Test
FeNdB-PSD-15	9	1.11	2625	235	AP	Anisotropic	Test	Test
FeNdB-C	3	NA	NA	NA	AP	Anisotropic	Test	Not considered
FeNdB-D	3	NA	NA	NA	AP	Anisotropic	Train	Not considered
FeNdB-Q-Iso	14	NA	NA	NA	TP	Quasi Isotropic	Test	Train-partially
FeNdB-Iso	6	NA	NA	NA	NA	Isotropic	Test	Test

The area of the FeNdB-A sample is 457x338 µm². The FeNdB-E and FeNdB-B each have an area of 462x363 µm², while FeNdB-C and FeNdB-D each have an area of 250x187 µm². Samples FeNdB-AP, FeNdB-TP, and FeNdB-IP each have an area of 500x402 µm². The FeNdB-PSD-01, FeNdB-PSD-05, FeNdB-PSD-11, and FeNdB-PSD-15 areas are 462x371 µm² each. The isotropic samples, FeNdB-Q-Iso and FeNdB-Iso, each have an area of 125x100 µm².

Kerr microscopy and EBSD investigations from SEM were used for microscopic examination. Using the correlative microscopy approach described in section 2.1.5, the multi-scale information was collected. Firstly, the KM images of the samples were acquired using a 4x4 tiling approach that enabled the capture of large sample areas, as shown in *Figure 14*. These images provided insights into grain size and domain orientation using polarized light. Later, the same samples were investigated for texture information using the EBSD technique, involving a correlative microscopy approach. Texture analysis using EBSD mandates the measurement of grain orientation for a specific minimum number of grains, a quantity contingent upon crystal symmetry and texture sharpness. However, optimal texture analysis involves the evaluation of approximately 10,000 grain orientations to match the efficacy of XRD based texture characterization [215]. Consequently, it is recommended that large sample areas be analyzed through EBSD investigations. Initial investigations also indicate that at least 1000 grains are needed for quantitative texture analysis. Thus, by reducing the SEM magnification to 450x, the investigation area of the sample reached dimensions of up to 400x400 µm² without compromising the quality of the EBSD patterns.

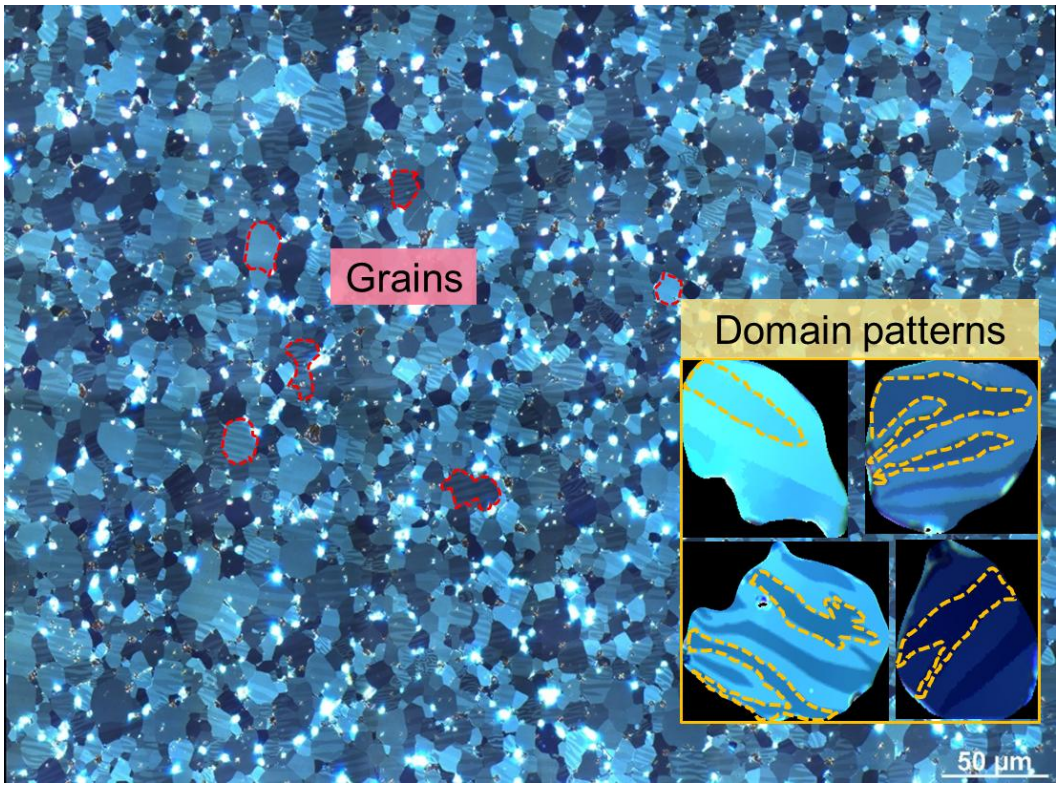


Figure 14: Kerr microscopy image of the FeNdB-A sintered permanent magnet showing grains in red and an example of the domain patterns present within the grains highlighted in yellow. The image has been acquired at 1000x magnification.

Figure 15 shows a KM image of a sintered FeNdB sample with different ECD and characteristics. The isotropic sample FeNdB-Iso has ECD 6 µm and the anisotropic sample FeNdB-E, which has been sectioned at 30 ° to the axis of anisotropy, has ECD 10 µm. The appearance of the domain patterns within the grain when observed under KM is indicative of the orientation of the grains in two dimensional space as shown in *Figure 15.*

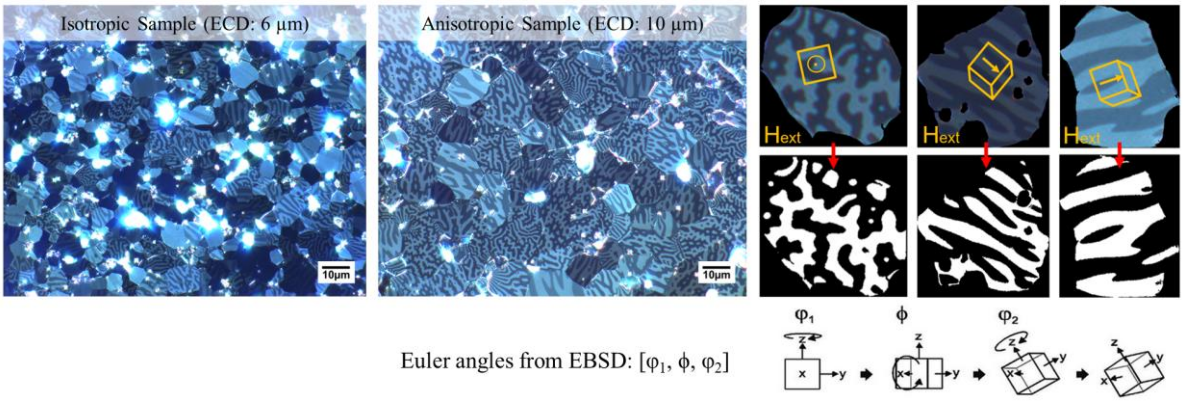


Figure 15: The KM image of sintered FeNdB magnet samples with different grain sizes and characteristic is shown. The crystallographic orientation of the grains from the domain structures in the KM image is visualized for reference. Using the EBSD approach, the grain orientation can be obtained as Euler angles [φ_1, ϕ, φ_2]. The isotropic sample shown is FeNdB-Iso, and an anisotropic sample is FeNdB-E, which has been sectioned at 30° to the axis of anisotropy [216]. Note that the cuboid representation is used here for better visualization of grain orientation. In general, it is tetragonal.

3.2. Dataset for the development of data-driven models

3.2.1. Dataset for intrinsic property prediction of 14:2:1 magnetic phases

The investigation of intrinsic property prediction is limited to the hard magnetic 14:2:1 phase found in sintered permanent magnets. This specific phase, characterized by its chemical composition ($TM_{14}RE_2B$), is crucial in generating exceptional magnetic properties closely linked to its microstructure. The samples comprise alloys spanning the ternary, quaternary, quinary, and senary classes, including various categories of RE substitutes. These categories have both light and heavy RE substitutes, and a mixed RE category combines both light and heavy RE elements.

The dataset used to predict the mass density in [g/cm^3], and intrinsic magnetic properties of the hard magnetic 14:2:1 phase has been compiled from various literature sources [42], [43], [44], [45], [46], [46], [47], [48], [124], [125], [130], [137], [138], [217], [218], [219], [220], [221], [222], [223], [224], [225], [225], [226], [227], [228], [229], [229], [230], [231], [232], [233], [234], [235], [236], [237], [237], [238], [239], [240], [241], [242], [243], [244], [245], [246], [247], [248], [249], [250], [251], [252], [253], [254], [255]. This includes intrinsic magnetic properties such as T_c in [K] and M_s in [μ_B/f.u.], along with the corresponding ρ_m, all specific to the 14:2:1 hard magnetic phase of various magnetic alloys. In materials science, datasets often tend to be relatively small compared to other fields due to the significant experimental costs of building comprehensive materials databases [256], [257], [258]. Furthermore, labels or ground truth in materials datasets can sometimes be biased due to the inclusion of outcomes from unsuccessful experiments [259]. Therefore, creating a generic predictive model applicable to various alloys presents a practical challenge. Nonetheless, ML can play a role in extrapolating within the existing materials database the specific candidate space of alloys within the solubility limits and feature space limitation that contains the number and type of elemental composition from which the model has been trained.

Given these considerations, the data collection effort has been confined explicitly to the 14:2:1 hard magnetic phase within the context of sintered permanent magnets, primarily of the FeNdB type. Additionally, it should be noted that intrinsic properties here are valid for Newtonian elemental space of chemical compositions. The experimental data regarding ρ_m, T_c, and M_s for various alloys is documented in the literature, often presented in tabular formats or graphical representations. These data points must be systematically extracted, subjected to data cleansing procedures, and organized to construct the dataset required for predictive modeling of the desired properties. A visual summary of the compiled information on material properties from the various literature sources is presented in *Figure 16*. The data collected from literature sources has been categorized into distinct alloy systems, with further divisions based on rare earth and intermetallic substitutes for various compounds. This classification process aids in structuring the dataset and segregating the information effectively for subsequent analysis and modeling.

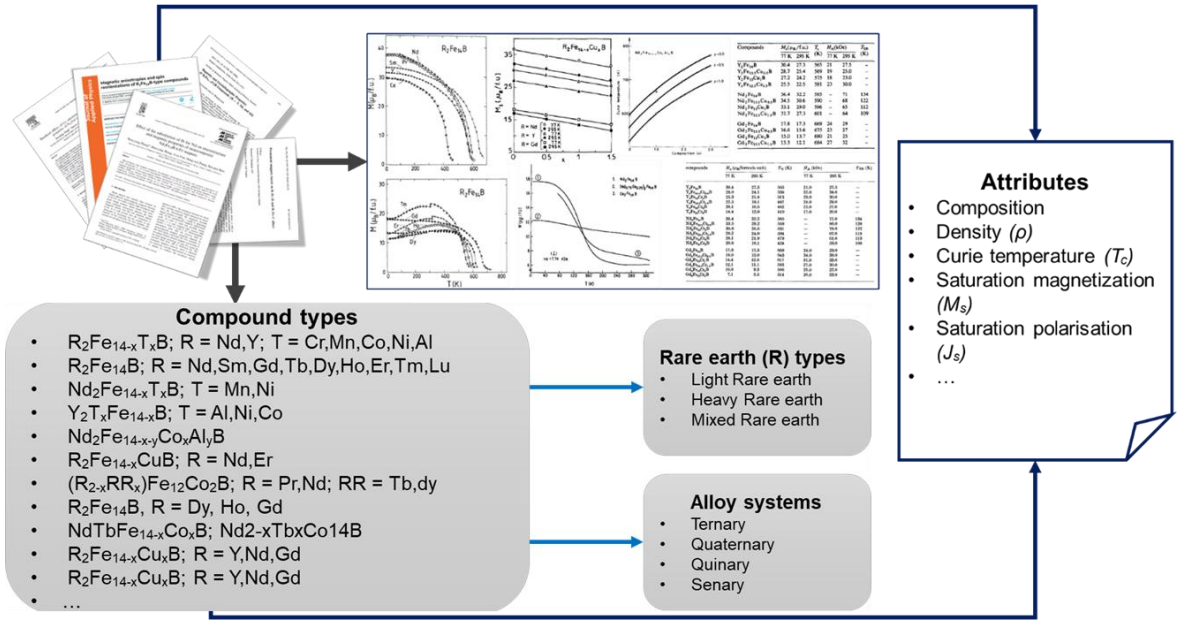


Figure 16: The magnet intrinsic property prediction model has been trained on the experimental data reported in different literature. The dataset includes 14:2:1 phases with a combination of rare earth types and other elements categorized into ternary, quaternary, quinary, and senary alloy systems. Attributes such as chemical composition and various intrinsic properties and density at ambient temperature have been included in the dataset.

The database of composition-based features of magnetic 14:2:1 phases and their corresponding target property measurements at room temperature was constructed and subsequently divided into sets for training, validation, and testing. The compositional features derived from the elemental chemical composition (CC) were extracted from the chemical formulas and appended to each specific magnetic phase. The atomic mass (AM) of each element in the composition was also extracted. In [98], [178], [185], this database of ρ_m, T_c and M_s for the experimental work on data-driven approaches for property prediction was used.

Additionally, a subset of alloys exists in the database for which magnetic properties have been reported differently across various literature sources. These alloys are classified as repeated phases. On the other hand, alloys whose magnetic properties are reported solely by a single literature source are categorized as unique phases. This disparity in reported values can be

attributed to varying measurement methodologies, equipment variation, measurement errors, and notable fluctuations in ambient temperature during experimentation (ranging from 293 K to 305 K). For instance, the reported M_s values in literature sources are obtained by applying saturation law or employing devices such as vibrating sample magnetometers or extraction magnetometers, often at distinct external magnetic field strengths, ranging between 1.0 T to 8.0 T [42], [43], [44], [45], [46], [47], [48], [124], [125], [130], [137], [138], [217], [218], [219], [220], [221], [222], [223], [224], [225], [226], [227], [228], [229], [230], [231], [232], [233], [234], [235], [236], [237], [238], [239], [240], [241], [242], [243], [244], [245], [246], [247], [248], [249], [250], [251], [252], [255], [247], [248], [260].

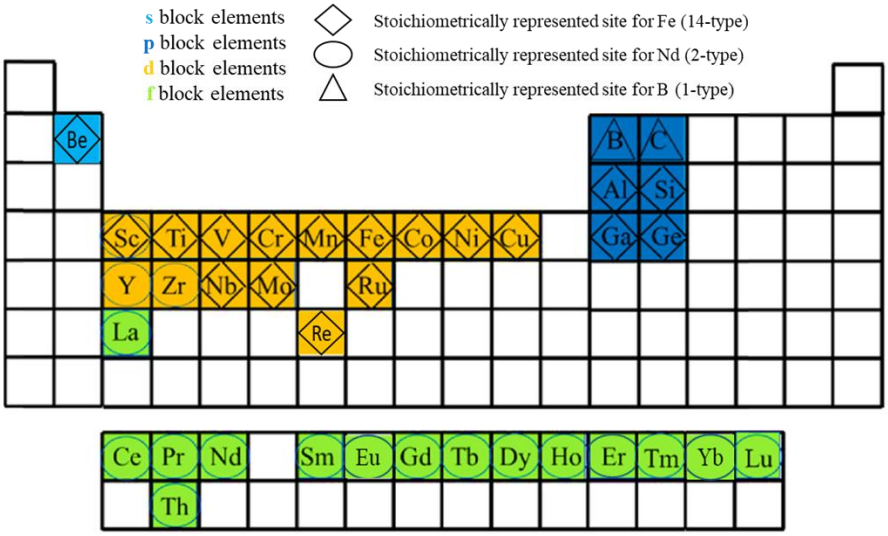


Figure 17: Periodic table showing the different elements in the intrinsic property prediction database and their elemental occupancy at three stoichiometrically defined sites for Fe, Nd and B in $Fe_{14}Nd_2B$ [98].

Figure 17 provides information on the element occupancy at one of the three stoichiometrically defined sites in the unit cell for phases in the intrinsic property prediction database. '14-type' for Fe, '2-type' for Nd, and '1-type' for B represented as diamond, circle, and triangle shapes respectively. The color represents the elements belonging to various blocks in the periodic table. The presence of unique and repeated phases within the dataset provides the model with realistic data, including the diversity of measurements obtained from various sources.

Table 7: Size of the dataset for the density, Curie temperature, and magnetic saturation prediction models. The dataset for each property has been classified into alloy systems. The fraction of unique and repeated phases for each property is listed here.

Property	Size of dataset	Unique phases	Repeated phases	Ternary alloys	Quaternary alloys	Quinary alloys	Senary alloys
Density	189	176	13	46	123	20	0
Curie temperature	449	270	179	283	120	44	2
Magnetic saturation	198	143	55	35	121	40	2

Table 7 gives an overview of the dataset with unique and repeated phases for each target property. Furthermore, the dataset is categorized on the basis of alloy systems, spanning from ternary to senary compositions. The frequency distribution of each target property for 14:2:1 phases from the literature is represented as a histogram in *Figure 18(a,c,e)*. *Figure 18(b,d,f)* shows the frequency distribution of elements in the dataset for model training.

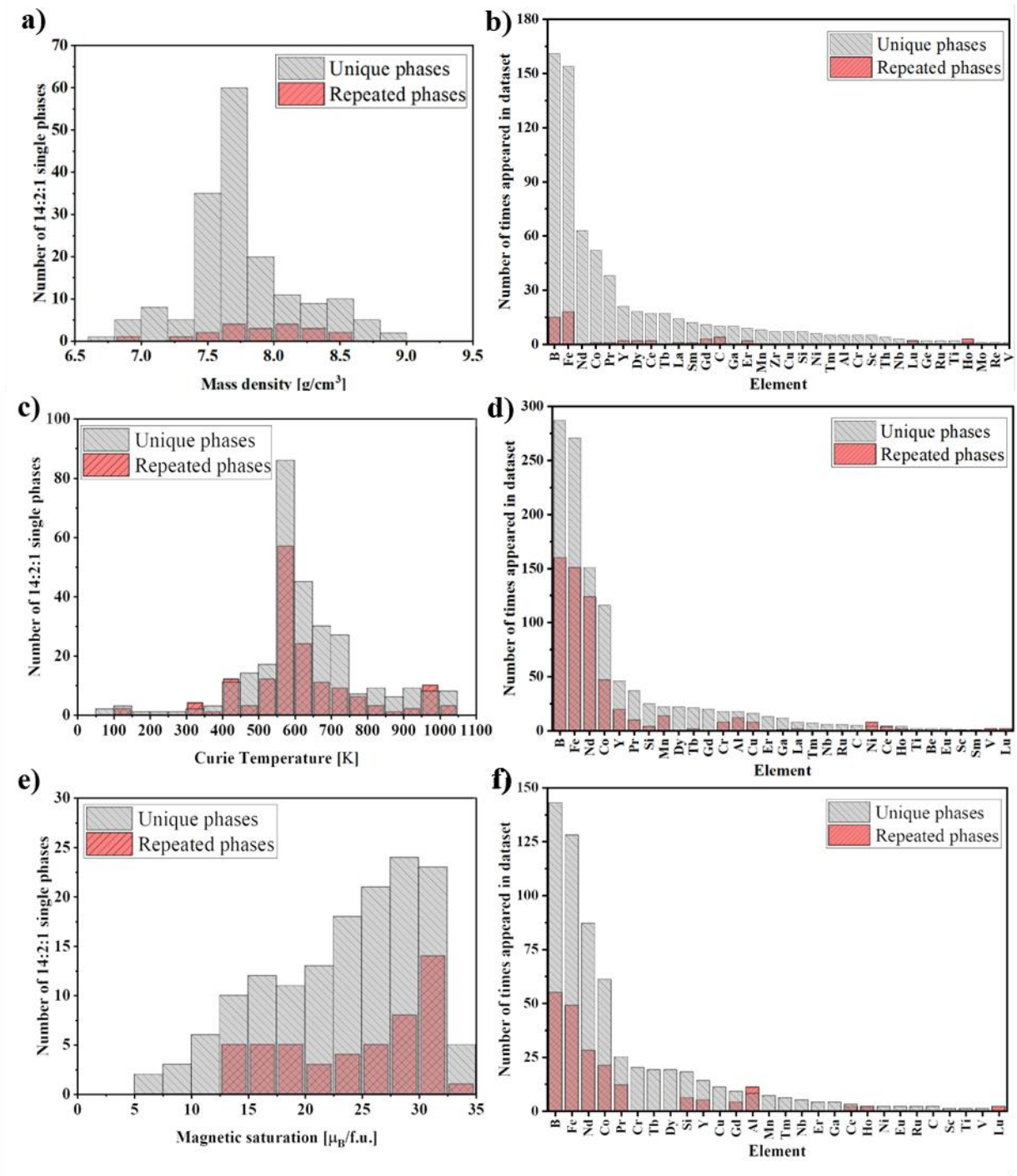


Figure 18: (a), (c) and (e) the frequency data distribution of the mass density, Curie temperature, and magnetic saturation. (b), (d) and (f) show the number of times chemical elements appear in the dataset for density, Curie temperature, and magnetic saturation.

The solubility of various substitution elements (M) in $Fe_{14-x}M_xNd_2B$ has been identified, with the 'x' value representing the maximum substitution level. This information is limited to the dataset used for the intrinsic property prediction task (T_c and M_s) within this thesis. Among the elements, Co exhibits the highest solubility, reaching x=14, while Mn and Ni also have significant solubility levels, with values of 'x' around 5.6 and 3 respectively. Elements such as Cr, Si, and Al show moderate solubility, with x values being 3, 2 and 3 respectively. Several elements, including scandium (Sc), titanium (Ti), vanadium, gallium, germanium (Ge), molybdenum (Mo), ruthenium (Ru), copper, and rhenium (Re) in the data, also indicate their minimum solubility limits. Specifically, Sc, Ti, Cu and V have solubility limits of approximately x=1.5, while Ga has a limit of x=1.8. Ge exhibits minimum solubility limits of x=0.5. For Mo and Re, the minimum solubility limits are approximately x=0.2. Ru exhibits a minimum solubility of 2.8.

Density of the 14:2:1 magnetic phase

For the mass density predictor model in [98], 33 chemical elements as the elemental compositional features were included. They are Fe, Co, Ga, Cu, Mn, Si, Ni, Zr, Al, Cr, Sc, Nb, Ge, Ti, Mo, Ru, V; RE metals Nd, Sm, Ho, Pr, Dy, Ho, Ce, Y, Tb, La, Gd, Er, Tm, Th, Lu; stabilizer B and C. The atomic mass of each element in the composition was also extracted for the density predictor model and the chemical elements. In *Figure 18(a & b),* the frequency distribution histogram of the mass density distribution and the distribution of each element across the dataset is shown. For model training and testing, a total of 123 data entries were allocated for the training, leaving 66 data aside for testing of the model's performance. It is worth highlighting that within the database, there were 123 unique phases considered for training, while the unseen test set includes a combination of 53 unique phases and 13 repeated phases. Interestingly, regardless of the specific combinations of features used, it has been observed that the Leave-One-Out Cross-Validation (LOOCV) approach [261] consistently outperformed the K-fold cross-validation approach (with K=5) in terms of achieving a more accurate and robust fit for the model. The LOOCV approach, which involves training the model on all but one data entry and testing the omitted data point, proved to be more effective in capturing the underlying relationships within the dataset, leading to improved prediction performance compared to the 5-fold method.

Curie temperature of the 14:2:1 magnetic phase

Similarly, for the elemental compositional features for the T_c predictor model in [178], 37 chemical elements were included. They are Fe, Co, Ga, Cu, Mn, Si, Ni, Zr, Al, Cr, Sc, Ho, Nb, Ge, Ti, Mo, Ru, V, Be, Re, Nd, Sm, Pr, Dy, Ce, Y, Tb, La, Gd, Er, Tm, Th, Lu, Eu, Yb, B, and C. In *Figure 18(c),* the frequency distribution histogram of the T_c distribution across the dataset shows that it is non-uniform and has a wide range from 62 K to 1100 K. *Figure 18(d)* presents a frequency distribution histogram plot, showing the distribution of each element found within the dataset. Notably, it has been observed that the reported T_c values for a specific composition can exhibit variations based on the source of the literature, generally falling within a range of 20 K. A notable exception is a single composition with a T_c value of 34 K.

Moreover, the dataset is split into training and validation sets, maintaining an 80:20 ratio. This splitting strategy serves the crucial purpose of evaluating the learning model's performance on new, unseen data before subjecting it to testing on an independent, unseen test set. As a result,

the dataset comprised 289 phases within the training set, 72 phases within the validation set, and 88 phases within the test set. Furthermore, among the 270 unique phases in the dataset, 209 unique phases were included in the training set, while the remaining 61 unique phases were used for the test set. This splitting process ensures the model's reliability and adaptability across various phases and compositions.

Magnetic saturation of 14:2:1 magnetic phase

For the elemental compositional features for the M_s predictor model in [185], 28 chemical elements were included. They are Fe, Co, Ga, Cu, Mn, Si, Ni, Al, Cr, Sc, Ho, Nb, Ti, Ru, V, Nd, Pr, Dy, Ce, Y, Tb, Gd, Er, Tm, Lu, Eu, B, and C. The range of M_s values within the dataset is shown in *Figure 18(e)*, which spans from 5 to 35 [μ_B/f.u.]. *Figure 18(f)* presents a frequency distribution histogram plot, showing the distribution of each element found within the dataset. The dataset of 198 magnetic 14:2:1 phases consists of 143 unique phases and 55 instances of repeated phases. Specifically, the training dataset has 103 unique phases and 43 repeated phases. This training data is further split using the K-fold (K=5) cross-validation technique, which facilitates the evaluation of the learning model's effectiveness. To ascertain the trained model's generalization ability, an unseen test set that includes 52 phases, of which 40 phases are unique, is used, thus providing a means to evaluate model performance under different scenarios.

Regardless of the target property under consideration, features 'CC' and AM are continuous variables. For instance, considering a $Fe_{14}Nd_2B$ phase, the feature vector (CC) has values of 14 for Fe, 2 for Nd, and 1 for B, with the remaining elements set to zero. Similarly, in the same $Fe_{14}Nd_2B$ phase, the AM features for Fe, Nd, and B are 781.9, 288.48, and 10.8, respectively, while the other elements are set to zero. This methodical approach comprehensively represents elemental contributions within each phase, facilitating accurate property prediction modeling.

3.2.2. Dataset for the microstructural characterization of sintered FeNdB magnets

Grain size and orientation information are obtained through EBSD measurements for FeNdB magnets. The KM image and the corresponding SEM-EBSD mapping are acquired using the correlative microscopy approach. The developed method for establishing a training dataset and quantifying grain orientation distribution is based on the resultant polar MOKE phenomena. This approach is termed Domain Pattern Analysis and Crystallographic Orientation Measurement (DoPACOM) [68].

Grain size analysis of the sintered FeNdB magnets

The detection of grain boundaries from KM images has been achieved using a supervised segmentation method [69]. The dataset includes KM images of five distinct non-commercially available FeNdB sintered permanent magnets. Some samples here were used for grain size detection in [69]. They are FeNdB-A, FeNdB-B, FeNdB-C, FeNdB-D and FeNdB-E, as in *Table 6*. The KM images were acquired at 1000x optical magnification obtained using a 100x objective and 10x magnification (lens) using an oil immersion objective, and the Kerr effect

was observed through a polarizer-analyzer pair set at an angle of 88° between their polarization planes.

In contrast, FeNdB-C and FeNdB-D are also anisotropic products of axial compaction, although their B_r and H_{ci} values are unknown because they were included in experiments aimed at evaluating the robustness of the learning models. KM images of the FeNdB-C and FeNdB-D samples were acquired using a Zeiss AxioImager.M2, at an exposure time of 10,000 ms under analyzer reflector light and the camera pixel resolution is 0.065 µm/pixel. Meanwhile, images of FeNdB-A and FeNdB-B samples were acquired using a Zeiss AxioImager.Z2 with an exposure time of 1800 ms under polarized reflector light and the camera pixel resolution is 0.045 µm/pixel. These samples consist of grains primarily aligned along the magneto-crystalline anisotropy axis of the grain. In addition, a high number of stripe domain structures within the grains is expected. To further expand the dataset, KM images of magnets that feature grains with closure domain structures were included. Additionally, KM images from 0° and 15° sections of the FeNdB-E samples, as depicted in *Figure 23*, were included to augment the dataset, given their higher abundance of grains exhibiting closure domain structures. The KM images of FeNdB-E were acquired using a Zeiss AxioImager.M2 microscope with an exposure time of 6 seconds acquired at 1000x optical magnification obtained using a 100x objective and 10x magnification (lens). The microscope was set-up for polar Kerr effect with 88° angle between the polarization planes using a Zeiss HXP 120 C as light source.

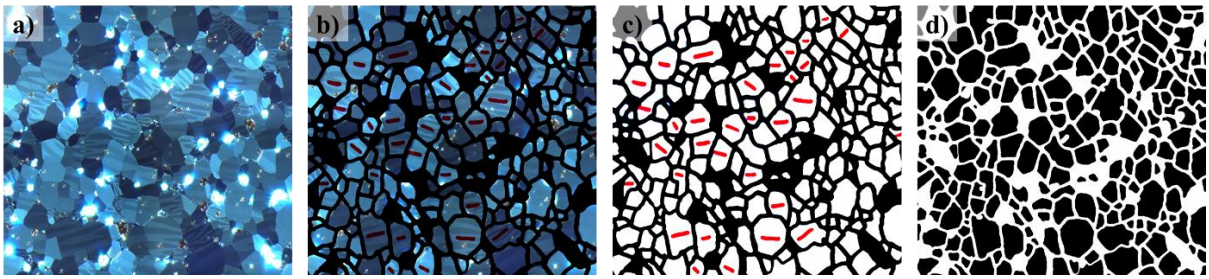


Figure 19: The workflow for detecting the grain boundaries in the FeNdB-A sintered permanent magnet sample from the Kerr microscopy images using the manual approach. (a) and (b) shows the hand-traced grain boundaries marked in black, and the grain orientation of grains with visible domain patterns are marked in red for the Kerr microscopy image. (c) and (d) are the post-processed images after multiple image processing steps involving threshold-based image segmentation and morphological operations to obtain a binary mask that can be used for grain size analysis.

The ground truth for this task initially is a binary mask showing the grain boundaries prepared manually by the three different subject experts that involves hand tracing the grain boundaries from the KM image of the magnet, as shown in *Figure 19*. Manually annotated images were processed through a series of image processing techniques to extract desirable grain boundaries, highlighted as black pixels. As shown in *Figure 20*, the experimental samples include large sample areas, which requires the division of acquired images into smaller tiles (1024 x 1024 pixels) to improve computational processing efforts during the image processing phase, which is essential for model training.

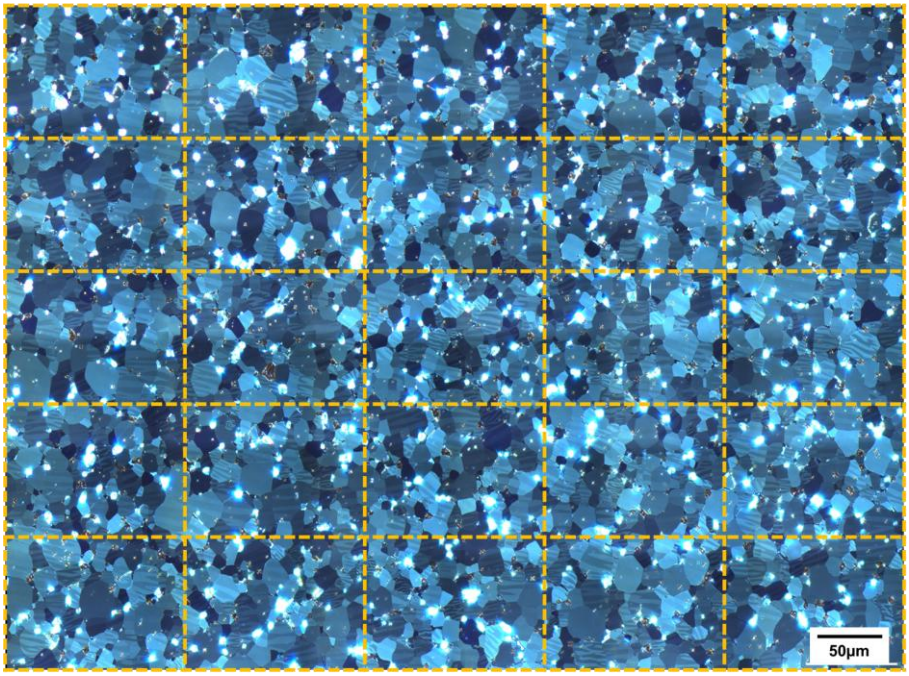


Figure 20: Kerr micrograph of FeNdB-A split into tiles based on the area of the sample. The area of the FeNdB-A sample is 457×338 µm^2. The tile images are subdivided into training, validation, and test datasets.

Table 8 provides an overview of the samples and the corresponding image count used for the experiment. The dataset was split into training, validation, and testing images. Furthermore, EBSD maps for the FeNdB-A and FeNdB-B samples were obtained using a Zeiss Sigma 300 VP, with an EDAX-AMETEK EDX detector, along with the EDAX Hikari OIM-Analysis software (v7.3.1), facilitating EBSD mappings. However, it is essential to note that the acquired EBSD maps do not correlate perfectly with the KM images. Post-processing steps applied to the EBSD maps often lead to the removal of small grains to refine the noisy data derived from the SEM. Thus, manually labeling data was used to establish a benchmark for evaluating the performance of the machine learning models instead of relying solely on EBSD as a reference or ground truth. The manual labeling task involved hand-tracing to annotate the grain boundaries by placing a transparent sheet over the KM images. The hand traced data are processed later to binarize them.

Table 8: Overview of the various FeNdB sintered permanent magnet samples used to train, validate, and test the pores and oxides and grain boundary detection model. The acquired Kerr micrograph of each sample is a large image with multiple tile images.

Sample ID	KM image (as tiles)	EBSD map	Hand labeled data	Training dataset	Validation dataset	Testing dataset
FeNdB-A	30	Yes	Yes	18	08	04
FeNdB-B	21	Yes	Yes	00	00	21
FeNdB-C	12	No	Yes	00	00	12
FeNdB-D	13	No	Yes	08	03	02
FeNdB-E	24	Yes	Yes	14	06	04

Area of each tile: FeNdB-A and FeNdB-B: 92×67 µm^2 and FeNdB-C and FeNdB-D: 78×58 µm^2.

Additionally, the pores and oxides are removed from the KM images as part of the pre-processing step using a supervised image segmentation approach. The presence of these in the magnet samples produces a scintillating effect in the KM image. They can be visualized using optical light microscopy, as shown in *Figure 21*. Due to the significant difference in the gray

values of the pores and oxides of the hard magnetic phase or grains in the optical light microscopic image of the magnet, using the Otsu thresholding approach a binary mask with regions of interest was obtained. This binary mask was used as a label to create a dataset for the pores and oxides removal model. The 40 image patches of 1024 x 1024 pixels from samples FeNdB-A and FeNdB-D were used to create the dataset for training, validation and testing.

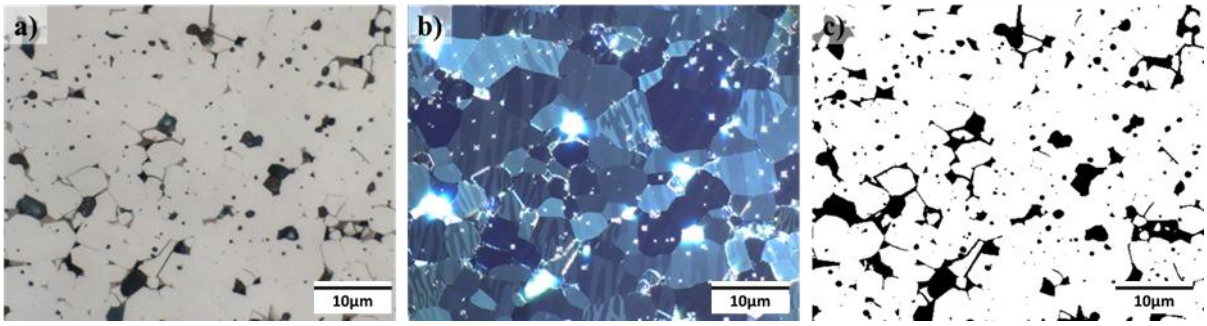


Figure 21: Correlative optical light microscopy (a) and Kerr microscopy (b) image of a FeNdB-E sintered permanent magnet sample. The pores, and oxides in KM image are obtained as a binary mask by applying a threshold-based image segmentation technique for the optical light microscopy image (c).

Crystallographic grain orientation analysis of the sintered FeNdB magnets:

The experimental dataset includes correlative Kerr and EBSD-SEM data from non-commercially available demagnetized FeNdB sintered permanent magnets. KM images of the FeNdB-E sample were acquired using a Zeiss AxioImager.M2 microscope with an exposure time of 6 seconds at 1000x optical magnification obtained using a 100x objective and 10x magnification (lens). The microscope was setup for the polar Kerr effect with an 88° angle between the polarization planes using a Zeiss HXP 120 C as a light source. The corresponding EBSD map was acquired using a Zeiss Sigma 300 VP scanning electron microscope, operated at 20 kV and equipped with a Hikari Super EBSD camera. An EDAX EBSD detector and the EDAX OIM-Analysis software facilitated EBSD mappings. The acquisition time for the EBSD mappings was about 2.5 hours, providing a mapping area of about 500 x 400 µm^2 at 0.7 µm step size and a magnification of 450x. Adding 30 minutes post-processing time gives 3 hours until the data are available. The acquisition time for the KM images in the experimental setup here was about 6 seconds per image. To cover the same area as in the EBSD mapping, 4 x 4 image tiles need to be acquired. The total acquisition time on the KM was 20 minutes for 16 image tiles, 27 minutes including the movement of the microscope xy-table and the adjustment of the optical focal plane for each tile. It is worth noting that the samples utilized for training and testing the trained model exhibit anisotropy, leading to grains in the samples being well-oriented in a single direction.

The grain orientation can be represented as a function of Euler angles (Bunge notation) as mentioned in Equation (43):

$$f(g) = f(\varphi_1, \phi, \varphi_2) \tag{43}$$

EBSD maps provide Euler angles, consisting of three continuous values (φ_1,ϕ,φ_2), depicting the crystal orientation of grains within the sample. These angles are then associated with the

corresponding grains in the KM images. As a result, the EBSD Euler angles serve as a reference (or ground truth) for training, validating, and testing the model. *Figure 22* shows the correlative microscopy setup for constructing the dataset used for the texture analysis of sintered FeNdB permanent magnets.

Figure 22: The overview of the correlative microscopy technique used for preparing the dataset for training and validation of the grain orientation prediction model. Images from Kerr microscopy and correlative EBSD data from the scanning electron microscopy are manually combined [68].

Table 9 shows the overview of the samples used for the experiments, their magnetic properties, and grain statistics, including the number of grains and the ECD values. The FeNdB-E(T) and FeNdB-E(V) samples are used for the training and validation of the model, respectively, and the FeNdB-AP, FeNdB-TP, FeNdB-IP, FeNdB-PSD-01, FeNdB-PSD-05, FeNdB-PSD-11, and FeNdB-PSD-15 samples are used for testing the performance of the trained model.

Table 9: Characteristics of the sintered FeNdB magnets used for the training, validation, and testing the grain orientation predicting models. The training and validation sets have the FeNdB-E sample, while the FeNdB-AP, FeNdB-TP, FeNdB-IP, FeNdB-PSD-01, FeNdB-PSD-05, FeNdB-PSD-11 and FeNdB-PSD-15 samples were used to evaluate the performance of the trained model. ECD [µm] Φ represent the equivalent circle diameter of the grains extracted from each sample. FeNdB-E(T) and FeNdB-E(V) represents training and validation sample sets.

Sample	Mean ECD Φ [µm]	B_r [T]	$(BH)_{max}$ [kJ/m^3]	H_{ci}[kA/m]	#Grains	Dataset
FeNdB-E(T)	10	1.11	235	2625	3610	Training set
FeNdB-E(V)	10	1.11	235	2625	880	Validation set
FeNdB-AP	10	1.34	340	1115	732	Test set
FeNdB-TP	10	1.41	385	1115	690	Test set
FeNdB-IP	10	1.44	400	1115	837	Test set
FeNdB-PSD-01	9	1.11	235	2625	792	Test set
FeNdB-PSD-05	9	1.11	235	2625	994	Test set
FeNdB-PSD-11	9	1.11	235	2625	900	Test set
FeNdB-PSD-15	9	1.11	235	2625	755	Test set

The FeNdB-E(T) and FeNdB-E(V) samples are manufactured through axial compaction, and they have an average grain size of 10 µm, measured as an ECD. These anisotropic samples have been carefully sectioned in 15° increments to generate a dataset with known grain orientations, as shown in *Figure 23*. In *Figure 23*, the structure of the 3D domain within a grain is illustrated. The appearance of the domain structure varies depending on whether it is sectioned parallel or perpendicular to the magnetic orientation direction of an anisotropic sintered magnet. From closure patterns to stripe patterns, variations in domain structure patterns are linked to changes in the sectioned angle from 0° to 90°. The pattern and contrast of domain structures are influenced by processing parameters and applied magnetic fields. Furthermore, the divergence in the KM images of the same sample, taken at different sectioned angles, is highlighted in *Figure 23*.

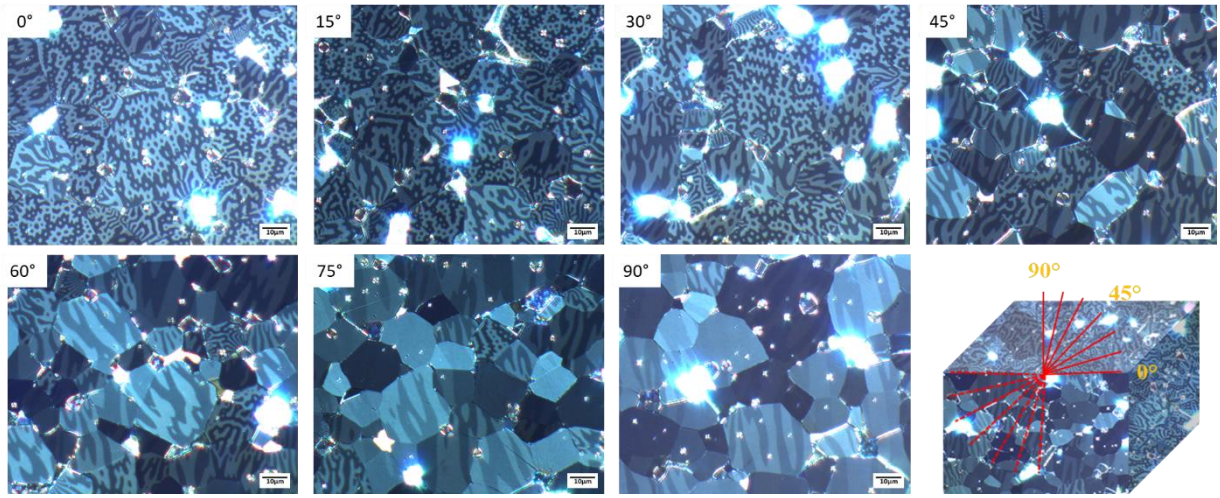


Figure 23: Displays a KM image of a cuboid sample of anisotropic sintered FeNdB-E magnet, sectioned at angles of 0°, 15°, 30°, 45°, 60°, 75° and 90°, observed at 1000x magnification. Notably, variations in the micromagnetic domain patterns are evident at different cutting angles. The sample's 3D simulation represents the dynamic changes in the domain structure along the grain, transitioning between closure and stripe patterns based on magnetic field direction and processing conditions or parameters.

Table 10 shows the number of grains from different sectioned FeNdB-E samples that were part of the training and validation set. This dataset is termed as dataset-1. The sectioned sample at 0° was not included in the dataset despite having a large number of grains with the closure domain structures due to limitations of the grain boundary detection (grain extraction) model. The grains from the sample section from 15° to 90° in steps of 15° were part of the dataset. The test samples FeNdB-AP, FeNdB-TP, and FeNdB-IP are sintered FeNdB magnets with different degrees of alignment, in which the orientation of the pressing direction to the direction of the applied external magnetic field was varied during pressing. The other test samples, FeNdB-PSD-01, FeNdB-PSD-05, FeNdB-PSD-11 and FeNdB-PSD-15 are sample with pole strength difference.

As the orientation φ_1 and ϕ changes, various image characteristics such as domain contrast, domain structure orientation, and grain entropy, which are extracted from KM, also transform. On the contrary, a rotation aligned with the φ_2 angle is not detectable through polar MOKE and does not impact magnetic properties. Furthermore, variations in the φ_2 value do not yield visible changes in the KM image of the magnet. Therefore, the training of the models is based on the reduced Euler angles φ_1 and ϕ.

Table 10: Overview of the number of grains extracted from the sintered FeNdB permanent magnet used to train the grain orientation model. The FeNdB-E sample has been sectioned in steps of 15°.

Sample code	Sample type	Sample section	Number of grains	Contribution in dataset [%]
FeNdB-E	Anisotropic	90°	1177	26
FeNdB-E	Anisotropic	75°	457	10
FeNdB-E	Anisotropic	60°	688	15
FeNdB-E	Anisotropic	45°	995	22
FeNdB-E	Anisotropic	30°	690	15
FeNdB-E	Anisotropic	15°	483	11

Figure 24 illustrates the distribution of 11 of the 13 Haralick parameters [91] calculated from each grain within the sample concerning the φ_1 orientation values. The Haralick parameters H1, H2, H4, H6, H7, and H10 refer to the angular second moment, contrast, sum of squares variance, sum average, sum variance, and difference variance, respectively. Furthermore, H5, H8, H9, H11, and H12 refer to the inverse difference moment, the sum entropy, the entropy, information measure of collection 1, and the information measure of collection 2, respectively. The features extracted from grains with φ_1 value ranging from [181°, 360°] exhibit the same distribution as those falling within φ_1 [0°, 180°]. Similar distribution patterns were observed when analyzing other image features, such as deep features extracted from the pre-trained VGG16 model [262] and color features [263], such as colorfulness and domain contrast. This symmetrical behavior between φ_1 [0°, 180°] and φ_1 [181°, 360°] suggests a consistent pattern. On the other hand, for the ϕ orientation, the values span [0°, 95°]. Consequently, a novel coordinate system was introduced in [68], resulting in reduced φ_1 and ϕ orientations, denoted as theta (θ) [0°, 180°] and rho (ρ) [0°, 90°] respectively, using Equation (44) and Equation (45).

$$\theta = \begin{Bmatrix} \varphi_1, & if\ \varphi_1 \geq 0° \text{ and } \varphi_1 \leq 180° \\ \varphi_1 - 180°, & if\ \varphi_1 > 180° \end{Bmatrix} \tag{44}$$

$$\rho = \begin{Bmatrix} \phi, & if\ \phi \geq 0° \text{ and } \phi \leq 90° \\ 90° - (\phi - 90°), & if\ \phi > 90° \end{Bmatrix} \tag{45}$$

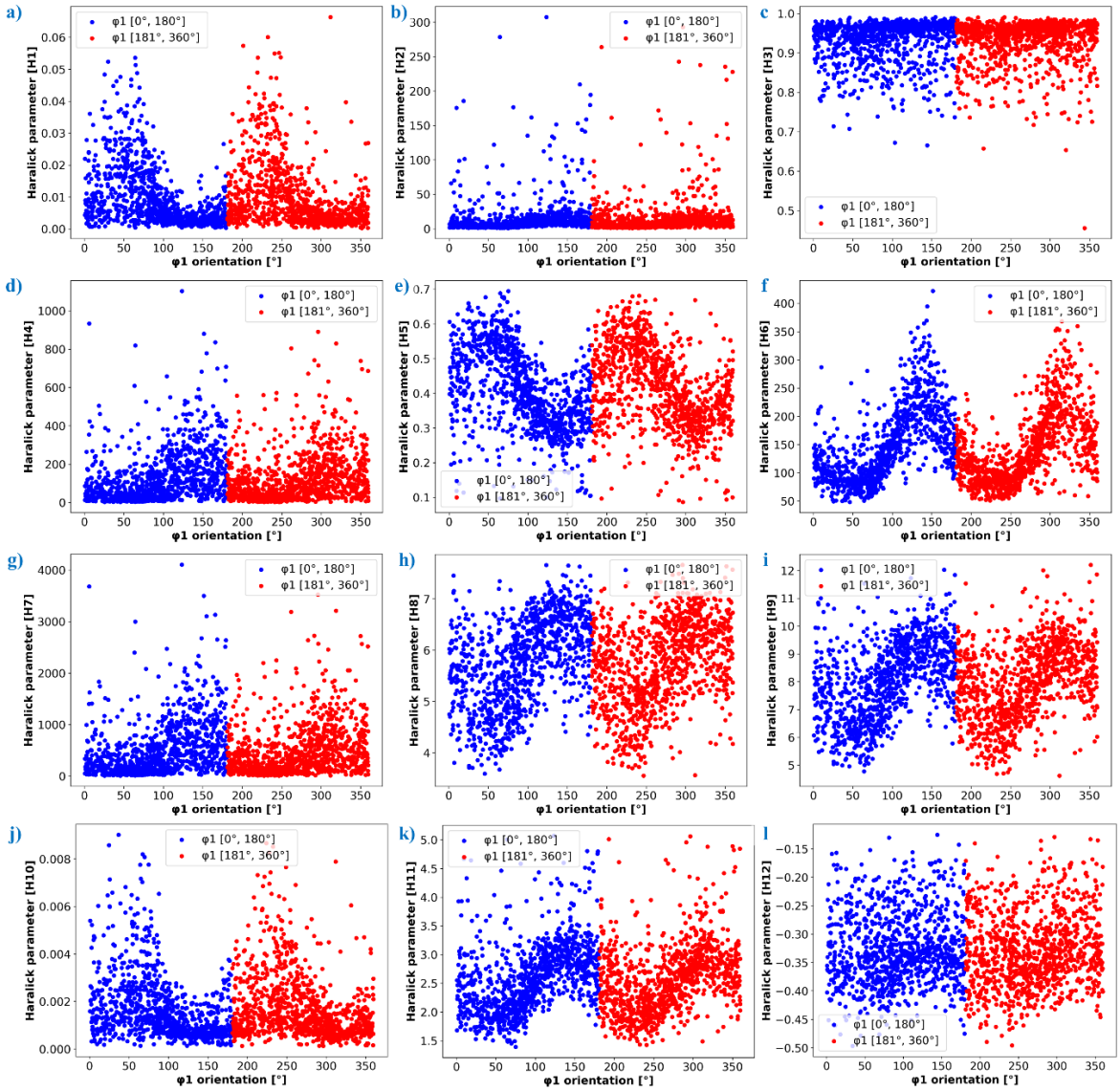


Figure 24: The distribution of texture features as the Haralick parameters (H1-H12) extracted from individual grain images extracted from the KM within the training sample. These features are compared against the φ_1 orientation derived from EBSD measurements. In these sub-figures, we observe the similarity in the features extracted from grains with φ_1 orientations falling within the ranges of [0°, 180°] and [181°, 360°]. The grains from KM do not distinguish the domain patterns with φ_1 orientations falling within the ranges of [0°, 180°] and [181°, 360°].

Figure 25 shows the coordinate system to quantify the reduced orientation from the KM image where the orientation of the grains around the x and y axes affects the orientation values θ and ρ, respectively.

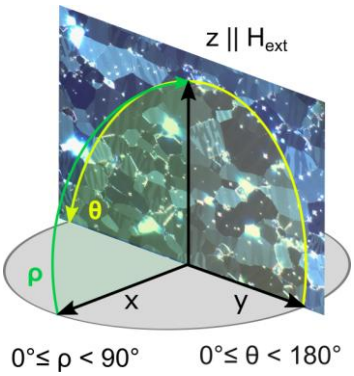


Figure 25: The DoPACOM coordinate system is utilized to quantify orientation from Kerr microscopy images. The θ value is influenced by the orientation around the x-axis, while the ρ value is impacted by its orientation around the z-axis [68].

Moreover, during the acquisition of both EBSD mappings and KM images for the training dataset, special attention was given to ensure accurate correlative positioning. This involved meticulous efforts to optimize the alignment of coordinate systems between the sample and the respective microscopes. Finally, the dataset included for training and evaluating the crystal orientation prediction model from the KM image incorporates the grains extracted from KM as input and the corresponding θ and ρ values as the ground truth. Of the 4490 grains extracted from the FeNdB-E sample, 3610 grains were allocated for training the learning models, while 880 grains were designated for validation. The grains of the AP, TP, and IP and PSD samples were then used to test the trained models.

Figure 26 illustrates the distribution of grain orientation values used for training the models. The dataset was split using a 5-fold cross-validation (k-Fold, k=5) approach. Notably, the dataset shows an imbalance since it is a non-uniform distribution, with the grains exhibiting θ values within the range of 40° to 130° and ρ values exceeding 30° are in more numbers when compared to other grains. In *Figure 26(a)*, it can be seen that the distribution is not uniform or normal with a wide range spread. Similarly in *Figure 26(b)*, it can seen that the distribution graph is skewed towards right and has grains with ρ value less than 30° lower in number suggesting the imbalance.

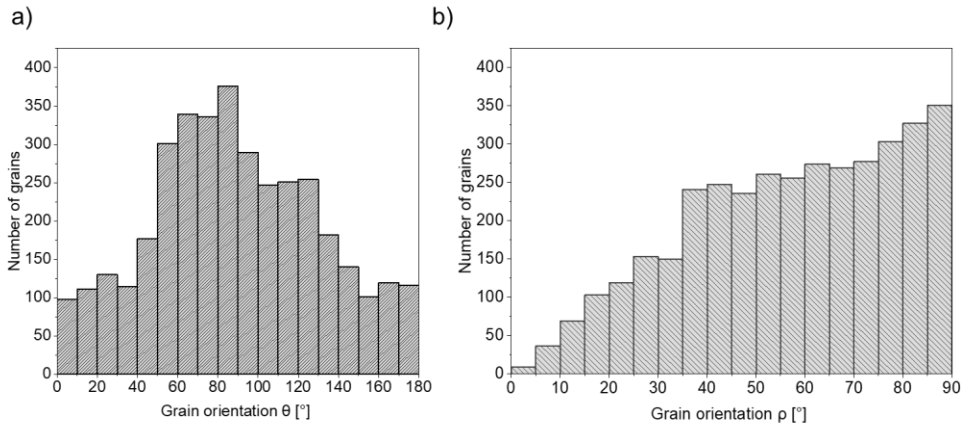


Figure 26: The frequency data distribution of the grain orientation used for training the ML and DL models. The (a) theta (θ) values and (b) rho (ρ) values. The choice of binning size is for visualization only, and it aligns with grain orientation tolerance values of 10° and 5° for θ and ρ, respectively [68].

Additionally, a dataset-2, which combines the grains from both anisotropic FeNdB-E and quasi-isotropic sample FeNdB-Q-Iso, was created to populate the grains that are less in number to reduce the effects of an imbalanced dataset. The quasi-isotropic sample used here is a transversal pressed magnet with a mean ECD of 14 µm. 1894 grains were extracted from the FeNdB-Q-Iso sample, and grains having θ values in the range of 40° to 130° and ρ values greater than 30° were not included in the existing dataset, which contained 4490 grains.

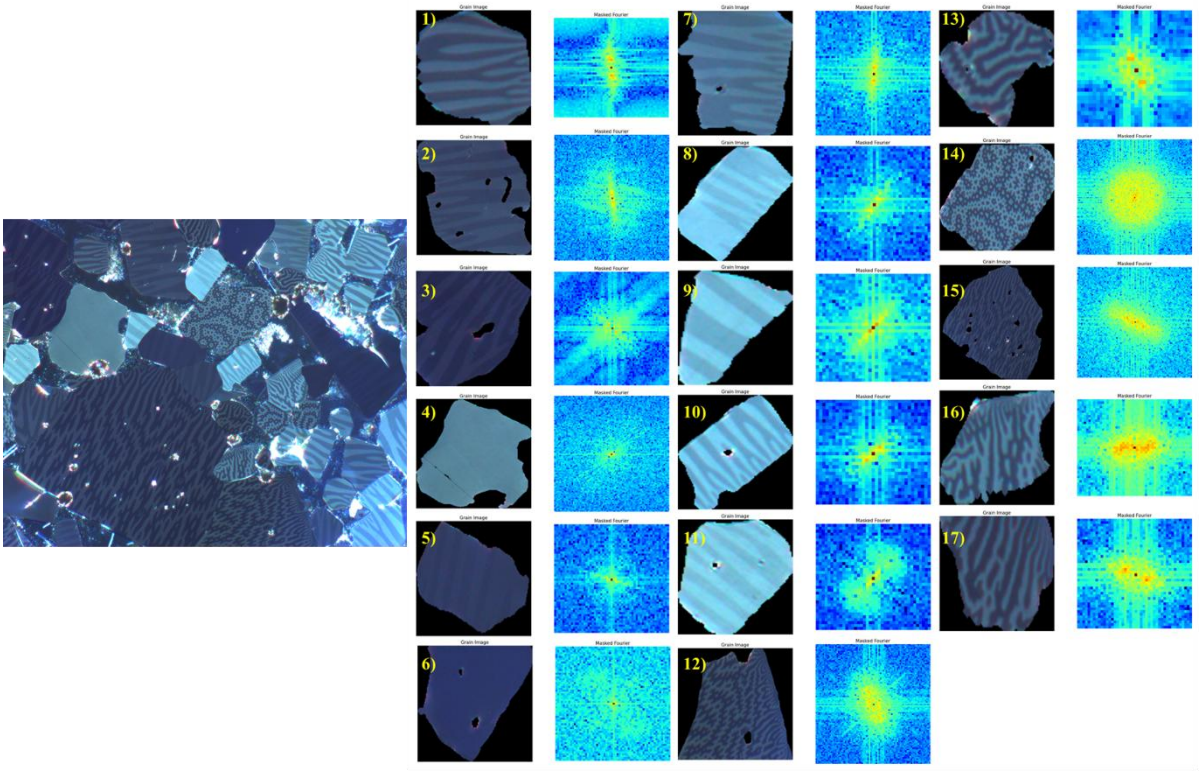


Figure 27: Overview of the grains extracted from the quasi isotropic sintered FeNdB magnet (FeNdB-Q-Iso) and corresponding discrete Fourier spectrum image obtained from the texture feature descriptor 2D fast Fourier transformation. The changes in the grain domain structure can be visualized from the discrete Fourier spectrum image.

Figure 27 shows an example image of the quasi-isotropic sintered FeNdB permanent magnet (FeNdB-Q-Iso) with grains having both stripe and closure domain structures and its corresponding Euler angles obtained from EBSD measurement. The texture feature descriptor 2D fast Fourier transformation (FFT) [145] is obtained for the grains extracted from the KM image corresponding to a discrete Fourier spectrum image. They provide information on how the domain pattern of the grains differs in relation to the measured Euler angle and the reduced Euler angles (θ, ρ).

Table 11 provides the grain orientation for the grains shown in *Figure 27*. The dataset with grains from an anisotropic sample is common for both θ values and ρ value prediction models. As the data imbalance is different for θ and ρ values, as shown in *Figure 26*, the dataset-2 that combines some grains of the quasi-isotropic sample is different for θ values and ρ values. Therefore, 267 grains having θ values less than 40° and 478 grains with θ values in the range of 130° to 180° were included in the existing dataset of θ values, and 189 grains with ρ values less than 30° were included in the existing dataset of ρ values.

Table 11: The grain orientation for the 17 images shown in Figure 27 as Euler angles (φ_1, ϕ, φ_2) obtained from the measured EBSD approach and the reduced Euler angles (θ, ρ).

Grain ID	φ_1 [°]	ϕ [°]	φ_2 [°]	θ [°]	ρ [°]
1	183	63	172	3	63
2	8	77	354	8	77
3	228	63	98	48	63
4	102	90	235	102	90
5	70	83	272	70	83
6	47	90	281	47	90
7	171	76	221	171	76
8	143	80	186	143	80
9	322	72	55	142	72
10	129	67	203	129	67
11	315	78	33	135	78
12	202	38	180	22	38
13	206	25	148	26	25
14	193	5	210	13	5
15	241	52	146	61	52
16	275	37	119	95	37
17	246	36	108	66	36

Table 12: Overview of the number of grains from the combined dataset of anisotropic samples FeNdB-E and isotropic sample FeNdB-Q-Iso that were used for the training and validation of the grain orientation prediction models.

Sample code	Sample type	Sample section	Number of grains	Contribution in dataset [%]
FeNdB-E	Anisotropic	90°	1177	22
FeNdB-E	Anisotropic	75°	457	08
FeNdB-E	Anisotropic	60°	688	13
FeNdB-E	Anisotropic	45°	995	18
FeNdB-E	Anisotropic	30°	690	13
FeNdB-E	Anisotropic	15°	483	09
FeNdB-Q-Iso	Quasi-isotropic	-	934	17

Nevertheless, the imbalance in the dataset still existed, which can be seen in *Table 12*, but it is comparatively lower than the dataset-1. Further, the dataset imbalance has been compensated using data augmentation on the grains that are less in number. Data augmentation techniques, such as adding noise and translational shifts, have been used to populate the undersampled regions in both dataset-1 and dataset-2. Initially, the grain images are subjected to dimension adjustments, introducing a 10% shift in both height and width. Additionally, these images undergo zooming and panning operations, effectively altering their scale within a range of 0.75 to 1. To enhance variability and mimic real-world imperfections, Gaussian noise is included, with intensities varying between 5% and 7%. The application of motion blur, achieved through a kernel size of 3, simulates the blurring effects caused by motion during image capture [264]. Moreover, the application of the developed workflow for grain size analysis and texture analysis was evaluated for feasibility in two use cases that included sintered FeNdB permanent magnet

with different degrees of alignments namely AP, TP and IP samples and samples with the pole strength difference exceeding 20% and a magnetic dipole moment misalignment angle α_Z ranging from 0° to 1° mentioned in the *Table 6* and discussed in section 2.1.6.

3.2.3. Ground truth and metrics for the evaluation of the model performance

Annotation of the dataset involves ground truth data for model training and evaluation in supervised learning approaches.

Ground truth for the intrinsic property predicting models for 14:2:1 magnetic phases

The experimental data values reported in the published literatures mentioned in section 2.2.2 are considered the ground truth for the property prediction models. The property values for different alloys of the magnet are reported in graphical or tabular format in the literature, as discussed in section 2.2.2. Information was collected in tabular format using the data-processing steps, followed by cleaning and sorting to remove the outliers and non-target properties. These steps were manually performed and evaluated by the subject expert. There are some phases for which the reported property values differ from one literature to another primarily due to differences in experimental setup and equipments. In the cases where literature reported values for repeated phases differ by large (> 15%), the reported value with the highest deviation from others is excluded from dataset and followed in an other research work. For T_c, ten phases and for M_s, five phases were identified as outliers as they had deviation more than 4% from mean value among repeated phases. The outliers are discussed in section 5.1. Moreover, the data collection here was manually carried out which resulted in manual errors such as incorrect placement of decimal points, presence of phases which are not 14:2:1 and property values at non-ambient temperatures. However, such errors were identified and corrected before being processed for model training and evaluation.

Ground truth for the grain size and crystallographic orientation predicting models

For the microstructure characterization of the sintered FeNdB magnet, obtaining accurate and unbiased size data and grain orientation is paramount.

For the grain size analysis task, the grain boundaries, pores, and oxides were manually traced by placing a transparent sheet over the KM images. The Fe_4NdB_4 (η) phase which exists in the form of small grains near the boundaries of the main phase ($Fe_{14}Nd_2B$) i.e, grains, are in a low volume fraction [265]. Moreover, this phase does not have domain contrast; therefore, it has been combined with main phase/grain for quantification and analysis. These tasks consume a lot of time and effort from the expert. The annotation results are highly subjective and induce fatigue errors if the goal is to create a large dataset. Therefore, the manually annotated grain boundaries from 3 different materials science experts and 2 scientific helpers under experts' guidance are cross-validated against the EBSD maps for consistency and compared statistically using the grain size values such as number of grains per μm^2. Additionally, the scientific assistants involved in the labeling task performed it digitally using the "Apple iPad tablet" device which was later validated by the subject experts. The thicknesses of the labeled mask were between 3-5 pixels and involved an Otsu auto-thresholding approach to detect and convert

the grain boundaries to 3 pixels. The entire process of the labeling task and validation for the samples in *Table 8* took approximately 70-80 hours. Note that the labeling task using the 'Apple iPad tablet" was 2x times faster than using hand-tracing on a transparent sheet. A variance of less than 3% was observed for the annotated grain boundaries of different experts and scientific helpers using number of grains per μm^2 as metric. The variance was calculated by quantifying the area fraction of the falsely annonated grain boundaries. For the grain size analysis task, an end-to-end grain boundary detection model has to be trained, which takes the KM image as input and outputs a binary mask with detected grain boundaries in it. Despite the use of correlative microscopy, aligning the grain boundaries of EBSD with those in KM images remains challenging. Post-processing steps applied to raw EBSD maps can alter grain boundaries and, in some cases, remove small grains and noise [266], further complicating the alignment process.

To streamline the annotation process, a semi-automated approach is introduced. Initially, a machine learning model was trained on a limited set of manually labeled KM images. Subsequently, the model's predicted grain boundaries were manually corrected by experts, saving considerable time and effort while ensuring consistency. Consequently, manually annotated masks were used to train the model for grain size analysis, improving efficiency and accuracy. Moreover, this saved time and effort required for labeling tasks, and consistency was evaluated against the EBSD maps. With this approach, the time spent for labeling was reduced by 3x times as the accuracy of the predicted grain boundaries on the initial set of labeled images was between 75% and 84%. The predicted images were manually corrected for missing and falsely detected grain boundaries only. Therefore, manually annotated masks are used to train the model for grain size analysis.

For the grain orientation task, the Euler angles obtained from the EBSD mapping is the ground truth. This task involves assigning each grain extracted from the KM image with an orientation value obtained through the correlative EBSD map.

Performance metrics and loss functions

Accuracy is the most common metric used to assess the performance of a learning system in classification tasks. Accuracy is determined by dividing the count of correctly predicted instances within the dataset by the total number of samples present. Nonetheless, in cases where the dataset exhibits a significant skew or is an imbalance in data distribution, the reliability of accuracy as an indicator of model performance can diminish. This is because the influence of the less prevalent but more significant instances is minimized compared to the dominant class [267]. For example, consider a dataset with a 90:10 class imbalance. Achieving a 90% classification accuracy could be as simple as predicting the majority class for all instances within the test set. This outcome is undesirable and can lead to misleading conclusions. Therefore, evaluating model performance requires using alternative metrics such as the F1-score and intersection over union. These metrics commonly address the issues from the imbalanced datasets for the classification tasks.

F1-score: The F1-score is a metric to assess prediction accuracy, and it combines precision and recall through a weighted average (harmonic mean). The value ranges from 0 (worst) to 1 (best). Notably, precision and recall have an equal influence over the F1-score's calculation. Precision represents the proportion of accurate positive predictions within the combined count of true and

false positive results (TP + FP) from the classifier. Meanwhile, recall represents the ratio of true positive results (TP) to the total instances that should have been classified as positive (TP + FN). It is important to recognize that the calculation of precision and recall excludes the use of true negatives, focusing solely on accurate predictions for each class. A reliable model should ideally yield elevated values for both precision and recall.

$$F1 - \text{score} = \frac{2\ x\ \text{precision}\ x\ \text{recall}}{\text{precision} + \text{recall}} \tag{45}$$

$$\text{precision} = \frac{TP}{TP + FP} \tag{46}$$

$$\text{recall} = \frac{TP}{TP + FN} \tag{47}$$

Intersection over Union (IoU): IoU is a performance metric that measures the effectiveness of segmentation tasks, particularly in datasets with sparse images characterized by an 80% - 90% background and a small proportion of positive labels or regions of interest [268]. Unlike simple metrics such as accuracy, IoU is not influenced by true negatives, rendering it more reliable. It quantifies the degree of overlap between the ground truth and the predicted images. The calculation for IoU can be observed in Equation (49). IoU scores range between 0 and 1, with a score of 1 denoting a precise match between the predicted and ground truth images. In contrast, a score of 0 signifies complete disagreement.

$$IoU = \frac{|Y \cap P|}{|Y \cup P|} \tag{49}$$

Binary cross-entropy (BCE): BCE is particularly suited for binary classification tasks, and its calculation equation is presented in Equation (50). In this equation, "Y" represents the ground truth image, "y" pertains to each pixel within that image, "P" denotes the predicted image, and "p" signifies each pixel in the predicted image produced by the trained model. Interpreting this, a loss score of 1 signifies a disparity between the predicted and ground truth images. In contrast, a score of 0 implies a precise alignment between the model's output and the ground truth.

$$L_{BCE} = -[y * \log(p) + (1 - y) * \log(1 - p)] \tag{50}$$

Weighted cross-entropy (WCE): WCE is a modification of the cross-entropy loss function in which positive instances are assigned different weights through coefficients, thus imposing greater penalties on the predictions of the minority class. This approach is commonly employed in cases where there are class imbalances. For example, if the ratio between the majority and minority classes is 90:10. The conventional cross-entropy, in this case, fails to rectify the imbalance and may skew the model's inclination toward the majority class.

WCE can be defined as follows in Equation (51):

$$WCE = -(\alpha y\ log(p) + (1 - y)\ log(1 - p)) \tag{51}$$

Where p is the prediction probability, and y is the ground truth label. α is the weight term for balancing the positive and negative examples. $\alpha > 1$ is used to decrease the number of false negatives, and $\alpha < 1$ is used to decrease the number of false positives.

Focal loss (FL): For imbalanced datasets, focal loss encourages the model to prioritize correcting misclassifications, and thus helps improve the model's performance on challenging examples while not being excessively affected by the abundance of easy examples. This makes it particularly useful for tasks with imbalanced datasets. It aims to alleviate this issue by down-weighting the contribution of well-classified examples (those with low loss values) and concentrating on the hard examples with which the model is struggling. It introduces a modulating factor called the "focusing parameter," which assigns higher weights to misclassified examples and decreases the weights for well-classified examples.

Thus, focal loss can be defined as follows in Equation (52):

$$FL = -(\alpha(1-p)^{\gamma} y \log(p) + (1-\alpha)p^{\gamma}(1-y)\log(1-p)) \tag{52}$$

α is a hyperparameter that balances the importance of positive and negative examples. Typically, it is set to a value between 0.5 and 0.99, with higher values giving more weight to the minority class (positive class). γ is another hyperparameter known as the focusing parameter. It controls the rate at which the loss decreases as the predicted probability p gets closer to the true label y. Higher values of γ make the loss decrease more slowly for confidently predicted examples. p is the predicted probability that a given example belongs to the positive class.

The key idea behind FL is that it significantly reduces the loss contribution of well-classified examples (where p is close to the true label) and gives more weight to misclassified examples with high confidence. This helps the model pay more attention to challenging examples and can improve performance, especially when dealing with imbalanced datasets or noisy labels.

Distance transform loss: For the semantic segmentation task such as edge detection, there is high class imbalance. It is important that pixels near the edges are identified with more weight and those far from the edges are given less weight [269]. The distance transformation maps measure the distance between edge and non-edge pixels. This map is obtained for both the ground truth edge map and the probability map from the model (after thresholding) to measure the regression loss called mean squared loss. This loss is combined with the classification loss (Dice loss) function of the segmentation network, as shown in Equation (55). This allows the model to learn effectively both the classification task and the distance transformation task [270], [271], [272].

$$\text{Distance transformation loss } (DTL) = \frac{1}{m}\sum_{i=1}^{m}(y_i - p_i)^2 \tag{53}$$

$$p_i = \eta \,.N(X_i) \tag{54}$$

$$\text{Combined loss} = L_{seg} + \lambda L_{DTL} \tag{55}$$

Where y_i and p_i are the pixel values at the i^{th} position in the true and predicted distance transformation maps, respectively. m denoted the total number of pixels in the predicted

probability map. X_i is the predicted probability map after applying thresholding. $N(X)$ is a sigmoid function. η is the distance field parameter to get the distance transformation map (ranges between 0-5), and λ is the assigned weight to the DTL loss.

Performance evaluation of regression tasks, such as grain orientation and property prediction tasks, requires metrics that can evaluate continuous target values. Among these, MAE, mean squared error (MSE), RMSE, MAPE and R^2 are extensively utilized measures for comparing the effectiveness of regression models. One significant difference between MAE and RMSE is their treatment of prediction errors. While MAE treats all errors equally, MSE and RMSE emphasizes more significant errors, which could be attributed to outliers or the model's struggle to generalize effectively across the entire dataset [273]. MAPE, on the other hand, expresses prediction errors as a percentage, making it useful for understanding relative errors especially when the target values vary widely. In evaluating the performance of regression models, both MAE and RMSE serve as important metrics to get insight into model accuracy, while MSE and MAPE are often used as a loss function. In cases where the dataset size is small, and to avoid the possibility of overfitting while improving generalization, the MSE loss function is regularized using the mean of squared weights (MSW) and it is termed as regularized MSE (MSE_{reg}) [256]. Furthermore, using the R^2 metric aids in understanding the relationship between predicted and actual values. Equation (56), Equation (57), Equation (58), Equation (59), Equation (60) and Equation (61) shows the function to calculate MAE, MSE, RMSE, MAPE, MSE_{reg} and R^2, respectively.

$$\mathrm{MAE} = \frac{1}{m}\sum_{j=1}^{m} \mid p_j - y_j \mid \tag{56}$$

$$\mathrm{MSE} = \frac{1}{m}\sum_{j=1}^{m} \left(p_j - y_j\right)^2 \tag{57}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{m}\sum_{j=1}^{m} \left(p_j - y_j\right)^2} \tag{58}$$

$$\mathrm{MAPE} = \frac{100}{m}\sum_{j=1}^{m} \left|\frac{p_j - y_j}{y_j}\right| \tag{59}$$

$$MSE_{reg} = \gamma \mathrm{MSE} + (1-\gamma)\frac{1}{n}\sum_{i=1}^{n} W_i^2 \tag{60}$$

$$R^2 = 1 - \frac{\sum_{j=1}^{m}\left(p_j - p_k\right)^2}{\sum_{j=1}^{m}\left(y_j - p_k\right)^2} \tag{61}$$

Here, m represents the number of samples in the dataset, y_j is the j^{th} target value, p_j is the predicted value for the j^{th} sample, and p_k is the mean target value of all samples in the dataset. W_i are model weights, n is the number of weights, and γ balances the trade-off between

prediction error and regularization, serving as the performance ratio used to optimize the weighting between MSE and MSW during training.

Concordance coefficient of correlation (R_{ccc}): R_{ccc} is a statistical measure for evaluating the level of correlation between two sets of ordered observations [274]. A value of 1 indicates strong concordance or agreement between the two sets, while -1 signifies strong disagreement. A value around 0 suggests little or no relationship between the rankings. Further, the concordance coefficient of correlation helps quantify the extent to which the observations in one set correspond to the order or ranking in another set, providing insights into their agreement or similarity. R_{ccc} is obtained using the Equation (62).

$$R_{ccc} = \frac{2 * \sigma_{12}}{(\mu_1 - \mu_2)^2 + \sigma_1^2 + \sigma_2^2} \tag{62}$$

Where (μ_1, σ_1) and (μ_2, σ_2) are the mean and standard deviation of actual and predicted values, σ_{12} is the covariance for the actual and predicted values.

3.3. Software and hardware specifications

All experiments were executed utilizing Python version 3.7+ and additional open source libraries. In the case of conventional machine learning, Scikit-learn version 0.22.1+ [275], while for deep learning purposes, Keras version 2.2.0+ [276], along with the Tensorflow [5] backend, was used. The deep learning model was trained on a system equipped with a 256 GB NVIDIA Quadro RTX 8000 graphics card.

3.4. Intrinsic magnetic property prediction model from chemical composition

The choice of regressors for each target property prediction use case is based on two main aspects that are performance and diversity. The screening of the algorithms was based on the performance of the different regressor models on the validation set, followed by pairwise correlation. A total of 40 different regressors were initially trained on the same dataset and validated using the K-fold (K=5) cross-validation strategy. Note that for these 40 regressors, the default hyperparameters from the Scikit-learn framework [275] were used. The metrics for the evaluation are MAE, RMSE, mean absolute percentage error (MAPE) and the R^2 score. The mean validation score for each metric for all 40 models is considered a baseline score. The regressors whose performance is below the baseline score are removed. A pairwise correlation of the rest of the models was obtained to reduce the redundancy and retain diverse models for better generalization. Models with correlation less than 0.4 were considered as the base learner for the VR model.

Density of the 14:2:1 magnetic phase

For the regressor model to predict the density in [98], a set of 3 algorithms were screened: SVR, RF, and Linear Regressor (LR). Additionally, a VR model with the base regressors SVR and RF with uniform weights was also designed. The optimal hyperparameters of the RF, SVM, and LR after grid search with a 5-fold cross-validation split are shown in *Table 13.*

Table 13: The optimized parameters of the different regressors used for the density predicting model. The optimized parameters are obtained after the grid search approach on a 5-fold cross-validation split.

Base regressor	Optimized parameters
RF	max_depth: 100; n_estimators: 80; criterion: mae; bootstrap: True
SVM	C: 1; gamma: scale kernel: rbf; degree: 3
LR	fit_intercept: true

Curie temperature of the 14:2:1 magnetic phase

For the T_c prediction in [178], LR, SVM, RF, extra tree regressor (ET), GB, XGB, adaptive boosting with RF as the base model (AB-RF), VR, and artificial neural network based MLP were evaluated. Based on the low error on the validation set, ET, XGB, RF, and AB-RF algorithms were shortlisted. These were used for constructing a VR model. Uniform weights were assigned to all the base regressors as they are tree-based models and homogeneous (seen in Equation (63)).

$$\text{Prediction } (T_c) = \frac{F_{ET}(f) * 1 + F_{XGB}(f) * 1 + F_{RF}(f) * 1 + F_{AB-RF}(f) * 1}{4} \quad (63)$$

Where $F_{ET}(f)$, $F_{XGB}(f)$, $F_{RF}(f)$, and $F_{AB\text{-}RF}(f)$ are base regressors ET, XGB, RF, and AB-RF, respectively, and f is the input feature vector corresponding to the alloy's chemical composition.

Table 14: The optimized parameters of the different regressors used for the Curie temperature prediction model. The optimized parameters are obtained after the grid search approach on a 5-fold cross-validation split.

Base regressor	Optimized parameters
RF	max_depth: 40; n_estimators: 30; criterion: mae; bootstrap: True
SVM	C: 1; gamma: scale kernel: rbf; degree: 3
LR	fit_intercept: true
ET	min_sample_leaf: 1; min_sample_split: 2; n_estimators: 10; criterion: mae
GB	n_estimator: 80; max_depth: 20; min_samples_leaf: 2; min_samples_split: 8
XGB	objective: reg:squarederror; learning_rate: 0.05; max_depth: 20; alpha: 0.5; n_estimators: 200
MLP	activation: relu; hidden_layer_sizes: (100,); max_iter: 500
AB-RF	learning_rate: 0.001; n_estimators: 50; estimator: RandomForestRegressor(max_depth: 50, n_estimators: 50)

The optimal hyperparameters of the RF, SVM, LR, ET, GB, XGB, MLP, and AB-RF after grid search with a 5-fold cross-validation split are shown in *Table 14*.

Saturation magnetization of the 14:2:1 magnetic phase

Different machine learning regression models were tested in [27] using the supervised learning approach. These included the linear Huber regressor (HR) [277], tree-based techniques like RF, GB, XGB, and adaptive boosting with decision tree as a base learner (AB-DT), and MLP that functions based on artificial neural networks. Additionally, a weighted VR was trained and evaluated for its performance on the validation set. Upon analyzing each model's performance on the validation set, HR, RF, MLP, and AB-DT were identified as potential base regressors for constructing a VR model. In this proposed approach, the determination of optimal weights for these four selected base regressors uses performance-based and rank-based approaches with metrics, specifically, the R^2 score and MAE measured on the validation set. The performance values for each base learner are normalized between minimum and maximum value and then average of it is taken for grid search. The grid search technique was adopted to obtain the optimized weight for each base regressor. Consequently, the base regressors HR, RF, MLP, and AB-DT were assigned weights of 0.89, 0.94, 0.95, and 0.96 (as seen in Equation (64)).

$$\text{Prediction } (M_s) = \frac{F_{HR}(f) * 0.89 + F_{RF}(f) * 0.94 + F_{MLP}(f) * 0.95 + F_{AB-DT}(f) * 0.96}{(0.89 + 0.94 + 0.95 + 0.96)} \quad (64)$$

Where $F_{HR}(f)$, $F_{RF}(f)$, $F_{MLP}(f)$, and $F_{AB\text{-}DT}(f)$ are base regressors of HR, RF, MLP, and AB-DT, respectively, f is the input feature vector corresponding to the chemical composition of the alloy. The optimal hyperparameters of the RF, HR, GB, XGB, MLP, and AB-DT after grid search with a 5-fold cross-validation split is shown in*Table 15*.

Table 15: The optimized parameters of the different regressors used for the magnetic saturation prediction model. The optimized parameters are obtained after the grid search approach on a 5-fold cross-validation split.

Base regressor	Optimized parameters
RF	max_depth: 100; n_estimators: 100; criterion: mae; bootstrap: True
HR	epsilon: 1.35; max_iter: 500; alpha: 0.0001
GB	n_estimator: 50; max_dpeth: 20; min_samples_leaf: 2; min_samples_split: 8
XGB	objective: reg:squarederror; learning_rate: 0.05; max_depth: 100; alpha: 0.8; n_estimators: 100
MLP	activation: relu; hidden_layer_sizes: (50, 50); max_iter: 600; solver: lbfgs
AB-DT	learning_rate: 0.01; n_estimators: 50; estimator: DecisionTreeRegressor(max_depth:10)

3.5. Grain boundary detection model for sintered FeNdB magnets

The KM image of the sintered FeNdB permanent magnet reveals details about the grains and domain patterns within each grain. However, during manufacturing, the sample can develop pores and oxides, causing them to appear as white scintillating in the KM images. To ensure accurate microstructural characterization, it is essential to eliminate these oxides and pores from the KM image as a pre-processing step. *Figure 28* shows the workflow for labeling and training the artefact removal model with KM images as input. As shown in *Figure 28*, the KM dataset and binary mask with regions of interest (ROI) are considered for training an artefact removal model (oxides and pores). The training and validation set has 40 and 17 images, respectively, as discussed in section 3.2.1. The test set contains 43 images from different samples. For the pores and oxides detection task, an end-to-end segmentation network such as the U-net architecture was chosen as they have been proven to perform well for the pixel-wise segmentation task and cover well on smaller datasets compared to the other CNN architectures [120]. The architecture of the model here is the U-net which is described in section 2.2.2. The architecture is synonymous with the encoder-decoder; feature extraction is carried out in the encoder or contracting path, and the decoder or expanding path does the localization to retain the spatial information despite down-sampling and max-pooling performed during the contracting path.

Usually, a DL architecture requires a large dataset so that feature extraction and distillation can be carried out effectively to develop the relationship between input and expected output. Over the decade, many deep learning models have been trained on large datasets such as ImageNet, COCO, etc. [278]. During the learning process, such models can extract local and global features from the image data. Using the transfer learning approach, such pre-trained models can achieve good accuracy on relatively small datasets by freezing the feature extraction layers of the pre-trained networks, followed by adding customized classification or regression layers and fine-tuning them. The encoder layer in U-net serves as the feature extraction component and can be substituted with pre-trained networks such as VGG16, ResNet, EfficientNets, InceptionV3, etc.

Figure 28 shows the labeling of pores and oxides in the KM images using the correlative LM images and training the U-net model to identify them directly from the LM images. For this task, the pre-trained EfficientNetB3 network was considered as an encoder for feature extraction. As a pre-processing step to suit the pre-trained network, the input images were rescaled between [0, 1] followed by normalizing with the mean [0.485, 0.456, 0.406] and the standard deviation [0.229, 0.224, 0.225] obtained from the ImageNet dataset. The batch size is set to 8 with Adam optimizer, and the initial learning rate is $1x10^{-4}$. Using the Reduce on Plateau as a learning rate scheduler [276], the learning rate is dynamically decreased by a factor of 0.1 if there is no improvement in the validation error for 10 epochs. The model is trained for 100 epochs, and dice loss combined with binary focal loss were loss functions. The performance metrics used are the IoU and the F1-score.

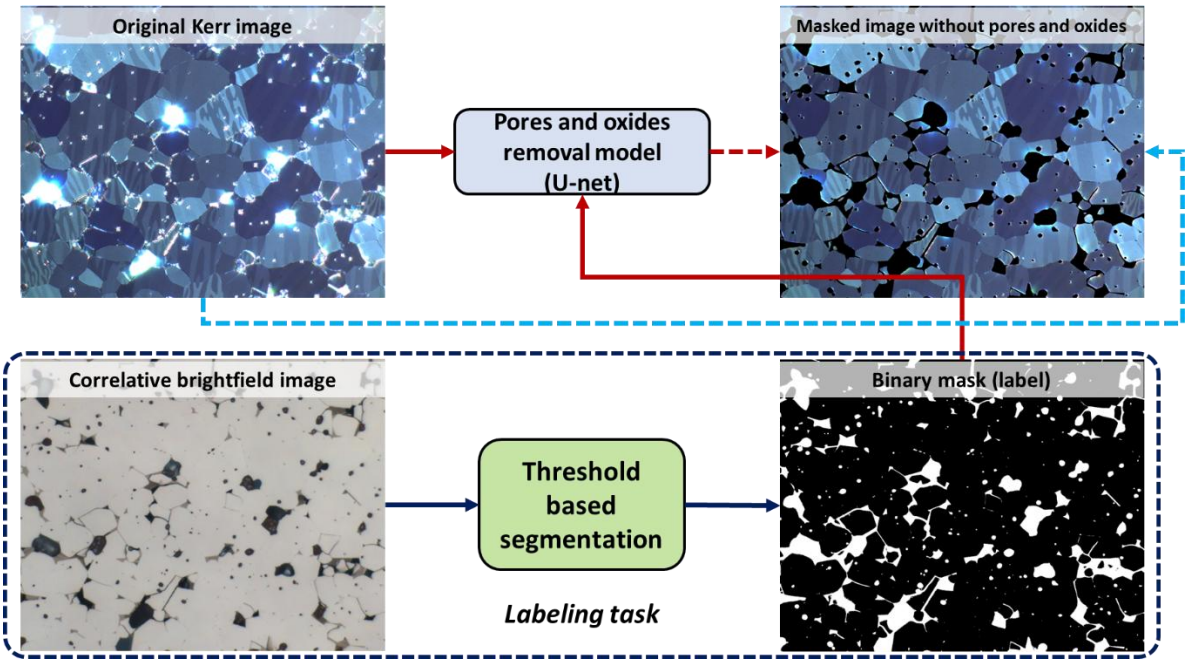


Figure 28: Workflow of the artefact removal model from the Kerr microscopy images of the FeNdB-E magnet. The binary mask used as labels for the learning model is obtained from the correlative approach that involves Kerr and brightfield images. The pores and oxides in the brightfield image is detected using the threshold-based image segmentation approach and using the labels.

Additionally, data augmentation techniques, both at the pixel and spatial levels were carried out, to populate the dataset and enhance the robustness of the model. These techniques involve making minor adjustments to the dimensions of the images, introducing a 5% shift in height and width to vary their aspect ratios. Additionally, application of zooming and panning operations within limits of 0.75 to 1, effectively altering the scale of the images. To simulate real-world imperfections, Gaussian noise with intensity levels ranging from 3% to 5% was introduced. Furthermore, Gaussian and motion blur effects were incorporated using a kernel size of 3, replicating the blurring that can occur during image acquisition. To diversify the dataset further, cropping of random parts of the input images, random pixel drops replaced with a pixel value of [0, 0, 0], and perspective transformations are included [264]. Collectively, these augmentation techniques contribute to a more adaptable and robust dataset for training the model.

The pores and oxides free KM images are used for the grain boundary detection task. Grain boundary detection from KM images of magnets involves edge detection tasks such that grains with varying domain patterns are separated. Edge detection is a binary segmentation approach and involves high class imbalance. The experimental steps involve developing the machine learning model using hand-crafted features and an advanced deep learning approach. *Figure 29* outlines the procedure for the supervised machine learning approach to magnet characterization, that requires the expected output as labels during training. In traditional machine learning, an approach involving a feature vector is utilized. This vector forms a matrix comprising extracted image features from the training dataset, merged with their corresponding labels. Feature extraction is important to help the learning model throughout its learning and generalization stages during training. On the other hand, the deep learning model has an architecture designed to automatically extract the features and filter the redundant ones during the learning process.

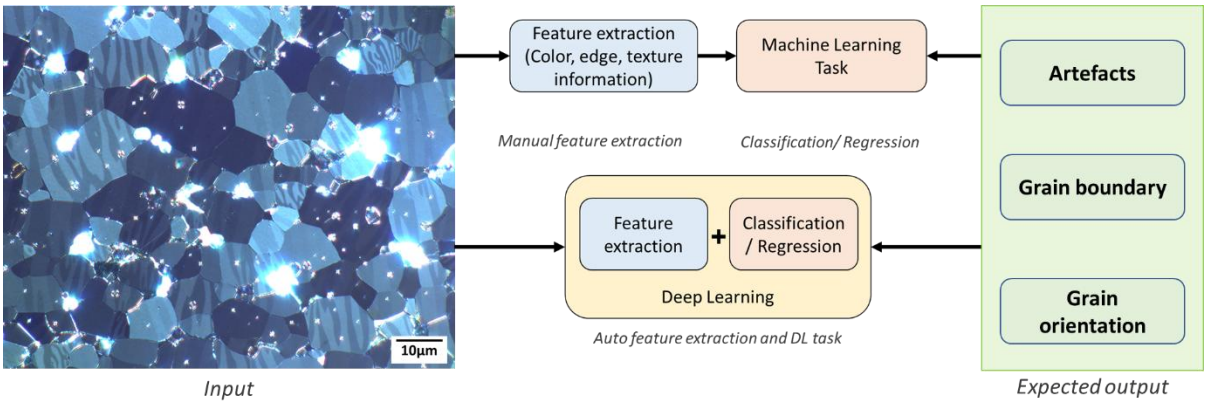


Figure 29: The difference in traditional feature-based machine learning and the deep learning approach involving the auto feature extraction for the pores and oxides, grain boundary detection, and grain orientation prediction from the Kerr microscopy image of the FeNdB-E magnet as the input data.

3.5.1. Feature based machine learning approach for grain boundary detection

The process of extracting the image features is strongly influenced by the expected output from the model and the appearance of the sample's microstructure or different regions of interest in the image. The features derived from the input data substantially impact the model's performance. As a result, one should be well-versed in various image processing methods to extract an appropriate feature vector containing relevant information. This ensures that the model can effectively build the relationship between the input and the desired output. During the feature extraction step, a 39-dimensional feature vector is generated, encompassing a variety of image information. These comprise six color-related features spanning RGB and HSV channels, 15 texture-based characteristics derived from operations such as median filtering and structure tensor calculation, and 18 edge-oriented features obtained by bilateral filtering, Gaussian difference, and Sobel filtering. This diverse set of features is computed across various degrees of Gaussian blur, specifically at sigma levels 1, 2, and 4. The sigma values are multiples of 2, and only 3 are considered here based on trial and error. The workflow unfolds with the creation of 39 feature maps, each corresponding to a distinct feature extraction method.

Consequently, for every pixel in the input image, 39 corresponding feature values emerge. These values are then aligned with the appropriate binary mask. For the training of the model, two distinct sets of features are utilized. First feature set includes only three color (C) features from RGB channels, and the second feature set includes all 39 features extracted (C,E,T). Tree-based machine learning models, such as RF and XGB, are considered to train the model with first and second feature sets.

RF and XGB classifiers are ensemble tree-based learning algorithms that decide the input class based on votes from multiple decision trees, as discussed in section 2.2.1. The hyperparameters of the optimized RF classifier for both feature sets used here have 100 trees, a Gini function to measure the quality of the split, a maximum depth of 100 for the tree, and out-of-bag samples to generalize the accuracy. The hyperparameters of the optimized XGB classifier on the (C,E,T) features here have a learning rate set to $1x10^{-2}$, a maximum depth of 80 for the tree, an objective is set to logistic, and area under the curve metric is for measuring the performance. The training

set includes 40 images, the validation set has 17 images, and the test set includes 43 images of size 1024 x 1024 pixels. Note that the input images here are KM images without pores and oxides. Such pre-processed KM image is obtained using the image masking approach using bitwise operation (AND). The output image from the trained RF and XGB classifier needs post-processing to remove the noise from the predicted mask and improve the precision of the grain boundary detection.

3.5.2. Deep learning approach for grain boundary detection

For permanent magnet samples, the visual representation of the grains in KM images emphasizes the subtle color distinctions between individual grains, as shown in *Figure 14*. As a result, a feature vector with higher dimensions becomes necessary to accurately classify each grain. Furthermore, extracting features from such complex microstructures proves to be a labor-intensive task. Creating labels for KM images, specifically grain boundaries, requires a significant investment of time, as domain experts must perform this task manually. Given these challenges, the adoption of advanced ML methodologies becomes important. These techniques should be able to effectively handle limited datasets and avoid the need for tedious manual feature extraction procedures [16].

The training, testing, and validation datasets are consistent across RF(C), RF(C,E,T), XGB, and U-net. The model architecture is detailed in Section 2.2.2. The U-net architecture with various pre-trained network encoders, including VGG16, ResNet50, EfficientNetB3, and InceptionV3, aiming to improve performance on the validation set, has been assessed. The input data is normalized between [0,1] or [-1,1] depending on the U-net encoder. To normalize the image between [-1,1], the image is rescaled to [0,2] and subtracted by -1. Similarly, to normalize between [0,1], the image is rescaled to [0,1] followed by subtracting by mean [0.485, 0.456, 0.406] and division by standard deviation [0.229, 0.224, 0.225] of the ImageNet dataset.

Additionally, a customized U-net without pre-trained network was also trained on the same training set. In the encoder path, the U-net architecture includes a series of Conv2D layers with 3x3 kernels. These 3x3 kernels are fundamental for capturing local image features efficiently. Each Conv2D layer has a stride of 1 to detect fine grained details and preserve spatial information, which moves one pixel at a time during the convolution operation. This allows for detailed feature extraction without significant downsampling. Following each convolution, a ReLU activation function is applied. ReLU introduces non-linearity, making the model capable of learning complex relationships between input features. To reduce the spatial dimensions of the feature maps and aggregate high-level features, the contracting path employs MaxPooling2D layers with a pool size of 2x2. A stride of 2 is applied, leading to a halving of the feature map size in both the horizontal and vertical directions. This down-sampling operation reduces the computational load and focuses on retaining the most salient features. The bottleneck layer is unique because it applies a dropout operation with a rate of 0.5. The 0.5 dropout rate means that, during training, 50% of the neurons in the layer are randomly set to zero during each forward and backward pass. Although U-Net primarily consists of convolutional layers rather than fully connected neurons, dropout is still applicable by randomly disabling elements within the convolutional feature maps. This dropout operation acts as a form of regularization, helping prevent overfitting by reducing the reliance on any single neuron,

thus improving model generalization. No additional convolution or pooling is performed in this layer. The decoder path is the reverse of the contracting path, where feature maps are up-sampled and concatenated. The UpSampling2D layers are used for upsampling, which doubles the spatial dimensions without introducing learnable parameters. The subsequent Conv2D layers maintain a 3x3 kernel size, 1 stride, and ReLU activation, which refine the segmented features.

In the decoder path, after each UpSampling2D operation, the feature maps are concatenated with the feature maps from the encoder path. This concatenation operation is achieved through the concatenate function and is called skip connections. The feature maps from the encoder path capture low-level features and fine details, while the up-sampled feature maps in the decoder capture high-level features. The concatenation effectively combines both types of information. The skip connections allow the network to recover spatial information lost during the encoder path. This helps to improve the accuracy of segmentation by combining low-level and high-level features. Thus, the network can leverage local and global context to make more accurate segmentation predictions. The final output layer is a Conv2D layer with a 1x1 kernel size, effectively performing a global average pooling operation. The activation function used in this layer is 'sigmoid'. The 1x1 kernel reduces the feature maps to a single channel, and the sigmoid activation scales the output to a probability score between 0 and 1. This probability represents the likelihood of a pixel belonging to the foreground class, making it suitable for binary segmentation tasks.

The batch size is 8, and the Adam optimizer with an initial learning rate is $1x10^{-4}$. Using the Reduce on Plateau as a learning rate scheduler, the learning rate is dynamically decreased by a factor of 0.05 if there is no improvement in the validation error for 10 epochs. The model is trained for the 300 epochs. Different loss functions, such as binary focal loss, dice loss, and distance transformation loss, were evaluated for performance. The binary focal loss with α = 0.25 and γ = 2, dice loss with β = 1 and a combined loss function incorporating binary focal loss and distance transform loss, with a weight parameter (λ) of 0.35 has been considered for experiments. These values are based on the best experimental results for the imbalanced dataset reported in [279]. The performance metrics for the model evaluation were the IoU and the F1-score. The sigmoid activation function was preferred over the softmax activation function because the model architecture has one neuron in the output layer, making it computationally more effective than the more generalized softmax function [280]. The model training is largely supported by data augmentation to create new training samples from the existing ones. Data augmentation is a technique to increase the size of the dataset and prevent the model from overfitting. The augmentation techniques and parameters adopted here are the same as those used for the pores and oxides detection model, including pixel-level and spatial level transformations.

Additionally, state-of-the-art general purpose edge detection models such as the HED algorithm trained on the BSD500 dataset [99] and the foundation model for prompt-based segmentation tasks such as the Segment Anything Model (SAM) trained on SA-1B dataset containing 11M images and 1B masks [281] were also evaluated for grain boundary detection in magnet KM images for baseline. Note that HED and SAM are trained on the large-scale general-purpose dataset but not on the specific magnet KM images. Nevertheless, they serve as foundational or general-purpose models for several computer vision tasks in different fields, including materials

science. The post-processing steps for the predicted masks by both the HED and SAM-trained models involve a sequence of steps to refine the final output. This process includes transforming the probability map into a binary mask, eliminating noise, and improving grain boundary detection. These operations are executed automatically through a series of steps, including thresholding the probability mask by setting a confidence value and applying morphological processes such as erosion and dilation, followed by skeletonizing the binary image. To identify the grain boundaries of large magnet samples using the trained model in KM images, an overlap-tiling strategy has been used. This strategy entails breaking down the large image into tiles measuring 1024 x 1024 pixels, with a 20-pixel overlap. Each of these tiles is then individually subjected to the trained model's processing. Following this, the processed tiled images are integrated using their spatial information to generate a comprehensive grain boundary map for the large image.

Additionally, to statistically compare the grain sizes of the magnet samples, the morphological characteristics of each grain within the sample were extracted. This involves measuring various parameters such as the area in µm², the ECD in µm, and the maximum diameter of the Feret in µm. For each sample area-weighted and number-weighted grain size distributions were obtained, using a semi-automated approach using an internally developed tool using the MS Excel Visual Basic for Applications, allowing us to compare these distributions with the reference distribution. In a number-weighted distribution, each grain contributes equally regardless of size, highlighting the abundance of smaller grains, as the distribution is based on grain count per size class. In contrast, an area-weighted distribution gives more weight to larger grains by factoring in their measured area, thus emphasizing their spatial or volumetric dominance. This tool requires measured grain morphological information to be fed as a CSV file, and it outputs cumulative plots and classification on grains into D10, D25, D50, D75, D90 values based on the ASTM E 1382 standard [282], [283].

3.6. Crystallographic grain orientation model for sintered FeNdB magnets

As discussed in section 3.2.1 and depicted in *Figure 23*, the grains have different domain patterns depending on their orientation, which can be measured as Euler angles using the EBSD method. The prediction models to determine the grain orientation in FeNdB magnet samples are developed as supervised learning-based regression models. Specifically, two distinct regression models were created, one for predicting θ and another for ρ. Both traditional ML techniques and advanced DL methodologies were explored for the orientation prediction task.

The KM images of FeNdB-E(T) and FeNdB-E(V) were processed to remove pores and oxides for training the regression models. Subsequently, grain boundaries are detected using the trained U-net model described in section 3.4, enabling the extraction of individual grains from the dataset. The extracted grain images as a gallery of individual grain images from the KM data, as shown in *Figure 30* are based on the deep learning segmentation model discussed in section 3.5.2. As part of the post-processing steps, grains with ECD smaller than 2 µm are ignored. The preprocessing and post-processing steps for both ML and DL approaches remain the same.

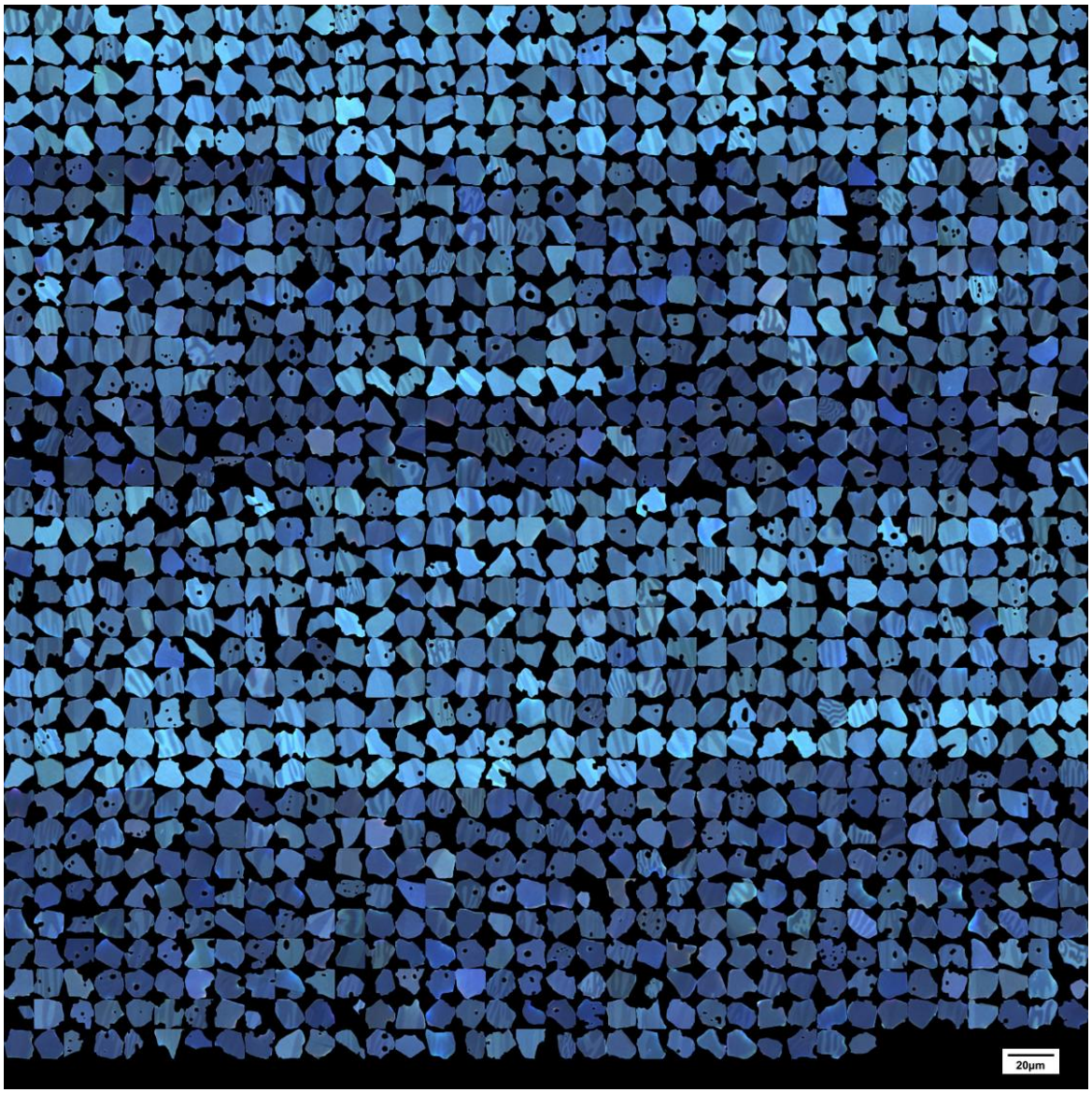


Figure 30: Picture gallery of the grains extracted from the Kerr microscopy image of the sintered FeNdB-E magnet sample. The prediction from the grain boundary model is post-processed using the morphological operations to extract each grain in the KM image. Note that the grains shown are randomly ordered.

3.6.1. Feature based machine learning model for grain orientation prediction

Various image features are manually extracted from grains to train the ML model for predicting θ and ρ based on KM data. *Table 16* presents the list of features acquired from each grain within the samples, including texture, color, shape, morphology, and domain statistics information. The overall feature set comprises 45 dimensions, with 13 Haralick parameters characterizing the grain texture. Included are three morphological attributes, three domain network statistics, an RGB colorfulness parameter [263], and 25 features derived from Zernike moments [284], effectively capturing and quantifying the shape of the domain structures. Colorfulness is a perceptual attribute describing the vividness or intensity of a color, indicating how much a color differs from a neutral gray of the same brightness. The colorfulness metric by Hasler et al. [263] quantifies colorfulness by combining the standard deviation and mean of the red–green and

yellow–blue color differences in an image. The standard deviation reflects the variation or spread of colors, while the mean captures the overall color bias, together providing a comprehensive measure of perceived colorfulness. Among the morphological characteristics, the mean orientation of domain patterns within the grain, the aspect ratio, and the area of the grain are considered. The remaining three attributes are domain statistics calculated from the skeletonized representation of the domain structures [285].

Table 16. Manually extracted features from the grains of the FeNdB magnet sample. The extracted features and its descriptors are shown in the table.

Features extracted	Description of extracted features from grain
Texture	Haralick parameters
Color	RGB colorfulness metric value
Shape	Zernike moments
Morphology	Orientation, aspect ratio, and area
Domain Statistics	Nodes, longest pattern, and standard deviation between patterns

Figure 31 shows examples of the skeletonized image representing the domain structure within a grain. This image is used for calculations involving domain statistics like the nodes' count, the longest pattern length, and the standard deviation between individual domain patterns within the grain. *Figure 31(c & f)* shows the nodes in the blue dots and the skeleton of the pattern referred to as branches in red. Grains with stripe patterns show a reduced number of nodes and branches. Additionally, in grains with stripe patterns, the branches are distributed more sparsely than in grains featuring closure domain patterns, as shown in *Figure 31(c & f)*.

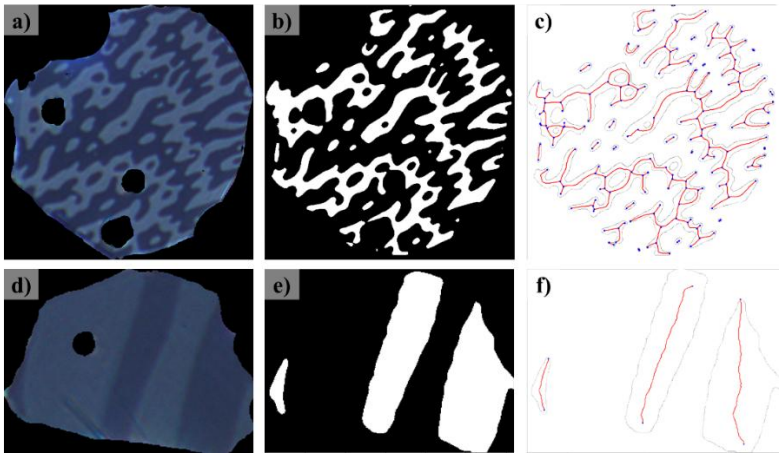


Figure 31: The workflow to obtain the descriptors of the domain statistics features extracted from the grains is shown. (a) and (d) original images of grains, (b) and (e) binary masks of domain structures in each grain, and (c) and (f) domain networks showing nodes in blue and branches in red, respectively.

The 45-dimensional feature for training the different traditional regression models that include tree-based models such as the AdaBoost regressor with random forest as base learner (A-RF), GB, and XGB, as well as simple ANN, which is an MLP. Based on the model's performance on the validation dataset, hyperparameters were tuned for better generalization and to avoid overfitting.

3.6.2. Deep learning model for crystallographic grain orientation prediction

The limited data availability for training a DL model poses constraints, making it challenging to achieve meaningful convergence and extract insights from the dataset. The transfer learning approach [216] has been adopted as a solution. This method offers the advantage of adapting well to comparatively smaller, custom datasets, even when the pre-trained model has been initially trained on different domains (such as ImageNet [286] or CIFAR-100 [287]). In the transfer learning process, the feature extraction layer of the pre-trained model remains frozen, while custom FC layers are introduced to adapt to the domain orientation datasets. Different pre-trained models' feature extraction layers, including VGG16 [262], ResNet50V2 [288], and EfficientNetB0 [289] were tested in combination with the custom FC layer. These pre-trained models, initially trained for classification tasks on ImageNet [31], had their final layers replaced by custom FC layers, including a regression layer to generate continuous output values instead of discrete ones.

The custom FC layer comprises three dense layers, with 32, 16, and 8 neurons in each respective layer. Additionally, batch normalization with weight decay of $1x10^{-3}$ and momentum of 0.9, dropout layers with dropout rate as 0.6, and L2 regularization with $1x10^{-4}$ decay parameter were used to prevent overfitting. The dense layers have ReLu as an activation function, and the weights are initialized with Xavier initialization [290]. In parallel, a simple baseline small CNN was trained, consisting of four convolutional layers, two 2D max-pooling layers, and three dense layers, culminating in a final output layer with a single output neuron. This baseline model served as a benchmark to evaluate the performance of the pre-trained models with enhanced feature extraction capabilities. The RGB image input for all the regression models comprises dimensions of 224 x 224 pixels. MAE and MSE_{reg} were considered the loss function and MAPE as performance metrics, with the Adam optimizer and a dynamically adjusting learning rate. For MSE_{reg} loss function, γ is obtained by the Bayesian regularization approach suggested by MacKay et al. [291]. To manage the learning rate, the Reduce on Plateau approach [292] was adopted, with a reduction of the learning rate by a factor of 0.1 when the model's learning progress stagnates over ten consecutive epochs. The learning rate minimum threshold was set at $1x10^{-9}$. Batch size is set to 24.

The hyperparameters considered for predicting both θ and ρ remained consistent, and the models were trained for 700 and 350 epochs, respectively. The DL model has been trained on two datasets with varying sample characteristics. As discussed in section 3.2.1, first, a dataset has grains from an anisotropic sample with a mean ECD of 10 µm. The second dataset includes additional grains from a quasi-isotropic sample with a mean ECD of 14 µm. The effect on the model performance with a combined dataset of different grain sizes and sample types were evaluated.

Additionally, an isotropic sample of ECD 14 µm was used to investigate the correlation between grain orientation (θ) and its brightness using the Goniometer [293]. The KM images were acquired at 500x magnification with exposure of 6.0s and the sample is rotated in the steps of 15°, representing the change in the transverse Kerr effect or θ value of the grains in the sample.

4. Results

4.1. Evaluation of the intrinsic property prediction model from chemical composition

For each property prediction task, the training dataset, cross-validation strategy, and learning regressors are different. In this section, the results from each regressor model on training, validation, and test set are mentioned.

4.1.1. Density of 14:2:1 magnetic phase

From 189 literature data on the density of 14:2:1 phases ranging between 6.62 [g/cm³] to 8.89 [g/cm³], two regression models were developed to predict the density of these phases from the chemical composition and atomic mass. Based on the screening strategy described in section 3.4, the initial assessment of the ρ_m prediction model using chemical compositional ('CC') features resulted in a baseline score of MAE: 0.22 [g/cm^3], MAPE: 2.8% and R^2: 0.72. Subsequently, all regression models with validation scores higher than the baseline score were subjected to pairwise correlation analysis. LR, RF, and SVM were shortlisted from this analysis for further tuning and experiments. The VR model, with the base regressors RF and SVM with uniform weights, was also compared against the shortlisted models. The next step involved fine-tuning all four models to optimize their hyperparameters. Similarly, the same set of four models was trained using an expanded feature set, which combined chemical composition with atomic mass (CC+AM). This extended feature set was used to improve the predictive capabilities of the models. In [98], the performance of the mass density predictor model and its performance with 'CC' and (CC+AM) are discussed.

Table 17 shows the performance comparison, measured in terms of MAE, RMSE, MAPE, and R^2, among different models using both the 'CC' and (CC+AM) feature sets. MAE provides a straightforward average of how far off the predictions are from the actual values, without considering the direction of errors. MAPE provides the accuracy of a model's predictions in terms of percentage error. RMSE is similar to MAE but penalizes larger errors more heavily compared to MAE, making it more sensitive to outliers [273]. The evaluations were performed using both LOOCV and a 5-fold cross-validation approach. Interestingly, the linear regression model performs better than the non-linear models. This outcome validates the selection of the LR model over alternatives such as RF, SVR and VR.

Table 17: The table presents a comparative analysis of different machine learning models based on performance metrics, MAE, MAPE, RMSE, and R^2 scores, across two feature sets: 'CC' and (CC+AM). The models evaluated include LR, RF, SVM, and VR. The table showcases how adding the feature set (CC+AM) generally improves model performance, as indicated by lower MAE and RMSE values and higher R^2 scores across all the models.

Metric	Feature set	LR	RF	SVM	VR
MAE [g/cm³]	CC	0.06	0.17	0.14	0.16
	CC+AM	0.04	0.16	0.12	0.15
MAPE [%]	CC	0.76	2.16	1.79	2.04
	CC+AM	0.51	2.04	1.53	1.91
RMSE [g/cm³]	CC	0.04	0.21	0.18	0.19
	CC+AM	0.05	0.18	0.17	0.19
R^2	CC	0.96	0.76	0.87	0.81
	CC+AM	0.98	0.81	0.89	0.85

The addition of AM features to the existing feature set of 'CC' has resulted in an improvement in the LR model. This improvement can be observed in *Figure 32(a)*, which illustrates the LR model's outcomes when trained only on 'CC'. The prediction results are further improved when the chemical composition and atomic mass are combined for model training, as shown in *Figure 32(b)*. It is important to note that regardless of the feature combinations, the LOOCV technique has consistently shown a better fit compared to the K-fold cross-validation approach (K = 5). The LR model showcases its ability to generalize effectively without overfitting. Furthermore, the LR model trained with 'CC' features using LOOCV exhibits an improved generalization capacity of 2% compared to the 5-fold cross-validation approach. Similarly, for the model trained on the (CC+AM) features, a comparable 1% improvement in generalization is observed.

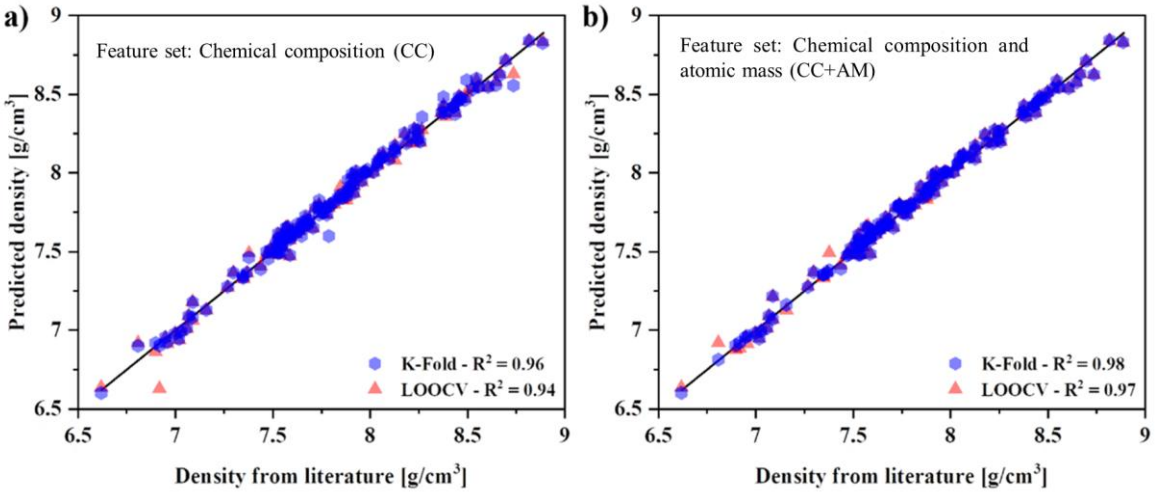


Figure 32: Comparison between the predicted and ground truth density values obtained from the various literature sources for the 14:2:1 magnetic phase using the LOOCV and the K-fold (K=5) cross-validation approach. The cross-validation results shown are from the LR model trained using (a) chemical composition and (b) chemical composition and atomic mass as combined features [98].

When evaluating the performance of trained LR model on an unseen test dataset (illustrated in *Figure 33*), the 'CC' feature set resulted in an MAE of 0.04 [g/cm³], RMSE of 0.06 [g/cm³], and an R^2 of 0.95. In comparison, improved results are achieved with the (CC+AM) features, with an MAE of 0.038 [g/cm³], RMSE of 0.04 [g/cm³], and an R^2 of 0.97. Additionally, the test set is categorized based on the number of chemical elements in the phase, such as ternary (3 elements), quaternary (4 elements), and quinary (5 elements) alloys. Quaternary alloys with the 'CC' feature as input resulted in the lowest MAE of 0.03 [g/cm³] and the quinary with a comparatively higher MAE of 0.05 [g/cm³]. However, with (CC+AM) features, the MAE decreased by 1.3% for ternary and quaternary alloys, and quinary, a predicted improvement of nearly 8% was observed as MAE reduced to 0.02 [g/cm³].

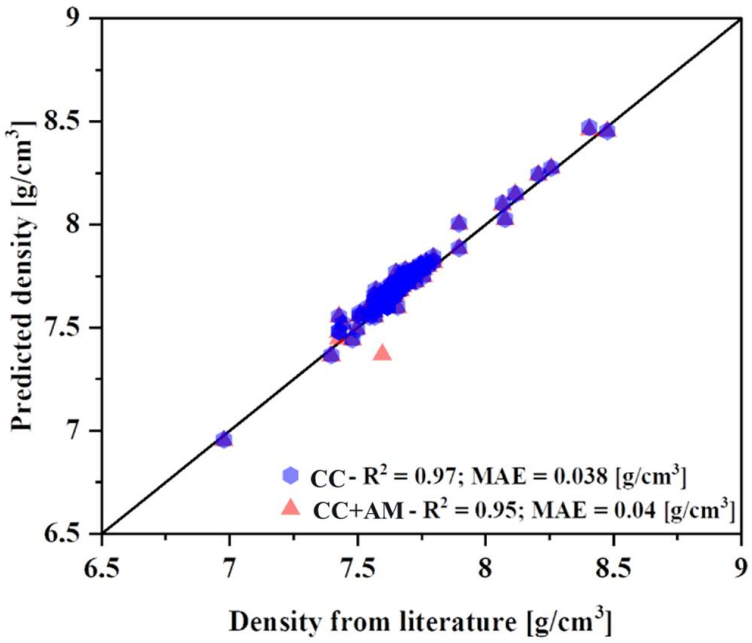


Figure 33: The comparison between the predicted and ground truth values for the test set obtained from the LR model trained using chemical composition only and another model with features such as chemical composition and atomic mass.

It is important to note that the training set consists of 123 unique phases, while the test set comprises 66 phases, with 13 of them being repeated phases or duplicates. The MAE for the 53 unique phases in the test set resulted in an MAE of 0.041 [g/cm³] using the LR model trained with 'CC' as features. The 13 repeated phases resulted in a slightly lower MAE of 0.035 [g/cm³]. Further, within the test set, there is one data, $Fe_{13.3}Cu_{0.7}Nd_2B$, which resulted in a comparatively higher absolute error of 0.24 [g/cm³].

4.1.2. Curie temperature of 14:2:1 magnetic phase

From 449 literature data on the Curie temperature of 14:2:1 phases ranging between 62 K to 1100 K, a regression model was developed to predict the Curie temperature of these phases from the chemical composition. Following the screening approach described in section 3.4, the baseline score obtained for the T_c predicting model is MAE of 48.8 K, MAPE of 18.4% and R^2 of 0.88 when trained on chemical compositional features. Further, after the pairwise correlation of the models, ET, XGB, RF, and AB-RF were shortlisted for further experiments as the validation scores for these models were better than the baseline score. Further, models such as LR, MLP, and SVR were excluded due to their relatively high validation errors. In addition, the VR model, with base regressors ET, XGB, RF, and AB-RF with uniform weights, was also evaluated against the shortlisted models. All models were fine-tuned to optimize their hyperparameters. In [178], the experimental results are explored for the T_c predictor model.

The ET, XGB, RF, and AB-RF models resulted in validation errors of 15 K, 17 K, 19 K, and 20 K, respectively, using MAE as the metric. When considering RMSE, the respective validation errors were 36 K, 40 K, 38 K and 39 K. To improve the generalization of the T_c predictions, a voting regressor was constructed with equal weightage to the predictions from the four chosen base regressor models. The trained voting regressor shows an MAE of 16 K

and a RMSE of 38 K on the validation dataset (seen in *Figure 34(a)*). The difference between the MAE and RMSE metric values suggests that the model performed well on most of the phases within the training (and validation) dataset while revealing relatively higher prediction errors for specific phases. Specifically, two phases, $Fe_{11}Cr_3Gd_2B$ and $Fe_{11}Ni_3Nd_2B$, resulted in relatively higher prediction errors that exceed 50 K, notably 63 K and 57 K, respectively. On the contrary, most phases resulted in an absolute error below 50 K, mostly around 20 K.

Furthermore, considering the partially biased frequency distribution of T_c values in the experimental dataset, as shown in *Figure 18(c)*, with a concentration of phases featuring the T_c values ranging from 450 to 750 K, an 80:20 split ratio was used for the training and validation sets. This split is aimed to overcome data leakage and prevent the model from overfitting. Moreover, the repeated phases in the training set have a prediction error of 7.8 K and 14 K in terms of MAE and RMSE, respectively. The repeated phases have T_c deviation within 20 K, except for one phase ($Fe_{12}Co_2Nd_2B$) where it is 34 K.

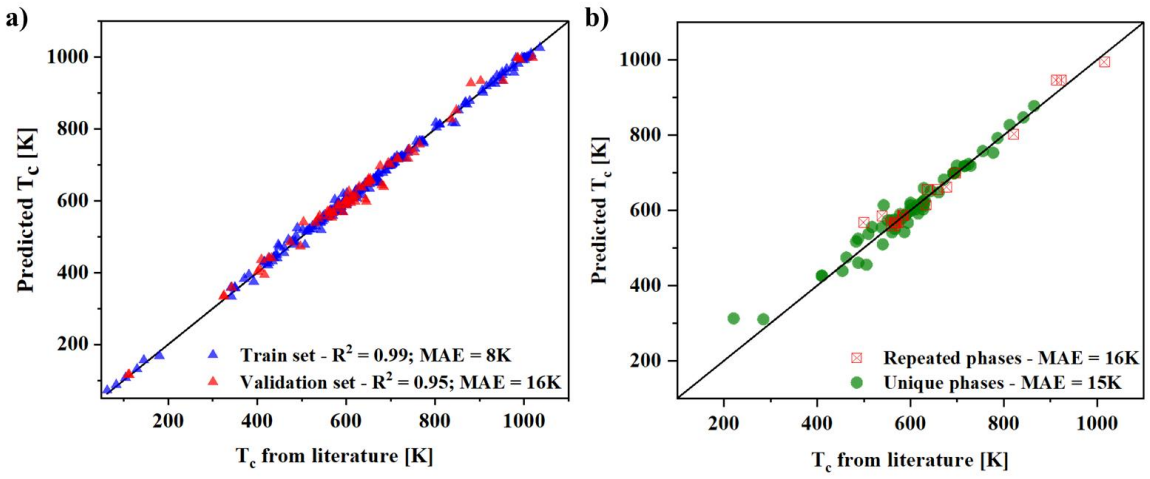


Figure 34: Comparison between the literature reported and the trained model predicted T_c [K] values (a) training and validation set and (b) test set showing unique and repeated phases (on the right) [178].

The T_c predictions on an unseen test set have low errors, with an MAE of 16 K, RMSE of 20 K, and a high R^2 value of 0.97, as shown in *Figure 34(b)*, affirming the model's strong generalization ability. Notably, this test set comprises phases with compositions obtained from literature references that differ from those in the training set. Further, the test set includes 61 unique and 27 repeated phases, whereas the latter includes 13 distinct phases in varying quantities. The predictions of the trained model for both the unique and repeated phases within the test set resulted in low prediction errors. For the unique phases, the MAE is 16 K and the RMSE is 22 K; for the repeated phases, the MAE is 15 K and the RMSE is 18 K. The slightly higher MAE of 15 K for the repeated phases can be due to variations in the reported T_c values across diverse literature sources. Similarly, the marginally higher RMSE of 22 K observed in a subset of the test set containing unique phases is due to the $Fe_{11}Mn_3Y_2B$ phase, which resulted in an absolute error of more than 50 K.

The predictions for the phases in the test set have been further split into distinct categories of phases, namely 14:2:1 phases containing heavy rare earth and light rare earth elements, which hold significance in the context of manufacturing magnets, as seen in *Figure 35*. The phases

with elements such as Gd, Y, Tb, Dy, Ho, Er, and Tm are categorized as heavy rare earth elements, resulting in an MAE of 16 K and an R^2 of 0.97. On the other hand, phases containing Ce, Pr, Nd, and Sm, referred to as light rare earth elements, resulted in an MAE of 13 K and an R^2 of 0.97. Among the heavy rare earth bearing 14:2:1 phases, $Y_2Fe_{11}Mn_3B$ is the only phase with an error exceeding 50 K, specifically 92 K. This error can be rationalized by considering the feature importance, given the sensitivity of T_c to the presence of Mn. On the contrary, among the light rare earth bearing 14:2:1 phases, none show a prediction error exceeding 50 K.

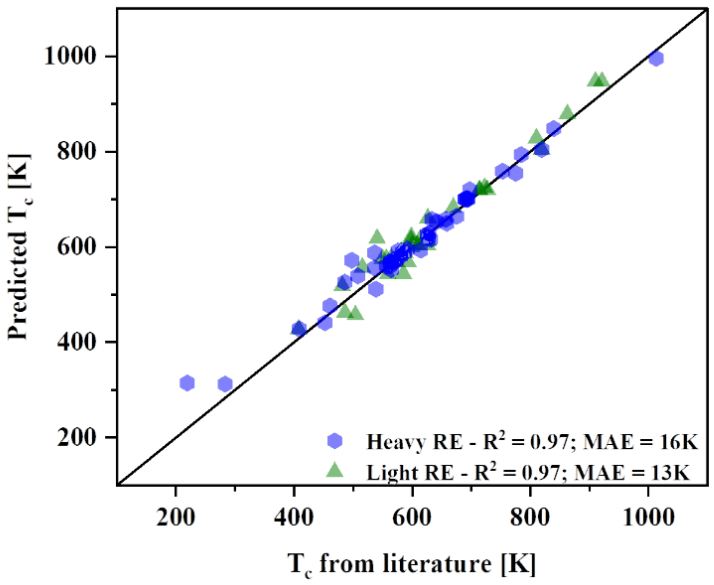


Figure 35: Comparison between the predicted and the literature reported T_c [K] for the phases in the test set which are categorized into light and heavy RE substitutes [178].

4.1.3. Magnetic saturation of 14:2:1 magnetic phase

From 198 literature data on the magnetic saturation of 14:2:1 phases ranging from 5 [μ_B/f.u.] to 35 [μ_B/f.u.], regression model was developed to predict the M_s of these phases from the chemical composition. Based on the screening strategy mentioned in section 3.4, the baseline score obtained for M_s predicting model is MAE of 3.3 [μ_B/f.u.], MAPE of 12.5% and R^2 of 0.87 when trained with chemical compositional features. Further, after the pairwise correlation of the HR, MLP, RF and AB-DT were shortlisted for further experiments as they resulted in more validation scores than baseline value. Note that the remaining models, such as GB and XGB, have not been considered because of a high validation error. The VR model with base regressors HR, MLP, RF and AB-DT having uniform and non-uniform weights is also compared against chosen models.

Table 18 shows the performance evaluation of the trained regression models on the validation dataset, using the RMSE, MAE, and R^2 metrics. The base regressors, HR, RF, MLP, and AB-DT were considered for the voting regressor, which resulted in validation errors of 1.08, 0.80, 0.71, and 0.45 [μ_B/f.u.] when measured by MAE, and 1.17, 1.17, 1.09, and 0.78 [μ_B/f.u.] when using RMSE, respectively. Other regression models, such as GB and XGB, resulted in validation errors of 1.50 and 1.16 [μ_B/f.u.] in terms of MAE, and 1.90 and 1.61 [μ_B/f.u.] based on RMSE, respectively. Constructing the voting regressor involves using the base regressors with uniform and non-uniform weights. In the uniform weighted VR, the four base regressors

are assigned equal weights of 1. In contrast, the non-uniform weighted VR assigns weights of 0.89, 0.94, 0.95, and 0.96 to the base regressors HR, RF, MLP, and AB-DT, respectively. These weightings are determined through a performance-based approach, where normalized MAE, RMSE, and R^2 metrics for each base regressor on the validation set are used to obtain optimal weights through the grid search.

Table 18: Performance comparison between the different base regressor models and the weighted regressor model for M_s in terms of R^2, RMSE, and MAE on the validation dataset having 29 phases. AB-DT has highest R^2 score, NU-VR and MLP has lowest RMSE and MAE values respectively, among all the regressor models.

Regression models	Model characteristic	R^2	RMSE	MAE
Huber regressor (HR)	Linear model - L2 regularized	0.97	1.17	1.08
Random forest (RF)	Tree-based – Bagging	0.97	1.17	0.80
Gradient boosting (GB)	Tree-based – Boosting	0.93	1.90	1.50
Extreme gradient boosting (XGB)	Tree-based – Boosting	0.91	1.61	1.16
Decision tree+adaptive boosting (AB-DT)	Tree-based - Adaptive boosting	**0.98**	1.08	0.75
Multi-layer perceptron (MLP)	Artificial neural network	0.96	1.09	**0.71**
VR - Uniform weight (U-VR)	Stacking - uniform weights	0.94	1.16	0.79
VR - Non-uniform weight (NU-VR)	Stacking - uniform weights	0.97	**1.07**	0.72

The validation performance of the uniform weighted VR model resulted in errors of 0.79 [μ_B/f.u.] and 1.16 [μ_B/f.u.] using MAE and RMSE, respectively. On the other hand, the non-uniform weighted VR resulted in errors of 0.72 [μ_B/f.u.] and 1.07 [μ_B/f.u.] with MAE and RMSE, respectively, for the same validation set. The uniform weighted VR resulted in an average deviation of 3.1%, whereas the non-uniform weighted VR had a comparatively lower deviation of 2.6%. Based on the superior performance of the non-uniform weighted VR (NU-VR), it was selected for further assessment.

Figure 36 shows the performance of the NU-VR model on a training set comprising 117 phases, combined with a validation set of 29 phases, compared against reference values from diverse literature sources. The model achieved a good fit, exhibiting an overall MAE of 0.72 [μ_B/f.u.] and a high R^2 score of 0.97 for the combined training and validation sets. This highlights the model's effectiveness in fitting the dataset, with absolute errors below 2 [μ_B/f.u.] for the most phases. However, a few phases, specifically $Fe_{10}CoAl_3Nd_2B$, $Fe_{14}Nd_{1.6}Er_{0.4}B$, and $Fe_{14}Nd_{1.5}Tb_{0.5}B$, resulted in a relatively higher absolute error of 5.16, 2.6, and 2.49 [μ_B/f.u.], respectively.

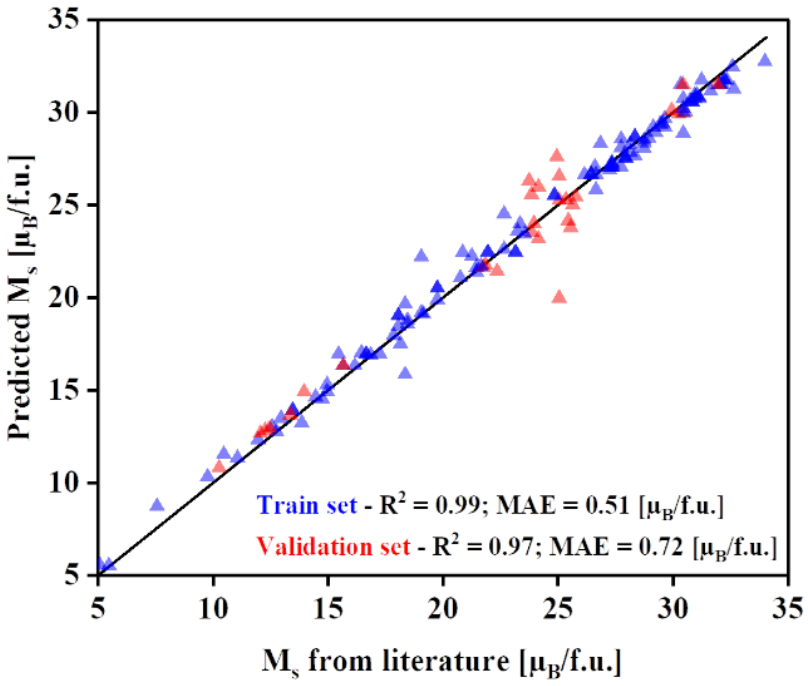


Figure 36: Comparison between the predicted saturation magnetization values from the non-uniform weighted voting regressor and literature reported values for the phases in the training set (in blue color) containing 117 phases in terms of MAE and R^2 score and on an independent validation subset (in red color) of 29 phases [185].

On the unseen test set, the trained model resulted in an error of 0.77 [μ_B/f.u.] and 0.88 [μ_B/f.u.] using MAE and RMSE, respectively. *Figure 37* shows the comparison between predicted and literature reported M_s values for the test dataset. This implies that an average deviation of 2.9% is obtained from the predictions of the trained model. The low error for the unseen test set suggests the generalization of the model across the dataset. Furthermore, the model's effectiveness on the repeated and unique phases within the unseen test set was evaluated. Among the 40 unique phases, the model achieved an MAE of 0.8 [μ_B/f.u.] and an RMSE of 0.9 [μ_B/f.u.]. For the 12 repeated phases within the total of 52 phases in the unseen test set, the model resulted in a low MAE of 0.7 [μ_B/f.u.] and an RMSE of 0.8 [μ_B/f.u.]. Further, the 12 repeated phases in the unseen test set are from different literature sources, which have reported varying M_s values which is within ±4% deviation from mean value. The only unique phase with an absolute error exceeding 2 [μ_B/f.u.] was $Fe_{12}Cu_2Y_2B$. The remaining 51 phases exhibited absolute errors below 2 [μ_B/f.u.], with a majority displaying an error below 1 [μ_B/f.u.].

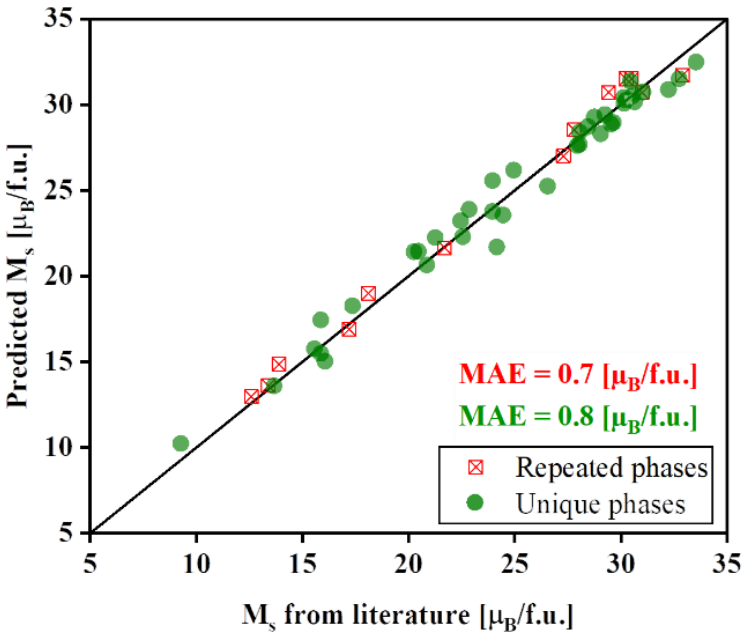


Figure 37: Comparison between the predicted saturation magnetization values from the non-uniform weighted voting regressor and the literature reported values for the phases in the unseen test set (52 phases). The test set contains 39 unique and 13 repeated phases taken from independent literature sources that were not part of the training set [140].

4.2. Evaluation of the grain boundary detection model for sintered FeNdB magnets

The pores and oxides removal model described in section 3.5 is a semantic segmentation model trained using the transfer learning technique. It is based on the U-net architecture with pre-trained EfficientNetB3 as encoder resulted in the F1-score of 0.96 and IoU of 0.97 on the validation set having 17 image patches of size 1024 x 1024 pixels. The same model resulted in a binary focal loss of 0.04 for the training set and 0.05 for validation set. The pores and oxides detection model has generalized well on training set and has high IoU on the test set that has KM images of various magnet samples. The IoU on KM images from different magnet samples by the trained model is more than 0.95. The obtained probability map from trained model is processed with threshold value of 0.3 to obtain a binary mask. The combined area fraction of the pores and oxides in the samples vary between 6.6% to 13.8% for different samples mentioned in *Table 6.*

In the experimental steps described in sections 3.5.1 and 3.5.2, the RF (C) model trained on color features from RGB channels resulted in a very low IoU of 0.36 when tested on the validation set. For the same validation set, the RF(C,E,T) and XGB trained on all 39 features resulted in comparatively better IoU of 0.81 and 0.67, respectively. Additionally, the model RF(VGG16) trained on the features extracted from the pre-trained VGG16 resulted in an IoU of 0.65.

Further, in the experimental test on the state-of-the-art edge and prompt-based semantic segmentation model described in section 3.5.2, the HED model resulted in the IoU of 0.86 on the validation set, and SAM resulted in the IoU of 0.82. The performance of HED and SAM is

better than the models trained on the hand-crafted features, and therefore, using the transfer learning approach U-net architecture with different pre-trained models as encoders are trained and evaluated for the performance. All U-net models for performance comparison are trained with dice, binary focal loss and a combined loss with distance transformation loss combined with equal weightage.

Among the U-net models with pre-trained encoders, EfficientNetB3 as encoder resulted in a low loss of 0.09, a high IoU of 0.92, and an F1-score of 0.90. U-net with encoder as pre-trained VGG16 resulted in IoU of 0.85, which is close to the HED model's performance. The U-net with pre-trained encoders InceptionV3 and ResNet50 resulted in similar IoU scores of 0.87 and 0.88, respectively, which is better than HED, SAM, and U-net with VGG16 as encoder but has lower IoU than U-net with EfficientNetB3 as encoder. The performance evaluation was obtained by considering manually annotated grain boundary maps as a reference mask. However, the custom U-net model resulted in the highest IoU of 0.94 and an F1-score of 0.91 when evaluated on a validation set of 17 image patches. *Table 19* shows the performance of all trained models on the validation set using the F1-score, mAP, and IoU metrics. Additionally, it shows the number of grains in the sample, average grain size, and number of grains found per unit area in the sample.

Table 19: Comparison between the different ML and DL models used for the performance evaluation of the grain boundary detection model using the F1-score, mAP, and IoU as metrics. Other measured values such as number of grains, average grain size and number of grains per μm^2 is also shown. Among all the models, U-net - Custom has the highest F1-score, mAP and IoU values.

Models	Model type	F1-score	mAP	IoU	No. of grains	Average grain size	No. of grains per μm^2
RF(C)	ML	0.38	0.33	0.36	3977	1.55	0.64
RF(C,E,T)	ML	0.72	0.69	0.81	2627	2.26	0.43
RF(VGG16)	Hybrid	0.52	0.51	0.65	2788	2.21	0.45
XGB	ML	0.49	0.47	0.67	2806	2.20	0.46
HED	DL	0.77	0.79	0.86	2550	2.42	0.41
SAM	DL	0.72	0.71	0.84	2611	2.36	0.42
U-net - Custom	DL	**0.91**	**0.90**	**0.94**	2289	2.69	0.37
U-net - EfficientNetB3	DL	0.90	0.87	0.92	2250	2.74	0.36
U-net - VGG16	DL	0.70	0.75	0.85	2515	2.45	0.41
U-net - InceptionV3	DL	0.88	0.82	0.87	2308	2.67	0.38
U-net - ResNet50	DL	0.89	0.84	0.88	2287	2.48	0.37

Figure 38 shows grain boundaries detected by the trained custom U-net, HED, and SAM models. For the input image shown in *Figure 38(a)*, the IoU score for the mask predicted by the HED and SAM is nearly the same, close to 0.80. However, the custom U-net model has a very high IoU of 0.96. In spite of HED and SAM being trained on general purpose dataset such as ImageNet and SA-1B, they were able to perform well on KM images.

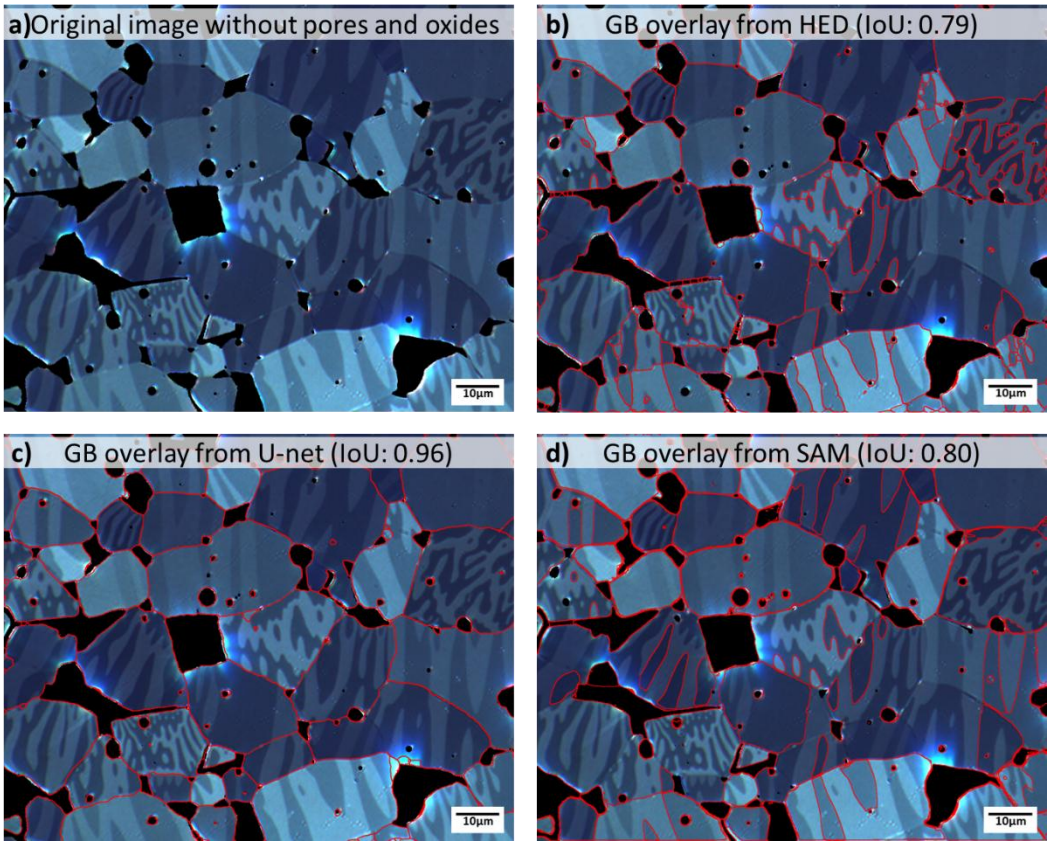


Figure 38: The detected grain boundary (GB) for the KM image overlaid on the original image without pores and oxides is shown using the different models and respective IoU scores. (a) original image, (b) GB predicted from the pre-trained HED model, (c) GB predicted from the trained custom U-net, and (d) GB predicted from the pre-trained SAM model. The sample shown here is FeNdB-E.

The custom U-net model has also been evaluated for different loss functions. The model has been trained with binary focal loss, dice loss, and distance transformation loss. The distribution-based loss function, such as binary focal loss, is widely used for segmentation tasks with imbalanced data distribution, such as edge detection, which has more pixels to be classified as background than foreground. Dice loss is a regional-based loss function similar to the IoU metric. The distance transformation loss function aims to reduce the loss by assigning more weightage on the pixels close to grain boundaries. It is often used in combination with the region-based loss functions.

Further, a combined loss function is also evaluated using the distance transformation loss and binary focal loss with the a weight distribution of 0.35 and 0.65, respectively. The combined loss function resulted in the F1-score of 0.93, better than the F1-score of 0.91 obtained from the combined loss function of binary focal loss and dice loss with uniform weightage. Therefore, the combined loss function of binary focal and distance transformation loss is considered suitable for the custom U-net model.

Figure 39 shows the precision-recall curve for all the trained models compared with each other, EBSD, and manually generated values. The area under the PR-curve for the RF(C) is the least; for custom, U-net is the highest among all trained models. The baseline precision of 0.38 has been adopted. The variation of the F1-score with different threshold values for the probability

map obtained from the top 5 models is shown in *Figure 39*. The threshold value of 0.4 is optimal for a custom U-net model.

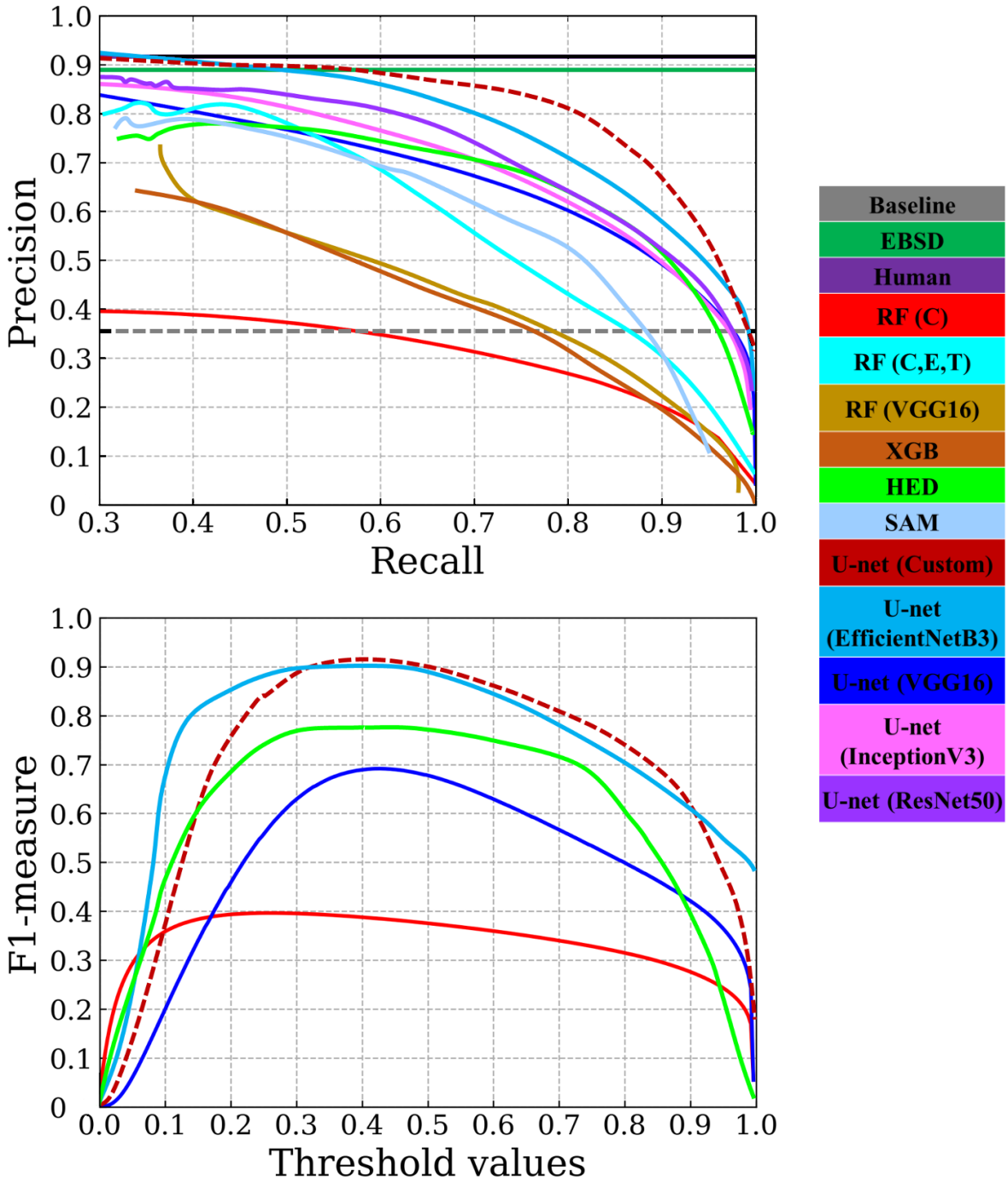


Figure 39: Precision-recall curve for the different ML and DL models on the combined grain orientation training and validation dataset is shown at top. The pre-trained edge detection and segmentation model are also included in the results for comparison. The flat line at the top represents human results, indicating consistently high precision across all recall values. The F1-measure or F1-score versus threshold values for specific models for obtaining the optimal threshold values is shown in the bottom.

The custom U-net model with a threshold value of 0.4 is then further evaluated for the performance and robustness of samples with different types of grains. As mentioned in section 3.2.2, the grain has stripe and closure domain patterns depending on the grain orientation.

Figure 40 shows the performance of the trained model on the KM images of the magnet samples, which have grains with varying domain patterns. *Figure 40(a)* has most grains with closure domain patterns and high contrast, whereas *Figure 40(c)* mostly has grains with stripe domain patterns with very low contrast. Further, a sample with grains having different degrees of stripe and closure domain patterns is shown in *Figure 40(b)*. The predicted and reference masks are highlighted in blue [0, 0, 255] and red [255, 0, 0], respectively. The grain boundary with an overlap of more than 2 pixels is white [255, 255, 255]. This overlap accounts for a minimum of 2 pixel overlap for the grain boundary of 5 pixels' width. The IoU for the predicted mask for the sample decreases from 0.94 to 0.90 for input images in *Figure 40(a- c)*. The sample with low-domain contrast grains cannot detect grain boundaries when neighboring grains have similar image textures. Meanwhile, grain boundaries are detected with greater accuracy for samples with closure domain patterns despite complex local patterns for each grain. However, it must be observed that the overlapping pixels have a high precision between 3-5 pixels.

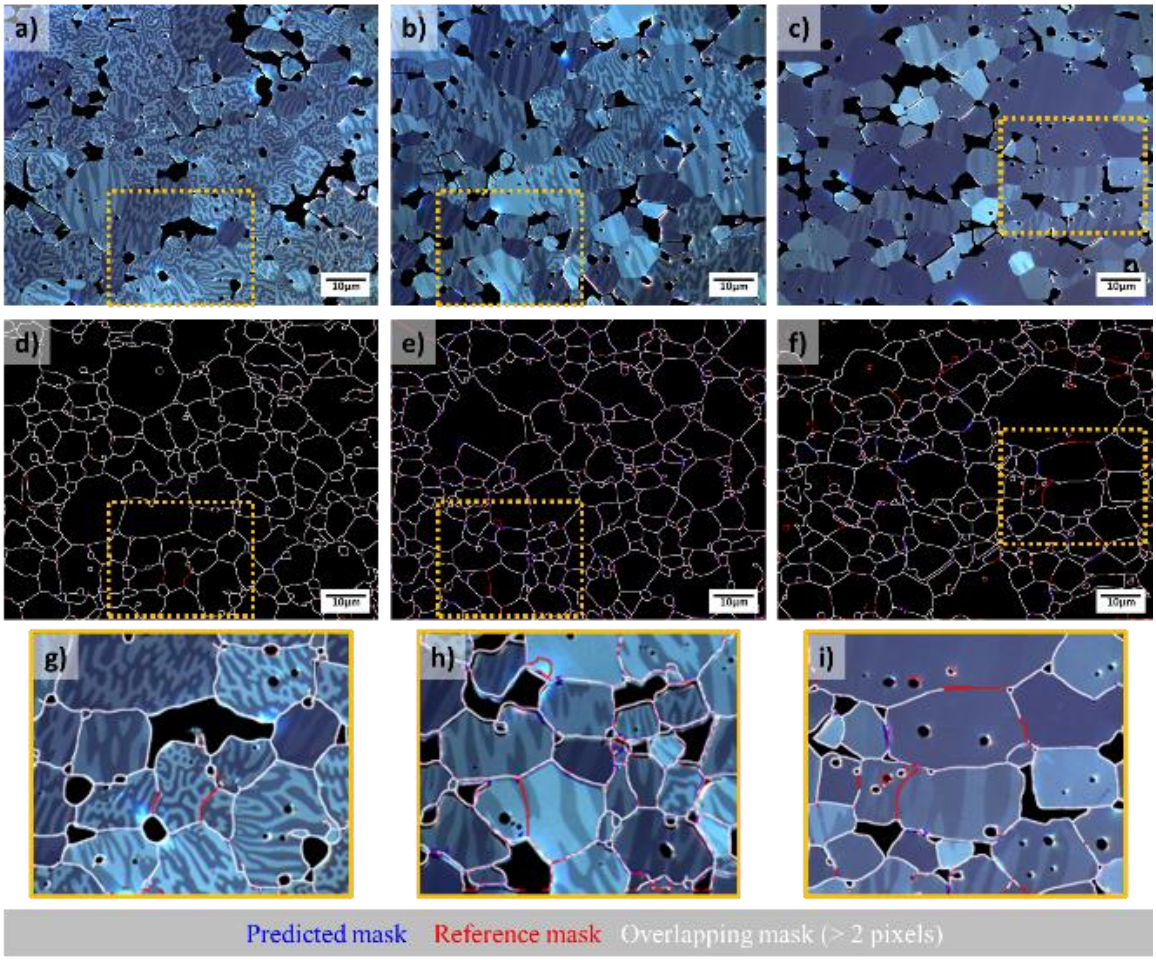


Figure 40: The grain boundary detected by the trained custom U-net model for the KM images from the sample FeNdB-E sectioned at different angles against the axis of anisotropy is shown. The reference mask is shown in red color, the predicted mask in blue color, and the overlapping grain boundary upto 2 pixels in white color. (a)-(c) Original KM images of sample sectioned at 15º, 45º, and 90º respectively, (d)-(f) predicted and reference map overlaid, and (g)-(i) predicted and reference masks overlaid on sub-section of the original KM image.

For the sample shown in *Figure 40(b)*, the IoU is 0.91, and the misclassification is observed mostly along the grain boundaries where the predicted mask and the reference mask do not overlap precisely in many places across the sample. The overlap is mostly between 2-3 pixels, accounting for less precision.

For the effective grain size analysis of the magnet, a large area over the sample is to be analyzed and using the overlapping tile strategy, the model trained on 1024 x 1024 pixels can be used on large size images. *Figure 41* shows a sliding window of size 1024 x 1024 pixels with an overlap of 10 pixels. The information present in the overlapping region of 10 pixels helps overcome the border artefacts and provides information about the neighboring grains to the model. Such approaches have proven effective for inference on large samples [120], [294].

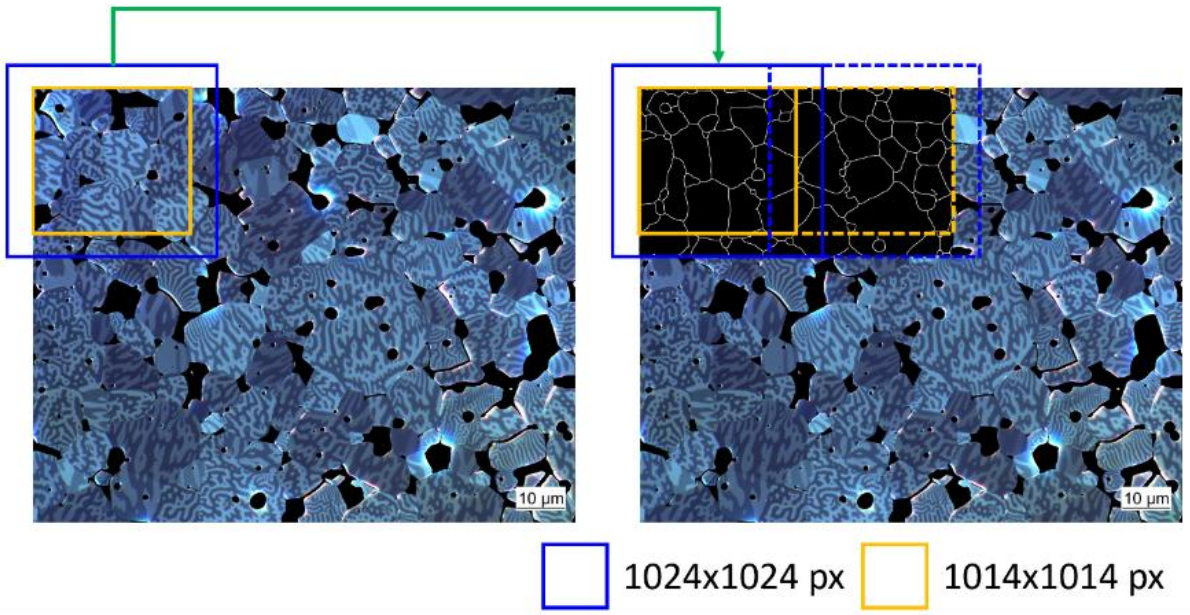


Figure 41: The overlapping tiling approach used for analyzing large areas of the sample using the trained model is shown. The sliding window of size 1024x1024 pixels with 10-pixel overlap has been used to avoid the border artefacts generated from the trained model prediction.

The grain boundary mask generated by the trained custom U-net model can be a foundation for extracting statistical information about the grain size distribution. The analysis of grain sizes can be obtained through two distinct approaches: the intercept-based method and the planimetric approach [282]. *Figure 42* shows the grain size analysis using the EBSD approach and from the grain boundary detection model that can be quantified using the planimetric and intercept-based approaches. The intercept measurements follow Abrams' three-circle procedure, a circular variant of the traditional Heyn line intercept method, designed to reduce orientation bias [295]. By utilizing the grain boundary mask, essential morphological parameters such as area, ECD, aspect ratio, maximum Feret diameter, and circularity for each grain can be derived. This morphological data, in turn, becomes instrumental in generating both number-weighted and area-weighted grain size distribution curves.

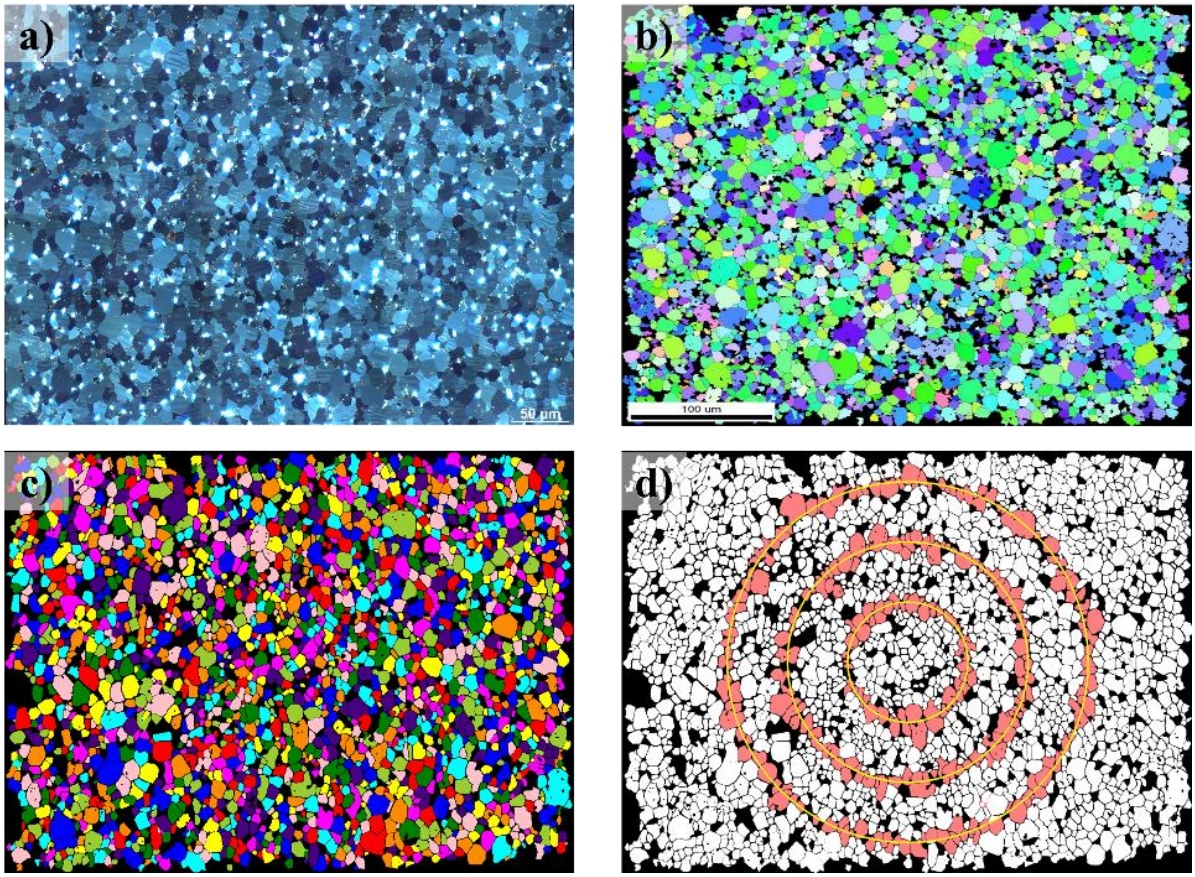


Figure 42: The KM image of the sintered FeNdB-A magnet image is shown. (a), SEM EBSD map, (c) detected grain boundaries in KM image by trained custom U-net model with random colors that can be used for planimetric analysis, and (d) binary mask of detected grains which can also be used for intercept-based grain size analysis such as Abrams three-circle procedure.

Figure 43 illustrates the area-weighted and number-weighted grain size distributions for the FeNdB-C and FeNdB-D samples. These distributions were obtained using the trained RF(C,E,T), a custom U-net, and a manual approach conducted by an expert in the field. As described in section 3.2.3, the results from the manual approach serve as the benchmark or reference data for performance assessment. Furthermore, the D25, D50 and D90 values are obtained from the grain distribution curves is computed to evaluate the performance. The RF(C,E,T) is an ML model, and custom U-net is a DL model here. Note that for the first experiments related to grain size distribution, the correlative EBSD data were not obtained and therefore *Figure 43* does not include EBSD results.

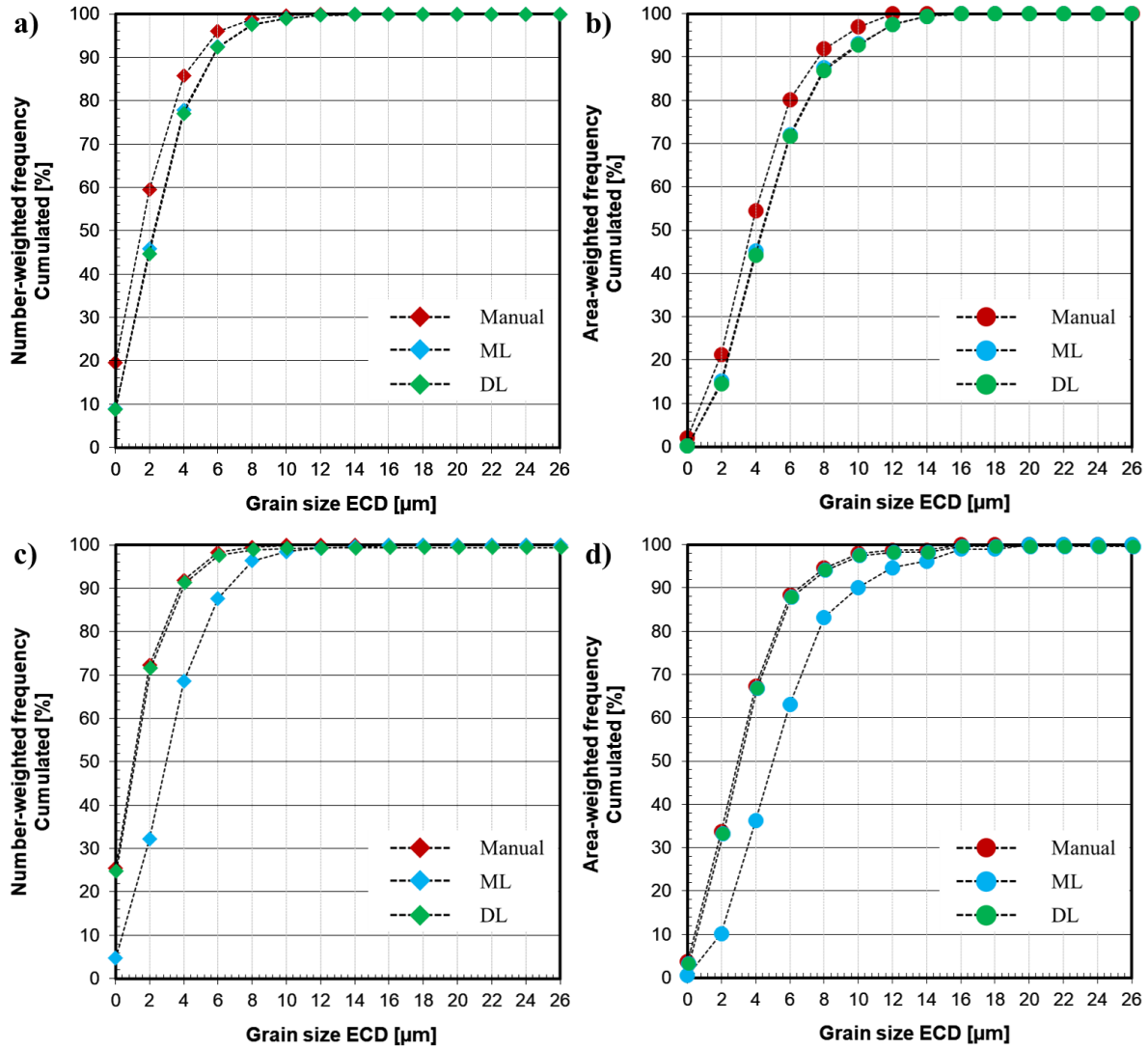


Figure 43: The area-weighted and number-weighted grain size distribution curve for the sintered FeNdB magnet samples obtained from the ML, DL, EBSD, and manual approaches. (a) and (b) results for the FeNdB-D sample and (c) and (d) results for the FeNdB-C sample [69].

The grain size distribution, area-weighted and number-weighted, for the FeNdB-C and FeNdB-D samples implies that the distributions from the trained custom U-net model (DL model) align closely with the reference or manual values. Specifically, when comparing the U-net model's area-weighted distribution curve to the reference distribution (manual), the deviation in the grain size value refers to the signed difference between the reference (manual) measurement and the model prediction. A negative value indicates that the model overestimates grain size, while a positive value indicates underestimation. The deviation in D50 is 0.6 µm for FeNdB-D and 0.1 µm for FeNdB-C. Similarly, in the case of the number-weighted distribution curves, a similar trend is observed, with deviation of 0.6 µm for FeNdB-D and 0.05 µm for FeNdB-C. The relative deviation in the D25 value (fine or small grains) for FeNdB-D and FeNdB-C from the area-weighted distribution curve are 2.0 µm and 0.1 µm respectively. Similarly, the relative deviation for coarse or large grains represented as the D90 value in area-weighted distribution curves for FeNdB-D and FeNdB-C are 1.1 µm and 0.4 µm respectively.

Deviations are noticeable when comparing the predictions from the trained random forest classifier with the reference distribution. For FeNdB-D, the relative deviation in D50 is 0.6 µm for the area-weighted distribution curve and 0.5 µm for the number-weighted distribution curve. However, for FeNdB-C, the deviation of the predicted area-weighted and number-weighted distribution curves from the reference curve is 2.1 µm and 1.6 µm, respectively. In terms of the D25 and D90 values for FeNdB-D, the relative deviation from the area-weighted distribution curves are 1.1 µm and 1.2 µm respectively. For FeNdB-C, it is 1.6 µm and 2.8 µm for D25 and D90 values, respectively. These deviations are relatively close to the reference values and the outputs from the trained custom U-net model. It is important to note that while the results obtained through the manual approach are considered a reference for evaluating model performance, assessing the accuracy of manual results is equally crucial. This is because manually annotated masks could introduce bias and vary based on subjective expertise. To address this, grain size distribution was obtained from the EBSD data for the FeNdB-A sample.

Figure 44 shows the area-weighted and frequency-weighted grain size distribution curves for FeNdB-A obtained using various approaches: EBSD, manual annotation, RF(C,E,T) and a custom U-net model. The signed difference in the D50 from the area-weighted distribution curve between the manual and EBSD methods is 0.6 µm, while for the frequency-weighted distribution curve it is 0.05 µm. The deviations for D25 and D90 from the area-weighted distribution curve are 0.3 µm and 2.5 µm respectively. This observation underscores the productivity of using a manual approach as a reference for evaluating trained models, particularly when compared to the EBSD approach. The manual approach is advantageous due to the reduced time and effort required for obtaining results, in contrast to the demands of EBSD. Furthermore, it is important to acknowledge that correlative microscopy involving EBSD and KM is exceptionally challenging. This complexity arises from the need for a high level of technical proficiency to acquire accurate correlative images in magnet samples, minimizing errors to the utmost extent.

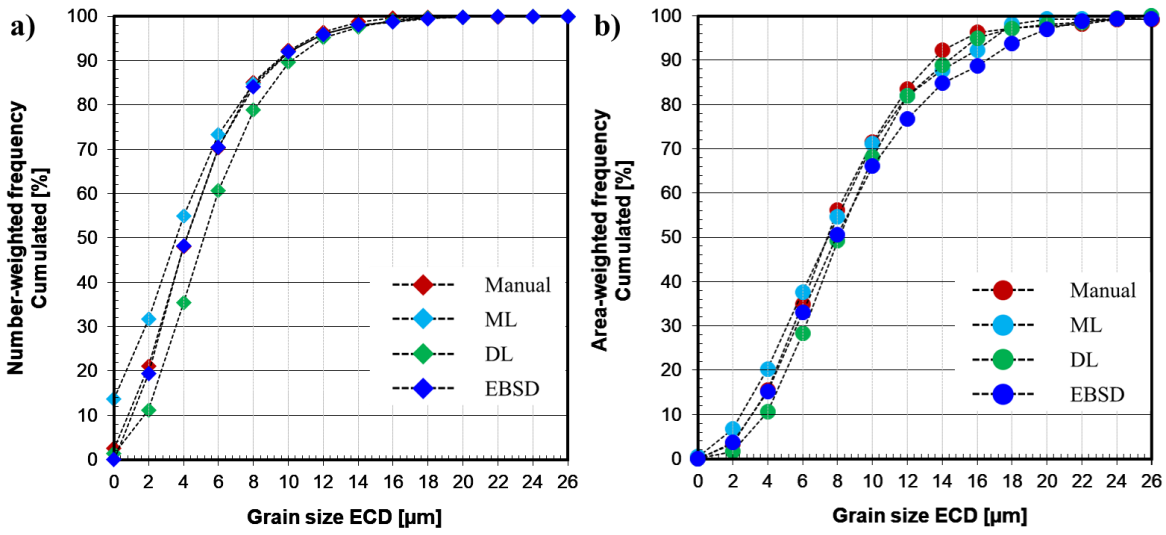


Figure 44: The number-weighted (a) and area-weighted (b) grain size distribution curve for the sintered FeNdB-A magnet samples obtained from the ML, DL, EBSD, and manual approaches [69].

The relative deviation in D50 value from the area-weighted grain size distribution curve from the EBSD and the RF(C,E,T) model is -0.4, while from the custom U-net model, it is 0.3 µm.

Similarly, the relative deviation of the number-weighted grain size distribution curve by the RF(C,E,T) model is -1.2 µm, and by the custom U-net model it is 2.2 µm. Both models perform almost equally well for this sample and exhibit results that closely align with the reference values. In terms of the D25 and D90 values from the area-weighted distribution curves, the signed difference between the EBSD and the RF(C,E,T) model are -0.5 µm and 1.0 µm respectively. The relative error between EBSD and custom U-net model based on D25 and D90 values from area-weighted distribution curves are -0.8 µm and -1.2 µm respectively.

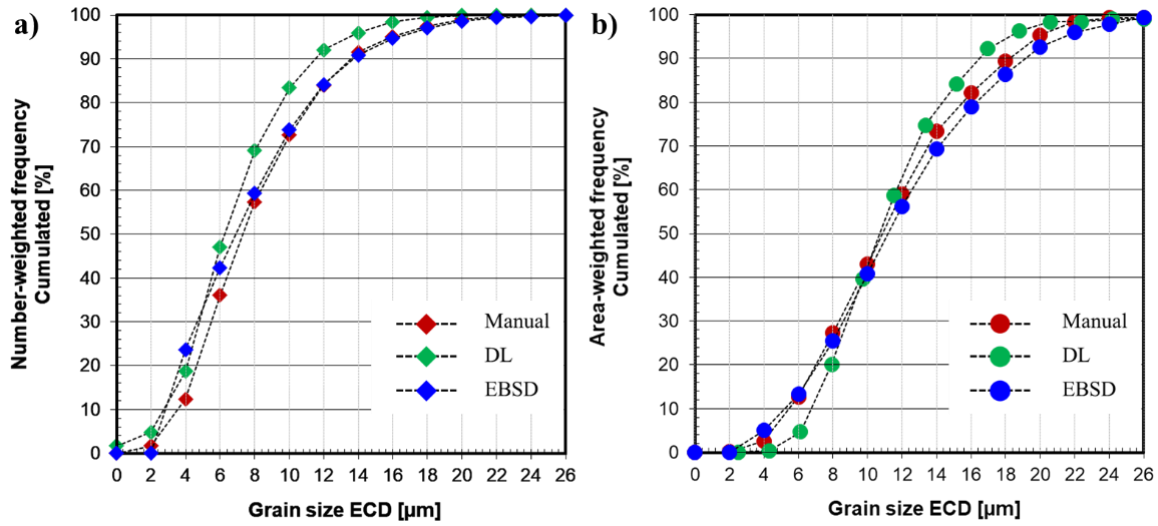


Figure 45: The number-weighted (a) and area-weighted (b) grain size distribution curve for the sintered FeNdB-B magnet samples obtained from the DL (custom U-net), EBSD, and manual approaches [69].

Figure 45 illustrates the distribution curve for the FeNdB-B sample, displaying results from the manual, EBSD and DL approaches. When comparing the area-weighted distribution curve from the DL (custom U-net) model with the EBSD curve, there is a relative deviation in D50 of -0.3 µm. In comparison, the number-weighted distribution curve shows a deviation of -0.8 µm from the EBSD. The slightly greater deviation in the number-weighted curve indicates that the U-net model identified more grains with larger equivalent diameters. The results from the DL model shown in *Figure 43*, *Figure 44*, and *Figure 45* are explored in [69]. In terms of the D25 and D90 values from area-weighted distribution curves, the relative error are 0.5 µm and -3.6 µm respectively.

4.3. Application of grain boundary detection model on different magnet samples

The trained custom U-net model for grain boundary detection is evaluated for robustness. Its application has been realized on a homogeneous magnet with different degrees of alignment and a heterogeneous magnet with pole strength difference. Note that these samples are not part of the training or validation set. The predicted results from the trained model are compared to the EBSD measured values statistically using the area-weighted histogram and cumulative area-weighted curves. Additionally, from these curves, D25, D50, and D90 values based on ASTM E 1382 standard are also obtained.

4.3.1. Grain size analysis for anisotropic AP-TP-IP samples

The FeNdB-AP, FeNdB-TP, and FeNdB-IP mentioned in *Table 6* represent AP, TP, and IP test samples, respectively, and they are homogeneous magnet samples with different degrees of alignment. *Figure 46* shows the area and number-weighted grain size distribution for the AP, TP, and IP samples predicted by the trained custom U-net model. Since the EBSD data is close to the manually generated grain size distribution, the results from the trained model were compared to the grain size values obtained from the EBSD, as shown in *Figure 47*. The EBSD grain size distribution always resulted in a comparatively lesser number of grains than predicted by the model.

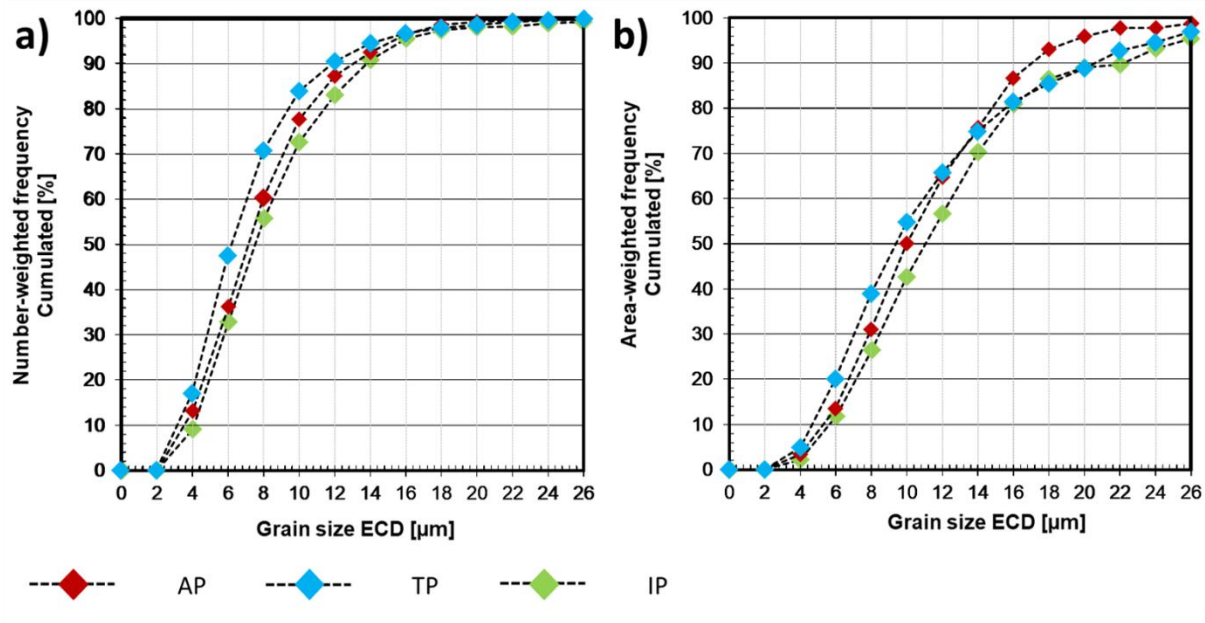


Figure 46: The cumulative grain size distribution obtained from the trained DL model (custom U-net) for the test samples AP, TP and IP is shown. (a) number-weighted frequency and (b) area-weighted frequency.

The D25, D50, and D90 values are obtained from area-weighted plots. *Table 20* shows the relative difference between the predicted and EBSD measured D25, D50, and D90 values. The relative error of 0.5% was lowest for the TP sample, followed by 3.7% for the AP and 5.3% for IP samples. The D90 values show the highest difference. Nevertheless, the average relative error for the D25, D50, and D90 values ranges between 3.1% - 7.7%.

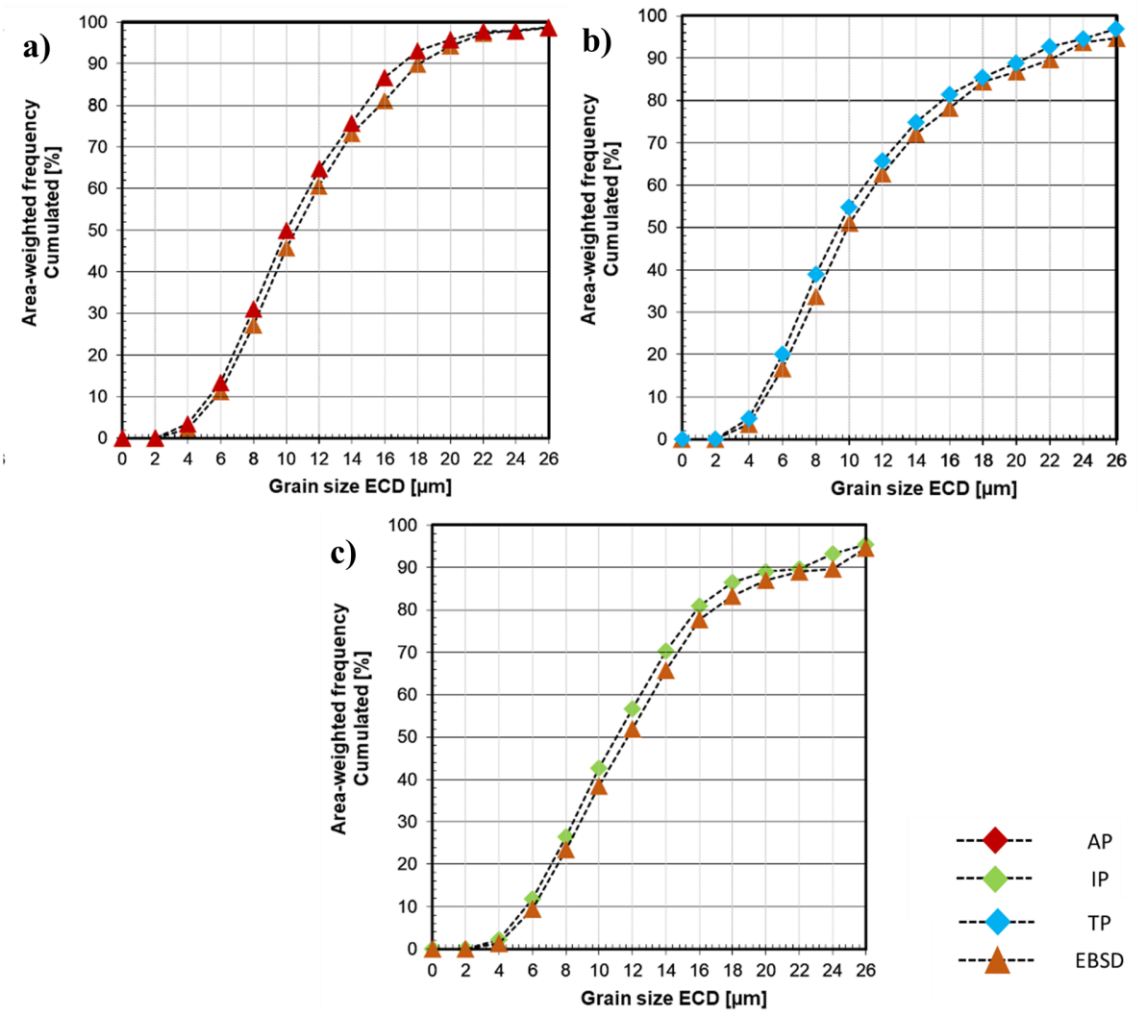


Figure 47: The grain size distribution obtained from the trained DL (custom U-net) model and compared to the measured EBSD data is shown for the AP, TP, and IP is shown. (a), (b), and (c) cumulative area-weighted frequency distribution for the AP, TP, and IP samples, respectively.

Table 20: Comparison between the predicted and EBSD values for the AP, TP and IP samples based on the D25, D50, and D90 values obtained from the area-weighted cumulative grain size distribution plots.

Sample ID	Relative error (EBSD-DL) [%]		
	Cummulative area-weighted		
	D25	D50	D90
AP	1.4	3.8	5.6
TP	5.4	3.2	9.1
IP	2.5	3.4	8.3

Table 20 provides information on the relative error in grain size measurements between the EBSD and DL methods for different percentile values (D25, D50, D90) in area-weighted grain size distributions. The percentages in the relative error column (EBSD-DL) show how much the DL approach diverges from the EBSD results regarding grain size measurements.

4.3.2. Grain size analysis for PSD samples

The FeNdB-PSD-01, FeNdB-PSD-05, FeNdB-PSD-11, and FeNdB-PSD-15 mentioned in *Table 6* represent the PSD-01, PSD-05, PSD-11, and PSD-15 test samples, respectively, and they are from heterogeneous magnet sample. *Figure 48* shows the area and number-weighted grain size distribution for the sample with pole strength difference. In *Figure 49*, comparison between model's prediction and EBSD measurement for area-weighted grain size distribution is shown. The difference in the grain size distribution predicted by the model and measured from EBSD can be obtained by evaluating the D25, D50, and D90 values. The EBSD grain size distribution always resulted in comparatively lesser grains than predicted by the model.

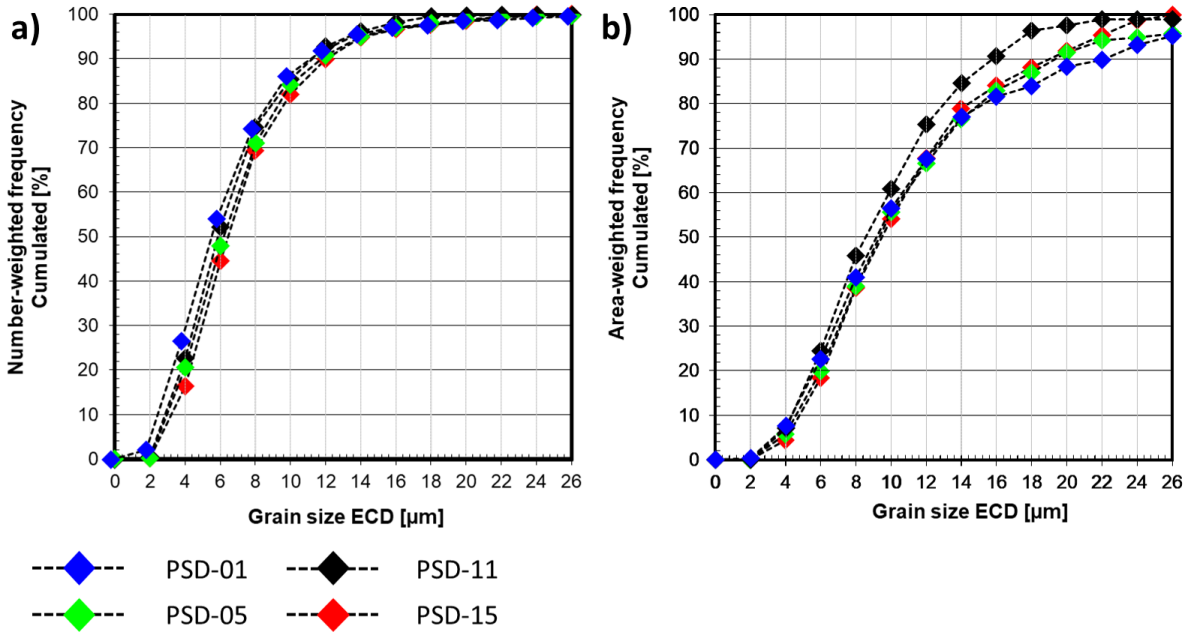


Figure 48: The cumulative grain size distribution obtained from the trained DL model (custom U-net) for the PSD samples (01, 05, 11, and 15) is shown. (a) number-weighted frequency and (b) area-weighted frequency.

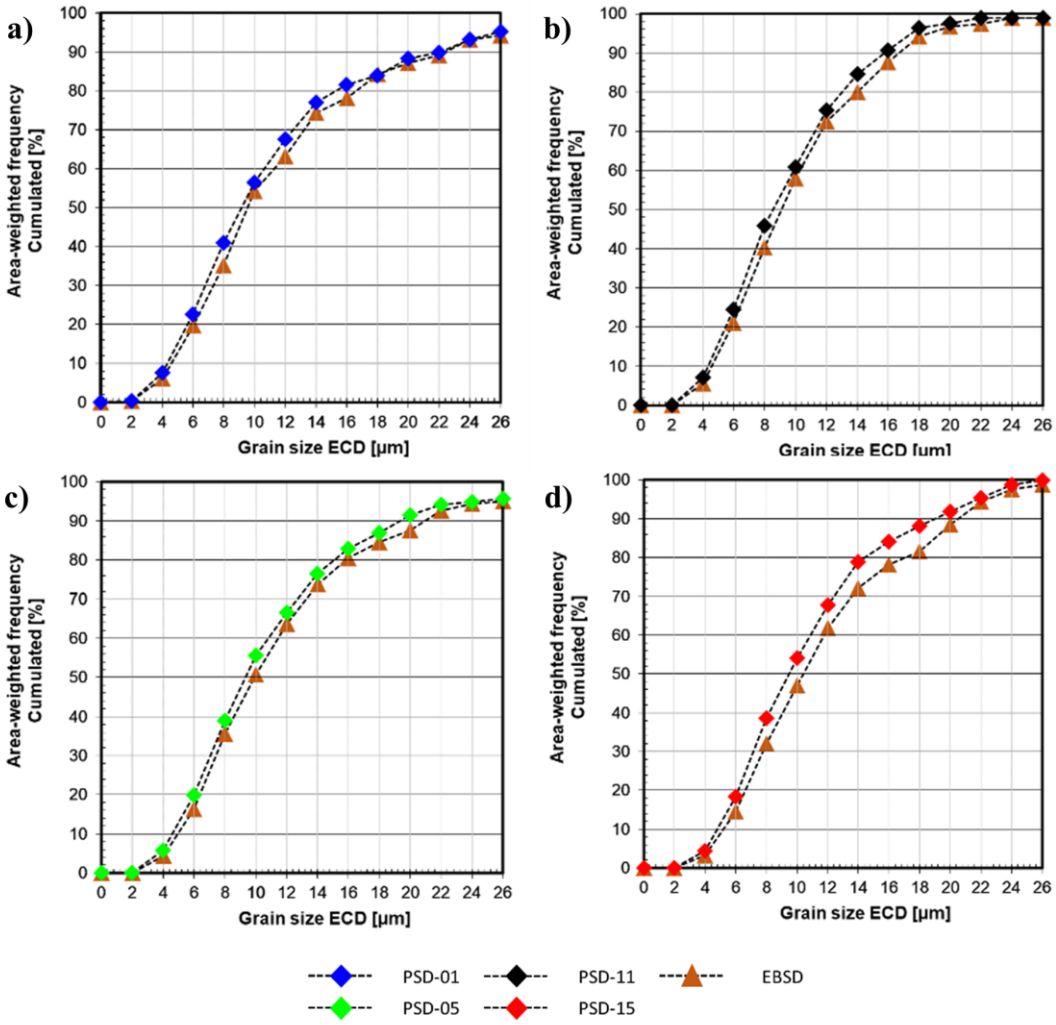


Figure 49: The grain size distribution obtained from the trained DL model (custom U-net) and compared to the measured EBSD data for the PSD samples (01, 11, 05, and 15) is shown. (a), (b), (c), and (d) cumumulative area-weighted frequency distribution for the PSD-01, PSD-11, PSD-05, and PSD-15 samples, respectively.

A similar behavior is observed when the D25, D50, and D90 values are obtained from area-weighted curves. *Table 21* shows the relative difference between the predicted and EBSD measured D25, D50, and D90 values. The average relative error for D25, D50, and D90 values ranges between 4.2% - 5.1%.

Table 21: Comparison between the predicted and EBSD values for the PSD-01, 05, 11, and 15 samples based on the D25, D50, and D90 values obtained from the area-weighted cumulative grain size distribution plots.

Sample ID	Relative error (EBSD-DL) [%]		
	Cummulative area-weighted		
	D25	D50	D90
PSD-01	1.6	4.0	3.3
PSD-05	5.9	1.5	6.7
PSD-11	6.3	3.3	3.4
PSD-15	6.7	5.9	3.2

Table 21 provides information on the relative error in grain size measurements between the EBSD and DL methods for different percentile values (D25, D50, D90) in area-weighted grain size distributions. The relative error percentages quantify the extent of variation between the two methods, with higher percentages indicating more substantial differences. The percentages provide insight into how well the DL approach aligns with the EBSD results for different grain size percentiles.

4.4. Evaluation of the crystallographic grain orientation model for sintered FeNdB magnets

Based on the experiments described in section 3.6, the performance evaluation on the validation dataset with 864 images of extracted grains from the FeNdB-E(V) sample was compared for the trained ML and DL models. ML and DL models are trained on the dataset with only anisotropic samples with a mean ECD of 10 µm. To avoid the DL model from overfitting, the dropout layer with a drop out of 0.6 has been used along with the L2 regularization of $1x10^{-4}$ at each of the last three dense layers added to the pre-trained frozen network. In [216], the initial results from the A-RF model for predicting the θ and ρ are explored.

Among the traditional ML models considered for the experiments, the A-RF model achieved the highest performance for both θ and ρ in terms of the MAE and R_{ccc}. The MAE for θ and ρ from the trained A-RF model are 7.0° and 5.0° respectively. Similarly, the R_{ccc} for θ and ρ are 0.84 and 0.81, respectively. Conversely, the DL models are trained on the higher dimensional feature set, which is automatically extracted by the base layers and captures local and global information in the data. The fine-tuned EfficientNetB0 model achieved the lowest MAE value for both θ and ρ, of 3.5° and 3.0° respectively, among all models. *Table 22* shows the performance of the trained ML and DL model for the validation set. The baseline model with small CNN predicted the θ and ρ values with a higher error than the other DL models.

Table 22. The table provides a comparative performance evaluation of various ML and DL regression models trained to predict grain orientations, measured in terms of θ [°] and ρ [°]. The evaluation includes metrics such as MAE and the R_{ccc}. The models are categorized into ML and DL, with the fine-tuned EfficientNetB0 model demonstrating noticeable accuracy with the lowest MAE and highest R_{ccc} scores [68].

	Models	θ [°]		ρ [°]	
		MAE	**R_{ccc}**	**MAE**	**R_{ccc}**
ML	**A-RF**	**7.0**	**0.84**	**5.0**	**0.81**
	GB	8.7	0.64	8.3	0.76
	XGB	9.2	0.73	5.1	0.80
	MLP	11.4	0.56	10.3	0.59
DL	Baseline model	9.6	0.61	9.1	0.67
	VGG16 fine-tuned	6.9	0.86	5.7	0.84
	ResNet50V2 fine-tuned	6.6	0.88	4.9	0.89
	EfficientNetB0 fine-tuned	**3.5**	**0.93**	**3.0**	**0.94**

The DL model performance for different pre-trained networks on the validation set using the MAE loss and MSE loss curves are shown in *Figure 50* and *Figure 51* for θ and ρ orientations respectively. The MAE and MSE loss curve for the prediction of theta for 350 epochs by

pretrained VGG16, ResNet50V2, EfficientNetB0, and small CNN suggests that compared to the small CNN which is baseline, the decrease in loss value is highest for EfficientNetB0. Similarly, for the ρ prediction, the loss curves suggest that EfficientNetB0 is comparatively better than VGG16 and ResNet50V2. It is to be noted that the models for ρ prediction converge faster when compared to the θ prediction model. The pre-trained EfficientNetB0 is considered for further tuning and evaluations among the DL models.

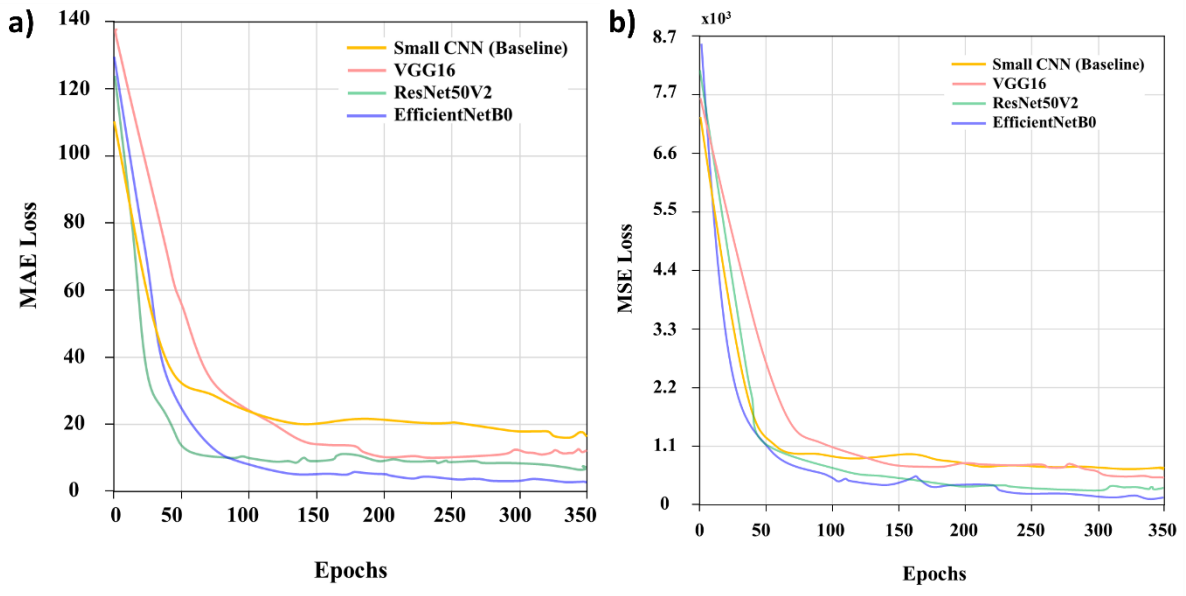


Figure 50: Comparison of the validation loss of the top 3 performing models for the θ prediction compared against the baseline model using the loss functions (a) MAE and (b) MSE.

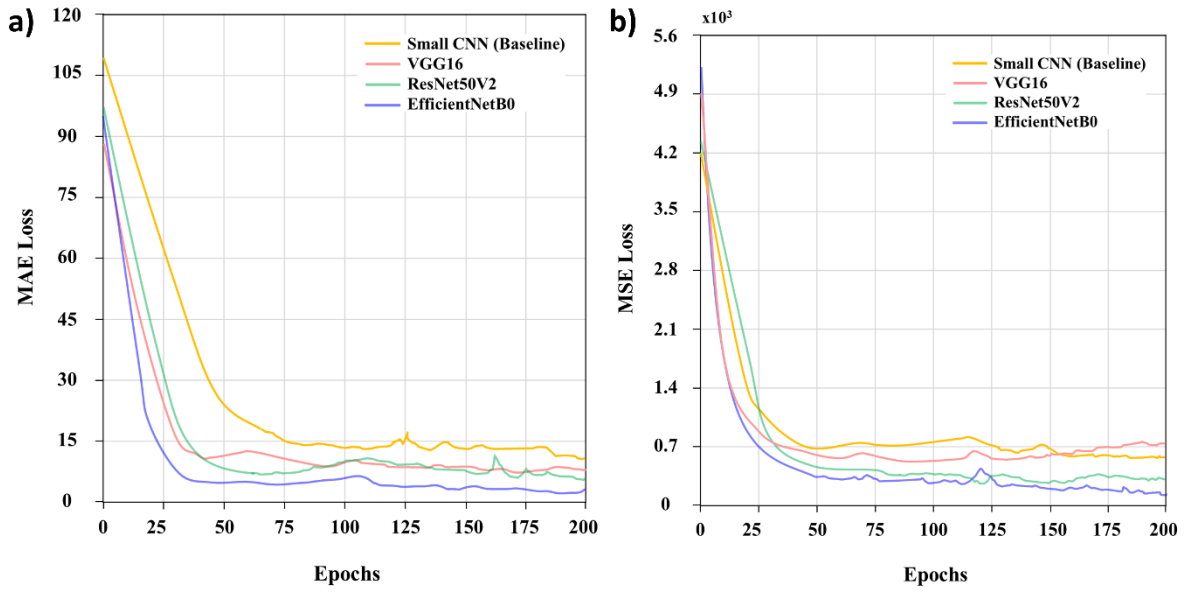


Figure 51: Comparison of the validation loss of the top 3 performing models for the ρ prediction compared against the baseline model using the loss functions (a) MAE and (b) MSE.

The learning curve for the trained and fine-tuned EfficientNetB0 using the MAE loss function for θ and ρ prediction is shown in *Figure 52*. The loss curves for the training and validation set

converge well, and there is no significant gap between the loss curves or stagnation of the curves. This suggests that the model has been generalized well on the dataset.

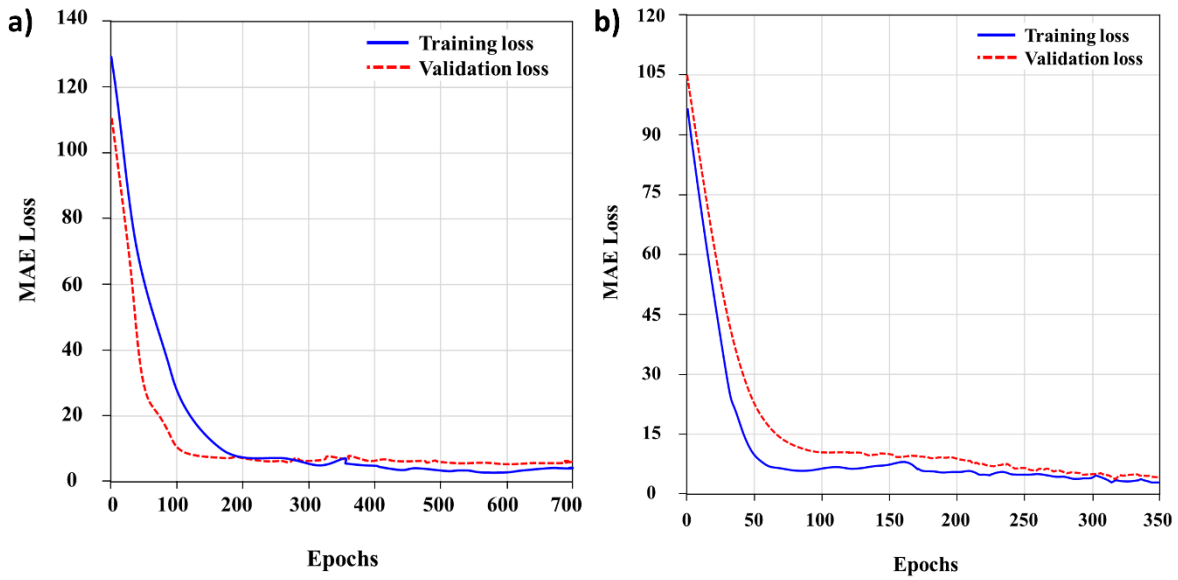


Figure 52: Loss curves of the fine-tuned EfficientNetB0 model on the training and validation set using the MAE loss function for predicting the (a) θ value and (b) ρ value. The convergence of the training and validation loss curves suggests that both models have generalized well on the dataset.

Figure 53 compares the performance of the trained A-RF model and the fine-tuned EfficientNetB0 model on the validation dataset. Notably, the A-RF model displays a substantial error in its predictions for grains with θ values below 60° or exceeding 120°, contrasting with grains ranging from 60° to 120° where its accuracy is relatively better. This suggests that the A-RF model struggles to generalize the complete data distribution while performing well on grains with orientation values primarily within the 60° to 120° range. Moreover, the fine-tuned EfficientNetB0 model resulted in higher prediction error for grains with θ values less than 10° or greater than 170°. Whereas for the ρ prediction performance, *Figure 53(a & b)* highlights the distinct behavior of the two models on the validation set. The A-RF model introduces greater variance, featuring an MAE of 5.0° and an R_{ccc} of 0.81. In contrast, the fine-tuned EfficientNetB0 model resulted in more accurate predictions, suggesting a lesser error and a better alignment with the measured EBSD values. However, the fine-tuned EfficientNetB0 model resulted in a low MAE of 3.0° and a high R_{ccc} of 0.94.

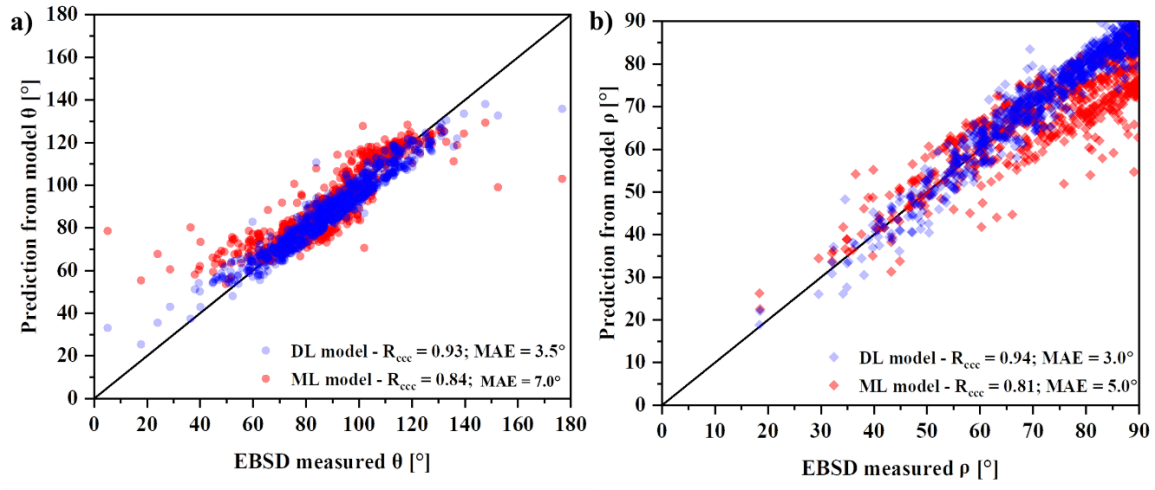


Figure 53: Comparison between the predictions for the validation set from the two models: a trained ML model (A-RF) and a fine-tuned DL model (EfficientNetB0). The predictions are evaluated against the ground truth EBSD measurements. The figures show the predictions from the ML model in red and in blue from the DL model for (a) θ and (b) ρ values.

Moreover, the pre-trained EfficientNetB0 network, with a custom FC layer, undergoes training using the second dataset. This dataset includes the additional grains from quasi-isotropic sample FeNdB-Q-Iso, characterized by an average ECD of 14 µm. Including grains from the quasi-isotropic sample contributes to an augmented variance within the training set. Additionally, introducing these grains enhances dataset balance by including a broader spectrum of grains featuring θ values spanning from 0° to 35° and ρ values below 40°, combined with grains exhibiting diverse orientation values. This addition of grains positively impacts the performance of the model, specifically the accuracy in θ prediction with an improvement of 1.8%. Similarly, an improvement of 2% is observed for the ρ prediction.

4.5. Application of grain orientation prediction model on different magnet samples

From the above results, it is evident that the trained DL model, particularly the fine-tuned EfficientNetB0, outperforms the traditional ML model A-RF in effectively predicting grain orientation within magnet samples derived from KM data. Applying such a developed grain orientation prediction model could involve determining the alignment of grains within the samples. Such alignment holds significance as it is influenced by the sintered magnet samples' processing conditions, which, in turn, affects the magnetic properties such as remanence.

4.5.1. Texture analysis for anisotropic AP-TP-IP samples

As discussed in section 2.1.6, the manufacturing process plays a pivotal role in grain alignment, forming AP, TP, or IP samples. It is important to note that these AP, TP, and IP samples can serve as crucial evaluation benchmarks for assessing the model's performance as they were not part of the training and validation sets. The evaluated samples include FeNdB-AP, FeNdB-TP and FeNdB-IP samples mentioned in *Table 6* and are represented as AP, TP and IP samples, respectively.

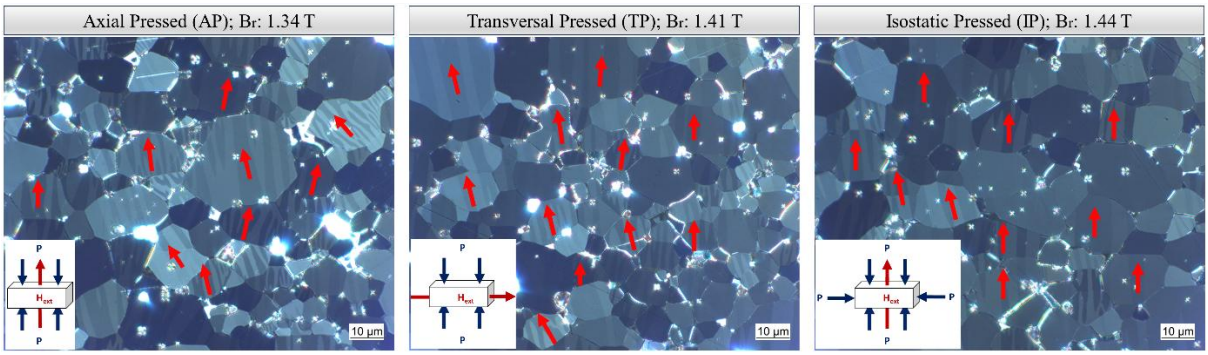


Figure 54: Kerr micrographs of the AP, TP and IP sintered FeNdB permanent magnet sample. The orientation of the grains in each sample obtained from domain structures is depicted in red. The magnetic remanence increases with an increase in the grain alignment from AP to IP due to the pressing (P) and magnetic field direction (H_{ext}). The images are observed at a 1000x magnification.

As shown in *Figure 54*, the KM image of the IP sample appears to have grains with domain patterns that exhibit better alignment, leading to elevated B_r values compared to the AP and TP sample. The KM image shows that the TP sample seems to have better alignment compared to the AP sample. This insight can be derived by obtaining the orientation of each grain in the sample, where samples with narrower distribution or a lower standard deviation in grain orientation values (θ and ρ) tend to exhibit a more pronounced grain alignment. Consequently, such samples tend to display higher remanence and higher grain alignment.

Moreover, one of the means to evaluate the statistical reliability of the predictions from the trained DL model is to establish a correlation between the predicted grain orientations within each sample and the magnetic properties, like magnetic remanence using Equation (5).

The statistical dispersion of grain orientation in each sample can be quantified using various methods, including standard deviation (SD) or interquartile range (IQR). The predicted grain orientations distribution when obtained using number and area-weighted approaches across the AP, TP, and IP samples, considering θ values, demonstrates a Gaussian-like distribution with 10º orientation class or bin width. However, for the ρ values, a right-skewed distribution is observed, as shown in *Figure 55(a & b)*. For ρ values, 5º orientation class or bin width has been considered to ensure that the number of classes are same as for θ value distribution. The extent of data spread holds statistical significance to grain misorientation.

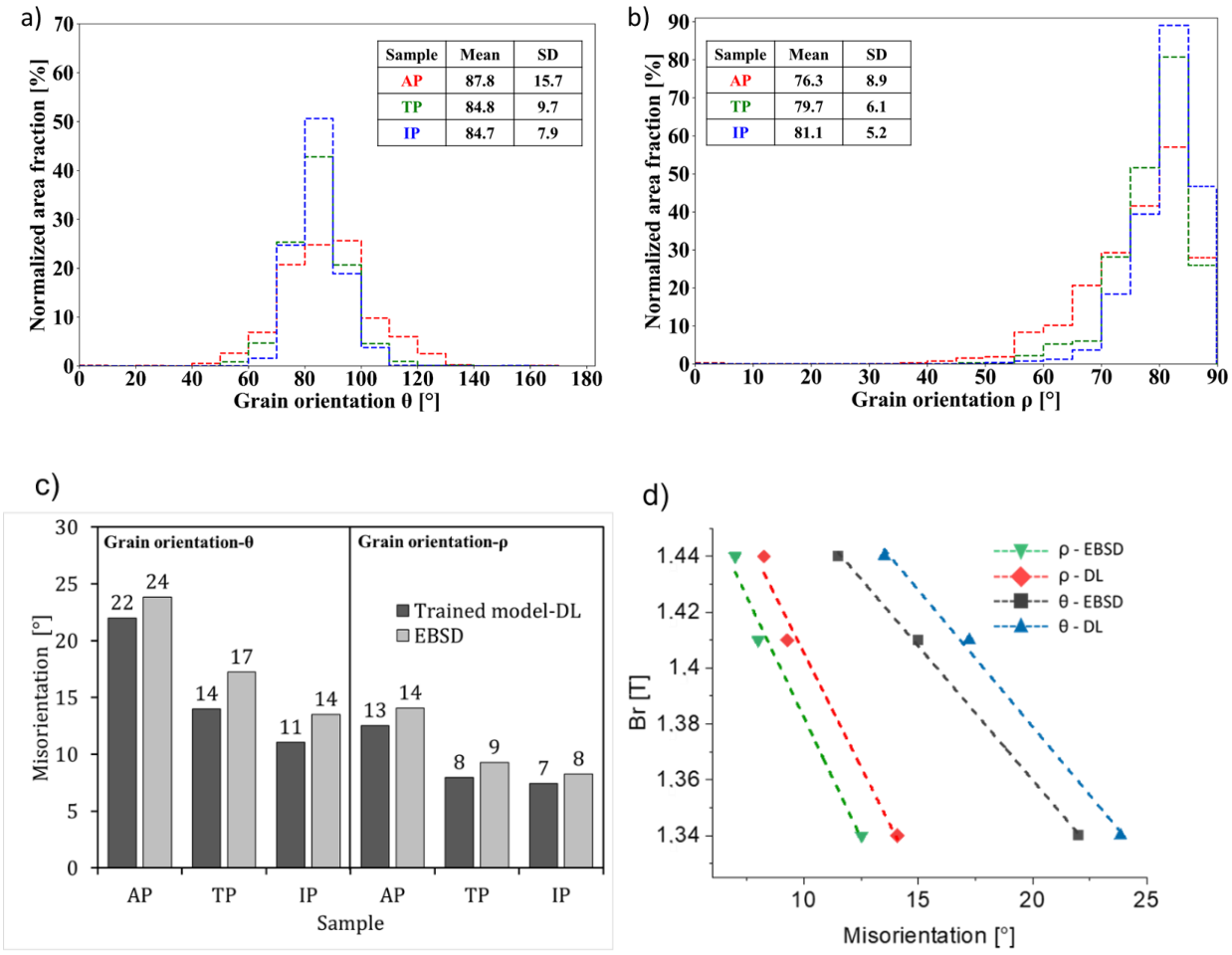


Figure 55: The predictions from the fine-tuned DL regression model evaluated for reliability based on statistics obtained from correlation of the magnetic remanence of the test samples with different degree of alignment. (a) and (b) area-weighted histogram showing the spread of the orientation θ and ρ predicted for grains of AP, TP and IP samples, respectively, (c) misorientation calculated as a function of the interquartile range of orientation histograms for predicted and measured EBSD, and (d) linear correlation between the magnetic remanence and misorientation in AP, TP and IP samples obtained from trained model predictions and measured EBSD [68]. Note that for (a) and (b) the class or bin width is 10º and 5º respectively.

The prediction error for the AP, TP and IP sample in terms of the grain misorientation values lies between 2º - 3º for θ values and for the ρ values it is 1º when calculated from the number-weighted grain orientation distribution. Additionally, when calculated using the area-weighted distribution, the prediction error for AP, TP and IP sample in terms of the grain misorientation values lies between 3.3º - 4.1º for θ values and for the ρ values it is between 1.1º-2.7º. Further, as shown in *Figure 55(c & d)*, a similar correlation is observed between the magnetic remanence and the misorientation values derived from the number-weighted orientation histogram of θ and ρ, whether through EBSD or the trained DL approach. Notably, the AP sample shows lower magnetic remanence than the TP and IP samples. In *Figure 55(d)*, each of the four lines corresponds to AP, TP, and IP samples, respectively, represented from left to right by three misorientation points. The degree of misorientation is linked to the data spread, quantified by the IQR values extracted from the distribution of the number-weighted histograms of deviation of orientation from the axis of easy magnetization for θ and ρ values within each sample.

4.5.2. Texture analysis for anisotropic PSD samples

In *Figure 56*, the PSD sample shows a pole strength exceeding 20% and a magnetic dipole moment misalignment angle α_Z ranging from 0° to 1°.The PSD sample has both convergent and divergent sides, and the pole strength gradually reduces from the convergent to the divergent region. The PSD samples here include four distinct positions, each characterized by varying global misalignment. Specifically, PSD-01 and PSD-05 are positioned closer to the c-axis, while PSD-11 and PSD-15 are situated further away from the c-axis. The appearance of the domain structures also differs across these four positions within the sample, enabling the quantification of misorientation.

The FeNdB-PSD-01, FeNdB-PSD-05, FeNdB-PSD-11 and FeNdB-PSD-15 mentioned in *Table 6* represents the PSD-01, PSD-05, PSD-11 and PSD-15 test sample respectively. Similar to the application to the samples with different degree of alignments in section 4.5.1, the trained DL model can be applied to samples with varying textures. Of the four positions in the PSD sample, two are convergent and the other two are divergent regions. Grains in the convergent region have domain patterns oriented toward the perpendicular c-axis. The divergent region has grains oriented away from the perpendicular c-axis. PSD-01 and PSD-11 are two positions in convergent regions of which PSD-11 has a higher degree of convergence than PSD-01. Similarly, PSD-05 and PSD-15 are two positions in divergent regions, PSD-15 having a higher degree of divergence than PSD-05. This can be visualized in *Figure 56*, which shows the positions of selected regions in the PSD sample.

The extent to which the grains in the sample converge or diverge can be obtained from the θ component of the grain orientation. As discussed in section 3.2.1, θ value depends on the orientation of the domain patterns in the grains and brightness. The measure of the mean from the θ value histogram distribution for a given sample can provide information on whether the sample has grains with convergence or divergence in nature. *Figure 57(a-d)* shows the θ and ρ values obtained from the area-weighted histogram plot for all four PSD samples. The mean and standard deviation for each sample is also shown. Note that the difference in standard deviation for all four samples for both θ and ρ prediction is very low, as they represent the degree of alignment as discussed in section 4.5.1. Since the PSD sample considered here is axially pressed, and four positions considered for evaluation are the AP sample's sub-sections, the distribution plot's standard deviation is nearly the same.

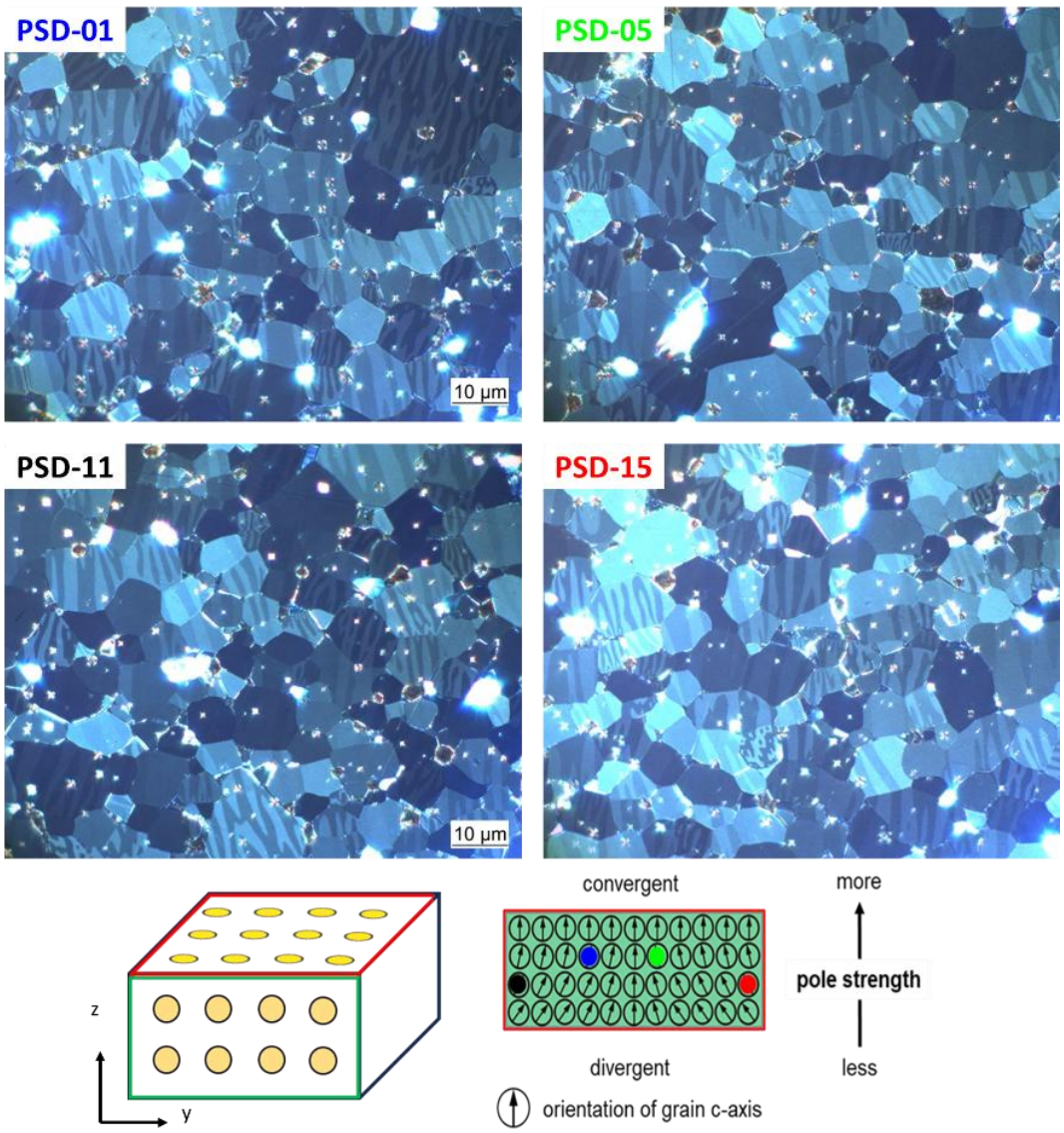


Figure 56: Example KM images of the sintered FeNdB magnet with pole strength difference observed at 1000x magnification. PSD-01 and PSD-11 have grain orientation pointed towards the c-axis. PSD-05 and PSD-15 have grain orientation pointing away from the c-axis. PSD-01 and PSD-05 are from the convergent region, whereas PSD-11 and PSD-15 are from the divergent region. (Figure unpublished Choudhary & Tvrtko, with permission from [296])

However, the mean values for the density plot for PSD-01 and PSD-11 are less than 90° and more than 90° for PSD-05 and PSD-15. The c-axis lies at 90°, and a mean value of less than 90° indicates that the test sample lies in the convergent region of the PSD sample. Similarly, the mean value of more than 90° represents that the test samples lie in divergent regions. The mean value of PSD-11 is less than PSD-01, suggesting that PSD-01 has higher degree of convergence due to a greater number of grains with domain patterns oriented towards the c-axis. On the other hand, PSD-15 has a higher degree of divergence due to the presence of a greater number of grains oriented away from the c-axis when compared to PSD-05.

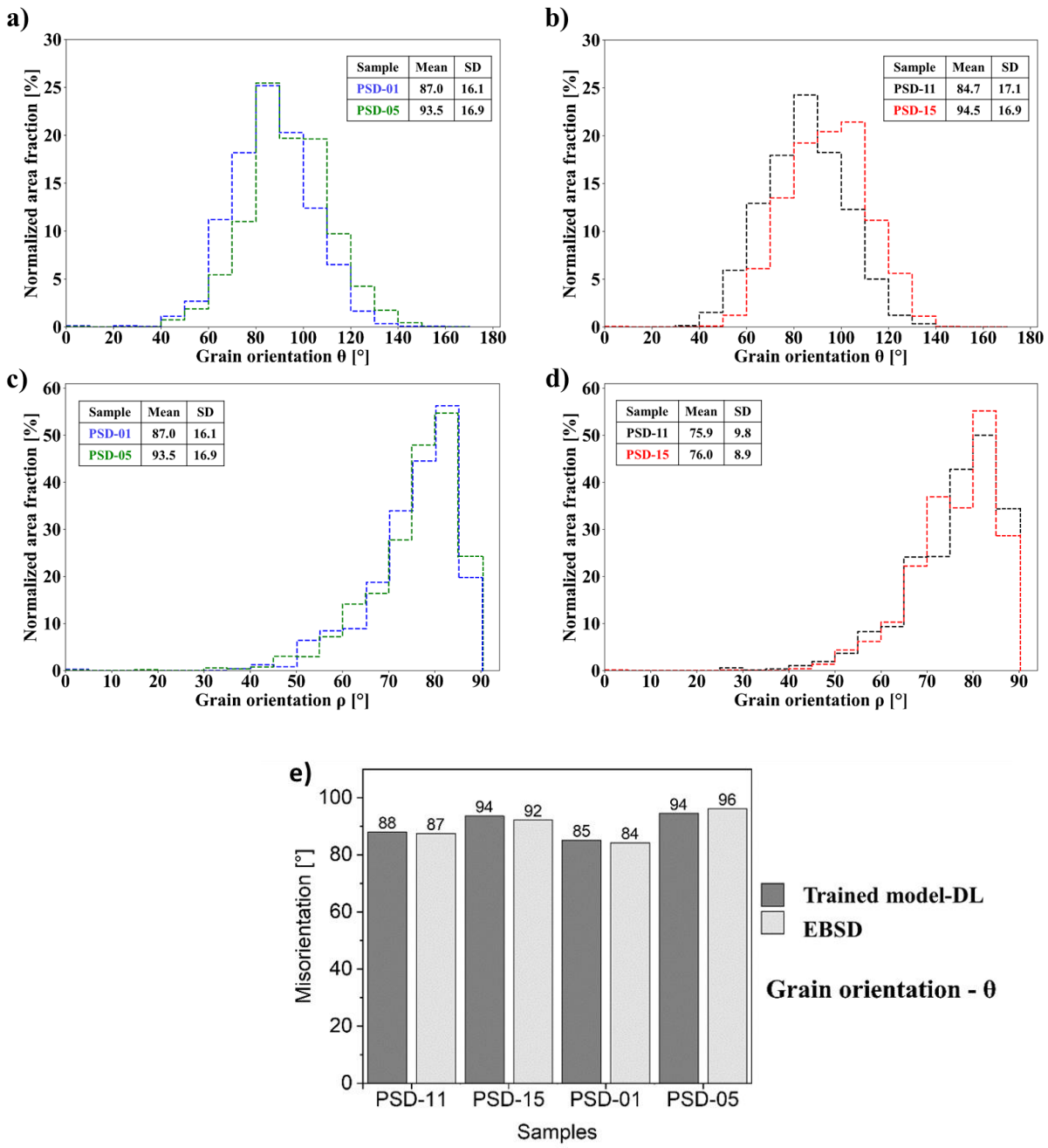


Figure 57: The area-weighted histogram plot of the predicted θ and ρ values of the trained DL model for PSD test samples showing the degree of convergence or divergence as a function of the measured mean values. (a) predicted θ for PSD-01 and PSD-05, (b) predicted θ value for PSD-11 and PSD-15, (c) predicted ρ value for PSD-01 and PSD-05, and (d) predicted ρ values for PSD-11 and PSD-15. (e) The calculated misorientation as a function of the mean value of the number-weighted cumulative density histogram plot for θ obtained from the prediction of the trained model and EBSD, respectively for all PSD samples is shown in the figure. Note that the class or bin width for plots in (a) and (b) are 10º and 5º respectively.

Further the performance evaluation of the predicted values by fine-tuned EfficientNetB0 model against the EBSD values is carried out. The grain misorientation values can be statistically linked to the mean value obtained from the distribution plots of θ. *Figure 57* shows the misorientation values for θ orientation predicted by the trained model for all four positions in the PSD sample and compares it against the statistics obtained from EBSD measurement. For ρ values, 5º orientation class or bin width and 10º for θ values has been considered to ensure

that the number of classes are same as for both distribution curves. From the EBSD measurement data, the mean for the distribution plot of θ is obtained for all four positions in the PSD sample. The difference between EBSD and DL model predicted misorientation values is less than 3.3% for θ values and less than 1.7% for ρ values.

The results of Goniometer test for isotropic same is also obtained such that for change in brightness or color of grain against rotated in steps of 15º transverse Kerr effect or θ value is noted for 10 grains. Each grain has different domain structure, and the colorfulness variation for each is obtained as shown in *Figure 57.*

5. Discussion

5.1. Evaluation of the intrinsic property prediction model from chemical compositions

In this section, the results of the experiments related to the intrinsic property prediction task using the ML regression model trained on chemical compositional features are discussed. Firstly, the challenges associated with the dataset for materials property prediction are addressed. Subsequently, the rationale of opting the tree-based models instead of ANNs were explored and the effect of repeated datapoints within the dataset were investigated. Later, a comprehensive performance evaluation of the developed regression models for mass density, Curie temperature, and saturation magnetization at room temperature is provided in detail with the support of feature importance plots. Finally, the impact of aggregating the repeated phases and the presence of correlated features in the dataset for T_c and M_s models are examined.

During the process of designing new alloys, a crucial step involves identifying the alloy composition within the vast composition space. This composition directly influences the properties exhibited by the resulting alloy. Hence, it is essential to make a well-informed choice of composition during alloy design to ensure that the intended properties match the requirements of the application environment. While material experiments are the most precise means of measuring the properties of newly formulated alloys, they are challenged by their time-intensive nature and the costs associated with their complexity. Furthermore, some inherent errors in experimental procedures necessitate repeated trials to ensure result accuracy, thereby leading to both higher cost and time invested in the design process. Addressing this challenge, there is a need to develop a predictive model that establishes the relationship between composition and properties. Such a model aims to reduce costs while maintaining the accuracy of alloy property predictions.

The major source of the dataset to train the property prediction model is the literature reported values in the form of tables, graphs, or physical formulas. This also stands true for the mass density, T_c, and M_s prediction models which are trained on data extracted from the 55 literatures mentioned in section 3.2.1 that includes data from the last 40 years. Most published papers narrowly focus on a handful of materials and only potentially important properties leading to the biased distribution of labels within material datasets [154] and this is also observed for the literature reported values for T_c where 64% of the data belong to ternary alloy systems and less than 10% for the senary and quinary alloy systems. For the M_s dataset, 61% of the data belong to quaternary alloy systems and less than 20% for the remaining. This leads to the domain and selection bias during the data splitting step needed to develop a well generalized model. Similar observations were observed with the presence of different types of rare earth elements in phases belonging to the T_c and M_s prediction models such as phases with light rare earth substituent account for 50% to 63% and less than 20% for heavy rare earth substituent. The systematic errors, the different experimental conditions and the underlying assumptions in the datasets from experimental test results make the materials dataset different from the image datasets available in massive numbers. Consequently, there is a gap to be filled for the missing information in material datasets which are not being experimentally measured or predicted using physical models due to their limitations. These obstacles in the material datasets limit the

efficacy of machine learning approaches in predicting the properties of alloy materials. Therefore, developing a universally applicable predictor model that covers a vast class of potential alloy candidates is practically very demanding and involves a high risk of bias.

In addition, the dataset to predict the property of the material is a tabular dataset that belongs to the structured dataset category. Unlike computer vision tasks, there is no standard benchmark for the tabular data. While there are general machine learning benchmarks, none of them specifically deals with the tabular data as far as the knowledge extends. For instance, benchmarks like CC-18 [297], CC-100 [298], and the AutoML Benchmark by Gijsbers et al. [299] incorporate tabular data, but they also include images and artificial datasets, which might explain why they have not been utilized in tabular deep learning research. Using DNN regression, diverse and scattered experimental data from the literature can be consolidated into straightforward quantitative expressions that define specific material properties based on chemical composition and processing parameters. These derived mathematical expressions can be effectively applied for the prediction of material properties, the development of new alloys, and comparing with experimental results. While large datasets are optimal for DNN, the issue with limited datasets can still use DNN, especially in combination with pre-training, when obtaining extensive datasets is not feasible. In material science, small datasets are common as seen for density, T_c and M_s prediction models here, which has less than 500 datapoints, and a smaller number of input variables which are less than 100 most of the tasks often characterizes the challenges compared to those encountered in computer recognition tasks involving image datasets [300], [301].

Therefore, developing a generic property prediction model that fits well on the multiple materials class poses significant difficulties. Considering the above obstacles, the focus for the property prediction has been limited to Curie temperature and saturation magnetization at room temperature for the 14:2:1 hard magnetic FeNdB phase. The developed models have been trained to fit well on the ternary, quaternary, quinary, and senary compositions of 14:2:1 phases and are limited to the elemental feature space of 37 elements for T_c prediction and 28 elements for M_s prediction. The focus has been to develop a data-centric model that is not overfitting despite being trained on small datasets. This has been achieved using bias-variance trade-off by combination of the diverse learning models to minimize the inductive bias, appropriate cross-validation technique, and collection of datasets from experimental results in literatures that includes duplicates due to difference in reported target property values for the same phase to compensate for domain and selection bias.

Furthermore, due to the lack of benchmark tabular data, the inability of ANNs to learn irregular target functions due to rotation invariance and presence of many uninformative features in the tabular data, tree-based models have often outperformed deep learning models [110].

Influence of the repeated datapoints in the property prediction dataset

The dataset for the mass density, Curie temperature and saturation magnetization of the 14:2:1 phase also had some of the challenges mentioned above. Of which the presence of repeated phases as a result of published experimental values for the same phase by different researchers over the years. The duplicates or repeated data differs in the case of structured and unstructured datasets. In case of structured data like tabular datasets similar to the property prediction in

magnet samples, the input features which is the number of elemental chemical composition is the same but the target value (property) is different as discussed in section 3.2.1. For the image dataset, this is uncommon and also problematic as they are considered as edge cases or outliers and are generally avoided. However, the presence of duplicates in tabular datasets is common and the decision to include or exclude them from the dataset depends on the the task, the data distribution, and the ML learning model. Further, one primary assumption while creating a dataset for training the ML model is that it should resemble the real-world dataset. Therefore, considering the domain and selection bias and the measurement error from the different equipment used for getting the material's property value, the dataset includes duplicates. However, the presence of outliers in the dataset that has been detected were excluded from the training set. *Figure 58* shows some of the repeated phases in the dataset used for the T_c model. The deviation of the reported T_c values across different literatures for the same phase is primarily within 4% when calculated from its mean. However, there are also some literatures which have the reported values which account for more than 4%.

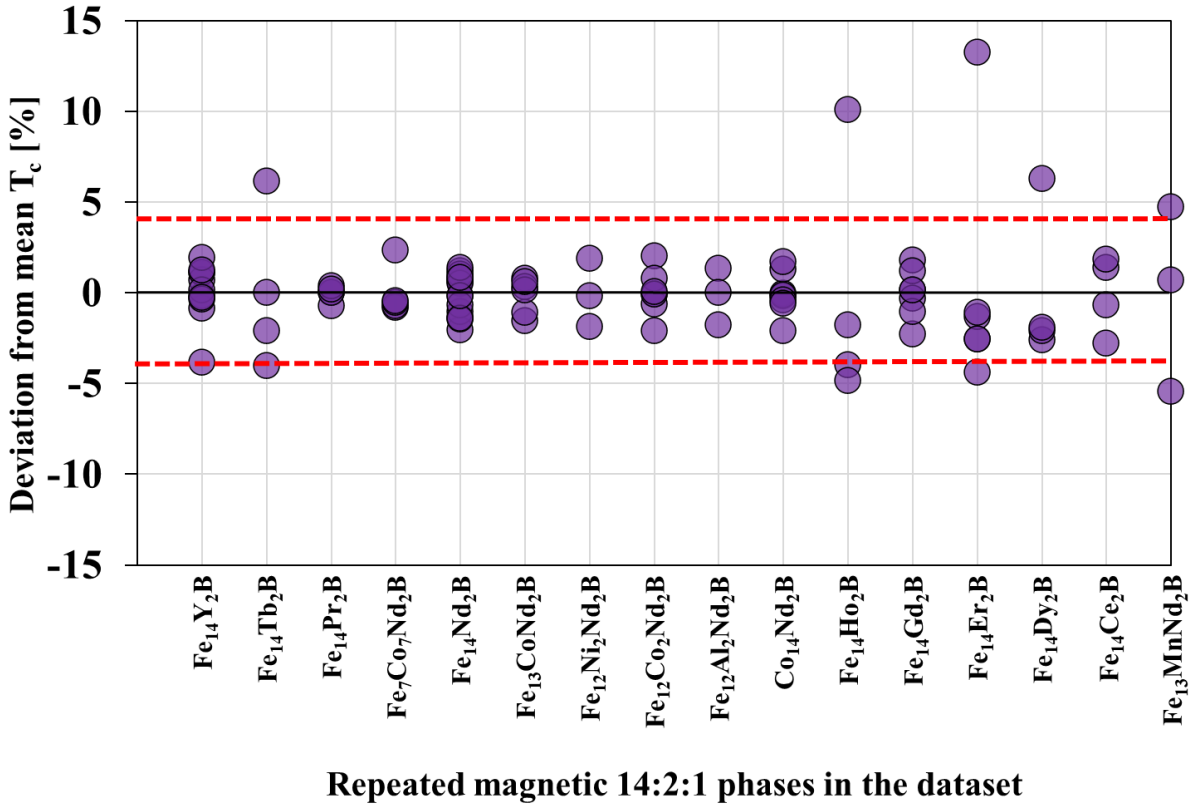


Figure 58: Plot shows the deviation of the repeated magnetic 14:2:1 phases in the T_c dataset. The deviation in % from the mean value for repeated phases is within [-15, 15]. Note that plot shows phases which are repeated more than twice in the dataset.

For the T_c use case, ten datapoints from the repeated phases category were identified as outliers. They are $Fe_{14}Dy_2B$, $Fe_{14}Er_2B$, $Fe_{14}Ho_2B$, $Fe_{14}Tb_2B$, $Fe_{12.6}Co_{1.4}Nd_2B$, $Fe_{12.6}Ni_{1.4}Nd_2B$, $Fe_{13.3}Cr_{0.7}Nd_2B$, $Fe_{13.3}Cu_{0.7}Nd_2B$, $Fe_{13.3}Mn_{0.7}Nd_2B$ and $Fe_{13}MnNd_2B$. *Figure 58* shows five of the detected outliers that have a deviation beyond ±4% from the mean. The remaining five are not shown in *Figure 58* as they appear only twice in the dataset. Each phase was individually analyzed to assess deviations in their reported T_c values and to understand the underlying reasons for the exclusion of certain datapoints. The analysis revealed significant deviations in

T_c values for repeated phases when compared to the literature reported data. The datapoints from the repeated phases $Fe_{14}Dy_2B$, $Fe_{14}Er_2B$, and $Fe_{14}Ho_2B$ were consistently identified as outliers, deviating by 50 K to 80 K from the mean value across the reported values by the literatures for repeated phases. For instance, $Fe_{14}Dy_2B$, which appears four times in the dataset with T_c values ranging from 538 K to 589 K, shows a 6.3% deviation from the mean of the T_c value of 538 K reported by Wallace et al. [249], suggesting potential calibration issues. Similarly, six entries for $Fe_{14}Er_2B$ exhibit deviations of 70 K to 80 K, with Wallace et al. [249] reported a T_c of 472 K, 13% lower than the mean value of the broader range (472 K to 550 K). Likewise, $Fe_{14}Ho_2B$, with four entries ranging between 499 K and 582 K, shows a similar trend. The T_c value of 499 K reported by Wallace et al. [249] deviates by 10% from the mean. $Fe_{14}Tb_2B$ appears 4 times in the dataset with the highest deviation of 6% from mean being the T_c value of 582 K reported by Wallace et al. [249]. The reported T_c values for $Fe_{14}Tb_2B$ from other literatures are 620 K [254], 633 K [233] and 645 K [130].

Additionally, the T_c values for phases such as $Fe_{12.6}Co_{1.4}Nd_2B$, $Fe_{12.6}Ni_{1.4}Nd_2B$, $Fe_{13.3}Cr_{0.7}Nd_2B$, $Fe_{13.3}Cu_{0.7}Nd_2B$ and $Fe_{13.3}Mn_{0.7}Nd_2B$, as reported by Lin et al. [255], consistently exhibit values 30 K to 50 K higher than those reported in other literatures [46], [229], [235], [240]. Notably, $Fe_{13}MnNd_2B$ demonstrates a large variation in T_c values, with reported values of 449 K, 468 K, and 497 K across sources [24], [41], [243] respectively. Consequently, the T_c values for these phases reported by Lin et al. [255] were excluded from the dataset due to significant inconsistencies. The higher deviations observed in some repeated phases may be due to calibration discrepancies, differences in sample preparation, measurement conditions, or potential systematic biases. The exclusion or labeling of these data points as outliers was based on their significant deviations from the mean and inconsistencies across different literature sources.

Similarly to the T_c, the dataset for the M_s prediction model is diverse, including phases from various alloy systems and phases featuring different rare earth elements. The dataset includes repeated phases with varying M_s values in [μ_B/f.u.] despite having the same chemical composition. Additionally, significant deviations in ambient temperature during the experimentation contribute to the observed discrepancies. Literature sources often report M_s values obtained through saturation law and measurements conducted using vibrating sample magnetometers or extraction magnetometers. These measurements were taken at different external magnetic fields, ranging from 1.0 T to 8.0 T, while the ambient temperature fluctuated between 295 K and 305 K.

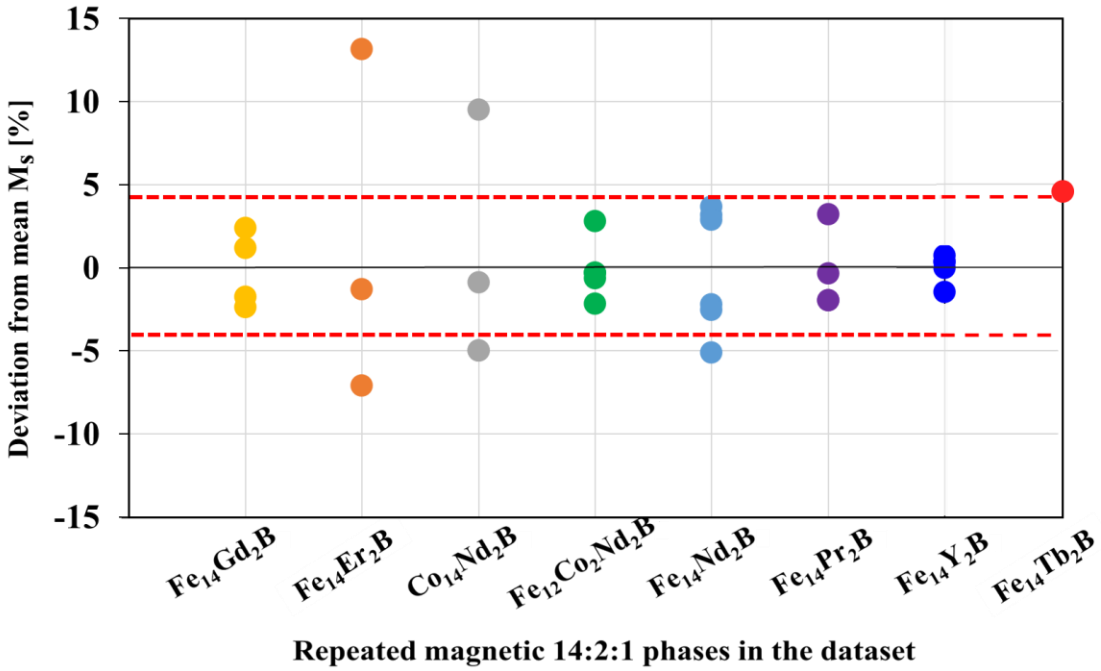


Figure 59: Deviation of the repeated magnetic14:2:1 phases in the M_s dataset. The deviation in % from the mean value for repeated phases is within [-15, 15]. The phase $Fe_{14}Er_2B$ has two reported values which are not shown in the plot but have the deviation of 23% [249] and -28% [130]. For $Fe_{14}Tb_2B$, two reported values deviation 17% [130] and -22% [233] are not shown in plot. Note that plot shows the repeated phases from the dataset with comparatively major deviation of more than 1%. The deviation from mean on y-axis is restricted to ±15% for better visualization of the deviations in each repeated phases.

Figure 59 shows the M_s in [μ_B/f.u.] for repeated magnetic 14:2:1 phases found in both the training and the test sets, including seven distinct phases. While there are additional repeated phases within the dataset beyond the seven shown in *Figure 59*, the focus here is on phases that appear multiple times in varying quantities. Specifically, the dataset includes $Fe_{14}Er_2B$, which appears six times, while $Fe_{14}Gd_2B$, $Fe_{14}Pr_2B$, and $Co_{14}Nd_2B$ are repeated four times each. $Fe_{14}Y_2B$ is repeated five times and $Fe_{14}Nd_2B$ is present in six instances. These repeated phases exhibit a range of standard deviations, with $Fe_{14}Er_2B$ showing the highest deviation of 5.6 [μ_B/f.u.], while $Fe_{14}Y_2B$ exhibits the least deviation of 0.4 [μ_B/f.u.]. The shared characteristic of these phases is their same chemical composition (input feature) despite having different M_s values, reflecting the influence of diverse literature sources. Calculating the mean value for each repeated phase allows us to quantify the deviation of each instance from its respective mean value. Notably, the $Fe_{14}Er_2B$ phase displays a maximum deviation of -28%, while the four repeated phases exhibit deviations lower than 5%.

For the M_s use case, five datapoints from the repeated phases category were identified as outliers. They are $Fe_{14}Tb_2B$, $Fe_{14}Er_2B$, $Co_{14}Nd_2B$, $Fe_{12}Cr_2Nd_2B$ and $Fe_{14}Nd_{1.56}Ce_{0.44}B$. Each phase was analyzed individually to assess deviations in their reported M_s values and the underlying reasons for excluding certain data points. For $Fe_{14}Tb_2B$, the M_s values reported range from 12.2 [μ_B/f.u.] by Abache et al. in [130] to 18 [μ_B/f.u.] by Abache et al. in [233], with an intermediate value of 14 [μ_B/f.u.] reported by Burzo et al. [41]. The measurement by Abache et al. [130] was conducted at 10kOe, while Abache et al. [233] recorded a significantly higher value at 85kOe. The deviation from the mean value exceeds 5%, with the least deviation being 4.9% and the highest reaching -22%. Due to significant inconsistency and the influence of the

higher measurement field (85kOe), the value of 18 [μ_B/f.u.] reported by Abache et al. in [233] was excluded from the dataset. For $Fe_{14}Er_2B$, the M_s values range between 12 [μ_B/f.u.] reported by Abache et al. in [233] and 19.6 [μ_B/f.u.] reported by Wallace et al. [249]. Among the six repeated datapoints, the highest deviation from the mean was 23% in the value from Wallace et al. [249], measured at 20kOe, while a deviation of -28% was noted for Abache et al. in [130]. The least deviation of 0.7%, was also reported by Abache et al. in [130] but at 10kOe. Given the large deviations from the mean, the values reported by Wallace et al. [249] and the outlier measurement by Abache et al. in [130] were excluded. In the case of $Co_{14}Nd_2B$, the reported M_s values range from 20 [μ_B/f.u.] in [130] to 23.2 [μ_B/f.u.] in both [249] and [237]. The largest deviation of 9.5% was observed in the measurement by Abache et al. in [130], which was conducted at 10kOe. Meanwhile, a deviation of -5% was observed for the values from Wallace et al. [249] and Jiang et al. [237], both measured at approximately 20kOe. Due to the significant deviation and inconsistency, the M_s value of 20 [μ_B/f.u.] reported by Abache et al. in [130] was excluded. For $Fe_{12}Cr_2Nd_2B$, two M_s values were reported: 17.6 [μ_B/f.u.] by Abache et al. in [124] at 10kOe and 21.9 [μ_B/f.u.] by Kowalczyk et al. in [125] at 26kOe. The difference in ambient temperature (275 K to 350 K) between these measurements and the difference in applied fields probably contribute to the observed variation. Despite this, the values do not exhibit large deviations and were therefore retained in the dataset.

Finally, for $Fe_{14}Nd_{2-x}Ce_xB$ (specifically x=0.44), Susner et al. [253] reported an M_s value of 32.3 [μ_B/f.u.] at 300K and 60kOe, while Abache et al. in [233] reported a significantly lower value of 27 [μ_B/f.u.] for $Fe_{14}Nd_{1.5}Ce_{0.5}B$. Additionally, Susner et al. in [253] reported M_s values of 31.7 [μ_B/f.u.] for Ce = 0.18 and 30.5 [μ_B/f.u.] for Ce = 0.76 at 300K and 60kOe, showing a more consistent trend with slight variations in composition. The sharp dip in M_s reported by Abache et al. in [233] for Cr addition was inconsistent with these trends. Consequently, the M_s value of 27 [μ_B/f.u.] reported by Abache et al. in [233] was excluded from the dataset for this phase. The exclusions were based on significant deviations from the mean and inconsistencies due to differences in experimental conditions such as field strength, temperature, or composition. The M_s values excluded from the dataset were the reported ones from [233] for $Fe_{14}Tb_2B$, the values reported in [249] and [130] are for $Fe_{14}Er_2B$, the values in [130] is for $Co_{14}Nd_2B$, and for the $Fe_{14}Nd_{1.56}Ce_{0.44}B$ values in [233]. These exclusions ensure the reliability of the training dataset by minimizing outlier influence.

Figure 60 shows the M_s values in [μ_B/f.u.] for the repeated magnetic 14:2:1 phases in both the training and test sets, focusing on six distinct phases. It should be noted that the M_s dataset has additional repeated phases beyond the six highlighted in *Figure 60*. Among these repeated phases without outliers, the highest standard deviation is observed for $Fe_{14}Nd_2B$ at 2 [μ_B/f.u.], while the least deviation is found in $Fe_{14}Ho_2B$ at 0.1 [μ_B/f.u.]. Consequently, the prediction from the trained model for these repeated phases remains the same, as indicated by the yellow highlights in *Figure 60*. In cases where the reported M_s value range for a particular phase is substantial and results in clusters, the model learns the optimal M_s value by leveraging the majority vote from the cluster that includes a higher number of repeated phases.

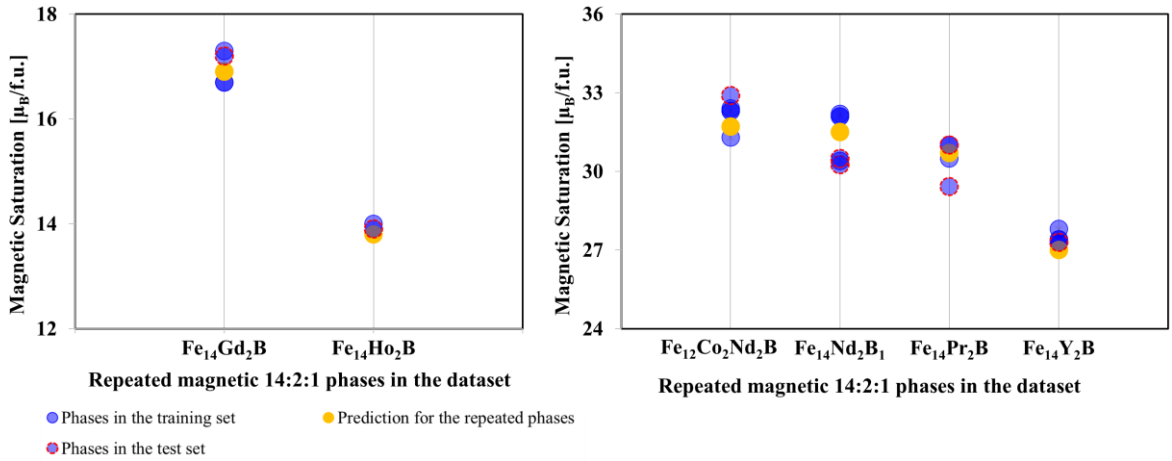


Figure 60: Example of the reported magnetic saturation [μ_B/f.u.] at room temperature for the repeated magnetic 14:2:1 phases in the dataset collected from the various literature sources. The repeated phases are split across training and test set, and the predicted magnetic saturation for these phases are included in the plot in yellow color.

In *Figure 60*, the phase $Fe_{14}Pr_2B$ appears four times, with three literature sources reporting M_s values exceeding 30 [μ_B/f.u.] and forming a cluster, while another source reports a value lower than 30 [μ_B/f.u.]. Using a majority vote approach, the ML model has effectively learned the optimal M_s value from the cluster with reported values that exceed 30 [μ_B/f.u.]. Consequently, the predicted value for $Fe_{14}Pr_2B$ is closely aligned with 30 [μ_B/f.u.], as highlighted in yellow in *Figure 60*. This scenario accounts for the variance in the prediction error among repeated phases. Similarly, $Fe_{14}Nd_2B$ appears six times with reported M_s values ranging from 30.2 to 32.5 [μ_B/f.u.], leading to a predicted value of 31.5 [μ_B/f.u.]. However, in cases where it is not feasible to form clusters based on the reported M_s values for a particular phase, the model optimizes by learning the optimal M_s value that minimizes errors in most of the repeated phases. Moreover, the inherent iterative learning nature of many ML models including the models used for M_s prediction ensures that outliers in repeated phases do not influence the model's decision-making process.

Density of 14:2:1 magnetic phase

The comprehensive comparison of the model performance using two different feature sets, 'CC' and (CC+AM), evaluated through multiple metrics is shown in *Table 17*. The LR model consistently outperforms the other models in terms of MAE, RMSE, and R^2. This is due to the inductive nature of LR model which has less parameters and is based on assumption of linearity that fits well for density dataset. Particularly, when using the (CC+AM) feature set, the LR model demonstrates superior predictive capabilities, achieving the lowest MAE of 0.04 [g/cm^3] and the highest R^2 score of 0.98. This accounts for a mean absolute percentage error of less than 0.5%. These results affirm the efficacy of the LR model when combined with the augmented feature set (CC+AM), showcasing its ability to provide accurate predictions for density. While non-linear models such as RF, SVM, and Weighted VR exhibit good performance, the consistently superior performance of the LR model underlines its suitability for this specific task. Additionally, the compositional feature set from the chemical formula of the 14:2:1 phase mentioned in literature, demonstrates strong predictive ability on the unseen test dataset with a low MAE error of 0.51%. Furthermore, with atomic mass as an additional feature, it resulted in comparable results with 0.50% difference in overall prediction accuracy in terms of MAE.

Thus, suggesting that the readily available composition-based features from the literature sufficiently capture the information of mass density.

In the present study, the trained linear ML model facilitates density predictions for the 14:2:1 phase, which includes combinations of 33 distinct elements (shown in *Figure 18(b)*). Within this 33-dimensional space, particularly for the (CC+AM) feature set, the mathematical complexity of the linear model is expected to be similar compared to the physical complexity of the mass density. Therefore, a linear model within this 33-dimensional space, excluding lattice parameter features, while achieving low prediction errors, is not straightforward. The current linear ML model, utilizing composition-based features, achieves a remarkably low prediction error of just 0.51% MAE on the test set.

Curie temperature of 14:2:1 magnetic phase

The T_c is required to exhibit continuity as a function of chemical composition [172], including substitutional compositions. In previous studies focusing on ternary and multi-component systems, such as those by Nelson and Sanvito [302] and Long et al. [177], a tree-based regressor, specifically a random forest regressor, was used to predict T_c. However, tree-based regressors can introduce discontinuous step functions [303], as highlighted by Dam et al. [174] and Nguyen et al. [173], who opted for kernel-based ridge regressions to achieve continuous functions with low prediction errors (32 K MAE), especially in binary systems. While these methods show efficacy in binary systems, their performance tends to decrease as the number of components increases, as seen in quaternary and quinary systems. Nevertheless, despite having tree-based regressors as the four screened base models, the developed model uses a strategy where the individual predictions are averaged to construct a voting regressor. This approach results in smoother functional approximations compared to individual tree-based predictions, overcoming the issue of functional discontinuity associated with tree-based regressors if used alone. By adopting the VR model, there is a reduction in the impact of function discontinuity and simultaneously achieve a low MAE on unseen data. Further, when establishing the T_c composition relation, the likelihood of multiple base models having identical step positions is low. Additionally, this approach has the benefits of generalization, thus lowering the risk of overfitting compared to relying solely on a standalone model.

The relatively higher prediction errors resulted in $Fe_{11}Cr_3Gd_2B$ and $Fe_{11}Ni_3Nd_2B$ from the validation set can be attributed to the substantial substitution of Fe by Cr and Ni, potentially reaching the solubility limits within their respective phases. The solubility limit for Cr and Ni for Fe substitution in $Fe_{14}Nd_2B$ is 3 as mentioned in Section 3.2.1. The prediction of compositions situated near the solubility boundary can be challenging due to the limited dataset available from which the model can learn. By refining the model hyperparameters, it is possible to achieve a lower validation error of around 8 K. However, such reduced errors might either approach the experimental temperature measurement error level or come close to the variation in T_c values for a specific composition across different literature sources incorporated in the training set. In such cases, there is a risk of overfitting the training data, which can lower the model's generalization capability. This is also the reason for opting the train test split based cross-validation scheme (with an 80:20 ratio) over other validation methods that were evaluated, such as LOOCV or K-fold cross-validation (K = 5), both of which resulted in MAE values of approximately 5 and 9 K, respectively. Similarly, in test set $Fe_{11}Mn_3Y_2B$ is the only

phase with prediction error of more than 50 K that is 23% absolute error, which results because there is only one data point with this chemical composition in the training set causing the model to develop uncertainty. Since there is only one instance of $Fe_{14-x}Mn_xY_2B$ [260] in the training set, the trained model has not been exposed to enough varied examples of phase with this particular composition to accurately learn and generalize its behavior.

The feature importance provides an insight into the extent to which an element influences T_c. Various feature importance measures to improve interpretability and deepen the understanding of model predictions. The mean decrease in impurity (MDI), a tree-based technique, to calculate feature importance by considering the reduction in node impurity weighted by the probability of reaching that node [304] was explored. This approach can exhibit bias towards features with high cardinality and continuous numerical features when categorical features are present. However, the dataset here comprises only continuous numerical features, therefore the issue of cardinality does not apply. Furthermore, a model-agnostic feature importance analysis based on permutation [305] was performed, including the use of SHapley Additive exPlanations (SHAP) values [306]. The results of MDI and the permutation importance were aligned, with the top five features being the same for the two methods. Due to difference in the approach to determine the feature importance values, the elements distribution varies. In case of T_c model, the top features in the feature importance plot from MDI and SHAP is nearly same except 4th element. The results from SHAP differs from MDI such that it includes Nd instead of Ru within top five features as seen in *Figure 61*. The color green signifies a positive impact on the model's prediction, whereas red indicates a negative impact. Moreover, the importance of features provides information on the model itself rather than the data on which it is trained.

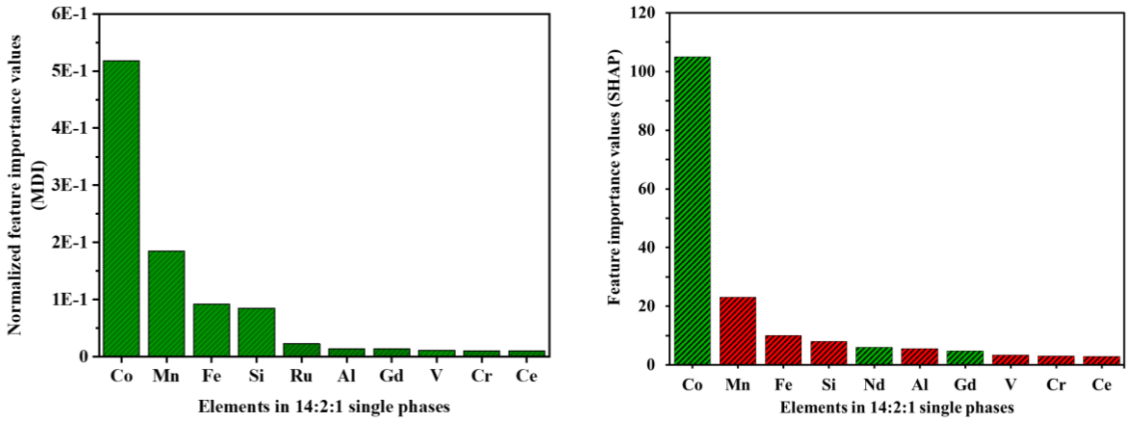


Figure 61: Feature importance plot for the voting regressor model showing the top 10 elemental features that have an effect on the model's T_c prediction. The normalized feature importance values determined using the MDI method is on the left [178]. The feature importance plot based on SHAP method is on the right. The mean absolute SHAP values are the global feature importance values where the positive and negative impact on the T_c prediction are shown in green and red color, respectively.

In *Figure 61*, the histogram on left represents the average normalized MDI feature importance across individual screened models and on right is the feature importance plot calculated using the SHAP method. Feature importance can be explained by two factors. First, there is the compositional sensitivity, indicating the slope in T_c composition curves. Second, the compositional range of an element over which this slope exists. A wider compositional range for an element with a corresponding slope results in higher qualitative feature importance. The

histogram of feature importance in *Figure 61* highlights cobalt as the most influential element in the model's predictions for T_c. This aligns with established findings indicating that $Co_{14}RE_2B$ phases (with RE being a rare earth element) exhibit the highest T_c among the 14:2:1 phase family, often approaching 1000 K. Furthermore, cobalt is often used as a substitution element for iron [41]. Moreover, cobalt exhibits a higher slope (approximately 80 K per unit elemental substitution) over a broader compositional range within T_c composition curves compared to other substitution elements in the Fe system. It is important to note that the magnitude of the slope estimate and the corresponding compositional interval provide insights into the strength of an element's impact on T_c. The $Fe_{14-x}Co_xNd_2B$ system [124], $Fe_{14-x}Co_xLa_2B$, and $Fe_{14-x}Co_xY_2B$ agrees well with this observation, as supported by literature sources [44], [238]. Among the substitution elements, manganese follows cobalt with a similarly high compositional interval for a given slope magnitude, spanning the 14:2:1 compositional space. However, Mn's influence on T_c differs from that of Co, as it leads to a decrease in T_c. This trend is consistent in systems such as $Fe_{14-x}Mn_xNd_2B$, $Fe_{14-x}Mn_xY_2B$, and $Fe_{14-x}Mn_xPr_2B$ [229]. On the other hand, elements such as silicon, ruthenium, and aluminum shows shorter compositional intervals for slope as they have low solubility limits, as evident in cases such as $(Co,Si)_{14}Nd_2B$ [42], $Fe_{14-x}Ru_xNd_2B$ [47], and $Fe_{14-x}Al_xNd_2B$. The relative positions of these elements compared to rare earth elements are in line with established understanding.

Saturation magnetization of 14:2:1 magnetic phase

The performance evaluation of various regression models based on three important metrics: R^2, RMSE, and MAE suggest that the HR and RF models, both achieving good R^2 of 0.97, demonstrate a remarkable ability to closely align predictions with actual data. However, the significant difference in MAE highlights that RF excels in minimizing prediction errors, making it an attractive choice when precision is crucial. On the other hand, the GB and XGB models, while showing good overall fits with R^2 values of 0.93 and 0.91 respectively, exhibit larger errors, which may limit their suitability for tasks demanding high precision. The MLP model, with its strong R^2 of 0.96 and the lowest MAE among all models, showcases consistent accuracy and reliability, making it a compelling choice for various applications. On the contrary, with its high R^2 of 0.98 and low RMSE and MAE values, the AB-DT model stands out as an ideal option when the accuracy is predominant.

Whereas, among VR models, NU-VR and U-VR, with same R^2 of 0.97, offer strong correlations between predictions and actual values. The relatively low differences in the RMSE and MAE values reflect their ability to provide accurate predictions with relatively small errors. These models present a robust choice for scenarios where ensemble learning is preferred to harness the collective predictive power of multiple regressors. The voting regressor model here combines diverse base regressors, comprising both parametric and non-parametric models. Within this ensemble of base regressors, RF and AB-DT are tree-based models, HR is characterized as a linear model, while MLP is an artificial neural network that is good at capturing non-linear patterns. Due to the distinct nature of these base regressors and their varying learning processes, allocating equal weightage to each becomes challenging. It is to be noted that there is no linear dependence of the degree of substitution element to the M_s values reported across various literatures. For example, the M_s values reported for the $Fe_{14-x}Co_xNd_2B$ phase for Co substitution ranging from 0 to 14 has non-linear trend with varying levels on non linearity [128], [137], [217]. Therefore, a VR model with base regressor that is capable of

identifying non linear functions is necessary. Therefore, an approach involving uneven weighting has been adopted to achieve a smoother and more continuous function. This strategy aims to create a generalized and robust model that exhibits a superior fit to the dataset compared to the use of a uniform voting regressor.

The low MAE and RMSE values of 0.72 [μ_B/f.u.] and 1.02 [μ_B/f.u.] observed for the phases in the validation dataset indicate the ability of the model to generalize across the diverse dataset that contains various alloy systems. However, despite the trained NU-VR model resulted in the low prediction errors, there is a significant difference between the MAE and RMSE values. RMSE penalizes the phases with higher errors, suggesting outliers or instances where the model struggled to achieve a good fit. The relatively higher RMSE value of 1.07 [μ_B/f.u.], in comparison to the MAE of 0.72 [μ_B/f.u.] on the validation set, is primarily influenced by phases, namely $Fe_{10}CoAl_3Nd_2B$, $Fe_{14}Nd_{1.6}Er_{0.4}B$, and $Fe_{14}Nd_{1.5}Tb_{0.5}B$. These phases have absolute errors exceeding 2 [μ_B/f.u.] which is more than 8% error, with $Fe_{10}CoAl_3Nd_2B$ contributing an absolute error of 5.16 [μ_B/f.u.] that is 20% error. Further, in the training set, the presence of phases in which the iron (Fe) element is substituted by varying proportions of two different elements (Co and Al) is limited to only three times.

As shown in *Figure 62*, an increased concentration of aluminum within the phase exhibits a negative influence, potentially resulting in a lower M_s value. Particularly notable is the instance of $Fe_{10}CoAl_3Nd_2B$, where aluminum substitutes the three atoms of iron. This substitution of aluminum significantly affects the prediction, resulting in a lower M_s value of 19.94 [μ_B/f.u.], whereas the M_s value for $Fe_{14}Nd_2B$ is 25.1 [μ_B/f.u.]. However, adding a greater number of such phases within the training set could prove beneficial to the learning model, helping it achieve better generalization. The inclusion of such phases presents a difficulty, given the solubility limit of Al, which has been reported to be 3 with substitution of 3 Fe atoms by Al in 14:2:1 formula. Moreover, higher prediction error for $Fe_{14}Nd_{1.6}Er_{0.4}B$ and $Fe_{14}Nd_{1.5}Tb_{0.5}B$ from training and validation and for $Fe_{12}Cu_2Y_2B$ from test set are due to the lack of sufficient training data. Addressing this challenge requires the inclusion of additional training phases, which allows the model to capture the underlying information of the data fit.

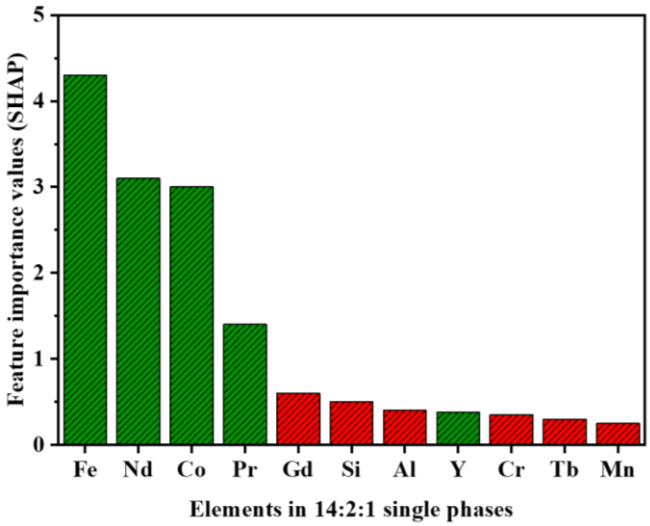


Figure 62: Feature importance plot for the trained M_s prediction model (weighted voting regressor) calculated using the SHAP method. The mean absolute SHAP values are the global feature importance values where the positive and negative impact on the M_s prediction are shown in green and red color, respectively [185].

Figure 62 shows the feature importance plot calculated using the SHAP method [306]. This approach is suitable due to the diverse nature of the base learners in the voting regressor model, that includes linear, non-linear, and tree-based models. Therefore, using a model-agnostic approach such as SHAP assists in obtaining a comprehensive feature importance plot. It is important to note that the plot showcases the top 11 of the 28 features utilized in the model's development. This plot provides information on the behavior of the trained model and the factors influencing its predictions. However, it does not provide insights into the distribution of the underlying data. The color green signifies a positive impact on the model's prediction, whereas red indicates a negative impact. For example, consider a phase containing Nd, Fe, and Si elements. In this case, the presence of Nd and Fe would lead the model to predict a higher M_s value, while the inclusion of Si would lead to the lowering of the M_s value prediction. This effect is directly proportional to the respective percentages of each element within the phase.

Effect of aggregated data for repeated phases on the model

In the investigation of 14:2:1 phases and their corresponding target properties, some common obstacles were encountered. These phases, associated with specific chemical compositions, often display varying ρ_m, T_c and M_s values across different literature sources, indicating discrepancies in the reported data for the same phases. Initially, an attempt to address this issue by aggregating the data composition-wise, using a single metric such as the mean, for the T_c model was carried out. Although this approach proved effective for certain compositions, it was unable to be used for others, leading to high errors. It became evident that a single metric for aggregation couldn't universally suit all compositions in our dataset. Moreover, the training data must reflect the real-world scenario, which includes the variability in measurements arising from diverse literature sources. Surprisingly, when the non-aggregated dataset was utilized, retaining each data point without aggregation, the validation results demonstrated significantly lower errors, with only a 16 K error on the validation set, compared to 81 K (or higher) with the aggregated dataset. In the test set, one composition, $Fe_{11}Mn_3Y_2B$, exceeded a prediction error of 50 K when using the non-aggregated dataset, reaching 92 K. However, this high MAE was primarily due to a single data point with an exceptionally high error. This single data point is a paramagnetic T_c obtained from the inverse susceptibility versus temperature curve reported by Sano et al [260]. In contrast, the aggregated dataset resulted in a MAE of 80 K on the test set, with most compositions showing predictions around this mean error.

The implication here is that the bias introduced by repeated compositions in the dataset is relatively insignificant compared to the error introduced by using a single metric for aggregation that does not account for composition dependence. Attempting to aggregate T_c values from distinct literature sources using a single metric leads to aggregation bias and overfits the learning model with high variance. Furthermore, the evaluation of the trained model's ability to represent the dataset universally is primarily based on its performance in predicting unseen test data [307], [308]. The low MAE of 16 K on unseen chemical compositions in the test set by the T_c model and the low MAE of 0.88 [μ_B/f.u.] for unique phases and 0.78 [μ_B/f.u.] for repeated phases in the test set by the M_s model suggest that the model generalizes effectively and is not biased toward specific compositions.

It is crucial to note that the hyperparameters of the models, regardless of whether the dataset is aggregated or not, are individually optimized and do not influence the inference process.

Additionally, the learning process of most ML algorithms, including the tree-based models used in this study, naturally gravitates toward learning target values that align with clusters of duplicate target values, minimizing errors during training. Therefore, aggregating repeated compositions within the dataset using a single metric such as the mean, may yield accurate results for some compositions, but introduce errors and bias for others. Furthermore, the model demonstrates effective generalization, avoiding bias towards repeated phases, and effectively capturing compositional dependencies on errors. The learning model can logically process duplicate or repeated phases, learning optimized target values that closely align with clusters of duplicate target values, thus minimizing errors during the training phase. For tree-based base regressor models like random forest and adaptive boosted decision trees, the impact of outliers or noise within repeated phases is partially mitigated through their internal learning patterns involving ensemble techniques.

Effects of correlated features on the model

While tree-based models are often robust to feature correlation when it comes to predictive performance, highly correlated features can still pose challenges for accurately interpreting feature importance. The features in the dataset for the T_c and M_s prediction model were also analyzed for the feature correlation, as it helps to better interpret the feature importance plots and to be aware of the potential effects on model's predictions. *Figure 63(a & b)* shows the Spearman rank based feature correlation plot for the T_c and M_s use-case respectively. Considering the distribution of the features in the dataset based on its cardinality, which is skewed, as seen in *Figure 18(d & f)*, the Spearman's rank correlation coefficient approach has been considered as it is robust to the skewed data [309]. Note that only a feature with cardinality of more than 4 is considered for visualization.

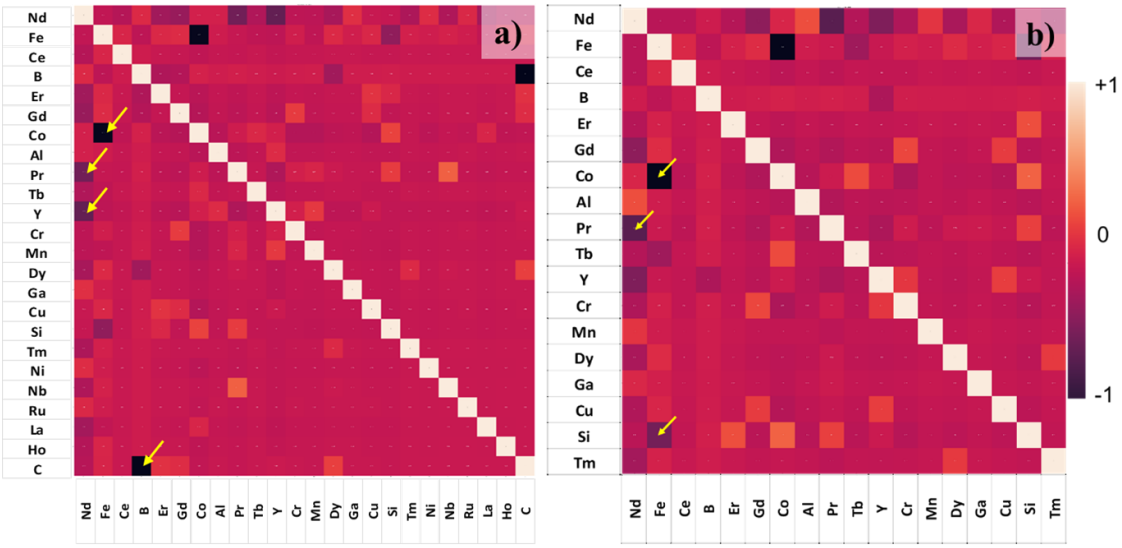


Figure 63: Correlation between the features in the intrinsic property prediction dataset calculated using the Spearman's rank correlation method. (a) T_c and (b) M_s. The feature shown in the plot is limited to the cardinality more than 4 for visualization. The color bar on ranges from -1 to 1 where -1 indicates strong negative correlation, 0 indicates a no correlation and 1 indicates strong positive correlation.

The *Figure 63(a & b)* suggest that there is a strong negative correlation between Fe and Co elements. As the concentration of Fe increases, the concentration of Co tends to decrease, and vice versa. A mild negative correlation can also be seen between Nd and Y for the T_c use-case

and between Nd and Pr for the M_s use-case. However, considering the size of feature set, which is more than 25 elements for both use-cases, the presence of a correlation between 2 elements does not have a larger impact on the model's performance.

5.2. Application of the intrinsic property prediction models

Material property prediction models excel in expanding material databases by either extrapolating known data or filling gaps with predicted values, thus enriching the understanding of material properties. These models enable high-throughput screening of potential candidates with desired properties, greatly accelerating the search for optimal materials for specific applications. The T_c and M_s predicting models find applications in expanding the existing database through interpolation, filling gaps within the solubility limits, and finding other intrinsic magnetic properties in combination with existing physical laws.

Application of the trained Curie temperature model for expanding databases

Figure 64 presents the T_c predictions for varying amounts of element substitution [24], [41] on an Fe base and an Nd base. The black data points are directly extracted from literature sources [41], [128], [130], [233] for the purpose of benchmarking predictions against the trained model. Meanwhile, the colored data points represent the model's predictions. The values obtained for the different substitution values for various elements can be used for expanding the existing database so that the training models can further improve the relationship between the chemical composition and the T_c value at room temperature. However, it has to be noted that the T_c prediction model here is a combination of four different tree-based models which resulted in inductive bias of learning a step function as hypothesis [103], [110] causing the sudden jump in prediction values. This step function pattern for tree-based models is an inherent nature of decision trees [310].

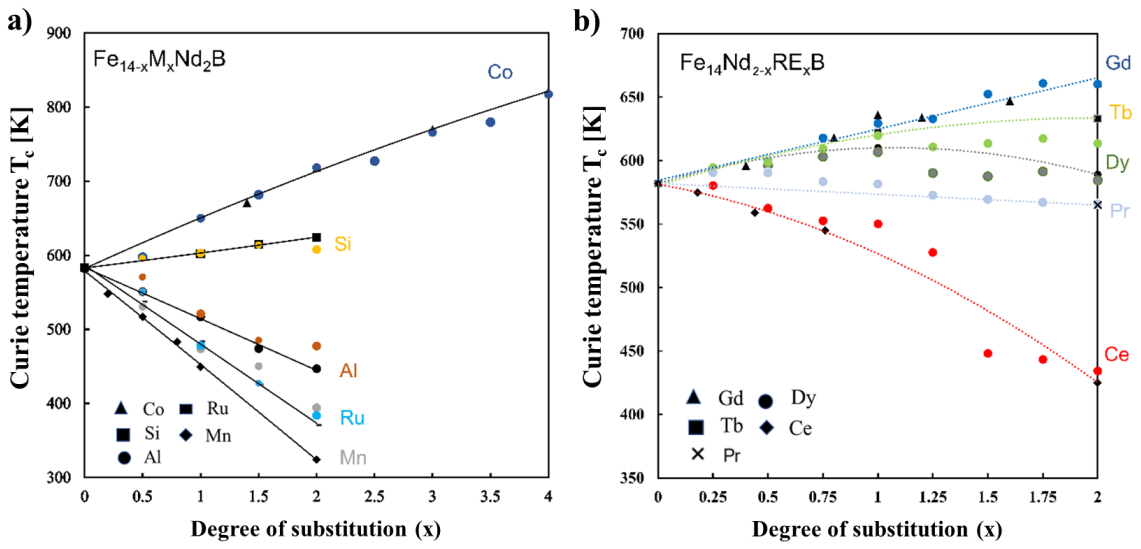


Figure 64: Predicted T_c values (colored data points) from the trained T_c model in comparison to reported data points (in black) from an external, unseen source. This evaluation is conducted for different types of substitutions: (a) $Fe_{14-x}M_xNd_2B$ (M = substitution elements), and (b) $Fe_{14}Nd_{2-x}RE_xB$ (RE = rare earth substitution) [178].

The prediction curve for the 'Ce' substitution shows that between substitution values of 1.25 and 1.5, there is a sudden drop in the T_c values. Tree-based models have the limitation that they add more uncertainty to predicted values (see *Figure 65*) when used for interpolating or extrapolating way beyond the boundary limits of the training dataset.

Figure 65 shows the predicted T_c values for 'Ce' substitution from x=0 to x=2 in steps of 0.02, the red error bars indicate the prediction uncertainty and it can be seen that the model has higher uncertainty for substitution the amount between 0.6 to 1.75 and between 1.3 to 1.6 it is very high. This is because the model has been trained up to the substitution limit of x=0.5 for 'Ce' and for x=2, while predictions are made up to x=2. There is large gap between the x=0.75 to x=1.98 where literature reported values are not available. The prediction uncertainty here is quantified by estimating the standard deviation of weak learner's prediction in tree-based models as described by Palmer et al. [311].. To overcome this limitation, the M_s prediction model includes both tree and non-tree based models so that the inductive bias of base learning models can be mitigated to a large extent. Additionally, it is important to include the phases with a substitution value more than 0.5 so that the uncertainty can be reduced when predictions are made for a higher substitution of 'Ce'. However, a more reliable approach would be to limit the model to interpolation, as this ensures that missing phases are obtained in a way that aligns with physical principles and can be evaluated meaningfully.

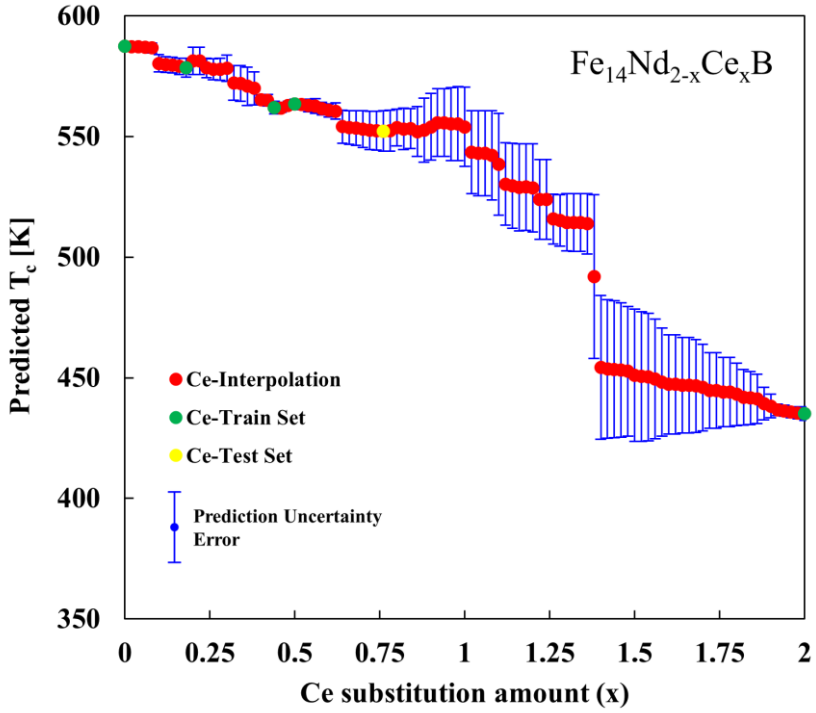


Figure 65: Predicted T_c values for 'Ce' substitution amount (x) in steps of 0.02 for $Fe_{14}Nd_{2-x}Ce_xB$. The prediction uncertainty for each data point is shown as an error bar. The substitution amount marked in red are the interpolated phases which are not present in the train and test set.

Application of the trained saturation magnetization model for expanding databases

The *Figure 66* aims to extend the insights derived from the existing plot introduced by Burzo et al. [41]. The extension involves predicting M_s values for higher substitution quantities while

remaining within the solubility limit except for Si. For Si, the solubility limit based on collected literature data is 2. However, the thermal stability of phases with Si content more than 2 that has been predicted by model is not known. It compares the M_s values reported in the literature [24], [41] and the values predicted by the model for various substitution elements, including Co, Al, Mn, and Si. The literature-reported values are shown in black, while the predicted values are represented using different colors. The elements Mn and Al exhibit steeper slopes, indicating a greater sensitivity compared to Co or Si. Using the developed model to predict M_s at room temperature, the saturation polarization can be calculated for phases with known density, either measured or reported in the literature. Additionally, based on the approach proposed by Kini et al. [98] for predicting density from the chemical composition, the density can be estimated for alloys with known M_s values in [μ_B/f.u.] from literature, allowing the calculation of J_s. It is important to note that the developed M_s prediction model and the existing density prediction model by Kini et al. [98] were trained on distinct chemical compositions, leading to varied chemical elemental feature vectors. Therefore, for predicting M_s or density of an unknown phase, its chemical composition must fall within the trained chemical elemental feature vector or space.

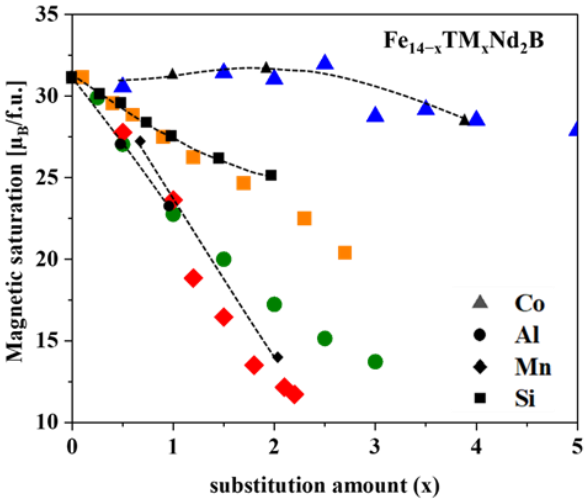


Figure 66: Comparison between the predicted (colored data points) and reported (black points) M_s at room temperature for the $Fe_{14-x}TM_xNd_2B$ (TM = transition metal) reported by Burzo et al. in [41]. Using the trained model, magnetic saturation for the higher substitution of the transition metals (Co, Al, and Mn) within the solubility limits has been predicted [185]. For Si, the solubility limit based on literature data is 2. However, the thermal stability of phases with Si content more than 2 that has been predicted by model is not known.

Among the 143 distinct phases within the M_s prediction model dataset, a subset of 24 phases have both their density and their M_s values documented in the literature. These 24 phases include ternary, quaternary, and quinary alloy systems. Further, the literature sources for the reported M_s and density of these 24 phases differ, introducing the possibility of variations due to experimental conditions or subjective discrepancies. Ideally, the calculated J_s values derived from the literature based M_s and density should align with the J_s values computed using the projected M_s and density values from the respective trained ML models. This comparison between literature-derived and predicted J_s values serves as a physics-driven or human-in-loop validation, confirming the trueness or fairness of the dataset used for training the ML model.

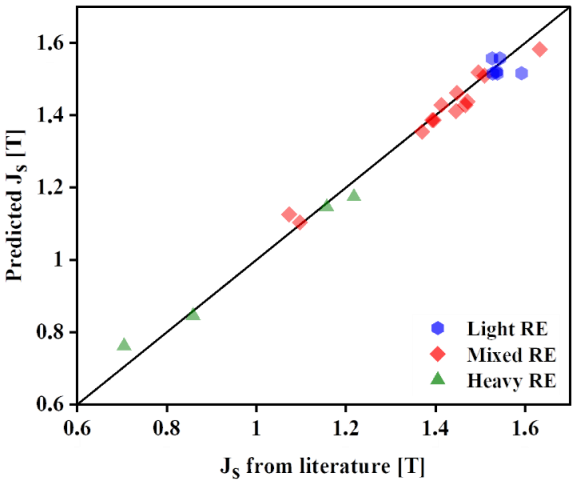


Figure 67: Comparison between the saturation polarization (J_s [T]) calculated through trained machine learning models and the corresponding values from the literature. The phases are classified into three categories: 14:2:1 phases containing heavy rare earth (depicted in green), light rare earth (depicted in blue), and a mixture of both heavy and light rare earth metals (depicted in red). The calculated J_s values in [T] are derived from the M_s in [μ_B/f.u.] and the mass density in [g/cm^3].

Figure 67 shows the comparison between the calculated J_s values in [T] using M_s in [μ_B/f.u.] and the density in [g/cm^3] derived from the literature sources against the results of the trained model. Based on the presence of RE elements in the alloy, the 24 phases are categorized into three groups: those with heavy RE, light RE, and mixed RE. The lowest value among the 24 phases is for $Fe_{14}Ho_2B$ (0.71 T) and the highest value is for $Fe_{12}Co_2Nd_2B$ (1.64 T). The analysis revealed an average absolute deviation of 2% for phases with light RE, 2% for mixed RE, and 3% for heavy RE. The highest deviation observed among the 24 tested phases amounted to 0.1 T and it is for $Fe_{14}Ho_2B$, which resulted in absolute deviation of 7.7%. The lowest absolute deviation was seen for $Fe_{12}Co_2Nd_{1.8}Tb_{0.2}B$ and it is 0.2%. Similarly, such trained models can obtain other properties such as anisotropy field and anisotropy constant and provide a basis for developing physics-informed ML models.

5.3. Strengths and limitations of the intrinsic magnetic property prediction model from chemical composition

The property prediction regression models developed are specifically for the 14:2:1 magnetic phase. These models are associated with some strengths, but also have limitations on the application due to the training model's inductive bias and prediction interval due to the datasets used.

Strengths of the developed approaches

1. **Data-driven solutions:** The models are data-driven, utilizing machine learning and statistical techniques to make predictions. This approach can reveal linear and non-linear relationships between composition and magnetic properties that may be challenging to ascertain through traditional methods.
2. **Composition-based predictions:** These models can accurately predict the ρ_m, T_c, and M_s of the 14:2:1 magnetic phase solely on the chemical composition. This is valuable

because it reduces the need for expensive and time-consuming experimental measurements or density functional theory calculations.

3. **Generalized model:** The developed models can be applied to various higher component systems upto senary alloys, containing various transition metals and rare earth elements. This versatility makes them valuable for applications with diverse materials. The applicable feature space of 33 elements for ρ_m, 37 elements for T_c, and 28 elements for M_s model ensures that model finds it's application on larger elemental space.
4. **High accuracy:** The models demonstrate high accuracy in predicting ρ_m, T_c, and M_s. The ρ_m prediction model achieves over 98% accuracy on unseen phases, the MAE for T_c prediction is low (16 K), and the average deviation of 3% for M_s predictions when compared to reference values indicates reliable results. The developed models can be used to expand the existing database and predict other intrinsic properties in combination with the existing physical laws, as explained for the J_s use case.

Limitations of the developed approaches

1. **Data limitations:** The reliability of the models heavily depends on the quality and quantity of the training data. The limited data in case T_c model for the 'Ce' substituted $Fe_{14}Nd_{2-x}Ce_xB$ shows that beyond training data boundary, the uncertainty in predicted value starts to increase. Limited data or discrepancies in reported values for the same compositions from different sources can impact the accuracy of predictions if a sanity check is not effectively done for potential outliers. However, additional data leads to the improvement of the model.
2. **Composition at solubility limits:** Phases with compositions at the solubility limit have comparatively higher prediction errors. This also implies that properties for the phases which do not exist as single 14:2:1 are also predicted by the model but with high uncertainty. Therefore, it is recommended to perform interpolation within training data limits and extrapolation upto 5% of training data limits followed by experimental evaluation if needed.
3. **Thresholds for element substitutions:** The accuracy of T_c predictions may vary depending on the amount of element substitutions, particularly in Fe or Co-rich bases. Substitutions involving elements such as Ni, Si, and Ga appear to yield results within a 20% margin, but this threshold can be different for each substitution element.

In conclusion, these predictive models offer an efficient and reliable way to predict intrinsic magnetic properties of the 14:2:1 magnetic phase. They have the potential to accelerate the development of materials. However, it is crucial to be aware of their limitations, particularly regarding data quality and materials' composition at solubility limits. One should use these models with experimental methods to validate their predictions and make informed decisions in materials research and development.

5.4. Evaluation of the grain boundary detection model for sintered FeNdB magnet

In this section, the complexities of grain size analysis for sintered FeNdB permanent magnets are discussed by examining topics that influence accuracy and reliability. These include the

challenges of microscopy-based analysis, ground truth data sources, feature engineering, the shift towards deep learning, the prominence of the U-net architecture, dataset and hyperparameter impacts, input image size effects, model robustness, and any discrepancies between ground truth and model predictions. This exploration offers insights into the diverse nature of grain size analysis, highlighting the challenges, methodologies, and innovations in materials science and engineering.

Complexity with microscopy-based grain size analysis for sintered FeNdB permanent magnet

Detecting the grain boundaries in sintered FeNdB magnets using microscopy techniques is a complex task, as it plays a pivotal role in understanding the material's properties and optimizing its performance. However, the task of obtaining grain boundary information depends on the imaging technique used, with KM and EBSD offering distinct advantages and challenges. KM primarily excels at capturing magnetic domain structures within FeNdB magnets. The obstacle in extracting grain boundary information using KM includes limitations in spatial resolution, making it difficult to distinguish individual grains and their boundaries, particularly in fine-grained magnets with ECD less than 3 μm for anisotropic samples and for isotropic samples it is 6 μm.

Furthermore, KM primarily relies on magnetic contrast. Although grain boundaries can influence magnetic domain patterns, they may not be directly visible, requiring advanced image processing techniques to infer their presence. Changes in the magnetic domain orientation at grain boundaries can suggest their existence, but they do not provide detailed structural information. This is primarily seen in anisotropic samples where crystallographic orientation between neighboring grains differs by a small margin, as seen in *Figure 23* for 0°, 15°, and 90° sectioned samples of FeNdB-E. EBSD, on the other hand, offers a high-resolution method to study grain boundaries in FeNdB magnets, detecting grains as small as 2 μm, identifying boundary, distinguishing phases, crystallographic orientation data for microstructure. Additionally, EBSD enables statistical analysis of grain properties, which is essential to assess material behavior.

While EBSD offers these advantages, it is not without its challenges, including the need for specialized equipment and sample preparation. In practice, a comprehensive approach may involve utilizing both KM and EBSD. KM can provide insights into magnetic domain behavior, while EBSD offers detailed structural information, including grain boundaries. EBSD provides more precise information about the orientation of the grains, but KM has advantages in the efficient analysis of larger areas or entire sample. Combining these techniques allows researchers to understand the complex microstructure of sintered FeNdB magnets, ultimately facilitating their optimization for various applications.

Ground truth for the training model: Manual annotation vs EBSD

EBSD and human annotations can serve as ground truth references for grain boundary detection to perform grain size analysis. However, each has its own strengths and limitations when used to evaluate the performance of machine learning or computer vision models. Several factors come into play when considering the choice between EBSD and human annotations as sources

of ground truth data. EBSD stands out for its exceptional accuracy and precision, delivering direct measurements of crystallographic orientations and capturing intricate structural details. It is an objective and consistent approach, but may require specialized equipment and be time-consuming for complex microstructures [54], [59], [209], [266].

Table 23: The table provides the comparison between the labels from the EBSD and manually annotated data from KM for grain boundary and orientation detection task. The values in table are based on correlative EBSD and KM images.

Evaluation criteria	**Metric**	**EBSD data**	**Manually labeled data based on KM**	**Observations are based on correlative EBSD and KM images**	**Conclusions**
Missed grains/ grain boundaries	Visual inspection and IoU	2% - 3.5%	2% - 7.5%	EBSD removes some grains during post-processing; Manual KM struggles with small grains (ECD < 3 µm) and closure domain patterns	EBSD performs better but misses some fine details, while manual KM struggles with closure domain patterns
False grain boundaries	Visual inspection and IoU	< 1.5%	2.5% - 5.5%	EBSD merges some grains into one large grain during post-processing; Manual KM has high error for samples having many grains with closure domain patterns	EBSD is more reliable but may oversimplify grain structures; manual KM is less precise in grains with closure domain patterns
Error in detection of small grains (ECD < 4 µm)	Grain size (D25)	upto 4%	upto 3%	Results are volatile, with some grains being detection and other being missed for both EBSD and manual KM	Both methods struggle with small grains, though manual KM is slightly better in detecting them
Variance in results	Number of grains per μm^2	< 2%	upto 3%	Data are obtained from different operators or users	EBSD is more consistent, while manual KM depends on user skill and interpretation
Grain morphology	Circularity and aspect ratio	High degree of alteration	Minimal effect	Shape of grain changes with post-processing steps in EBSD	EBSD alters grain shape, while manual KM preserves the grain morphology
Grain orientation information	None	$[\varphi_1,\phi,\varphi_2]$	$[\varphi_1]$	Partial information in KM is annotated. Orientation for the grains with stripe domain structures are determined in manual KM	EBSD provides complete grain orientation data, whereas manual KM only offers partial information
Accuracy of grain orientation	Visual inspection	High	Low	Orientations of grains with stripe domain patterns only are detected by manual KM	EBSD is highly accurate, while manual KM is only useful for specific grain patterns

On the other hand, human annotations offer flexibility, extending beyond crystallography to capture complex semantic information. However, they can introduce subjectivity and variability among annotators, potentially leading to inconsistencies. Human annotations may have advantages in terms of simple sample preparation and faster data acquisition. Moreover, evaluating a model's performance using both EBSD and human annotations as separate ground truth references can be valuable. This approach offers a comprehensive assessment that reveals where the model excels or struggles with each reference, providing a detailed understanding of its capabilities and limitations. Such considerations are essential when choosing the most suitable ground truth source for a given research or analysis task.

Table 23 compares the performance of the EBSD data and the manually labeled KM data for grain analysis across various criteria. EBSD misses 2% - 3.5% of grains or boundaries as shown in *Figure 68(e)* with the marked region in red, while the manual approach misses 2% - 7.5% (see *Figure 68(b)*), with both methods struggling to detect small grains. The absence of grains in EBSD is primarily due to post-processing effects carried out to remove noise from raw EBSD maps. The manual approach leads to high error in case of the samples with large amount of grains with closure domain patterns as seen in *Figure 68(a-c)* and *Figure 75(a)*. However, EBSD performs better overall in this regard.

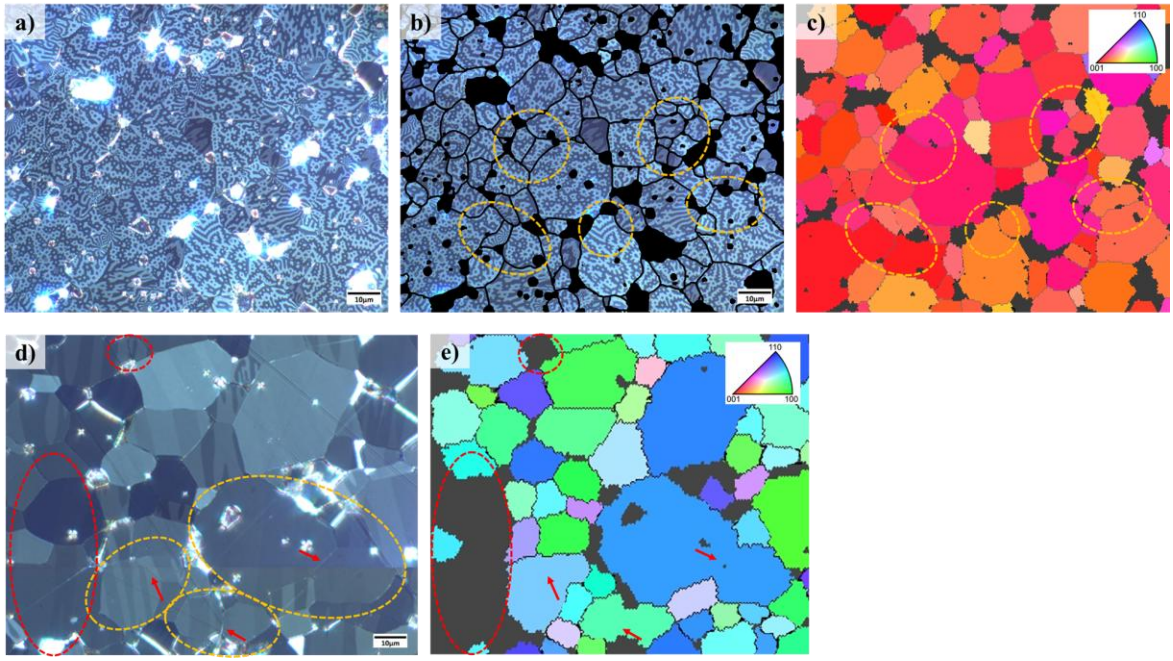


Figure 68: (a) and (c) Correlative KM and EBSD map of the anisotropic FeNdB-E sintered magnet sectioned at 0°. (b) KM image with manually labeled grain boundaries, pores and oxides. The marked grains shows the error in the manually labeled grain boundaries when compared to EBSD map. (d) and (e) shows correlative KM and EBSD map of anisotropic FeNdB-TP sintered magnet. The marked region in orange with red arrow highlights the presence of grain boundary between two grains which are missing in EBSD map. The absence of grains in the EBSD map due to post-processing is marked in red ellipse.

For false grain boundaries, EBSD has a lower error rate (<1.5%) compared to the manual approach (2.5% - 5.5%), though EBSD sometimes merges grains, as seen marked regions in orange in *Figure 68(d-e)*, simplifying structures. Both methods face challenges in detecting small grains (ECD < 4 µm), with errors from the EBSD approach reaching up to 4% and from the manual approach up to 3%, where manual approach slightly outperforms EBSD.

In terms of result consistency, EBSD shows less variance (<2%) than the manual approach (up to 3%), which depends highly on the subjective knowledge of the annotator. Grain morphology is significantly altered by EBSD post-processing as seen in *Figure 5(b,c,e & f)*, while the manual approach better preserves the original shapes.

For grain orientation, the EBSD provides complete data, whereas the manual approach offers partial information about the φ_1, particularly for grains with stripe domain patterns as shown in *Figure 19*. Lastly, EBSD demonstrates high accuracy in grain orientation, while the manual approach clearly struggles to extract information on Euler angles [φ_1, ϕ, φ_2]. Overall, EBSD is more consistent, accurate, and comprehensive, although it may miss fine details and alter grain shapes, while the manual approach is more reliable in preserving morphology but struggles with grains having the closure domain patterns.

The first ground truth or the reference data for performance evaluation of the grain boundary detection models is the manually annotated mask by the 3 different subject experts and 2 scientific helpers under expert guidance; therefore, the ground truth is subjective and includes uncertainty. The annotated mask could vary and induce bias from one subject expert to another. A variance of less than 3% was observed for the annotated grain boundaries from different experts and scientific helpers when calculated by quantifying the number of grains per μm^2. The EBSD map provides information on the grain size and orientation of each grain. The difference in the area-weighted grain size distribution results when calculated using the EBSD and the manually annotated mask is between 2% to 4.1%. In terms of D25, D50 and D90 values from area-weighted distribution curves, the relative error between manual and EBSD approach are 0.3, 0.6 and 2.5 respectively. Such a low error would imply that the manually annotated masks are in a very close range to the EBSD measurements for grain size. A comparatively higher error of 2.5 is for coarse grains because the EBSD map has some grains merged into one due to multiple post-processing steps as shown in *Figure 68(d & e)*. However, one should also note that obtaining an EBSD map and data for large sample areas is a time-consuming process. Hence, having a manual approach as a reference for comparing the performance of the models is productive when compared to EBSD because the amount of time and effort involved in generating the results from the manual approach is less than from the EBSD approach.

Further, it should be noted that it is very challenging to perform correlative microscopy involving EBSD and Kerr microscopy. This is primarily because obtaining the correlative image with low error in magnet samples requires a high degree of technical skill. For a EBSD map to be used as the reference mask, the grain boundaries visible in the KM images should overlap with the EBSD maps and have the least possible displacement between them.

Importance of feature engineering in traditional ML based grain size analysis

For experiments with traditional ML, the choice of features to be extracted from the KM images of the magnets included color, edge, and texture information. The features included local and global features, so the learning model is fed with rich information to build a relationship between input and output. The KM images have a highly complex microstructure with grains having different domain patterns in them, and thus require higher dimensional features to train the ML models. This is the primary reason for the better performance of the RF(C,E,T) model over the RF(C) model. The RF model 39 features input that includes color, edge, and texture

information and has better generalization ability than the RF model trained on RGB color features. The increase in IoU from 0.36 to 0.81 is observed with more hand-crafted features. However, the quality of the features matters more than their quantity. When trained on the 4096 features extracted by the pre-trained VGG16 network, the RF model resulted in lower F1 and IoU than RF(C,E,T) despite having more features than hand-crafted ones. This model proved to be under-fitting and had high training and validation errors. In such cases, the sheer number of features does not compensate for their quality, highlighting the importance of thoughtful feature selection and engineering to achieve optimal model performance.

Importance of DL-based approach for grain size analysis

First and foremost, ML typically necessitates manual feature extraction. This involves human-defined features being extracted from the data, a process constrained by the predefined features chosen manually, and it is highly subjective. This approach might not sufficiently capture the complexity and variability inherent in data, such as those present in KM. DL, in contrast, automatically learns hierarchical features from the data. This inherent capability allows the DL model to efficiently capture complex patterns and subtle variations that traditional ML methods might struggle to grasp. The second factor revolves around the inherent variability in the KM images of the magnet samples. These images can involve a wide spectrum of contrasts, colors, textures, and noise levels originating from variations in the manufacturing process, material characteristics, and imaging conditions. Conventional ML models, often trained on a specific set of features, can find it challenging to generalize effectively in the case of such diverse and complex variations. In contrast, DL models exhibit a remarkable ability to manage this variability, primarily by their ability to learn from a more extensive and varied dataset.

Among the DL-based models, HED and SAM were evaluated as they are state-of-the-art models for edge detection and prompt-based image segmentation tasks, respectively. Remarkably, the performance of both models was better than that of RF(C,E,T) even though these models were trained on natural images that do not include KM images of magnets. An improvement of 3%, based on reference values from manual approach is seen when comparing RF(C,E,T) to the HED and SAM models. However, these models struggled to differentiate grains with domain patterns that have high contrast, as seen in *Figure 38* in section 4.1. Nevertheless, the superiority of DL models over fine-tuned ML models for computer vision tasks is evident. Moreover, using transfer learning approach, different pre-trained networks such as VGG16, EfficientNetB3, ResNet50, and InceptionV3 were trained and evaluated for performance. The performance of the U-net with EfficientNetB3 as an encoder is better than that of other pre-trained networks. EfficientNet networks are superior due to their exceptional balance between model complexity and performance. They leverage a compound scaling method that systematically adjusts network width, depth, and resolution to optimize efficiency while maintaining high accuracy. This approach enables EfficientNet models perform better with fewer parameters than other architectures.

The custom U-net model consistently demonstrates close alignment with the reference or ground truth values across all samples. For FeNdB-D, the custom U-net model exhibits minimal relative errors for D50 and D90 values which is less than 1.2 µm, though the error of 2.0 µm for D25 is comparatively larger. Similarly, for FeNdB-C, the custom U-net model has strong agreement with ground truth values, with low relative errors (less than 0.4 µm) for D25, D50,

and D90. These low deviation highlight the effectiveness of the custom U-net model in accurately capturing grain size distributions.

In contrast, the RF(C,E,T) shows larger deviations, particularly for FeNdB-C, with a significant deviation in D50 and D90 being more than 2.5 µm. However, its performance for FeNdB-C remains with low error, with relative errors for D25, D50, and D90 at 1.1 µm, 0.6 µm, and 1.2 µm, respectively. For FeNdB-A, the custom U-net model and RF(C,E,T) model resulted in varying levels of accuracy when compared with the EBSD values. The custom U-net model performs better overall, with errors of -0.8 µm for D25, 0.3 µm for D50 and -1.2 µm for D90, while the RF(C,E,T) model has errors of -0.5 µm for D25, -0.4 µm for D50, and 1.0 µm for D90. Notably, the RF(C,E,T) model shows negative deviations for both D25 and D50, suggesting a tendency to slightly underestimate these metrics relative to the EBSD data.

The performance of the custom U-net model for FeNdB-B demonstrates its robustness, with errors of 0.5 µm for D25, -0.3 µm for D50, and -3.6 µm for D90. The significant deviation for D90 suggests is because some grains in EBSD are merged into one grain, as shown in *Figure 68(d & e)*. Overall, the custom U-net model consistently outperforms the RF(C,E,T) model in aligning with EBSD values, especially for fine grains (D25) and median grains (D50). However, the discrepancies in D90 highlight the inherent variability in the EBSD data.

Importance of the U-net architecture for grain boundary detection task

DL methods, specifically CNNs and U-net have been extensively involved in image segmentation tasks, but it is worth noting that CNN mainly focuses on classification tasks, takes an image as input, and produces a unique class label [312]. Whereas U-net, one of the important semantic segmentation frameworks of CNN, works well for pixel-level prediction tasks [120]. The fundamental difference between traditional CNNs and U-net lies in their approach to handling the output. In classification tasks, CNNs convert an image into a feature vector that is then used for classification. However, in image segmentation, the task goes beyond feature extraction; it involves reconstructing an image from these features. This reconstruction task is notably more challenging, as converting a feature vector back into an image while preserving structural details is a complex problem [313]. The innovation of U-net revolves around this challenge. Instead of learning a feature mapping solely for classification and then trying to reconstruct an image from it, U-net utilizes the same feature maps for both encoding and decoding. The encoder, similar to a conventional CNN, captures essential features of the input image. However, unlike traditional CNNs, U-net's decoder uses these feature maps to expand a vector into a segmented image. This ingenious approach ensures that the structural integrity of the image is maintained during the segmentation process, significantly reducing distortion. For example, *Table 19* shows that U-net with pre-trained networks achieved IoU greater than 0.84 and F1-score greater than 0.87 for grain boundary detection, outperforming the traditional ML approaches by ~5% to 20% in performance when measured using IoU and F1-score. Furthermore, its robustness to limited training data when combined with augmented data makes it practical for applications where labeled datasets are small.

The improved predictive performance of the RF(C,E,T) by ~71% for FeNdB-D compared to FeNdB-C in terms of D50 value from area-weighted distribution curves can be because a nearly 53% area of the FeNdB-D sample was present in the training dataset and 20% area in validation dataset for both the RF(C,E,T) and the custom U-net model. As a result, the RF(C,E,T) resulted

in higher accuracy for the FeNdB-D sample that had been part of its training data as shown in *Table 8*. Grain boundaries often exhibit subtle contrast variations, irregular geometries, and fragmented morphologies, posing challenges for conventional ML methods such as RF(C,E,T). However, the custom U-net model performed well for both samples, due to its ability to combine multi-scale contextual information through deep layers with localized details extracted using the skip connections allows it to delineate boundaries even in low-contrast or noisy regions. It is also necessary to note that the trained model performance in terms of precision is low for the samples having higher number of grains with stripe domain patterns with low domain contrast, as seen in *Figure 40(i)* in section 4.2. This is primarily because the KM image does not have a visible grain boundary in such samples. Such samples have grains with image features such as color and texture in close range to each other, and therefore, model prediction is with low confidence for grain boundary. Nevertheless, such issues with precision have negligible effect on the grain size analysis of large samples having grains of more than 1200.

Effect of the dataset and model hyperparameters on the model performance

The model's architecture and parameters significantly influenced its performance for grain boundary detection in Kerr microscopy images of FeNdB magnets. Key considerations included the choice of encoder networks (VGG16, ResNet50, EfficientNetB3, InceptionV3), data normalization methods, and training setup (batch size, optimizer, learning rate). Loss functions and performance metrics (IoU, F1-score) played a critical role. A combined loss function was considered, incorporating binary focal loss and distance transform loss, with a weight parameter (λ) of 0.35 for the latter.

Further, the sigmoid activation function was chosen over softmax in the output layer to enhance computational efficiency. Data augmentation techniques were employed to enhance robustness and prevent overfitting. The results highlighted the impact of encoder choice, with EfficientNetB3 yielding impressive results. Notably, a custom encoder and decoder network achieved the highest performance, with an IoU of 0.94 and an F1-score of 0.91 on the validation set. These findings underscore the importance of thoughtful architecture selection, normalization, loss function choice, and data augmentation in optimizing the model for grain boundary detection. To better understand the performance and stability of the present model based on the U-net architecture, the segmentation process is visualized using feature maps and [250], as shown in *Figure 69* and *Figure 70*.

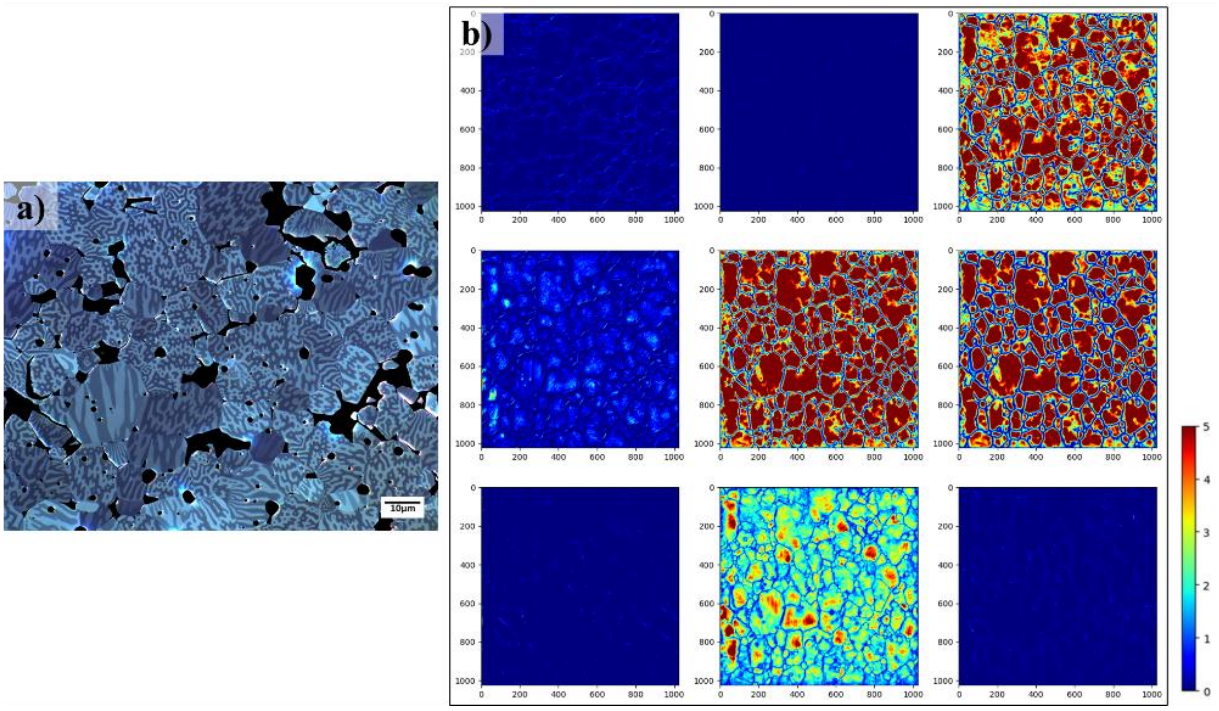


Figure 69: The activation map for input KM image to the trained model is shown. The color scale bar indicates the strength of the feature information in the KM image that contributed to the prediction of the grain boundary for the respective feature map. (a) Pre-processed KM image of the sintered FeNdB-E sample sectioned at 30° to c-axis of anisotropy (without pores and oxides) and (b) 9 feature maps from the penultimate layer (top left to bottom right) of the decoder path of the trained custom U-net model. The KM image of the sample has grains mostly with closure domain patterns. The x and y-axis for feature maps in (b) represents the size in terms of pixels (1024 x 1024 pixels).

Figure 69 visualizes the feature map for the sample with grains that mostly have closure domain patterns, whereas *Figure 70* shows feature maps for the sample with most grains that have stripe domain patterns. The activation map of the penultimate decoder layer of the trained U-net is shown. The feature strength at a given location within the feature map is shown with colormap and higher value from colormap indicates its importance for boundary detection in KM images. In the decoder part, detailed features of the input KM image are obtained, and the morphology of the grains is continuously enriched. The activation maps show that the trained model has learned to identify different grains within KM images as it gives high importance (red color) to grains. The effective generalization ability of the current model is attributed to the well-balanced complexity of the designed U-net architecture, enabling it to extract insights from the dataset fully.

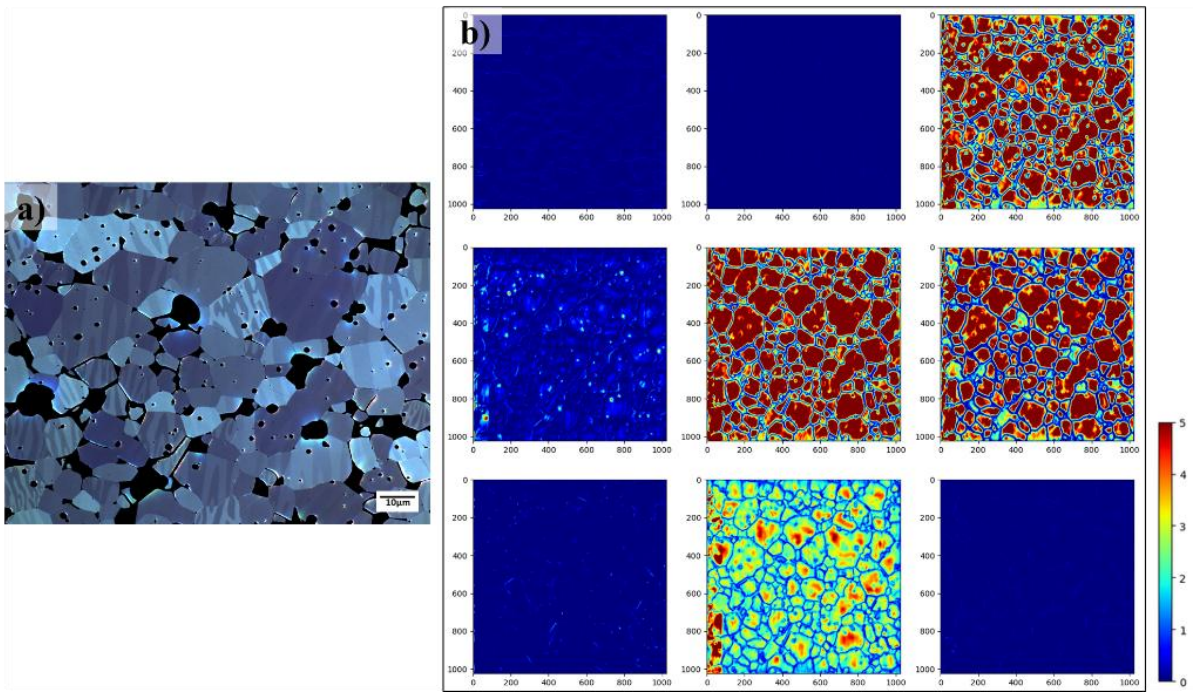


Figure 70: The feature activation map for input KM image to trained model is shown. The color scale bar indicates the strength of the feature information in the KM image that contributed to the prediction of the grain boundary for the respective feature map. (a) Pre-processed KM image of the sintered FeNdB-E sample sectioned at 90° to c-axis of anisotropy (without pores and oxides) and (b) 9 feature maps from the penultimate layer (top left to bottom right) of the decoder path of the trained custom U-net model. The color scale bar indicates the strength of feature information in the segmentation process. The KM image of the sample has grains mostly with stripe domain patterns. The x and y-axis for feature maps in (b) represents the size in terms of pixels (1024 x 1024 pixels).

Additionally, the model's accuracy hinges on the availability of trustworthy ground truth, serving as the accurate reference from which the deep learning model derives its knowledge. The importance of skip connections/layers in U-net can be seen in *Figure 71*, which directly maps the convolution layer feature to the corresponding decoder layers so that detailed information is not lost in the segmentation process. Utilizing "skip layers" for processing feature maps results in improved resolution than the original. This enhancement is attributed to the incorporation of more detailed information. These "skip layers" serve a dual purpose: they facilitate the training of network parameters using ground truth data, and they enable the morphological information initially extracted in the contracting path to be employed once more in constructing the output in the expansive path.

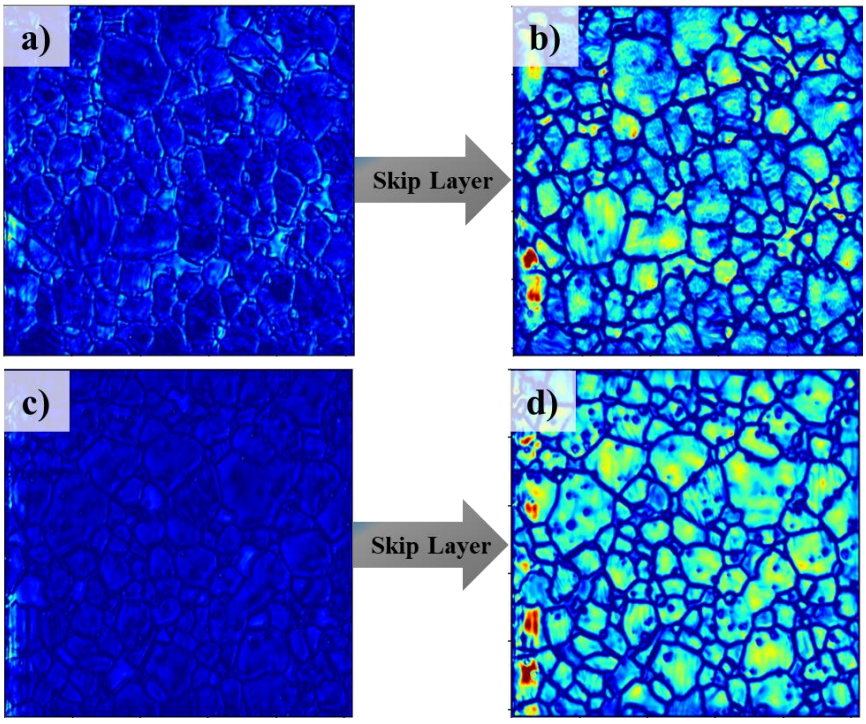


Figure 71: The visualization of the feature map before and after the skip connection is shown for one of the decoder paths in the trained custom U-net model. (a) & (b) are for the sintered FeNdB-E sample sectioned at 30° and (c) & (d) for the sintered FeNdB-E sample sectioned at 90°. The colormap is the same as in figure 68.

Effect of input image size on the learning model

The input image to the U-net should be divisible by 32, and the training images for grain boundary detection are 1024 x 1024 pixels. These image patches of size 1024 x 1024 pixels are obtained from the KM images acquired at 1000x magnification, resulting in each tile image of size 2752 x 2208 pixels. The average ECD of grains in the magnet samples used for training and evaluating the models is 10 µm. The input image size of 512 x 512 pixels reduces the receptive field, and also the spatial information needed for the U-net model to learn the appearance of grains with different domain patterns is lost. During the experiments, it was seen that the model trained on input images of size 512 x 512 pixels resulted in higher misclassification than the model trained with images of 1024 x 1024 pixels. The misclassification when compared to EBSD accounted for nearly 7% when tested on FeNdB-B images. *Figure 72* shows the tile image of magnet samples acquired at 1000x magnification with 512 x 512 pixels and 1024 x 1024 pixels patch extracted from it. In the case of microscopy images of materials, the distribution of the different regions of interest has significant spatial information appended to the material's characteristics. The input image for training a grain boundary detection model from KM images of magnet requires the image to have information on its neighboring grain appearance so that it learns to classify one grain from another to determine the optimal grain boundary. *Figure 72(c)* is a 512 x 512 pixels patch from the original tile image of 2752 x 2208 pixels and lacks the morphological features related to each grain, as two grains are truncated.

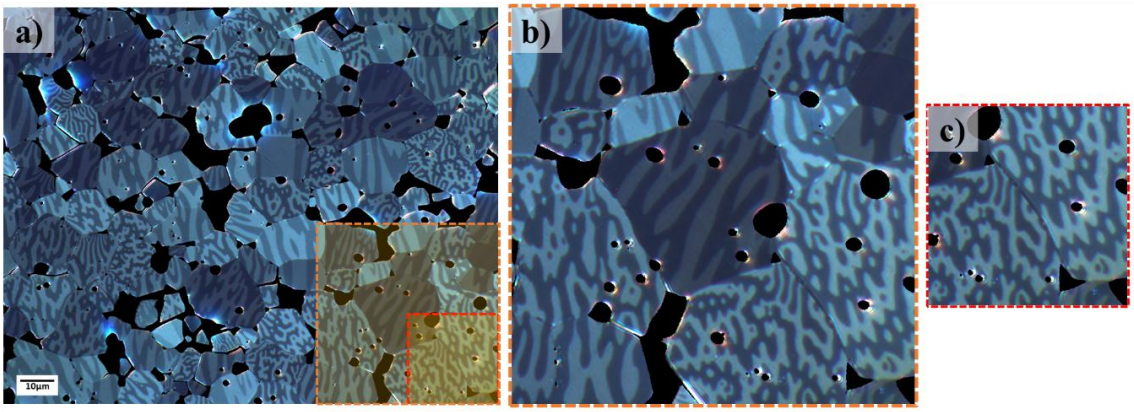


Figure 72: The tile image of the sintered FeNdB magnet samples acquired using a 100x objective and 10x magnification (lens) in (a) has a size of 2752 x 2208 pixels. The patch shown in (b) has size 1024 x 1024 pixels, and patch size of 512 x 512 pixels is shown in (c). The patch in (c) lacks the morphological features concerned with each grain, as two grains are truncated.

Robustness of the trained custom U-net model for grain boundary detection

To determine its robustness, the trained model has been extensively tested on magnets with different sample characteristics. The application of the trained model was tested on KM images of samples with different degree of alignment and also on the samples with heterogenous texture such as PSD. The D50 results from the trained model when compared to EBSD shows that the deviation is less than 4.3% when measured using the area-weighted grain distribution curves shown in section 4.3. This shows that the model has generalized well on the dataset and is not affected by the change in the appearance of the grains or domain patterns in terms of color and contrast. The KM images of the AP, TP, IP, and PSD samples vary in terms of domain contrast and brightness compared to the training images. This robustness in the trained model has a major contribution from the augmented data produced during the training step, including populating the existing training images with manipulated images by changing color, brightness, contrasts, etc. This step prevents the model from overfitting and fits well on different variations of the training images so that it is robust enough to predict unknown test images with good accuracy.

Additionally, the trained model fits well on the isotropic samples with finer and coarser grain size than the training samples. *Figure 73* shows the example image from two isotropic samples and the predicted grain boundary from the trained model overlaid on it. *Figure 73(a)* is a quasi-isotropic sample FeNdB-Q-Iso with coarse grain and has a mean ECD of 14 µm whereas *Figure 73(c)* is an isotropic sample FeNdB-Iso with fine grains and has a mean ECD of 6 µm. On the visual inspection, the trained model performance is good as it has detected most of the grain boundaries from KM images.

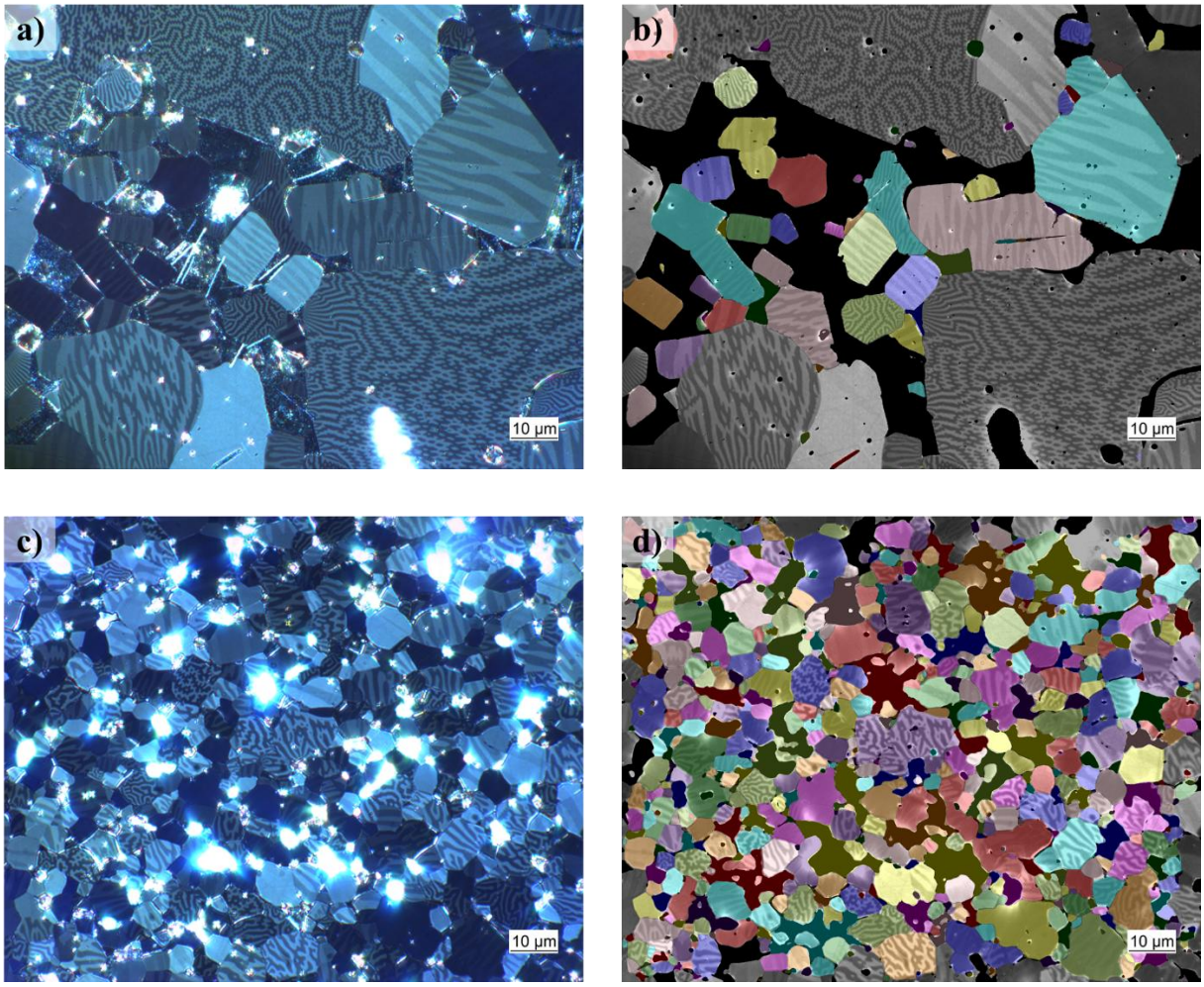


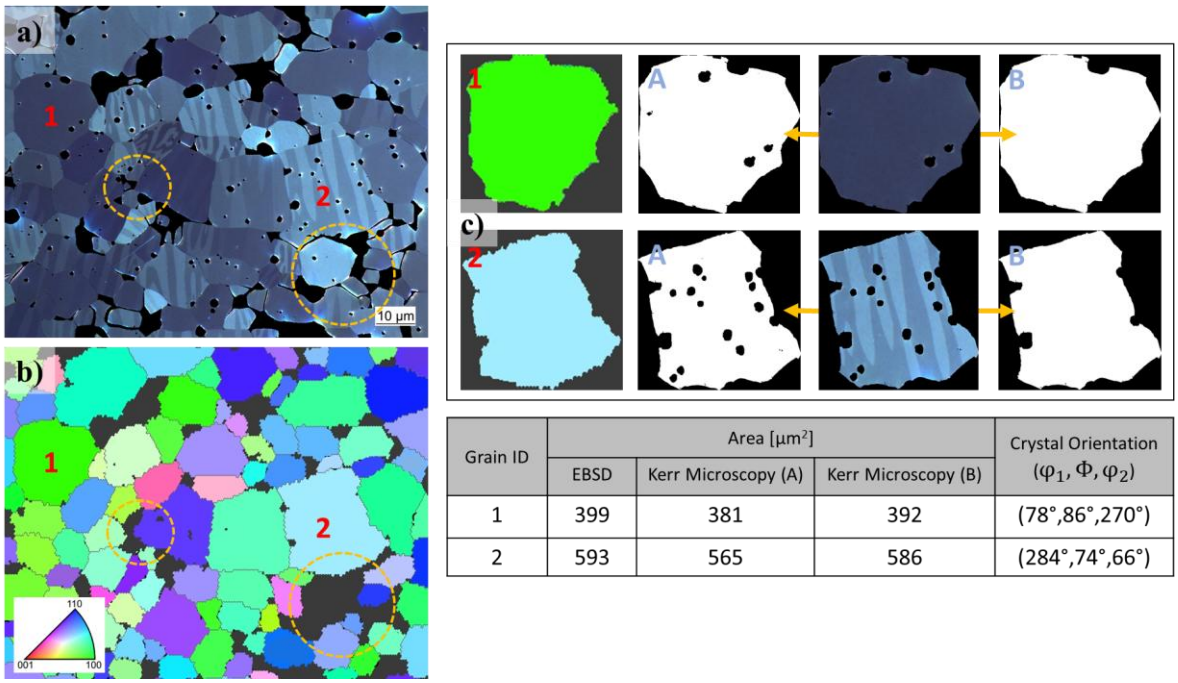
Figure 73: The grain boundary prediction for the isotropic sintered FeNdB magnet was obtained by the trained DL (custom U-net) model. (a) KM image of the quasi-isotropic magnet FeNdB-Q-Iso with grains size 14 µm, (b) predicted grain boundary overlaid on original KM image (a) with random color for each grain, (c) KM image of the isotropic sample FeNdB-Iso with grains size 6 µm and (d) predicted grain boundary overlaid on original KM image (c) with random color for each grain. The grains that touch the edges are excluded.

In comparison, the developed model's performance on isotropic samples with higher ECD value of 14 µm (coarse grains) is better than for samples with lower ECD value of 6 µm (fine grains). The coarse grain sample here offers a better perceptive field, thus assisting in a better decision-making process for the trained model. Nevertheless, the predicted result for sample with smaller ECD value is still good enough for analysis as trained model produces it in less than a minute and has detected most of the grains when tested visually.

Discrepancies between EBSD and predictions from a custom U-net model

Figure 74 shows the observed discrepancies between EBSD values and the predictions from U-net model, consistently revealing EBSD values to be lower by 0.6 µm for D25 and D50, for D90 it is lower by 3.05 µm, when measured for the area-weighted curve, can be attributed to several factors mainly to the methodologies employed by these two approaches. D50 offers better agreement since small grains disproportionately affect results and vary in number, while unresolved large or coarse grains impact D90 values. Area-weighted grain size analysis has the advantage that it reduces the effects on mean grain size from a large number of a small grain

sizes, even though their contribution to total area is only a small fraction. An important factor to consider is the propagation within the EBSD methodology. Each step involved in EBSD processing, including signal-to-noise issues, calibration inaccuracies, and orientation measurement errors, has the potential to introduce errors. These errors, although often minor individually, can accumulate and influence the final measurements. During the post-processing steps to clean the raw EBSD map, the morphology of the grains changes [266] such that the pores and oxides inside the grains are filled with the nearest pixels. *Figure 74(c)* shows that the pores are replaced by neighboring pixels for the grains marked 1 and 2. Further, the morphology of the grains is also influenced by the applied post-processing step in EBSD [266]. Another factor to consider is the error in the grain size measurement using manual approach, EBSD based approach, and DL-based approach from the KM images. Using the EBSD approach, depending on the sample characteristics to obtain a minimum accuracy of 10%, 6-8 pixels are needed. For materials such as single phase Al-Mg alloy [287] and equiaxed titanium sample [288] with mean ECD of 30 µm, atleast 5 pixels per grain are needed to obtain 10% accuracy. Similarly, for magnesium alloy, pixel clusters of less than 3 pixels were not considered for analysis to minimize the error [315]. In the case of the manual approach from KM images, the source of measurement error is prone to user bias and complexity of grain geometries that under represents small grains. During the labeling task, grains which are less than ~ 12 pixels are not recognized accurately, mainly because the thickness of the labeled grain boundaries is between 3-5 pixels and when using 5 pixels grain boundary, this error is comparatively higher than using the 3 pixel grain boundary. This error in grain size measurement from manual approach also gets extended to the DL-based approach as it has been used for training the model. However, further investigation of the grain size comparison between the results from KM (trained models) and EBSD suggests that the KM provide flexibility to carry out precise grain size analysis by eliminating the after effects of the post-processing step carried out as part of EBSD.



Grain ID	Area [µm²]			Crystal Orientation ($\varphi_1, \Phi, \varphi_2$)
	EBSD	Kerr Microscopy (A)	Kerr Microscopy (B)	
1	399	381	392	(78°,86°,270°)
2	593	565	586	(284°,74°,66°)

Figure 74: (a) and (b) Correlative KM and EBSD map of the anisotropic FeNdB-E sintered magnet. The effect of processing raw EBSD data effects the grain area and shape. The absence of grains in the EBSD map due to post-processing is marked in orange circles (b). The difference in the area of two grains using the measured EBSD and the calculated value from KM using the approaches (A) and (B) are shown in (c) [68].

Nevertheless, using the KM image for the grain size analysis approach offers more flexibility in that the measurement of grain structure can be carried out via two methods. The Kerr microscopy approach (A) involves measuring the grain with actual internal structures, so pores and oxides are not included. Kerr microscopy (B) involves filling the regions containing pores and oxides within the grains with the nearest pixels and bringing them closer to the EBSD measurements. The grain size measured using all three approaches is shown in *Figure 74*, and the difference in the results is due to the change in the morphology of the grains in the EBSD map. This highlights the importance of Kerr microscopy for sample characterization involving grains, as it provides information that is close to reality.

The grains which are visible in the KM image (*Figure 74(a)*) are not there in the EBSD map (*Figure 74(b)*) due to the filtering step to clean the raw EBSD map. The grains with pseudo-orientation or noise are eliminated. Therefore, some difference in the grain size analysis results is expected between the developed approach and the EBSD measurements. However, this difference is between 3% and 5% for the various samples tested for performance evaluation.

5.5. Application of the trained grain boundary detection model on different magnet samples

In this chapter, grain size analysis with a focus on two distinct applications are explored: anisotropic AP-TP-IP samples and PSD samples. The results from the trained U-net model for both samples provided in section 4.3 are analyzed using the grain size curves.

Grain size analysis in anisotropic AP-TP-IP samples

In terms of area-weighted grain sizes (D25, D50, D90), the relative error percentages are generally within a comparable range (< 5%) for D50, but for D25 and D90 varies largely. The main source of differences in results would appear to be the choice of the smallest grain size (D25) to be included in the calculation. Choice of a relatively large cut off size (D50) is likely to give the best agreement because the smallest grains can have a significant effect (disproportionate to the area they occupy) and are also the grains most likely to vary in number. With large grains, we have the problem of coherent, unresolved grains, which effects the D90 values. The absolute prediction error for AP-TP-IP samples from trained U-net model suggests that there are no large variations in the MAE when calculated against EBSD. The MAE for AP to IP has been 5.4 μm^2, 5.0 μm^2 and 4.5 μm^2 respectively. Overall, the relative errors suggest that the degree of alignment did not have a significant effect on the grain size measurement using the developed DL-based approach.

Grain size analysis in PSD samples

Similarly, the variation in the grain size values obtained form cumulative grain size plots, the focus should be on D50 values from area-weighted grain sizes, which accounts for relative error percentage between 4.4% to 7.3%. When we calculate the average relative error across all percentiles (D25, D50, and D90), we find that it ranges from 4.2% to 5.1%. Additionally, the absolute prediction error for PSD samples in convergent and divergent positions differs from PSD-01 and PSD-11 samples, resulting in higher MAE compared to PSD-05 and PSD-15. The

prediction error in terms of MAE for convergent position samples PSD-01 and PSD-11 are 4.7 μm^2 and 3.5 μm^2 respectively. For divergent position samples PSD-05 and PSD-15, the prediction errors are 5.3 μm^2 and 5.0 μm^2 respectively. In the divergent position samples here, the grains have a higher θ value of more than 120º and the grain boundaries and grains have fine differences in gray values, making it difficult to detect the grain boundary. This suggests that, on average, the DL approach exhibits a minor level of relative error compared to EBSD across the different samples and percentiles. These differences in results suggest that the performance of the DL approach in estimating grain size distributions can vary depending on the specific characteristics of the sample and the selected percentile values.

The experimental results of the DL-based model for detecting grain boundaries from KM images of FeNdB suggest that it is 10 times more time-efficient, more reliable with lower measurement errors than both EBSD and the manual approach, and a robust method that fits various samples, with the mean ECD ranging between 6 µm and 14 µm for grain size analysis. Further, developed model here serves as a basis for obtaining the crystallographic grain orientation information from KM images and micro-magnetic domain analysis.

5.6. Strengths and limitations of the grain boundary detection model for sintered FeNdB magnets

The grain boundary detection model developed for the sintered FeNdB magnets from the KM images is associated with some strengths and also has a limitation on application due to the inductive bias of the training model and the prediction interval due to the datasets used.

Strengths of the developed approaches

One of the significant strengths of the grain size analysis using the KM image is that a large sample area can be captured in less time compared to the EBSD or other diffraction-based approaches. However, since grain boundary detection was manually traced to calculate the grain size based on international standards, it is time-consuming and induces fatigue error. On the other hand, EBSD, a state-of-the-art approach, is also time-consuming for samples that requires large areas to be analyzed. However, the developed U-net model for grain size analysis from KM images overcomes the challenges of the manual and EBSD approaches by conducting analysis on large areas of the samples in less time and having accuracy close to the manual and EBSD approaches. For example, for the magnet of area 92x67 μm^2, the developed model needs less than 3 minutes to generate a grain boundary mask that can be later processed both by planimetric and intercept-based methods automatically using external platforms such as Fiji [316] or ZenCore [317], etc. In comparison, it takes more than 30 minutes to obtain the grain boundary mask using a manual or EBSD approach for the same sample area. Therefore, there is nearly 10x faster analysis and more robust to carry out recurring quality assurance using the developed approach.

Moreover, the discrepancies in the EBSD results due to the post-processing steps leading to changes in the size and shape of grains that are used for quantification of morphology can be overcome with the flexibility of the developed U-net model as discussed in section 5.4, which highlights the variation in results from EBSD and developed model for grain size analysis.

Further, the developed model is robust and has generalized well, which is evident from applying the trained model on samples with different manufacturing processes and characteristics as described in section 4.3.1. The developed model is also one of the major reasons for exploring research on obtaining the crystallographic grain orientation information from the KM images. This is mainly due to the effective approach for obtaining the picture gallery with individual grains in the KM images, which is the basis for developing a grain orientation prediction model and for micro-magnetic domain analysis to obtain the intrinsic properties from the KM images.

Limitations of the developed approaches

Some factors might affect the performance of the trained DL model. The trained model's performance would decrease or fail to detect grain boundaries if the KM image resolution is too low, has strong preparation artifacts such as deep scratches passing through the grains, and the acquired image's magnification is less than 500x for comparable grain sizes. Further, the developed DL approach exhibits certain limitations when applied to magnet samples with a significant number of grains with the closure domain structure. *Figure 75(a)* shows an example image where a grain boundary detection task using traditional methods proves ineffective, and resorting to a manual approach becomes very challenging. *Figure 75(b)* shows the grain boundaries the trained DL model detected. In comparison, the EBSD map is generated for the same sample in the same position using correlative microscopy, as shown in *Figure 75(c)*. It is evident that the DL model does not detect the grain boundaries in specifically marked sample areas, as observed in *Figure 75(b & c)*. However, the EBSD map separates the grains depicted in *Figure 75(d)*.

One of the primary explanations for the model's inability to detect certain grain boundaries in the sample could be attributed to the composition of the training dataset. The training data primarily comprise images of samples with more grains featuring stripe domain structures or a combination of grains with both stripe and closure domain structures. Additionally, manually labeling images of samples with a greater number of grains exhibiting closure domain structures poses a formidable challenge, given the high probability of false positives.

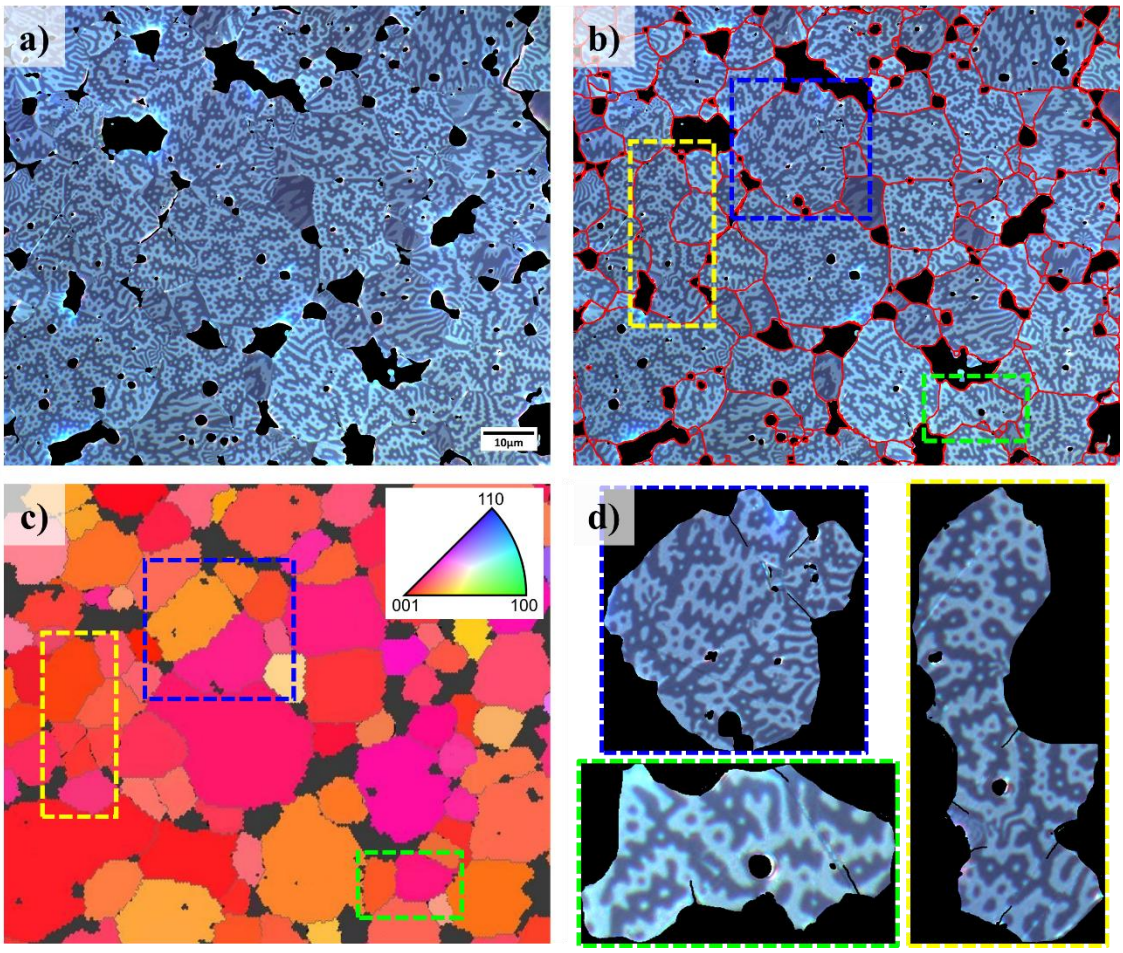


Figure 75: Visual comparison of grain boundary detection methods in a FeNdB sample with many closure domain structures. (a) the original image, (b) grain boundaries detected by the trained DL model highlighted in red, (c) the EBSD map displaying grain boundaries, and (d) individual images of grains extracted from the DL model where grain boundaries were not detected in comparison to the EBSD map.

A deeper insight into the grain and its misorientation and Euler angles [φ_1, ϕ, φ_2] representing crystallographic grain orientation were obtained using the EBSD approach [59]. When the relative difference in the φ_1 component of grain orientation is substantial, the grain boundaries become distinctly visible, leading to improved grain boundary detection by the trained model. This effect is noticeable as the ϕ value increases, explaining why the trained model successfully identifies the grain boundaries in certain grains within the sample shown in *Figure 75(a)*, but struggles to do so in the grains shown in *Figure 75(d)*. In the cases where the trained model fails to detect grain boundaries, the relative difference in φ_1 values of the grains is minimal, and the relative misorientations ($\Delta\phi$) between the grains fall within the range of 2° to 8°. In contrast, the grains the model identifies grain boundaries exhibit a substantial difference in φ_1 values, even though they have low ϕ values. Furthermore, it becomes a formidable task for the trained model to identify grain boundaries when the relative misorientation is low. An examination of the EBSD inverse pole figure map, which exclusively illustrates the ϕ value, reveals that most of the grains in the sample in *Figure 75(a)* possess a low ϕ orientation.

5.7. Evaluation of the crystallographic grain orientation prediction model for sintered FeNdB magnet

In this chapter, the complexities of grain orientation analysis using KM images are discussed. The challenges in extracting grain orientation, the relationships between θ and domain brightness, and ρ and image contrast are unraveled. Emphasizing the significance of deep learning, the discussion involves dataset and model impact, strategies to avoid overfitting, explainability, and the intriguing concept of domain branching. These discussions provide an insight into the grain orientation analysis from KM, its challenges, and the developed approaches.

The study demonstrated that variations in the orientation angles φ_1 and ϕ induce changes in image features such as domain contrast, domain structure orientation, and grain entropy extracted from KM images. However, a rotation corresponding to the φ_2 angle remains undetectable through polar MOKE and may not impact magnetic properties. Additionally, variations in the φ_2 value do not yield noticeable changes in the KM image of the magnet. As discussed in section 3.2.2, the training of models involves utilizing the reduced EBSD Euler angles φ_1 and ϕ due to these considerations.

Therefore, the developed DL-based approach for grain orientation analysis measures the reduced φ_1 and ϕ components of the Euler angles and is superior to the manual approach which measures only φ_1 but has limitations that it does not detect or measures the φ_2. *Figure 76* graphically shows the distinction between the manual approach, EBSD and DL-based approach for texture analysis in magnet samples.

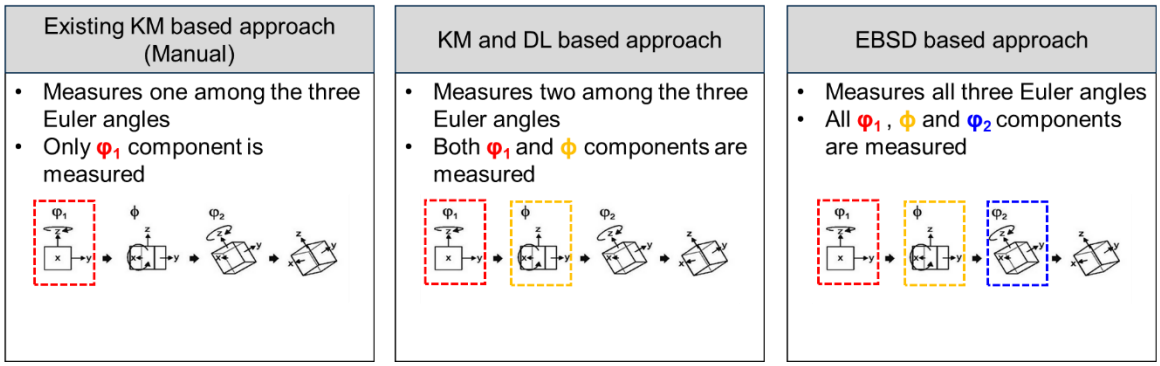


Figure 76: Comparison between the manual, developed DL-based and EBSD approach for the grain orientation measurement as Euler angles. Note that the cuboid representation is used here for better visualization of grain orientation. In general, it is tetragonal.

Analytical difficulties in obtaining the grain orientation from KM images

Correlation between the theta value and domain brightness/colorfulness

The colorfulness or brightness of the grain has a correlation to the θ value as seen in the *Figure 77*. It is seen that the colorfulness or brightness of the image increases with increase in the θ value. However, this does not hold true for the complete range of θ value which is between [0°, 180°]. For the 1000 grains analyzed randomly for an anisotropic sample, the grains with θ value between 50° to 130° shows positive correlation. The behavior for grains with θ values less than 50° and more than 130° has a negative correlation.

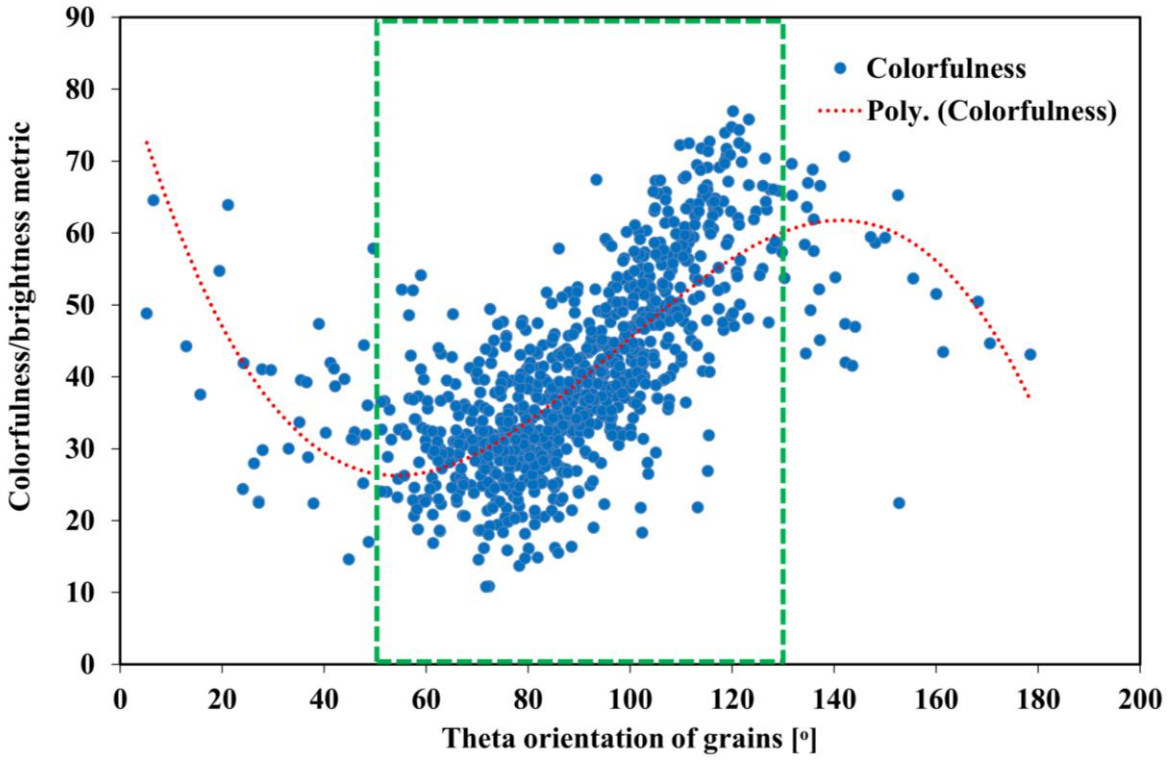


Figure 77: Plot showing the relationship between the EBSD measured θ value and the corresponding measured colorfulness value for the grains from KM for the FeNdB sample. The green marked region shows that the grains with θ value between 50° to 130° shows a positive correlation.

Further, a detailed investigation on this using the Goniometer was carried out which revealed that the θ and colorfulness correlation is dependent on other factors such as type of domain pattern. The transverse Kerr component (θ value) leads to a change in the intensity of the reflected light depending on the angle of rotation. The acquired images show that for each grain the magnitude of change in the colorfulness or brightness is different and also the range for which correlation is positive or negative is varying. Further, it can be seen that the colorfulness value varies differently depending on the structure of the domain pattern. For example, the plots for the grains without domain patterns in *Figure 78(a & b)* have non-cyclic trend with maximum colorfulness (> 90) at higher Goniometer rotation (130º - 150º). For grains with closure domain structure, the variation in the color value with the rotation of the Goniometer is low and such that maximum value is less than 70. On the other hand, for grains with stripe domain structure, it has cyclic trend with change in the Goniometer rotation.

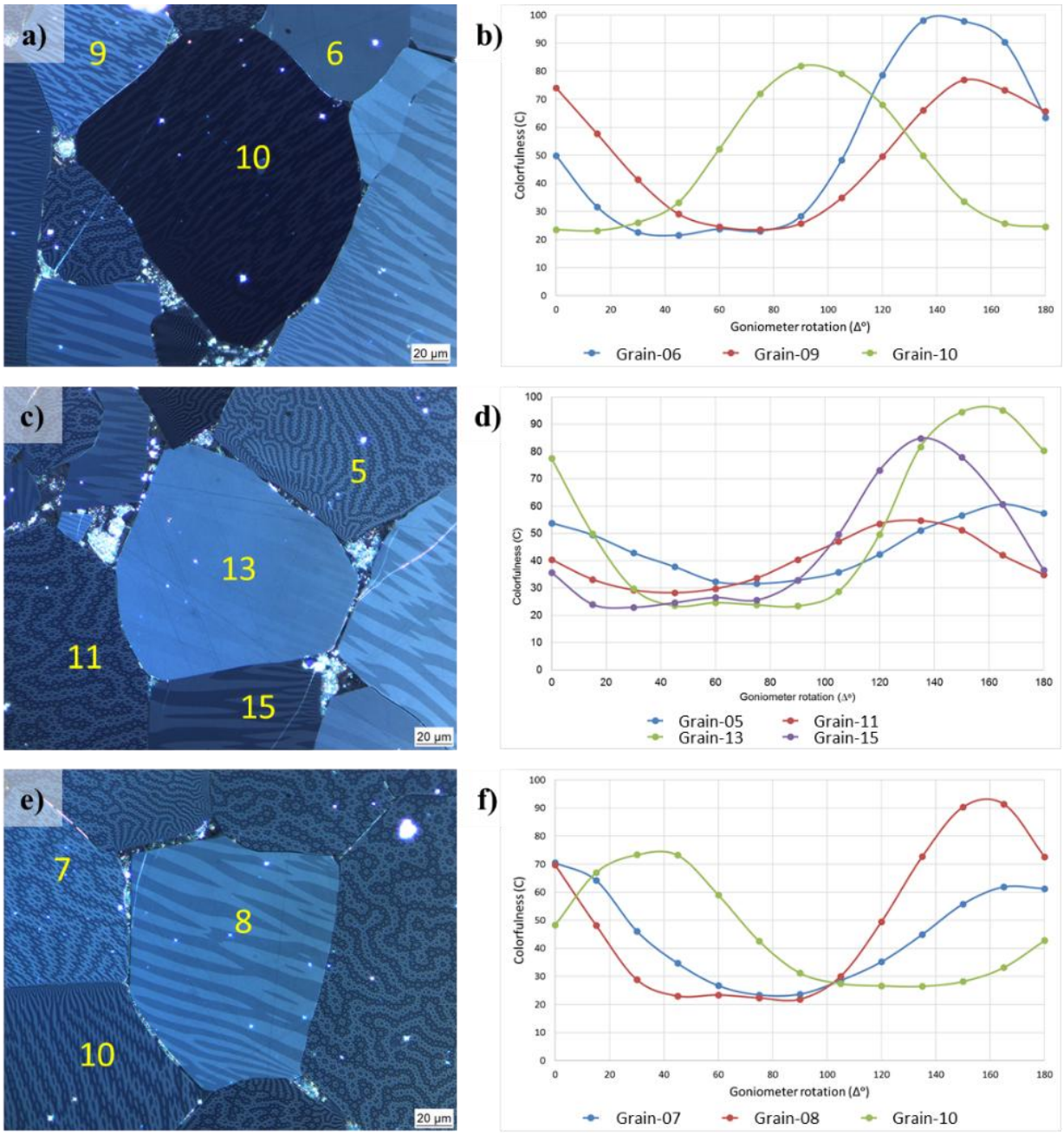


Figure 78: The isotropic sample with 6.0 s exposure time acquired using KM at 500x magnification using the Goniometer is shown. The sample is rotated in steps of 15°, which represents the change in the transverse Kerr effect or θ value of the grains in the sample. (a), (c), and (e) are original KM images with assigned grain position in yellow, and (b), (d) and (f) shows colorfulness of the numbered grains against the transversal Kerr effect. The colorfulness value changes its behavior with regards to the domain pattern in the grain. (Figure unpublished Choudhary & Tvrtko, with permission from [296]).

Correlation between rho and contrast of domain structures

Additionally, *Figure 27* and *Table 11* in section 3.2.2 shows that the θ and ρ values have correlation with the image features such as brightness, domain structure pattern, orientation of domain patterns and domain pattern contrast. *Figure 79* shows the measured domain contrast for each grain and its corresponding ρ values for the anisotropic sample. With the increase in the ρ value the measured contrast value decreases such that the grains with highest ρ value have either no visible domain patterns or they have domain patterns with very low contrast. Similarly,

one can see that the grains with high contrast have low ρ values. Also, one has to note that there could be other features of grain images from KM that it getting affected by change in ρ values such as the image texture represented as the Haralick parameters.

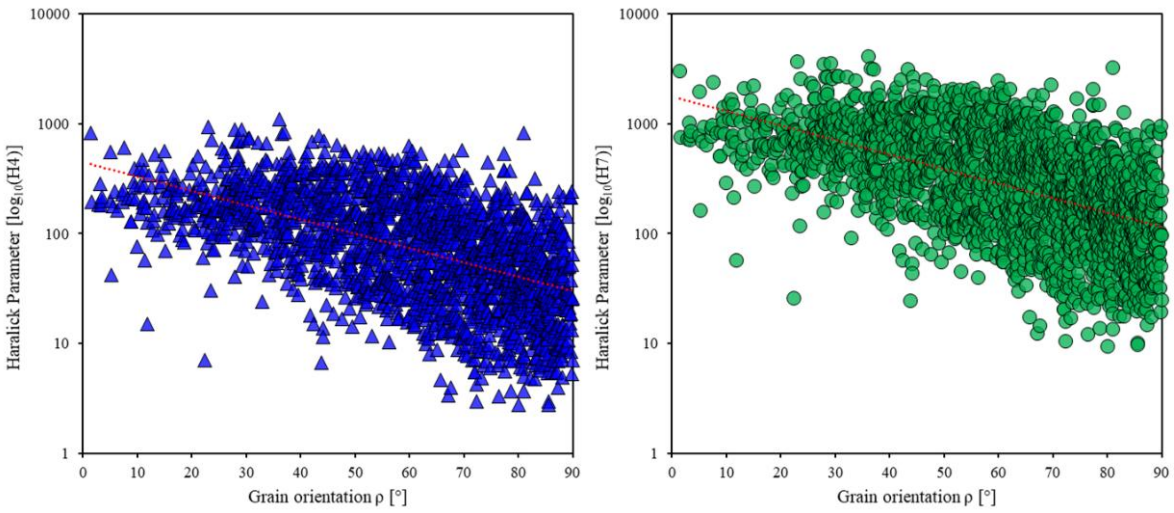


Figure 79: Plot showing the relationship between the EBSD measured ρ value and the corresponding calculated contrast value of the domain structures within the grains from KM for the FeNdB sample. The calculated contrast value here is calculated using the Haralick parameters H4 (sum of squares: variance) and H7 (sum variance) [91].

Such is the complexity of the correlation between grain orientation and its appearance in the KM images. Therefore, different image features are extracted to train the ML model such as color, shape, texture, morphology, and statistics from the domain structures. The performance of the ML model is highly dependent on the hand-crafted features extracted from input data, and it is subjective in nature and requires prior knowledge of the correlation between input data and expected output along with the physics behind it. For θ value prediction, the contribution of texture features characterized by Haralick Parameters emerge as the most influential contributors, with 64% contribution. Followed by the color feature, as RBG colorfulness having a positive impact of 10%. However, the shape features, extracted as the Zernike moments, had a negative contribution of -14%, indicating a potential limitation within the A-RF model. Morphology features, such as orientation, aspect ratio, and area, collectively contribute positively with 12%. Meanwhile, domain statistics features contributed positively with 14%. For ρ value prediction, the contribution of texture features characterized by Haralick Parameters emerge as the most influential contributors, with a 68% contribution. Following closely are the color feature, with RBG colorfulness making a modest but positive impact of 9%. The shape features, represented by Zernike moments, exhibit a negative contribution of -23%, indicating a potential limitation or counter productive influence within the A-RF model. Morphology features, such as orientation, aspect ratio, and area, collectively contribute positively with 8%. Meanwhile, domain statistics features, contributed positively with 15%. The role of domain statistics has mainly been in providing A-RF model with information on domain pattern type associated with grain.

In the cases where there is no previous work that has been reported, such as in the crystallographic grain orientation case, the features to be extracted were unknown and DL approaches can be helpful as they are proven to work well when there are limitations with feature engineering step such as type and number of features to be extracted is uncertain or in

some cases extraction of features is not possible. Therefore, considering the uncertainty in different image features that affect grains in KM images with changes in θ and ρ values, DL models were also tested for performance evaluation.

Effect of dataset and model hyperparameters on the learning model

One of the pre-requisites for training a DL model is the large dataset. The dataset includes the grains with θ values ranging between 0° to 180° and ρ values between 0° to 90° and their contribution for DL model (EfficientNetB0) varies accordingly. The prediction error of less than 5.7% was observed for the grains with θ values between 70° to 160°. For grains with θ value less than 30° and more than 170°, prediction error is between 18% to 40% with exception being that for grain with θ value less than 10° error is 88% precisely. For ρ values, the prediction error of more than 10% was observed for the grains with values less than 40° with maximum error of 12.8% for grains between 0° - 10°. On other hand, the observed prediction error of between 3% - 6% was seen for grains with ρ values more than 40° with least error being for grains between 70° - 90°. However, for the grain orientation task we noticed that geometrical transformations such as rotation and color space transformations such as contrast variation, brightness change, and color variation are not suitable as this effect the corresponding grain orientation values as described earlier. Therefore, augmentation involving adding different noises, translational shift, and random cropping were used.

Additionally, the dataset has grains with stripe, closure, and patterns which are in transition phase as shown in *Figure 27*. The samples cut at 90º have the highest number of grains with stripe domain and samples sectioned at 30º and 15º have mostly closure domain patterns. It has been observed that the trained DL-model (EfficientNetB0) has varying performance for grains with different domain patterns for θ predictions. The presence of closure domain patterns in large numbers in a sample (15º sectioned) results in a high MAE of 12.3º and the error reduces as to 2.7º for a sample with predominantly stripe domain patters (90º sectioned). However, such a trend in varying prediction error for different sectioned samples was not seen for the ρ predictions as MAE for 15º sectioned samples was between 2.2º to 4.1º with the highest being for 30º sectioned sample.

Steps to avoid overfitting in spite of using small dataset for training the model

The size of the dataset-2 for training the DoPACOM model is 5424 grain images, with 83% of grains belonging to the anisotropic sample sectioned at steps of 15° and the remaining 15% from the quasi-isotropic sample. Further, the dataset is imbalanced with a greater number of grains having θ values in the range of 40° to 130° and ρ values greater than 30°. Therefore, to avoid the possibility of overfitting and improve generalization, regularized MSE loss function as shown in Equation (60) has been adopted. This inclusion strikes a balance between fitting the training data effectively and controlling the model's complexity. During training, MSW plays a pivotal role in regularization by influencing the minimization of the MSE. Further, incorporated dropout layers with a 0.6 dropout rate to prevent overfitting by randomizing neuron activations during training and batch normalization to stabilize the training process by maintaining activation values within a reasonable range, effectively reduced the risk of overfitting.

Additionally, it is important to note that the number of the training data points used for developing regression models for both θ and ρ values are less than 5000 which is not high for DL-based approaches, it is crucial that both systematic and random errors in the form of bias and variance are quantified correctly. R_{ccc} considers both accuracy (measures bias) and precision (measures variance), offering a more comprehensive evaluation of the performance of the orientation prediction model against reference data. Unlike R^2, which primarily quantifies the variance in a model, R_{ccc} is applicable to diverse data types and distributions, including non-normal and skewed data such as in for grain orientation values (θ and ρ).

Influence of domain patterns on trained model's predictability

The KM image of a magnet sample has grains with different sizes, and other phases such as pores and oxides are present in it. During the pre-processing step, they are removed from the KM images followed by grain extraction using the developed U-net model discussed in section 4.1. They are present not only along the grain boundaries, but also within the grains which creates holes in 2D images of Kerr as shown in *Figure 28*. This hole would imply that some information loss in the input grain image to the DoPACOM model could impact the prediction results.

Based on the approach described by Ribeiro et al. in [318], the DoPACOM model's behavior to the loss of pixel data in the input grain image was determined. The approach involved segmenting the grain into multiple regions by clustering, feeding each segmented region to the trained model, and evaluating the model prediction error. Anchors aim to identify the most influential features or conditions that lead to a particular prediction, helping users understand and trust the model's behavior. Anchors generate decision rules in the form of IF-THEN statements. These rules define conditions in the feature space that lead to specific predictions. Using the super-pixel based image segmentation (SLIC) [319] approach, the input image has been segmented into 3 to 4 regions. *Figure 80* shows the grain images and corresponding masks from the SLIC algorithm. The prediction error from the trained model for each segmented region is different for θ and ρ values. In case of the θ value prediction, the prediction error for grains with stripe patterns is the same as for grain images shown in *Figure 80* with the anchor image. In the case of the θ value of the grains with closure domain patterns, the difference in the prediction error is between 2% and 4%. This suggests that for grains with closure domain patterns, the loss of pixels due to pores or oxides removal is significant for large pores (or oxides).

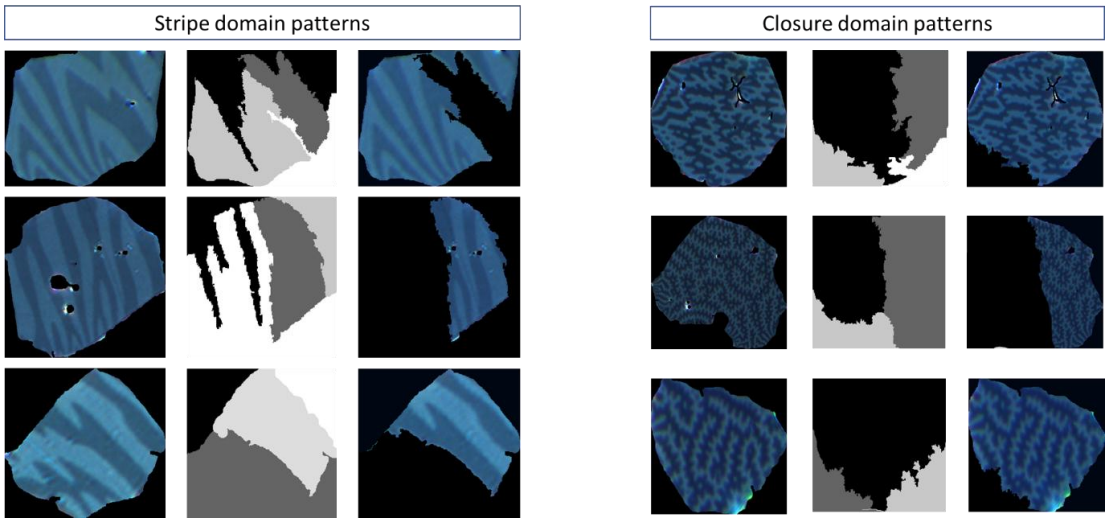


Figure 80: Example of KM images of the grains with stripe domain patterns and closure domain patterns with the corresponding anchor generated using the super-pixel based image segmentation approach. The anchors are the local sufficient information within the image which a trained model requires to predict with at least 90% confidence [318].

Additionally, for the grains with a closure domain pattern and branching effect, the prediction error for θ value differs significantly. Due to domain branching effect [320], [321], the domain pattern in the grain at the bulk region or the center has different width when compared to the surface region or edge of the grains. The appearance of the closure domain patterns at the surface or at edge region resembles the stripe domain patterns which the model fails to recognize effectively, leading to comparatively higher error. However, the domain branching effect is more severe in magnets with large grain sizes (stripe and closure domain patterns) such as quasi-isotropic magnet samples shown in *Figure 81*. In magnets with small grain size, the effect of domain branching is widely seen in grains with closure domain patterns and it is much lesser than in samples with larger grain size. *Figure 81* shows the domain branching for different grains in the quasi-isotropic sample with a mean ECD of 14 µm and the anisotropic sample with a mean ECD of 10 µm. Therefore, for such grains, the true information for the θ value prediction lies in the bulk region or center. However, detecting grains with domain branching automatically in samples with smaller grain size is challenging and has not been explored in this research.

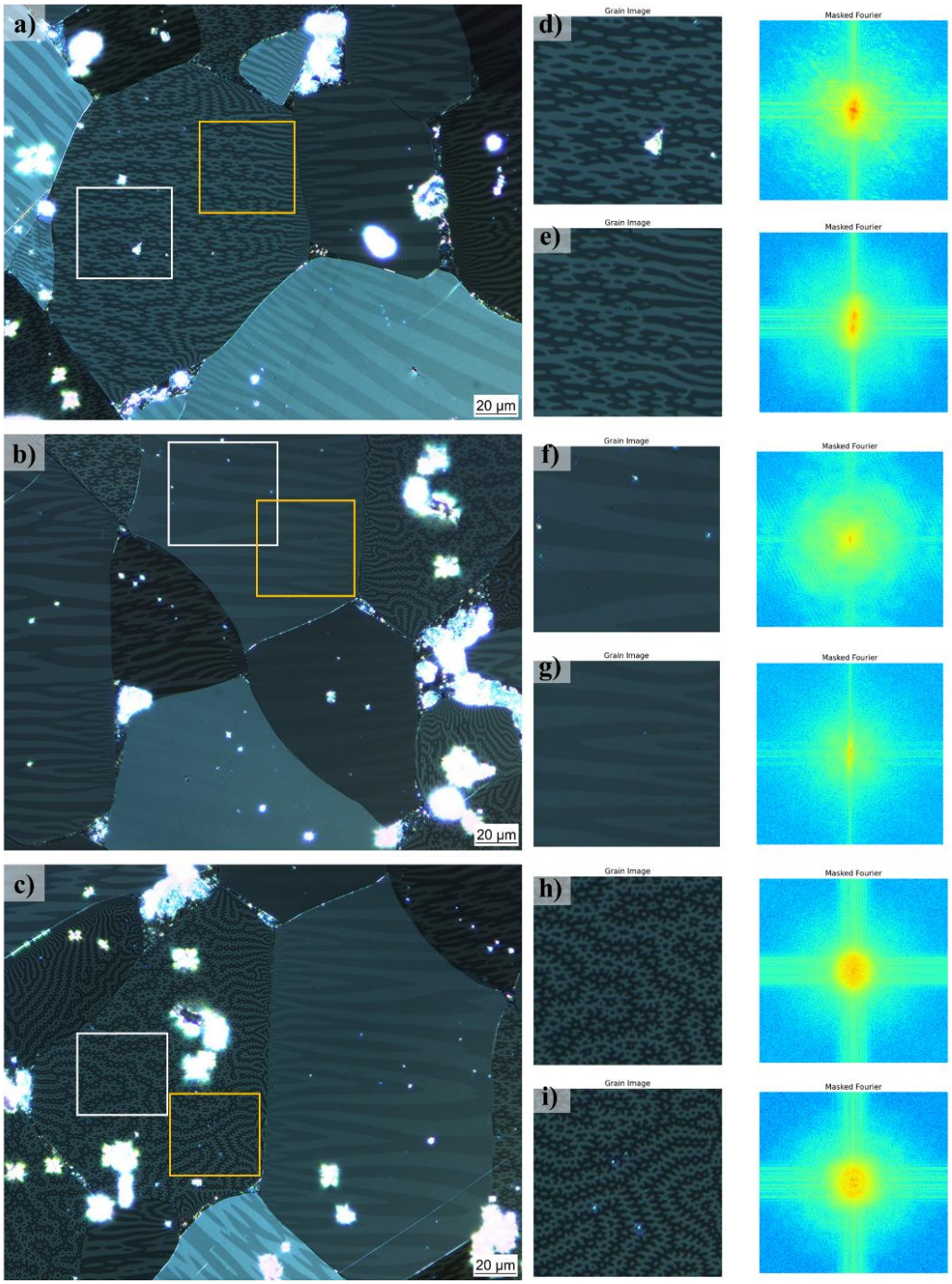


Figure 81. Domain branching effect in the quasi-isotropic sample. (a), (b) and (c) KM images with regions of interest marked to show the change in domain structure within the same grain and (d) to (i) cropped images from the marked regions in the grain and its corresponding masked FFT image to show the change in image texture.

A similar test for ρ value prediction suggests that pixel loss due to pores and oxides removal increases the prediction error which varies between 1.8% and 3% for grains with both stripe and closure domain patterns. One of the reasons for the discrepancy in the results could be that the domain contrast information, which affects the ρ value, is affected by the loss of pixels when pores and oxides in the grain are removed. Further, the domain branching phenomenon in grains did not have a measurable effect on the model's prediction for ρ value. Among the machine learning models, varying levels of performance were noted. Among these models, the A-RF model demonstrated the lowest MAE for both θ and ρ predictions, showcasing its superior

accuracy. On the other hand, the GB and XGB models showed moderately competitive results, with respectable MAE and R_{ccc} scores. However, the MLP model was less accurate, exhibiting higher MAE value and a relatively lower R_{ccc} score than other ML models, indicating less reliable predictions. When compared to MLP, tree-based models resulted in an improvement of more than 19% and more than 50% for θ and ρ predictions respectively. One of the primary reasons for tree-based models (A-RF, GB, and XGB) to outperform MLP is that it being an ANN is prone to overfitting when dealing with small tabular datasets (less than 1000 data points) and tends to produce overly smooth solutions [113].

When it comes to deep learning models, the baseline model delivered intermediate results, suggesting a moderate level of accuracy in predicting orientations. In contrast, the VGG16 fine-tuned, ResNet50V2 fine-tuned, and EfficientNetB0 fine-tuned models improved over time, as evidenced by decreasing MAE values and increasing R_{ccc} scores. Among these DL models, the fine-tuned EfficientNetB0 consistently outperformed the others, resulting in the lowest MAE value and the highest R_{ccc} score for both θ and ρ predictions. This demonstrates its remarkable predictive capacity, making it a standout choice. In summary, the DL models, especially EfficientNetB0 fine-tuned, showcase superior predictive accuracy for orientations (θ and ρ) compared to the ML models. This underlines the advantage of using deep learning techniques, particularly when fine-tuned with appropriate.

The adaptive boosting applied to the random forest regressor resulted in superior performance compared to the other conventional machine learning methods tested. In contrast to ensemble or tree-based models such as A-RF, GB, and XGB, MLP struggled to capture the information in the training dataset effectively. Additionally, manually extracted features, the input to the learning model, are a structured dataset stored in the tabular format. Additionally, artificial neural networks are generally less effective at handling structured data and are often favored for processing unstructured data, such as images [110]. Furthermore, the performance optimization of XGB has been better than that of GB and RF when applied to tabular data. However, given the size of the training set in this context, fine-tuning the A-RF model proved more effective than tuning XGB, which involves tuning hyperparameters that are very sensitive and result in large changes in model performance. The effectiveness of the adaptive boosted random forest model indicates a clear correlation between the manually extracted grain features and the actual θ and ρ values. The feature importance values suggest the substantial contribution of the texture features, color features, morphology, and domain statistics among all the extracted features used for model training. In contrast, the impact of shape features was minimal, suggesting their limited significance in improving the A-RF model's generalization. As a result, the shape features were discarded from the feature set.

However, compared to traditional ML models, fine-tuned DL models exhibited superior performance. The baseline model, which featured a basic DL architecture comprising only fully connected layers, struggled to generalize the training data. The prediction error from the baseline model was nearly 2.5 times greater than that of the EfficientNetB0 model and nearly 2 times higher than that of the fine-tuned VGG16 and ResNet50V2 models. This highlights the importance of the transfer learning strategy and suggests that DL models directly trained on the FeNdB-A magnet grain dataset face challenges in achieving adequate generalization to capture the relationship between grain features and their orientations. Therefore, using pre-trained models has an advantage in extracting meaningful features from input data, even though they

have been trained on completely different knowledge domains, as such as ImageNet dataset [286].

Realizing the complexity of the grain orientation task, it is important that the pre-trained models chosen for fine-tuning later should be able to identify the regions of interest from input image. Therefore, fifteen different pre-trained networks were tested on the grains with different domain patterns and activation maps were obtained. The pre-trained networks whose activation map resulted in having more more focus on the grain and its domain patterns as shown in *Figure 82* were short-listed for further tuning the θ and ρ prediction models. The heatmaps shown in the *Figure 82* are generated using Grad-CAM (Gradient-weighted Class Activation Mapping), a technique used for visualizing and interpreting the decisions made by CNNs. Grad-CAM works by computing the gradients of the target value (the class that CNN predicts) with respect to the feature maps of the last convolutional layer. These gradients are then weighted by the average pooling of the gradients. By doing so, Grad-CAM highlights the regions of the input image that are most important for CNN's prediction. The jet color bar has been used for visualization, representing the intensity of the activation in the heatmaps. The colors like blue and green typically indicate lower activation, while colors like yellow and red indicate higher activation.

The pre-trained networks which failed to identify the grains and domain patterns were removed. VGG16, ResNet50V2 and EfficientNetB0 were shortlisted for further fine-tuning and, based on the validation error, EfficientNetB0 was considered for optimized tuning for θ and ρ prediction tasks. Note that despite the popularity of the DL, there is no one-size-fits-all solution. A single universal architecture, set of parameters, loss function, training approach, or initialization cannot seamlessly address all kinds of existing problems. It is not practical to employ it as a ready-made classifier for diverse tasks. Although there is an approach to the application problems with a neutral perspective, it is crucial to acknowledge that extensive data-driven exploration is still essential to tailor the learning model to the specific characteristics of each individual problem.

Since the target for the learning model is a real number and a continuous value, it is a regression task; therefore, the classification layer from the pre-trained network is replaced by the regression layer. A typical practice involves employing multiple FC layers after several iterations of convolution. The arrangement of the last convolutional layer is flattened before being linked to the subsequent FC layer. In the case of a regression network, the ultimate layer is consistently fully connected, featuring a single output unit designed for the regression target.

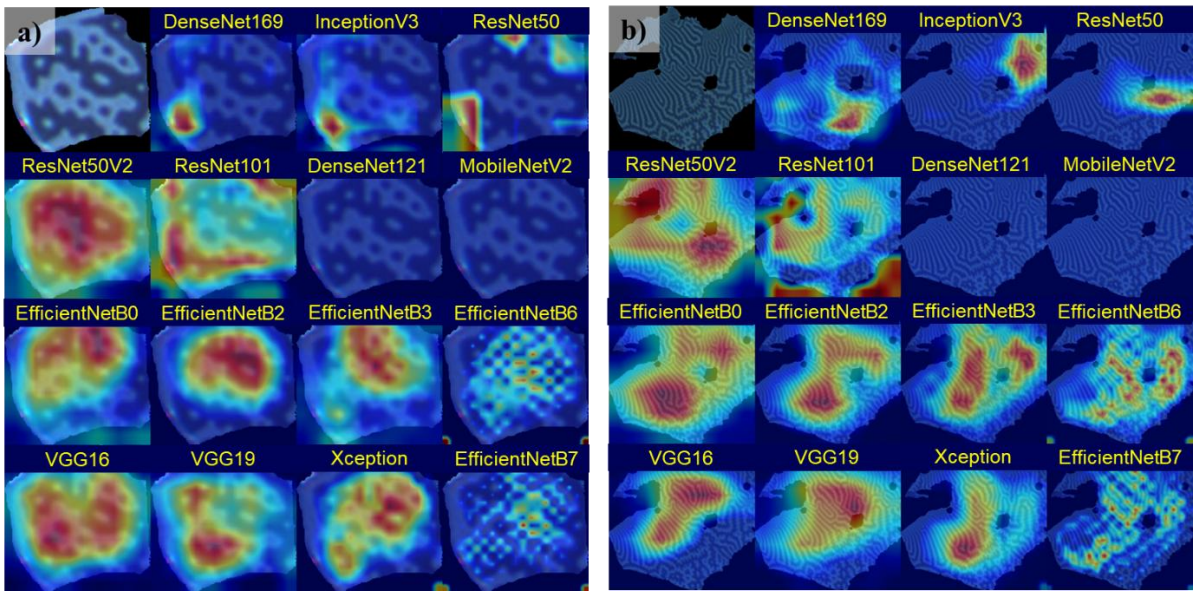


Figure 82: Activation map on the KM image of the grain for the 15 different pre-trained models. (a) Models trained on ρ values and (b) models trained on θ values. The color bar has been used for visualization and represents the intensity of the activation in the heatmaps with lower activation indicated by blue and green, while higher activation indicated by yellow and red color.

One possible explanation for the fine-tuned EfficientNetB0 model's higher error in predicting θ values for grains with orientations below 10° and above 170° could be due to the limited representation of such grains in the training dataset, as mentioned in section 4.1. In the context of ρ prediction, the A-RF model's performance reduces for grains with ρ values exceeding 70°, despite the training dataset containing a substantial number of grains within the 60° to 90° range. Note that, ρ values have a considerable impact on grain domain contrast [322]. Grains with higher ρ values exhibit low domain contrast compared to those with lower values. While the texture feature provides some aspect of this contrast information (the 2nd Haralick parameter), the A-RF model's inability to generalize over the entire dataset distribution might be due to the skewed training set distribution, as shown in *Figure 26(a)*, and the model's limitation to learn from the existing feature vector to identify the relationship between the influencing domain contrast in grains. In contrast, the fine-tuned EfficientNetB0 model shows a different prediction behavior, resulting in a superior generalization compared to the ML model. This underlines the significance of the DL approach's automatic feature extraction, which can capture a wide range of features locally and globally from input images to improve learning process.

5.8. Application of the trained grain orientation prediction model on different magnet samples

In this section, the results from the application of the trained grain orientation prediction model on homogeneous AP-TP-IP and heterogeneous PSD FeNdB samples are discussed in detail.

Grain orientation prediction for anisotropic AP-TP-IP samples

In this application, the fraction of grains that are not well oriented is quantified for AP-TP-IP samples, as they influence the magnetic remanence [323]. A sample with a large number of

grains oriented along the axis of easy magnetization have a high remanence which is being measured from the grain orientation prediction model. When grain orientation of individual grains does not have a large deviation among themselves, it suggests that they are well oriented along one direction and this information is being used to correlate magnetic remanence of the sample. The AP, TP, and IP plots in *Figure 55* show good agreement between the prediction and the measured EBSD values. The mean difference between both methods for AP and IP samples in the case of the θ is less than 4.5°, and less than 3° in the case of the ρ orientation, suggesting that there is some degree of consistency bias in the DL model prediction. However, this bias is very small (~ 2%) for both θ and ρ. The upper and lower limits of agreement for θ in both samples are less than 10° and less than 5° for the ρ orientation values, even though the samples used for testing here have very high remanence when compared to the sample used for the training as shown in *Table 6*.

The area-weighted histogram distribution of the predicted orientation values for the AP, TP, and IP samples in *Figure 55(a & b)* reveals insight into the spread of these values. Further, the AP sample shows a larger spread compared to the TP and IP samples. Conversely, the IP sample has a narrower spread, with a SD of 7.9° for θ orientation and 5.2° for ρ orientation. This indicates that the IP sample is characterized by a higher proportion of grains aligned in a single direction. Furthermore, the progression towards enhanced grain alignment from the AP to the IP sample is characterized from the sample's attributes, as evidenced by the fact that the B_r value for the IP sample is significantly higher at 1.44 T, compared to the 1.34 T B_r value for the AP sample. This trend underlines the influence of the material's magnetic properties on grain orientation. Furthermore, as one moves from the AP sample to the IP sample, we notice that the misorientation value from the trained model decreases. This trend is similar to what is seen in the values obtained from the EBSD approach, as shown in *Figure 55(c)*. This difference can be due to the higher misorientation seen within the grains of the AP sample in contrast to the TP and IP samples. This alignment of findings is also evident in the outcomes of the predicted values of the trained DL model, as seen in *Figure 55(c)*.

Similar observations were seen when the number-weighted histogram distribution of predicted orientation values for the AP, TP, and IP samples were used to obtain the data spread value like in case of area-weighted distribution. The SD of samples from AP to IP decreased from 16.3° to 8.6° for θ orientation and from 9.5° to 6.0° for ρ orientation indicating the IP sample is characterized by a higher proportion of grains aligned in a single direction.

Both the EBSD data and the predictions of the developed DL model's predictions show a that lower the misorientation is, higher is the B_r [T]. The difference between the predicted and EBSD values based on the misorientation data is quite small, within the range of 1° to 4°. This indicates that the developed DL model's predictions align well with the established EBSD method for determining grain orientation.

Grain orientation prediction for PSD samples

Unlike the AP, TP and IP samples, which have homogeneous texture throughout the sample, the samples with PSD have heterogeneous texture distribution. The four positions in the sample have different degrees of grain alignment and are anisotropic in nature. The grains are extracted

from the KM images using the developed grain boundary detection model and each grain is then fed to the DoPACOM model to predict the θ and ρ values.

The mean grain orientation of PSD-05 is higher than that of PSD-01 by 6.5°. This indicates that PSD-05 has a greater average grain orientation compared to PSD-01. Similarly, PSD-15 has a higher mean grain orientation than that of PSD-11 by 9.8°. This represents the largest deviation among all PSD samples, indicating a significant difference in the average grain orientation between these two samples. Nevertheless, the predicted values for grains in all four samples are in a close range to the EBSD measured values. The deviation between the predicted and measured θ values is between 6.1% and 7.9% and for ρ predictions it is between 4% and 5.3%. A comparatively higher prediction error for θ values is seen in PSD-15 which lies in the divergent region. The difference in the brightness of the grains in all four PSD samples can be seen in *Figure 56* in section 4.5.2. Further, under KM, the brightness of the grains gets affected with change in θ value and it increases with an increase in θ value. Therefore, the KM images of the sample PSD-15 have more brightness than PSD samples in other regions.

Nevertheless, the statistical results obtained from the histogram plot agree with the characteristics of the analyzed positions in the PSD sample. The area-weighted distribution plot for θ and ρ values predicted by the DoPACOM model is compared based on the mean and standard deviation obtained from it (*Figure 57*). The standard deviation reflecting the degree of alignment for all four histograms is nearly the same. The SD for θ and ρ values among the PSD samples differ by ~ 1°. This is expected because the PSD sample as a whole is axially pressed during the manufacturing process and has the same remanence. This would imply that for PSD-01, PSD-05, PSD-11, and PSD-15 sample are AP in nature and have the same amount of grain orientation dispersion. But the texture strength of each of the four PSD samples is different. The mean value from the θ plot suggests that PSD-01 and PSD-11 belong to convergent regions because each histogram's mean value is less than 90° which is c-axis direction.Similarly, the mean value for PSD-05 and PSD-15 is more than 90° suggesting that they belong to the divergent region. Additionally, the PSD-01 and PSD-05 have a lower degree of convergence and divergence than PSD-11 and PSD-15, respectively. This is also the outcome of the histogram plots of θ which shows that the mean values for PSD-01 and PSD-05 are lower than PSD-11 and PSD-15 respectively. Similar observations were seen for the number-weighted distribution plot for all the 4 PSD samples, where the deviation between the predicted and measured θ values is between 6% and 9% and for ρ predictions it is between 4% and 5%.

The deviation in the predicted ρ values for all four positions is nearly the same, indicating that this model is much more robust than the θ value prediction model. Finally, the misorientation value calculated as a function of mean from histogram of θ value shown in *Figure 57* in section 4.5.2 suggests that the predicted and EBSD measured values are in close range.

The DL-based model trained to predict grain orientation from KM data has the potential to emerge as a reliable and practical alternative to the EBSD approach for characterizing magnetic samples based on their crystallographic orientation or texture. However, it is essential to acknowledge certain limitations associated with this approach. The model performs less accurately and with lower confidence in cases where the sample contains numerous grains exhibiting fully developed closure domain structures. Such instances often occur when the samples are sectioned in a direction perpendicular to the applied magnetic field during pressing.

Finally, as discussed above, the developed DL method utilizing KM data can serve as a valuable and cost-effective tool for quickly analyzing the microstructure across large samples including heterogeneous, providing initial insights into the sample characteristics. If a more detailed analysis is required, a subsequent EBSD examination can be pursued. This approach is also well-suited for scenarios where a reasonable balance between accuracy and increased measurement efficiency is sought. This proposed approach saves time and resources and overcomes the challenge of analyzing numerous small subsets of large-scale samples, as typically done with representative EBSD measurements. This is particularly relevant in cases of non-uniformly oriented samples, such as those with PSD.

5.9. Strengths and limitations of the crystallographic grain orientation model for sintered FeNdB magnets

The crystallographic grain orientation prediction model developed for the sintered FeNdB magnets from the KM images is associated with some strengths and also has a limitation on the application due to the training model's inductive bias, prediction interval, and confidence interval as a result of datasets used.

Strengths of the developed approaches

The primary strength of developed approach is that the texture analysis of the magnet sample based on KM data for a large sample area can be analyzed in less time when compared to the EBSD approach. Further, the prediction error of less than 5° by the fine-tuned EfficientNetB0 model combined with a shorter analysis time becomes crucial for large coherent samples.

When compared to the existing SEM-EBSD approach for anisotropy analysis of sintered FeNdB magnet, the developed approach based on KM images offers two significant benefits. Firstly, the data acquisition and processing time is significantly lower using the DL model when compared to EBSD, as it is based on KM images. For instance, a standard EBSD mapping of an area like 500x400 μm^2 used for the experiment (see section3.2.2), as is the case in the study, typically requires around 3 hours. On the contrary, KM images can be acquired and processed in nearly 20 minutes. This time savings includes the entire process, which involves initializing the distinct trained model for θ and ρ values and applying it to KM images from 16 tile images, which have 4317 grains, all at 1000x magnification, along with the pre- and post-processing steps, and it takes close to 16 minutes. Also, it is to be note that the capital investment for EBSD setup is 8x to 10x times more expensive than KM.

Secondly, using the developed approach allows users of Kerr microscopes to perform anisotropy quantification on sintered FeNdB magnets without the requirement for SEM-EBSD equipment, which is typically both capital-intensive and more operationally complex than microscopes equipped with MOKE functionality. Additionally, the process of capturing images over a wide sample area is more straightforward when using a KM as opposed to SEM-EBSD.

Limitations of the developed approaches

The developed model only predicted the reduced Euler angles φ_1 and ϕ components, as KM images do not provide information on the φ_2 component. Furthermore, the model predicts

orientation comparatively less accurately (MAE between 9º to 12.3º) and with lower confidence when the sample contains more grains with closure domain structures. These structures are common when samples are sectioned perpendicular to the applied magnetic field. Additionally, it has been observed that for samples with stripe domain structures, the difference in the appearance of the domain patterns is very fine. Due to the limited number of such grains in the training dataset, the confidence value of the predicted grain orientation value for θ is significantly less. To improve prediction accuracy, it is advisable to obtain samples for analysis with fewer fully developed closure domain structures and subsequently acquire a greater number of grains with θ values less than 20° and more than 160°.

6. Conclusions

In summary, this thesis is focused on developing data-driven models for intrinsic property prediction from chemical composition for 14:2:1 phases and for characterizing the FeNdB type magnet using the grain size and orientation analysis from KM data. First, the thesis focuses on the development of ML regression models to predict properties of the 14:2:1 magnetic phase, specifically mass density, Curie temperature, and saturation magnetization. These properties are essential for understanding magnet's magnetic behavior and performance. Through a systematic approach, each of these properties was individually addressed, and the results showcased the effectiveness and potential of ML in materials science research.

The data-driven model approach discussed in this thesis takes the chemical composition of the 14:2:1 hard magnetic phase as input feature to predict ρ_m, T_c, and M_s at room temperature. The developed models are single generalized models applicable for ternary to senary alloys of 14:2:1 phases. Furthermore, the trained model has generalized well on a dataset such that it is not affected by the presence of different heavy and light RE substituted 14:2:1 phases. The models are developed to perform on a feature space of 33 elements for ρ_m, 37 elements for T_c, and 28 elements for M_s with the dataset including experimentally reported values across various literature sources.

The results for ρ_m emphasize the superior performance of the LR model over other ML models. This superiority is most effective when using the (CC+AM) as feature set over 'CC'. The combination of LR and the feature set (CC+AM) resulted in more accurate predictions, with a low MAPE of 0.51%, RMSE of 0.05 [g/cm^3] and a high R^2 score of 0.98 on the validation set with 39 phases. On the unseen test set with 66 phases, the trained LR model using the 'CC' feature only resulted in a low MAE of 0.038 [g/cm^3] of which 13 repeated phases had a slightly higher MAE of 0.041 [g/cm^3]. The combined low MAE for validation and test set accounts for an average deviation of less than 1.5%. This demonstrates the effectiveness of trained LR in providing accurate ρ_m predictions. The study shows that, despite their simplicity, LR models can be highly suitable for predicting properties, especially in systems where the linear correlation between composition and property holds.

For the prediction of T_c and M_s, a VR model with four different base learners was used. This ensemble approach offers smoother functional approximations compared to the individual base learner. For the T_c model, the base learners for the VR model are ET, XGB, RF, and AB-RF with uniform weights. When evaluated on an unseen test set, the T_c prediction model effectively predicts with a low MAE of 16 K. The performance was also assessed for 14:2:1 phases incorporating both heavy and light rare earth metals, revealing that the predictions consistently exhibited an MAE below 20 K. Additionally, for unique phases that accounts for 60% of the T_c dataset, a model resulted in the low MAE of 16 K.

Conversely, for the M_s prediction task, the trained model NU-VR is a VR model that includes HR, RF, MLP, and AB-DT as base learners with assigned weights of 0.89, 0.94, 0.95, and 0.96 respectively. These weights are based on performance-based and rank-based approaches of base learners. The predictions from the NU-VR model closely match the literature reported values of M_s with a low MAE of 0.8 [μ_B/f.u.] on the unseen test set, accounting for the average

deviation of 3%. Furthermore, the M_s prediction model resulted in MAE of 0.75 [μ_B/f.u.] on unique phases that account for 72% of dataset. It is observed that specific compositions, particularly those near solubility limits, namely $Fe_{11}Cr_3Gd_2B$ and $Fe_{11}Ni_3Nd_2B$, had higher prediction errors. This challenge arises from the limited training dataset that the model has for learning.

In addition, the developed models were used to expand existing databases and predict other intrinsic properties, such as J_s. Using the ρ_m and M_s models, J_s [T] was obtained for 24 phases with a low MAE of 0.1 T which accounts for average deviation of 2.3% compared against the literature reported in the literature. It is expected that the developed model can serve as a basis for developing a 14:2:1 phase with an optimized chemical composition with desired magnetic properties. Nevertheless, it is to be noted that the phases with compositions at the solubility limit have comparatively higher prediction errors and the prediction uncertainty is high.

For microstructure characterization of the sintered FeNdB type permanent magnet, the grain boundary and crystallographic orientation of the grains were predicted using the models developed based on the correlative KM and SEM-EBSD data. KM excels at capturing magnetic domain structures within FeNdB magnets, while EBSD provides valuable insight into grain size and orientation information. It is to be noted that the EBSD data have been considered as reference for performance evaluation of developed models, despite the deviation among the manually annotated masks from 3 subject experts and 2 scientific helpers for grain boundary detection task being less than 3%, indicating high agreement despite potential subjective bias.

For the grain boundary detection task, various supervised ML models were first trained on hand-crafted features, and the RF(C,E,T) model trained on a 39 dimensional feature set that includes color, edge, and texture information outperformed and resulted in IoU of 0.81 when compared to reference data. Additionally, it was also observed that when RF was trained on additional features of the pre-trained VGG16 network of 4096 features, the performance dropped by 14% and over-fitting occurred. This highlights the importance of quality hand-crafted features, as demonstrated by the under-fitting of models with excessive features. For trained RF(C,E,T), the overall prediction deviation based on IoU metric is between 3% to 11% when compared to the reference data. In the next steps, various supervised DL models with pre-trained networks such as VGG16, ResNet50, EfficientNetB3, and InceptionV3 were evaluated against the custom U-net model. The custom U-net model that includes a combined loss function of binary focal loss with $\alpha = 0.25$ and $\gamma = 2$, along with the distance transform loss with $\lambda = 0.35$, achieved the highest performance, resulting in an IoU of 0.94, suggesting that the predicted results differ by 6% compared to reference data from EBSD. This model included 0.5 dropout rate in bottleneck layer to avoid overfitting. Compared to reference values, based on IoU metric, the overall deviation from the trained custom U-net model is less than 4%. U-net, a deep learning architecture specifically designed for image segmentation, has proven to be highly effective in preserving the structural integrity of KM images for grain boundary detection task. To evaluate the robustness of the trained grain boundary detection model, the comparison between measured EBSD values and the grain size from predicted grain boundary masks for anisotropic AP-TP-IP and PSD samples was obtained. The relative errors in terms of D25, D50, and D90 grain sizes typically range from 3.1% to 7.7% in area-weighted grain sizes for anisotropic AP-TP-IP samples. In PSD samples, the developed model exhibited moderate relative errors, with average relative errors ranging from 4.2% to 5.1%. The robustness of the

trained model is evident in its ability to handle variations in sample characteristics, including samples with ECD of 7 µm and 10 µm. Furthermore, the model analyzes isotropic samples with a fine grain size of ECD 6 µm and a coarse grain size of ECD of 14 µm with 83% accuracy when measured against EBSD values, demonstrating its versatility and adaptability.

The research also highlights the impact of post-processing steps in EBSD, which can influence the morphology of grains and introduce deviations in grain size measurements between 3% and 5% for various test samples. KM, on the other hand, offers flexibility in grain size analysis by providing options to measure grains with or without domain structures, ultimately resulting in measurements that are closer to reality. Finally, the KM-based developed approach is 10x faster and 8x to 10x times less cost-intensive than the SEM-EBSD approach, which enables the analysis of much larger sample dimension that exceed 400x400 μm^2. The comparison of time-efficiency does not account for the effort and time required for sample preparation in KM and EBSD approaches, and it should also be noted that the image resolution differs, with EBSD offering 450x magnification and KM providing 1000x magnification. The developed model has the limitation that, when the relative difference in the φ_1 component of grain orientation is minimal and the relative misorientation ($\Delta\phi$) lies in the range of 2° to 8°, the model struggles to predict the grain boundary accurately. This work serves as the basis for exploring texture analysis or crystallographic grain orientation analysis, which is the next part of microstructure characterization in this thesis.

For the grain orientation task, using the supervised learning approach, regression models were trained to predict the reduced Euler angles for each grain extracted using KM images. Reduced Euler angles includes φ_1 and ϕ components obtained from SEM-EBSD. The φ_2 component of the Euler angle remains undetectable through polar MOKE and is presumed to have a limited influence on magnetic properties. Furthermore, based on the experimental results, the range of φ_1 was reduced from [0º, 360º] to [0º, 180º] and for ϕ range has been limited to [0º, 90º]. The Euler components of grain orientation φ_1 and ϕ after reduced range are indicated as θ and ρ, respectively. The developed model's evaluation against the measured EBSD data revealed that the A-RF model outperformed the others among traditional ML models, resulting in the lowest MAE of 7º and 5º for θ and ρ predictions respectively. DL models showed more robustness, with the EfficientNetB0 fine-tuned model consistently resulting in the best performance with MAE of 5º and 3º for θ and ρ predictions respectively. To prevent the EfficientNetB0 model from overfitting, batch normalization with weight decay of $1x10^{-3}$ and 0.9 as momentum, dropout layers with 0.6 dropout rate, and L2 regularization with $1x10^{-4}$ decay parameter were used. Additionally, the MSE loss function has been regularized with MSW. Further, since the number of training data for both θ and ρ prediction model is less than 5000, the model evaluation against was carried out using R_{ccc} metric which considers both accuracy and precision.

Similarly to the grain boundary detection model, the robustness of the developed grain orientation model's robustness was evaluated on AP-TP-IP and PSD samples. The prediction from the model resulted in an error of 1° - 4° when compared against EBSD measurements for AP, TP, and IP samples. Additionally, the developed model has been evaluated on the sample with PSD exceeding 20% and a magnetic dipole moment misalignment angle ranging from 0° to 1°. The model prediction error (MAE) is 3% for θ values and less than 1.7% for ρ values when compared against EBSD measurements. Moreover, for large sample analysis of an area of 500x400 μm^2, the time required for data acquisition, processing of KM images, and

prediction of orientations using a developed model is decreased by a factor of 10x and is also 8x to 10x times less expensive compared to SEM-EBSD. It is important to note that similar to the grain size analysis task, here for time-efficiency comparison, the effort and time needed for sample preparation are not considered, and also image resolution differs, with EBSD offering 450x magnification and KM providing 1000x magnification. However, it is essential to note that there are certain limitations. The fine-tuned EfficientNetB0 has a prediction error between 9º to 12.3º for samples having a large number of grains with fully developed closure domain structures, especially in samples sectioned perpendicular to the applied magnetic field.

Lastly, the fine-tuned EfficientNetB0 model offers a valuable and efficient tool for the initial characterization of magnetic samples based on their crystallographic orientation or texture, potentially streamlining the research process. While it may not replace more detailed methods such as EBSD in all cases, it provides a promising alternative for researchers seeking a balance between time-efficiency and accuracy. While EBSD excels at providing accurate crystallographic information, KM's capability to visualize domains efficiently makes it suitable for large-scale observations to be able to detect inhomogeneities within sample in less time. Considering the growing database of grains from FeNdB magnets with different grain sizes, along with rapid developments in the DL architecture, it should be possible to overcome the current limitations of the developed model.

In future research, a multi-dimensional approach can be adopted to advance the capabilities of models in materials science and engineering. Developing methods to estimate uncertainty for predicted intrinsic property values and correlating them with physics-based models could be crucial, particularly since these models are often developed using small datasets. Addressing limitations in grain boundary detection and crystallographic grain orientation prediction remains essential.

References

[1] N. R. Council, *Materials Science and Engineering for the 1990s: Maintaining Competitiveness in the Age of Materials*. Washington, DC: The National Academies Press, 1989. doi: 10.17226/758.

[2] A. Agrawal and A. Choudhary, 'Perspective: Materials informatics and big data: Realization of the "fourth paradigm" of science in materials science', *APL Materials*, vol. 4, no. 5, p. 053208, May 2016, doi: 10.1063/1.4946894.

[3] W. Xiong and G. B. Olson, 'Cybermaterials: materials by design and accelerated insertion of materials', *npj Computational Materials*, vol. 2, no. 1, Art. no. 1, Feb. 2016, doi: 10.1038/npjcompumats.2015.9.

[4] J. Schmidt, M. R. G. Marques, S. Botti, and M. A. L. Marques, 'Recent advances and applications of machine learning in solid-state materials science', *npj Computational Materials*, vol. 5, no. 1, Art. no. 1, Aug. 2019, doi: 10.1038/s41524-019-0221-0.

[5] A. Géron, *Hands-On Machine Learning with Scikit-Learn and TensorFlow*, 2nd ed. O'Reilly Media, Inc., 2019. [Online]. Available: https://www.oreilly.com/library/view/hands-on-machine-learning/9781492032632/

[6] Ian Goodfellow, Yoshua Bengio, Aaron Courvile, *Deep Learning*. The MIT Press, 2016. Accessed: Jul. 29, 2024. [Online]. Available: https://mitpress.mit.edu/9780262035613/deep-learning/

[7] R. Sarath Kumar Boddu, S. Ahamad, K. V. Pradeep Kumar, M. Ramalingam, L. Kirana Pallathadka, and F. Peniero Tupas, 'Analysis of Robotics, Artificial intelligence and Machine learning in the field of healthcare sector', *Materials Today: Proceedings*, vol. 56, pp. 2323–2327, Jan. 2022, doi: 10.1016/j.matpr.2021.12.016.

[8] S. A. Damiati, 'Digital Pharmaceutical Sciences', *AAPS PharmSciTech*, vol. 21, no. 6, p. 206, Jul. 2020, doi: 10.1208/s12249-020-01747-4.

[9] J. T. McCoy and L. Auret, 'Machine learning applications in minerals processing: A review', *Minerals Engineering*, vol. 132, pp. 95–109, Mar. 2019, doi: 10.1016/j.mineng.2018.12.004.

[10] A. Luckow, M. Cook, N. Ashcraft, E. Weill, E. Djerekarov, and B. Vorster, 'Deep learning in the automotive industry: Applications and tools', in *2016 IEEE International Conference on Big Data (Big Data)*, Dec. 2016, pp. 3759–3768. doi: 10.1109/BigData.2016.7841045.

[11] R. Gauß, C. Burkhardt, F. Carencotte, M. Gasparon, O. Gutfleisch, I. Higgins, M. Karajić, A. Klossek, M. Mäkinen, B. Schäfer, R. Schindler, and B. Veluri, 'Rare Earth Magnets and Motors: A European Call for Action', European Raw Materials Alliance (ERMA), Berlin, Sep. 2021. [Online]. Available: https://erma.eu/european-call-for-action/

[12] J. M. D. Coey, 'Perspective and Prospects for Rare Earth Permanent Magnets', *Engineering*, vol. 6, no. 2, pp. 119–131, Feb. 2020, doi: 10.1016/j.eng.2018.11.034.

[13] K. Hono and H. Sepehri-Amin, 'Strategy for high-coercivity Nd–Fe–B magnets', *Scripta Materialia*, vol. 67, no. 6, pp. 530–535, Sep. 2012, doi: 10.1016/j.scriptamat.2012.06.038.

[14] J. Cui, M. Kramer, L. Zhou, F. Liu, A. Gabay, G. Hadjipanayis, B. Balasubramanian, and D. Sellmyer, 'Current progress and future challenges in rare-earth-free permanent magnets', *Acta Materialia*, vol. 158, pp. 118–137, Oct. 2018, doi: 10.1016/j.actamat.2018.07.049.

[15] T. G. Woodcock, Y. Zhang, G. Hrkac, G. Ciuta, N. M. Dempsey, T. Schrefl, O. Gutfleisch, and D. Givord, 'Understanding the microstructure and coercivity of high performance NdFeB-based magnets', *Scripta Materialia*, vol. 67, no. 6, pp. 536–541, Sep. 2012, doi: 10.1016/j.scriptamat.2012.05.038.

[16] K. Uestuener, M. Katter, and W. Rodewald, ‘Dependence of the Mean Grain Size and Coercivity of Sintered Nd–Fe–B Magnets on the Initial Powder Particle Size’, *IEEE Transactions on Magnetics*, vol. 42, no. 10, pp. 2897–2899, Oct. 2006, doi: 10.1109/TMAG.2006.879889.

[17] A. Di Gianfrancesco, ‘8 - Technologies for chemical analyses, microstructural and inspection investigations’, in *Materials for Ultra-Supercritical and Advanced Ultra-Supercritical Power Plants*, A. Di Gianfrancesco, Ed., Woodhead Publishing, 2017, pp. 197–245. doi: 10.1016/B978-0-08-100552-1.00008-7.

[18] E. Zueco, W. Rave, R. Schäfer, A. Hubert, and L. Schultz, ‘Combined Kerr-/magnetic force microscopy on NdFeB crystals of different crystallographic orientation’, *Journal of Magnetism and Magnetic Materials*, vol. 190, no. 1, pp. 42–47, Dec. 1998, doi: 10.1016/S0304-8853(98)00268-6.

[19] Y. Zhu, *Modern Techniques for Characterizing Magnetic Materials*, 1st ed. Springer New York, NY, 2022. Accessed: Aug. 11, 2022. [Online]. Available: https://link.springer.com/book/10.1007/b101202

[20] S. Ebnesajjad, ‘Chapter 4 - Surface and Material Characterization Techniques’, in *Surface Treatment of Materials for Adhesive Bonding (Second Edition)*, S. Ebnesajjad, Ed., Oxford: William Andrew Publishing, 2014, pp. 39–75. doi: 10.1016/B978-0-323-26435-8.00004-6.

[21] D. Jha, S. Singh, R. Al-Bahrani, W. Liao, A. Choudhary, M. D. Graef, and A. Agrawal, ‘Extracting Grain Orientations from EBSD Patterns of Polycrystalline Materials Using Convolutional Neural Networks’, *Microscopy and Microanalysis*, vol. 24, no. 5, pp. 497–502, Oct. 2018, doi: 10.1017/S1431927618015131.

[22] X. B. Liu and Z. Altounian, ‘The partitioning of Dy and Tb in NdFeB magnets: A first-principles study’, *Journal of Applied Physics*, vol. 111, no. 7, p. 07A701, Apr. 2012, doi: 10.1063/1.3670054.

[23] B. Peng, T. Ma, Y. Zhang, J. Jin, and M. Yan, ‘Improved thermal stability of Nd-Ce-Fe-B sintered magnets by Y substitution’, *Scripta Materialia*, vol. 131, pp. 11–14, Apr. 2017, doi: 10.1016/j.scriptamat.2016.12.013.

[24] J. F. Herbst, ‘R2Fe14B materials: Intrinsic properties and technological aspects’, *Rev. Mod. Phys.*, vol. 63, no. 4, pp. 819–898, Oct. 1991, doi: 10.1103/RevModPhys.63.819.

[25] D. Goll, R. Loeffler, J. Herbst, R. Karimi, U. Pflanz, R. Stein, and G. Schneider, ‘High-Throughput Methods for Searching New Permanent Magnet Materials’, *IEEE Transactions on Magnetics*, vol. 50, no. 11, pp. 1–7, Nov. 2014, doi: 10.1109/TMAG.2014.2331458.

[26] D. Goll, R. Loeffler, D. Hohs, and G. Schneider, ‘Reaction sintering as a high-throughput approach for magnetic materials development’, *Scripta Materialia*, vol. 146, pp. 355–361, Mar. 2018, doi: 10.1016/j.scriptamat.2017.05.004.

[27] D. Goll, R. Loeffler, J. Herbst, R. Karimi, U. Pflanz, R. Stein, and G. Schneider, ‘Novel Permanent Magnets by High-Throughput Experiments’, *JOM*, vol. 67, no. 6, pp. 1336–1343, Jun. 2015, doi: 10.1007/s11837-015-1422-8.

[28] D. Goll, R. Loeffler, J. Herbst, R. Karimi, and G. Schneider, ‘High-throughput search for new permanent magnet materials’, *J. Phys.: Condens. Matter*, vol. 26, no. 6, p. 064208, Jan. 2014, doi: 10.1088/0953-8984/26/6/064208.

[29] ‘NdFeB Rare Earth Magnets’, Geomega Resources Inc. Accessed: Aug. 08, 2023. [Online]. Available: https://geomega.ca/about-magnets/

[30] R. Griissinger, ‘Novel Magnetism in Permanent Magnets’, *JUSAMI*, vol. 5, no. 1, p. 487963, Oct. 2003, doi: 10.17146/jusami.2003.5.1.5195.

[31] J. Fischbacher, A. Kovacs, M. Gusenbauer, H. Oezelt, L. Exl, S. Bance, and T. Schrefl, ‘Micromagnetics of rare-earth efficient permanent magnets’, *J. Phys. D: Appl. Phys.*, vol. 51, no. 19, p. 193002, Apr. 2018, doi: 10.1088/1361-6463/aab7d1.

[32] Jimenez-Villacorta F., Lewis, Laura H., 'Advanced Permanent Magnetic Materials', in *Nanomagnetism*, J. M. Gonzalez, Ed., Manchester , UK: OCP Publishing Group, 2014, pp. 160–185.

[33] M. Sagawa and Y. Une, 'Chapter 5 - The status of sintered NdFeB magnets', in *Modern Permanent Magnets*, J. Croat and J. Ormerod, Eds., in Woodhead Publishing Series in Electronic and Optical Materials. , Woodhead Publishing, 2022, pp. 135–168. doi: 10.1016/B978-0-323-88658-1.00010-8.

[34] H. Okazaki, D. Billington, N. Tsuji, W. Ueno, Y. Kotani, S. Kawaguchi, K. Sugimoto, K. Toyoki, T. Fukagawa, T. Nishiuchi, K. Hono, S. Hirosawa, and T. Nakamura, 'Quantitative identification of constituent phases in a Nd-Fe-B-Cu sintered magnet and temperature dependent change of electron density of Nd2Fe14B studied by synchrotron X-ray diffraction', *Acta Materialia*, vol. 181, pp. 530–536, Dec. 2019, doi: 10.1016/j.actamat.2019.10.004.

[35] G. Schneider, E.-T. Henig, F. P. Missell, and G. Petzow, 'Microstructure of Sintered Fe-Nd-B Magnets', *International Journal of Materials Research*, vol. 81, no. 5, pp. 322–329, May 1990, doi: 10.1515/ijmr-1990-810503.

[36] B. E. Davies, R. S. Mottram, and I. R. Harris, 'Recent developments in the sintering of NdFeB', *Materials Chemistry and Physics*, vol. 67, no. 1, pp. 272–281, Jan. 2001, doi: 10.1016/S0254-0584(00)00450-8.

[37] J. Sugiyama, K. Miwa, H. Nozaki, Y. Kaneko, B. Hitti, D. Arseneau, G. D. Morris, E. J. Ansaldo, and J. H. Brewer, 'Magnetic moment of rare-earth elements in $R2\times Fe14\times B$ estimated with $\mu+\times SR$', *Phys. Rev. Mater.*, vol. 3, no. 6, p. 064402, Jun. 2019, doi: 10.1103/PhysRevMaterials.3.064402.

[38] J. P. Gavigan and D. Givord, 'Intrinsic and extrinsic properties of rare earth-transition metal compounds and permanent magnets', *Journal of Magnetism and Magnetic Materials*, vol. 84, no. 3, pp. 288–298, Mar. 1990, doi: 10.1016/0304-8853(90)90107-2.

[39] H. Hiroyoshi, N. Saito, G. Kido, Y. Nakagawa, S. Hirosawa, and M. Sagawa, 'High-field magnetization of R2Fe14B single crystals', *Journal of Magnetism and Magnetic Materials*, vol. 54–57, pp. 583–584, Feb. 1986, doi: 10.1016/0304-8853(86)90717-1.

[40] S. Hirosawa and M. Sagawa, 'Spin reorientation and magnetization anomaly in Er2Fe14B and Tm2Fe14B', *Solid State Communications*, vol. 54, no. 4, pp. 335–337, Apr. 1985, doi: 10.1016/0038-1098(85)90009-2.

[41] E. Burzo, 'Permanent magnets based on R-Fe-B and R-Fe-C alloys', *Rep. Prog. Phys.*, vol. 61, no. 9, p. 1099, Sep. 1998, doi: 10.1088/0034-4885/61/9/001.

[42] A. T. Pedziwiatr and W. E. Wallace, 'Magnetic properties of Y2Fe14−xMxB compounds where M=Si OR Cu', *Solid State Communications*, vol. 61, no. 1, pp. 61–64, 1987.

[43] E. Burzo and N. Plugaru, 'Magnetic Properties of R2Fel4-xCuxB Compounds with R=Nd OR Er', *Journal of Magnetism and Magnetic Materials*, vol. 86, pp. 91–101, 1990.

[44] E. Burzo, A. T. Pedziwiatr, and W. E. Wallace, 'Magnetic properties of Nd2Fe14−x−yCoxAlyB alloys', *Solid State Communications*, vol. 61, no. 1, pp. 57–59, Jan. 1987, doi: 10.1016/0038-1098(87)90015-9.

[45] M. Jurczyk and W. Wallace, 'Magnetism of Nd2Fe12-xMnxCo2B alloys', *IEEE Transactions on Magnetics*, vol. 22, no. 5, pp. 755–756, Sep. 1986, doi: 10.1109/TMAG.1986.1064333.

[46] H. C. Ku and L. S. Yen, 'Magnetic properties of the new permanent magnet compounds Nd2(Fe0.2M0.1)14B (M ≡ Sc, Ti, V, Cr, Mn, Co, Ni)', *Journal of the Less Common Metals*, vol. 127, pp. 43–48, Jan. 1987, doi: 10.1016/0022-5088(87)90356-0.

[47] H. C. Ku, J. C. Chen, and L. S. Yen, 'Magnetic properties of pseudoternary compounds Nd2 (Fe1-xRux)14B', *Journal of Magnetism and Magnetic Materials*, vol. 54–57, pp. 533–534, Feb. 1986, doi: 10.1016/0304-8853(86)90696-7.

[48] A. Kowalczyk and A. Wrzeciono, 'Structural and magnetic characteristics of R2Fe14−xCuxB systems (R = Y, Nd and Gd)', *Journal of Magnetism and Magnetic Materials*, vol. 74, no. 3, pp. 260–262, Oct. 1988, doi: 10.1016/0304-8853(88)90196-5.

[49] R. Grössinger, X. K. Sun, R. Eibler, K. H. J. Buschow, and H. R. Kirchmayr, 'The Temperature Dependence Of The Anisotropy Field in R2Fe13B Compounds (R = Y, La, Ce, Pr, Nd, Gd, Ho, Lu)', *Journal of Physics Colloques*, vol. 46, no. C6, pp. C6-221-C6-224, Sep. 1985, doi: 10.1051/jphyscol:1985638.

[50] J. D. I. Morgan and W. Kutzelnigg, 'Hund's rules, the alternating rule, and symmetry holes', *J. Phys. Chem.*, vol. 97, no. 10, pp. 2425–2434, Mar. 1993, doi: 10.1021/j100112a051.

[51] S. NamKung, T. S. Jang, Y. H. Lee, M. K. Moon, T. H. Lim, and D. H. Kim, 'Effect of Particle Size Distribution on the Microstructure and Magnetic Properties of Sintered NdFeB Magnets', in *Proceedings of the Korean Magnestics Society Conference*, Seoul, Korea: The Korean Magnetics Society, 2009, pp. 129–130. Accessed: Aug. 11, 2022. [Online]. Available: https://koreascience.kr/article/CFKO200935751419684.page

[52] R. w. Gao, D. h. Zhang, H. Li, and J. c. Zhang, 'Effects of the degree of grain alignment on the hard magnetic properties of sintered NdFeB magnets', *Applied Physics A: Materials Science & Processing*, vol. 67, no. 3, p. 353, Sep. 1998, doi: 10.1007/s003390050783.

[53] X. Yuan and J. Zhu, 'Misorientations across boundary planes in a sintered NdFeB permanent magnet', *RSC Advances*, vol. 12, no. 31, pp. 20412–20422, 2022, doi: 10.1039/D2RA01670B.

[54] T. B. Britton, J. Jiang, Y. Guo, A. Vilalta-Clemente, D. Wallis, L. N. Hansen, A. Winkelmann, and A. J. Wilkinson, 'Tutorial: Crystal orientations and EBSD — Or which way is up?', *Materials Characterization*, vol. 117, pp. 113–126, Jul. 2016, doi: 10.1016/j.matchar.2016.04.008.

[55] F. J. Humphreys, 'Quantitative metallography by electron backscattered diffraction', *Journal of Microscopy*, vol. 195, no. 3, pp. 170–185, 1999, doi: 10.1046/j.1365-2818.1999.00578.x.

[56] G. Nolze and A. Winkelmann, 'About the reliability of EBSD measurements: Data enhancement', in *IOP Conference Series: Materials Science and Engineering*, Trondheim, Norway: IOP Publishing Ltd, Jul. 2020, p. 012018. doi: 10.1088/1757-899X/891/1/012018.

[57] R. A. Schwarzer, D. P. Field, B. L. Adams, M. Kumar, and A. J. Schwartz, 'Present State of Electron Backscatter Diffraction and Prospective Developments', in *Electron Backscatter Diffraction in Materials Science*, A. J. Schwartz, M. Kumar, B. L. Adams, and D. P. Field, Eds., Boston, MA: Springer US, 2009, pp. 1–20. doi: 10.1007/978-0-387-88136-2_1.

[58] Í. Carneiro and S. Simões, 'Recent Advances in EBSD Characterization of Metals', *Metals*, vol. 10, no. 8, Art. no. 8, Aug. 2020, doi: 10.3390/met10081097.

[59] Dominic Hohs, Tvrtko Grubesa, David Schuller, Timo Bernthaler, Dagmar Goll, Gerhard, 'EBSD - a Powerful Tool for the Analysis of Magnetic Materials | Microscopy and Microanalysis | Cambridge Core', *Microscopy and Microanalysis*, vol. 23, no. S1, pp. 596–597, doi: ttps://doi.org/10.1017/S143192761700366X.

[60] T. G. Woodcock and O. Gutfleisch, 'Multi-phase EBSD mapping and local texture analysis in NdFeB sintered magnets', *Acta Materialia*, vol. 59, no. 3, pp. 1026–1036, Feb. 2011, doi: 10.1016/j.actamat.2010.10.033.

[61] C. G. Hauegen, F. R. F. da Silva, F. A. S. da Silva, J. F. C. Lins, and M. F. de Campos, 'EBSD Texture Analysis of NdFeB Magnets', *Materials Science Forum*, vol. 727–728, pp. 135–139, 2012, doi: 10.4028/www.scientific.net/MSF.727-728.135.

[62] L. Lopes, E. Santos, T. Hartwig, and P. Wendhausen, *On the Texture Evaluation of Nd-Fe-B Magnets with Different Alignment Degrees*. 2014. doi: 10.13140/RG.2.1.4781.7767.

[63] S. Sawatzki, T. G. Woodcock, K. Güth, K.-H. Müller, and O. Gutfleisch, 'Calculation of remanence and degree of texture from EBSD orientation histograms and XRD rocking curves in

Nd–Fe–B sintered magnets', *Journal of Magnetism and Magnetic Materials*, vol. 382, pp. 219–224, May 2015, doi: 10.1016/j.jmmm.2015.01.046.

[64] K. Khlopkov, O. Gutfleisch, D. Eckert, D. Hinz, B. Wall, W. Rodewald, K.-H. Müller, and L. Schultz, 'Local texture in Nd–Fe–B sintered magnets with maximised energy density', *Journal of Alloys and Compounds*, vol. 365, no. 1, pp. 259–265, Feb. 2004, doi: 10.1016/S0925-8388(03)00636-4.

[65] Z. Wang, J. Ju, J. Wang, W. Yin, R. Chen, M. Li, C. Jin, X. Tang, D. Lee, and A. Yan, 'Magnetic Properties Improvement of Die-upset Nd-Fe-B Magnets by Dy-Cu Press Injection and Subsequent Heat Treatment', *Sci Rep*, vol. 6, no. 1, p. 38335, Dec. 2016, doi: 10.1038/srep38335.

[66] K. Opelt, T. Ahmad, O. Diehl, M. Schönfeldt, E. Brouwer, I. Vogel, J. D. Rossa, J. Gassmann, S. Ener, and O. Gutfleisch, 'Upscaling the 2-Powder Method for the Manufacturing of Heavy Rare-Earth-Lean Sintered didymium-Based Magnets', *Advanced Engineering Materials*, vol. 23, no. 10, p. 2100459, 2021, doi: 10.1002/adem.202100459.

[67] R. L. Snyder, J. Fiala, and H. J. Bunge, 'Defect and Microstructure Analysis by Diffraction', *Oxford University Press*, p. 10, 1999.

[68] A. K. Choudhary, T. Grubesa, A. Jansche, T. Bernthaler, D. Goll, and G. Schneider, 'Deep learning and correlative microscopy for quantification of grain orientation in sintered FeNdB-type permanent magnets by domain pattern analysis', *Acta Materialia*, p. 119563, Nov. 2023, doi: 10.1016/j.actamat.2023.119563.

[69] A. K. Choudhary, A. Jansche, T. Grubesa, F. Trier, D. Goll, T. Bernthaler, and G. Schneider, 'Grain size analysis in permanent magnets from Kerr microscopy images using machine learning techniques', *Materials Characterization*, vol. 186, p. 111790, Apr. 2022, doi: 10.1016/j.matchar.2022.111790.

[70] A. K. Choudhary, A. Jansche, T. Bernthaler, and G. Schneider, 'Machine Learning for Microstructure Quantification of Different Material Classes', *Practical Metallography*, vol. 57, no. 7, pp. 475–501, Jul. 2020, doi: 10.3139/147.110632.

[71] O. Furat, M. Wang, M. Neumann, L. Petrich, M. Weber, C. E. Krill, and V. Schmidt, 'Machine Learning Techniques for the Segmentation of Tomographic Image Data of Functional Materials', *Frontiers in Materials*, vol. 6, 2019, Accessed: Sep. 17, 2023. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fmats.2019.00145

[72] J. J. Bailey, A. Wade, A. M. Boyce, Y. S. Zhang, D. J. L. Brett, and P. R. Shearing, 'Quantitative assessment of machine-learning segmentation of battery electrode materials for active material quantification', *Journal of Power Sources*, vol. 557, p. 232503, Feb. 2023, doi: 10.1016/j.jpowsour.2022.232503.

[73] C. Shu, Z. Xin, and C. Xie, 'EBSD Grain Knowledge Graph Representation Learning for Material Structure-Property Prediction', in *Knowledge Graph and Semantic Computing: Knowledge Graph Empowers New Infrastructure Construction*, B. Qin, Z. Jin, H. Wang, J. Pan, Y. Liu, and B. An, Eds., in Communications in Computer and Information Science. Singapore: Springer, 2021, pp. 3–15. doi: 10.1007/978-981-16-6471-7_1.

[74] X. Song, J. Luo, Z. Shi, L. Zhuang, Y. Qiao, and J. Zhang, 'Quantitative prediction of grain boundary misorientation effect on twin transmission in hexagonal metals', *Materials & Design*, vol. 192, p. 108745, Jul. 2020, doi: 10.1016/j.matdes.2020.108745.

[75] C. Kittel and J. K. Galt, 'Ferromagnetic Domain Theory', *Solid State Physics*, vol. 3, pp. 437–564, doi: https://doi.org/10.1016/S0081-1947(08)60136-8.

[76] J. Mohapatra and J. P. Liu, 'Rare-Earth-Free Permanent Magnets: The Past and Future', in *Handbook of Magnetic Materials*, vol. 27, Elsevier, 2018, pp. 1–57. doi: 10.1016/bs.hmm.2018.08.001.

[77] R. Schäfer, 'Investigation of Domains and Dynamics of Domain Walls by the Magneto-optical Kerr-effect', in *Handbook of Magnetism and Advanced Magnetic Materials*, John Wiley & Sons, Ltd, 2007. doi: 10.1002/9780470022184.hmm310.

[78] J. McCord, 'Progress in magnetic domain observation by advanced magneto-optical microscopy', *Journal of Physics D: Applied Physics*, vol. 48, no. 33, p. 333001, Jul. 2015, doi: 10.1088/0022-3727/48/33/333001.

[79] P. Braun, J. Laukart, U. Golla-Schindler, R. Löffler, D. Goll, and G. Schneider, 'Analysis of microstructure evolution during heat treatment of CoSm permanent magnets using high-resolution scanning electron microscopy', *Practical Metallography*, vol. 59, no. 4, pp. 188–198, Apr. 2022, doi: 10.1515/pm-2022-0020.

[80] M. Takezawa, H. Ogimoto, Y. Kimura, and Y. Morimoto, 'Analysis of the demagnetization process of Nd–Fe–B sintered magnets at elevated temperatures by magnetic domain observation using a Kerr microscope', *Journal of Applied Physics*, vol. 115, no. 17, p. 17A733, May 2014, doi: 10.1063/1.4866894.

[81] R. Schäfer, 'Magnetic Domains', in *Handbook of Magnetism and Magnetic Materials*, J. M. D. Coey and S. S. P. Parkin, Eds., Cham: Springer International Publishing, 2021, pp. 391–434. doi: 10.1007/978-3-030-63210-6_8.

[82] R. Schäfer, 'Magnets, Soft and Hard: Domains', in *Encyclopedia of Materials: Science and Technology*, K. H. J. Buschow, R. W. Cahn, M. C. Flemings, B. Ilschner, E. J. Kramer, S. Mahajan, and P. Veyssière, Eds., Oxford: Elsevier, 2001, pp. 5130–5141. doi: 10.1016/B0-08-043152-6/00893-7.

[83] J. Caplan, M. Niethammer, R. M. Taylor, and K. J. Czymmek, 'The Power of Correlative Microscopy: Multi-modal, Multi-scale, Multi-dimensional', *Curr Opin Struct Biol*, vol. 21, no. 5, pp. 686–693, Oct. 2011, doi: 10.1016/j.sbi.2011.06.010.

[84] J. Cui, J. Ormerod, D. Parker, R. Ott, A. Palasyuk, S. Mccall, M. P. Paranthaman, M. S. Kesler, M. A. McGuire, I. C. Nlebedim, C. Pan, and T. Lograsso, 'Manufacturing Processes for Permanent Magnets: Part I—Sintering and Casting', *JOM*, vol. 74, no. 4, pp. 1279–1295, Apr. 2022, doi: 10.1007/s11837-022-05156-9.

[85] R. Hilzinger and W. Rodewald, *Magnetic materials: fundamentals, products, properties, applications*. Germany: Publicis, 2013.

[86] Y. Kaneko, K. Kitajima, and N. Takusagawa, 'Preparation of Sr-Ferrite Mgnets by the New Annealing-Remilling Process', *Journal of the Ceramic Society of Japan*, vol. 101, no. 1176, pp. 905–912, 1993, doi: 10.2109/jcersj.101.905.

[87] G. Martinek, U. Wyss, D. Maybury, and S. Constantinides, 'Optimizing Magnetic Effects through Shaped Field Magnets', 2014. Accessed: Mar. 22, 2024. [Online]. Available: https://www.semanticscholar.org/paper/Optimizing-Magnetic-Effects-through-Shaped-Field-Martinek-Wyss/e11d9576cb7c5dd239d356932cd88f0735903a2a

[88] C. Gao, X. Min, M. Fang, T. Tao, X. Zheng, Y. Liu, X. Wu, and Z. Huang, 'Innovative Materials Science via Machine Learning', *Advanced Functional Materials*, vol. 32, no. 1, p. 2108044, 2022, doi: 10.1002/adfm.202108044.

[89] 'Worldwide Global DataSphere Forecast, 2021–2025: The World Keeps Creating More Data — Now, What Do We Do with It All?', IDC: The premier global market intelligence company. Accessed: Aug. 16, 2022. [Online]. Available: https://www.idc.com/getdoc.jsp?containerId=US46410421

[90] I. Arganda-Carreras, V. Kaynig, C. Rueden, K. W. Eliceiri, J. Schindelin, A. Cardona, and H. Sebastian Seung, 'Trainable Weka Segmentation: a machine learning tool for microscopy pixel classification', *Bioinformatics*, vol. 33, no. 15, pp. 2424–2426, Aug. 2017, doi: 10.1093/bioinformatics/btx180.

[91] R. M. Haralick, K. Shanmugam, and I. Dinstein, 'Textural Features for Image Classification', *IEEE Transactions on Systems, Man, and Cybernetics*, vol. SMC-3, no. 6, pp. 610–621, Nov. 1973, doi: 10.1109/TSMC.1973.4309314.

[92] E. Miyamoto and T. M. Jr, 'Fast Calculation of Haralick Texture Features', Technical Report, 2009. [Online]. Available: http://users.ece.cmu.edu/~pueschel/teaching/18-799B-CMU-spring05/material/eizan-tad.pdf

[93] C. Tomasi and R. Manduchi, 'Bilateral filtering for gray and color images', in *Sixth International Conference on Computer Vision (IEEE Cat. No.98CH36271)*, Jan. 1998, pp. 839–846. doi: 10.1109/ICCV.1998.710815.

[94] I. Sobel and G. Feldman, 'A 3x3 Isotropic Gradient Operator for ImageProcessing', presented at the Stanford Artificial Intelligence Project, Stanford, US, 1968. Accessed: Mar. 21, 2024. [Online]. Available: http://rgdoi.net/10.13140/RG.2.1.1912.4965

[95] A. R. Rao and B. G. Schunck, 'Computing oriented texture fields', *CVGIP: Graphical Models and Image Processing*, vol. 53, no. 2, pp. 157–185, Mar. 1991, doi: 10.1016/1049-9652(91)90059-S.

[96] L. Kabbai, M. Abdellaoui, and A. Douik, 'Image classification by combining local and global features', *The Visual Computer*, vol. 35, no. 5, pp. 679–693, May 2019, doi: 10.1007/s00371-018-1503-0.

[97] Y. Su, S. Shan, X. Chen, and W. Gao, 'Hierarchical Ensemble of Global and Local Classifiers for Face Recognition', in *2007 IEEE 11th International Conference on Computer Vision*, Oct. 2007, pp. 1–8. doi: 10.1109/ICCV.2007.4409060.

[98] A. Kini, A. Kumar Choudhary, D. Hohs, A. Jansche, H. Baumgartl, R. Büttner, T. Bernthaler, D. Goll, and G. Schneider, 'Machine learning-based mass density model for hard magnetic 14:2:1 phases using chemical composition-based features', *Chemical Physics Letters*, vol. 811, p. 140231, Jan. 2023, doi: 10.1016/j.cplett.2022.140231.

[99] S. Xie and Z. Tu, 'Holistically-Nested Edge Detection', in *IEEE International Conference on Computer Vision*, Santiago, Chile: Institute of Electrical and Electronics Engineers, 2015, pp. 1395–1403. Accessed: Sep. 17, 2023. [Online]. Available: https://openaccess.thecvf.com/content_iccv_2015/html/Xie_Holistically-Nested_Edge_Detection_ICCV_2015_paper.html

[100] Y. Zhang and X. Xu, 'Predicting Magnetic Remanence of NdFeB Magnets from Composition', *J Supercond Nov Magn*, vol. 34, no. 11, pp. 2711–2715, Nov. 2021, doi: 10.1007/s10948-021-05901-2.

[101] E. Debie and K. Shafi, 'Implications of the curse of dimensionality for supervised learning classifier systems: theoretical and empirical analyses', *Pattern Anal Applic*, vol. 22, no. 2, pp. 519–536, May 2019, doi: 10.1007/s10044-017-0649-0.

[102] A. Jansche, A.-K. Choudhary, O. Badmos, H. Baumgartl, T. Bernthaler, R. Büttner, and G. Schneider, 'Künstliche Intelligenz in der Materialmikroskopie', *Jahresmagazin Werkstofftechnik: Ingenieurwissenschaften 2019*, pp. 14–17.

[103] D. H. Moore II, 'Classification and regression trees', *Cytometry*, vol. 8, no. 5, pp. 534–535, 1987, doi: 10.1002/cyto.990080516.

[104] Breiman, Leo, 'Random Forests', *Machine Learning*, vol. 45, pp. 5–32, 2001, doi: doi.org/10.1023/A:1010933404324.

[105] J. H. Friedman, 'Greedy function approximation: A gradient boosting machine.', *The Annals of Statistics*, vol. 29, no. 5, pp. 1189–1232, Oct. 2001, doi: 10.1214/aos/1013203451.

[106] K. C. Lee, 'The Evolution of Trees-Based Classification Models', Medium. Accessed: Aug. 17, 2022. [Online]. Available: https://towardsdatascience.com/the-evolution-of-trees-based-classification-models-cb40912c8b35

[107] R. Freund, P. Grigas, and R. Mazumder, 'AdaBoost and Forward Stagewise Regression are First-Order Convex Optimization Methods', *ArXiv*, 2013.

[108] C. Cervellera and D. Macciò, 'Gradient Boosting with Extreme Learning Machines for the Optimization of Nonlinear Functionals', *AIRO Springer Series*, 2019, doi: 10.1007/978-3-030-34960-8_7.

[109] G. Biau and B. Cadre, 'Optimization by gradient boosting', *ArXiv*, 2017, doi: 10.1007/978-3-030-73249-3_2.

[110] L. Grinsztajn, E. Oyallon, and G. Varoquaux, 'Why do tree-based models still outperform deep learning on tabular data?', Jul. 18, 2022, *arXiv*: arXiv:2207.08815. doi: 10.48550/arXiv.2207.08815.

[111] Z. Ullah, N. Yoon, B. K. Tarus, S. Park, and M. Son, 'Comparison of tree-based model with deep learning model in predicting effluent pH and concentration by capacitive deionization', *Desalination*, vol. 558, p. 116614, Jul. 2023, doi: 10.1016/j.desal.2023.116614.

[112] R. Shwartz-Ziv and A. Armon, 'Tabular data: Deep learning is not all you need', *Information Fusion*, vol. 81, pp. 84–90, May 2022, doi: 10.1016/j.inffus.2021.11.011.

[113] N. Rahaman, A. Baratin, D. Arpit, F. Draxler, M. Lin, F. Hamprecht, Y. Bengio, and A. Courville, 'On the Spectral Bias of Neural Networks', in *Proceedings of the 36th International Conference on Machine Learning*, PMLR, May 2019, pp. 5301–5310. Accessed: Sep. 29, 2023. [Online]. Available: https://proceedings.mlr.press/v97/rahaman19a.html

[114] Ravid Shwartz-Ziv, Amichai Painsky and Naftali Tishby, 'Representation Compression and Generalization in Deep Neural Networks', 2019. Accessed: Sep. 30, 2023. [Online]. Available: https://arxiv.org/pdf/2202.06749#page=56

[115] T. Chen and C. Guestrin, 'XGBoost: A Scalable Tree Boosting System', in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco California USA: ACM, Aug. 2016, pp. 785–794. doi: 10.1145/2939672.2939785.

[116] C. Shorten and T. M. Khoshgoftaar, 'A survey on Image Data Augmentation for Deep Learning', *Journal of Big Data*, vol. 6, no. 1, p. 60, Jul. 2019, doi: 10.1186/s40537-019-0197-0.

[117] I. K. Hassani, T. Pellegrini, and T. Masquelier, 'Dilated convolution with learnable spacings', Dec. 07, 2021, *arXiv*: arXiv:2112.03740. doi: 10.48550/arXiv.2112.03740.

[118] S. Jadon, 'A survey of loss functions for semantic segmentation', *2020 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)*, pp. 1–7, Oct. 2020, doi: 10.1109/CIBCB48159.2020.9277638.

[119] F. Chollet, *Deep learning with Python*. Shelter Island, New York: Manning Publications Co, 2018.

[120] O. Ronneberger, P. Fischer, and T. Brox, 'U-Net: Convolutional Networks for Biomedical Image Segmentation', in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds., in Lecture Notes in Computer Science. Cham: Springer International Publishing, 2015, pp. 234–241. doi: 10.1007/978-3-319-24574-4_28.

[121] M. E. Deagen, L. C. Brinson, R. A. Vaia, and L. S. Schadler, 'The materials tetrahedron has a "digital twin"', *MRS Bulletin*, Feb. 2022, doi: 10.1557/s43577-021-00214-0.

[122] A. Samin, M. Kurth, and L. R. Cao, 'An analysis of radiation effects on NdFeB permanent magnets', *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms*, vol. 342, pp. 200–205, Jan. 2015, doi: 10.1016/j.nimb.2014.10.006.

[123] H. Sun, H. Zhang, G. Ren, and C. Zhang, 'A Knowledge Transfer Framework for General Alloy Materials Properties Prediction', *Materials*, vol. 15, no. 21, p. 7442, Oct. 2022, doi: 10.3390/ma15217442.

[124] C. Abache and H. Oesterreicher, 'Structural and magnetic properties of R2Fe14−xTxB (R=Nd, Y; T=Cr, Mn, Co, Ni, Al)', *Journal of Applied Physics*, vol. 60, no. 3, pp. 1114–1117, Aug. 1986, doi: 10.1063/1.337353.

[125] A. Kowalczyk and A. Wrzeciono, 'Magnetic and crystallographic properties of R2Fe14–xCrxB compounds (R = Y, Nd, and Gd)', *Physica Status Solidi (a)*, vol. 110, no. 1, pp. 241–245, Nov. 1988, doi: 10.1002/pssa.2211100125.

[126] A. Kowalczyk, 'Magnetic and crystallographic properties of substituted Pr2Fe14−xMxB compounds (M = Si, Ga, Cr andCu)', *Journal of Magnetism and Magnetic Materials*, vol. 82, no. 1, pp. L1–L4, Nov. 1989, doi: 10.1016/0304-8853(89)90054-1.

[127] A. Kowalczyk and A. Wrzeciono, 'Structural and magnetic characteristics of R2Fe14−xCuxB systems (R = Y, Nd and Gd)', *Journal of Magnetism and Magnetic Materials*, vol. 74, no. 3, pp. 260–262, Oct. 1988, doi: 10.1016/0304-8853(88)90196-5.

[128] C. Abache and H. Oesterreicher, 'Structural and magnetic properties of R2Fe14−xTxB (R=Nd, Y; T=Cr, Mn, Co, Ni, Al)', *Journal of Applied Physics*, vol. 60, no. 3, pp. 1114–1117, Aug. 1986, doi: 10.1063/1.337353.

[129] N. V. Kostyuchenko, I. S. Tereshina, E. A. Tereshina-Chitrova, L. A. Ivanov, M. Paukov, D. I. Gorbunov, A. V. Andreev, M. Doerr, G. A. Politova, A. K. Zvezdin, S. V. Veselova, A. P. Pyatakov, A. Miyata, O. Drachenko, and O. Portugall, 'Drastic reduction of the *R*-Fe exchange in interstitially modified (Nd,Ho)2Fe14B compounds probed by megagauss magnetic fields', *Physical Review Materials*, vol. 5, no. 7, p. 074404, Jul. 2021, doi: 10.1103/PhysRevMaterials.5.074404.

[130] C. Abache and H. Oesterreicher, 'Magnetic properties of compounds R2Fe14B', *Journal of Applied Physics*, vol. 57, no. 8, pp. 4112–4114, Apr. 1985, doi: 10.1063/1.334635.

[131] G. K. Marasinghe, O. A. Pringle, G. J. Long, W. J. James, W. B. Yelon, and F. Grandjean, 'A neutron diffraction and Mössbauer effect study of the magnetic structure of Y2(Fe1− *x* Mn *x*)14B', *Journal of Applied Physics*, vol. 70, no. 10, pp. 6149–6151, Nov. 1991, doi: 10.1063/1.350025.

[132] R. Grössinger, X. C. Chou, R. Krewenka, G. Wiesinger, R. Eibler, X. K. Sun, and Y. C. Chuang, 'The effect of substitution in Nd2Fe14-xZxB (Z = Al, Si, Ga, Co, Ni) compounds', *J. Phys. Colloques*, vol. 49, no. C8, pp. C8-600, Dec. 1988, doi: 10.1051/jphyscol:19888273.

[133] O. A. Pringle, J. Fu, G. J. Long, W. J. James, D. Xie, W. B. Yelon, and F. Grandjean, 'Neutron-diffraction and Mössbauer-effect studies of Pr2(Fe1−xMnx)14B', *Journal of Applied Physics*, vol. 67, no. 9, pp. 4762–4764, May 1990, doi: 10.1063/1.344778.

[134] J. J. Bara, B. F. Bogacz, A. Kowalczyk, and A. Wrzeciono, 'Crystal and magnetic properties of Y2(Fe1−xMx)14B(M = Cu *or* Cr) compounds', *Journal of Magnetism and Magnetic Materials*, vol. 111, no. 1, pp. 39–46, Jun. 1992, doi: 10.1016/0304-8853(92)91051-T.

[135] O. Moze, L. Pareti, M. Solzi, F. Bolzoni, W. I. F. David, W. T. A. Harrison, and A. W. Hewat, 'Magnetic structure and preferential site occupation in manganese- and chromium-substituted Y2Fe14B compounds', *Journal of the Less Common Metals*, vol. 136, no. 2, pp. 375–383, Jan. 1988, doi: 10.1016/0022-5088(88)90440-7.

[136] G. Wiesinger, R. Grössinger, R. Krewenka, X. C. Chou, X. K. Sun, and Y. C. Chuang, 'The effect of F substitution on the magnetic properties of Nd−Fe−B based compounds', *Hyperfine Interact*, vol. 50, no. 1, pp. 693–699, Jun. 1989, doi: 10.1007/BF02407709.

[137] Y. Matsuura, S. Hirosawa, H. Yamamoto, S. Fujimura, and M. Sagawa, 'Magnetic properties of the Nd2(Fe1− *x* Co *x*)14B system', *Applied Physics Letters*, vol. 46, no. 3, pp. 308–310, Feb. 1985, doi: 10.1063/1.95668.

[138] M. Jurczyk and O. D. Chistjakov, 'Crystallographic and magnetic properties of R2(Fe, Co, Nb)14 B-based systems (R = Pr, Pr-Dy)', *Journal of Magnetism and Magnetic Materials*, vol. 78, no. 2, pp. 279–282, Feb. 1989, doi: 10.1016/0304-8853(89)90279-5.

[139] R. Ramprasad, R. Batra, G. Pilania, A. Mannodi-Kanakkithodi, and C. Kim, 'Machine learning in materials informatics: recent applications and prospects', *npj Computational Materials*, vol. 3, no. 1, Art. no. 1, Dec. 2017, doi: 10.1038/s41524-017-0056-5.

[140] G. L. W. Hart, T. Mueller, C. Toher, and S. Curtarolo, 'Machine learning for alloys', *Nature Reviews Materials*, vol. 6, no. 8, Art. no. 8, Aug. 2021, doi: 10.1038/s41578-021-00340-w.

[141] X. Huang, H. Wang, W. Xue, A. Ullah, S. Xiang, H. Huang, L. Meng, G. Ma, and G. Zhang, 'A combined machine learning model for the prediction of time-temperature-transformation diagrams of high-alloy steels', *Journal of Alloys and Compounds*, vol. 823, p. 153694, May 2020, doi: 10.1016/j.jallcom.2020.153694.

[142] C.-T. Wu, H.-T. Chang, C.-Y. Wu, S.-W. Chen, S.-Y. Huang, M. Huang, Y.-T. Pan, P. Bradbury, J. Chou, and H.-W. Yen, 'Machine learning recommends affordable new Ti alloy with bone-like modulus', *Materials Today*, vol. 34, pp. 41–50, Apr. 2020, doi: 10.1016/j.mattod.2019.08.008.

[143] J. D. Evans and F.-X. Coudert, 'Predicting the Mechanical Properties of Zeolite Frameworks by Machine Learning', *Chem. Mater.*, vol. 29, no. 18, pp. 7833–7839, Sep. 2017, doi: 10.1021/acs.chemmater.7b02532.

[144] G. Costanza, M. E. Tata, and N. Ucciardello, 'Superplasticity in PbSn60: Experimental and neural network implementation', *Computational Materials Science*, vol. 37, no. 3, pp. 226–233, Sep. 2006, doi: 10.1016/j.commatsci.2005.06.009.

[145] H.-K. Park, J.-H. Lee, J. Lee, and S.-K. Kim, 'Optimizing machine learning models for granular NdFeB magnets by very fast simulated annealing', *Scientific Reports*, vol. 11, no. 1, p. 3792, Feb. 2021, doi: 10.1038/s41598-021-83315-9.

[146] J. J. Möller, W. Körner, G. Krugel, D. F. Urban, and C. Elsässer, 'Compositional optimization of hard-magnetic phases with machine-learning models', *Acta Materialia*, vol. 153, pp. 53–61, Jul. 2018, doi: 10.1016/j.actamat.2018.03.051.

[147] L. Q. Jiang, J. K. Guo, H. B. Liu, M. Zhu, X. Zhou, P. Wu, and C. H. Li, 'Prediction of lattice constant in cubic perovskites', *Journal of Physics and Chemistry of Solids*, vol. 67, no. 7, pp. 1531–1536, Jul. 2006, doi: 10.1016/j.jpcs.2006.02.004.

[148] A. S. Verma and V. K. Jindal, 'Lattice constant of cubic perovskites', *Journal of Alloys and Compounds*, vol. 485, no. 1, pp. 514–518, Oct. 2009, doi: 10.1016/j.jallcom.2009.06.001.

[149] M. G. Brik and I. V. Kityk, 'Modeling of lattice constant and their relations with ionic radii and electronegativity of constituting ions of A2XY6 cubic crystals (A=K, Cs, Rb, Tl; X=tetravalent cation, Y=F, Cl, Br, I)', *Journal of Physics and Chemistry of Solids*, vol. 72, no. 11, pp. 1256–1260, Nov. 2011, doi: 10.1016/j.jpcs.2011.07.016.

[150] S. Dimitrovska, S. Aleksovska, and I. Kuzmanovski, 'Prediction of the unit cell edge length of cubic A22+BB′O6 perovskites by multiple linear regression and artificial neural networks', *Open Chemistry*, vol. 3, no. 1, pp. 198–215, Mar. 2005, doi: 10.2478/BF02476250.

[151] K. K. Bharathi, R. S. Vemuri, and C. V. Ramana, 'Dysprosium-substitution induced changes in the structure and optical properties of nickel ferrite (NiFe2O4) thin films', *Chemical Physics Letters*, vol. 504, no. 4, pp. 202–205, Mar. 2011, doi: 10.1016/j.cplett.2011.02.003.

[152] Z. Wan, Q.-D. Wang, D. Liu, and J. Liang, 'Discovery of ester lubricants with low coefficient of friction on material surface via machine learning', *Chemical Physics Letters*, vol. 773, p. 138589, Jun. 2021, doi: 10.1016/j.cplett.2021.138589.

[153] Y. Zhang and X. Xu, 'Machine learning lattice constants for spinel compounds', *Chemical Physics Letters*, vol. 760, p. 137993, Dec. 2020, doi: 10.1016/j.cplett.2020.137993.

[154] N. Fujinuma, B. DeCost, J. Hattrick-Simpers, and S. E. Lofland, 'Why big data and compute are not necessarily the path to big materials science', *Commun Mater*, vol. 3, no. 1, Art. no. 1, Aug. 2022, doi: 10.1038/s43246-022-00283-x.

[155] K. Thirunavukkarasu, A. S. Singh, P. Rai, and S. Gupta, 'Classification of IRIS Dataset using Classification Based KNN Algorithm in Supervised Learning', in *2018 4th International Conference on Computing Communication and Automation (ICCCA)*, Dec. 2018, pp. 1–4. doi: 10.1109/CCAA.2018.8777643.

[156] H. Yamada, C. Liu, S. Wu, Y. Koyama, S. Ju, J. Shiomi, J. Morikawa, and R. Yoshida, 'Predicting Materials Properties with Little Data Using Shotgun Transfer Learning', *ACS Cent. Sci.*, vol. 5, no. 10, pp. 1717–1730, Oct. 2019, doi: 10.1021/acscentsci.9b00804.

[157] T. G. Dietterich, 'Ensemble Methods in Machine Learning', in *Multiple Classifier Systems*, in Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, 2000, pp. 1–15. doi: 10.1007/3-540-45014-9_1.

[158] J. Large, J. Lines, and A. Bagnall, 'A probabilistic classifier ensemble weighting scheme based on cross-validated accuracy estimates', *Data Mining and Knowledge Discovery*, vol. 33, no. 6, pp. 1674–1709, 2019, doi: 10.1007/s10618-019-00638-y.

[159] A. Krogh and J. Vedelsby, 'Neural network ensembles, cross validation and active learning', in *Proceedings of the 7th International Conference on Neural Information Processing Systems*, in NIPS'94. Cambridge, MA, USA: MIT Press, Jan. 1994, pp. 231–238.

[160] M. Shahhosseini, R. A. Martinez-Feria, G. Hu, and S. V. Archontoulis, 'Maize yield and nitrate loss prediction with machine learning algorithms', *Environmental Research Letters*, vol. 14, no. 12, p. 124026, Dec. 2019, doi: 10.1088/1748-9326/ab5268.

[161] M. Shahhosseini, G. Hu, and H. Pham, 'Optimizing ensemble weights and hyperparameters of machine learning models for regression problems', *Machine Learning with Applications*, vol. 7, p. 100251, Mar. 2022, doi: 10.1016/j.mlwa.2022.100251.

[162] V. C. Osamor and A. F. Okezie, 'Enhancing the weighted voting ensemble algorithm for tuberculosis predictive diagnosis', *Scientific Reports*, vol. 11, no. 1, Art. no. 1, Jul. 2021, doi: 10.1038/s41598-021-94347-6.

[163] Y. Li and Y. Luo, 'Performance-weighted-voting model: An ensemble machine learning method for cancer type classification using whole-exome sequencing mutation', *Quant Biol*, vol. 8, no. 4, pp. 347–358, Dec. 2020, doi: 10.1007/s40484-020-0226-1.

[164] B. Erdebilli and B. Devrim-İçtenbaş, 'Ensemble Voting Regression Based on Machine Learning for Predicting Medical Waste: A Case from Turkey', *Mathematics*, vol. 10, no. 14, p. 2466, Jul. 2022, doi: 10.3390/math10142466.

[165] Y. Zhao, Z. Shi, J. Zhang, D. Chen, and L. Gu, 'A novel active learning framework for classification: Using weighted rank aggregation to achieve multiple query criteria', *Pattern Recognition*, vol. 93, pp. 581–602, Sep. 2019, doi: 10.1016/j.patcog.2019.03.029.

[166] A. R. M. Castro, E. Galego, N. B. Lima, and R. N. Faria, 'X-ray pole figure analysis of sintered PrFeB magnets', *Journal of Magnetism and Magnetic Materials*, vol. 278, no. 1, pp. 39–45, Jul. 2004, doi: 10.1016/j.jmmm.2003.11.371.

[167] T. Kawai, B. M. Ma, S. G. Sankar, and W. E. Wallace, 'Effect of crystal alignment on the remanence of sintered NdFeB magnets', *Journal of Applied Physics*, vol. 67, no. 9, pp. 4610–4612, May 1990, doi: 10.1063/1.344828.

[168] S. M. Rathod, A. R. Chavan, S. S. Jadhav, K. M. Batoo, M. Hadi, and E. H. Raslan, 'Ag+ ion substituted CuFe2O4 nanoparticles: Analysis of structural and magnetic behavior', *Chemical Physics Letters*, vol. 765, p. 138308, Feb. 2021, doi: 10.1016/j.cplett.2020.138308.

[169] J. F. Herbst and J. J. Croat, 'Neodymium-iron-boron permanent magnets', *Journal of Magnetism and Magnetic Materials*, vol. 100, no. 1, pp. 57–78, Nov. 1991, doi: 10.1016/0304-8853(91)90812-O.

[170] Y. H. Hou, Y. L. Wang, Y. L. Huang, Y. Wang, S. Li, S. C. Ma, Z. W. Liu, D. C. Zeng, L. Z. Zhao, and Z. C. Zhong, 'Effects of Nd-rich phase on the improved properties and recoil loops for

hot deformed Nd-Fe-B magnets', *Acta Materialia*, vol. 115, pp. 385–391, Aug. 2016, doi: 10.1016/j.actamat.2016.06.015.

[171] S. Curtarolo, G. L. W. Hart, M. B. Nardelli, N. Mingo, S. Sanvito, and O. Levy, 'The high-throughput highway to computational materials design', *Nature Mater*, vol. 12, no. 3, pp. 191–201, Mar. 2013, doi: 10.1038/nmat3568.

[172] J. Nelson and S. Sanvito, 'Predicting the Curie temperature of ferromagnets using machine learning', *Phys. Rev. Materials*, vol. 3, no. 10, p. 104405, Oct. 2019, doi: 10.1103/PhysRevMaterials.3.104405.

[173] D.-N. Nguyen, T.-L. Pham, V.-C. Nguyen, A.-T. Nguyen, H. Kino, T. Miyake, and H.-C. Dam, 'A regression-based model evaluation of the Curie temperature of transition-metal rare-earth compounds', *J. Phys.: Conf. Ser.*, vol. 1290, p. 012009, Oct. 2019, doi: 10.1088/1742-6596/1290/1/012009.

[174] H. C. Dam, V. C. Nguyen, T. L. Pham, A. T. Nguyen, H. Kino, K. Terakura, and T. Miyake, 'A regression-based feature selection study of the Curie temperature of transition-metal rare-earth compounds: prediction and understanding', *arXiv:1705.00978 [cond-mat]*, May 2017, Accessed: Jul. 06, 2021. [Online]. Available: http://arxiv.org/abs/1705.00978

[175] Q. Han, T. Birol, and K. Haule, 'Phonon Softening due to Melting of the Ferromagnetic Order in Elemental Iron', *Physical Review Letters*, vol. 120, no. 18, p. 187203, May 2018, doi: 10.1103/PhysRevLett.120.187203.

[176] N. Drebov, A. Martinez-Limia, L. Kunz, A. Gola, T. Shigematsu, T. Eckl, P. Gumbsch, and C. Elsässer, '*Ab initio* screening methodology applied to the search for new permanent magnetic materials', *New Journal of Physics*, vol. 15, no. 12, p. 125023, Dec. 2013, doi: 10.1088/1367-2630/15/12/125023.

[177] T. Long, N. M. Fortunato, Y. Zhang, O. Gutfleisch, and H. Zhang, 'An accelerating approach of designing ferromagnetic materials via machine learning modeling of magnetic ground state and Curie temperature', *Materials Research Letters*, vol. 9, no. 4, pp. 169–174, Apr. 2021, doi: 10.1080/21663831.2020.1863876.

[178] A. K. Choudhary, A. Kini, D. Hohs, A. Jansche, T. Bernthaler, O. Csiszár, D. Goll, and G. Schneider, 'Machine learning-based Curie temperature prediction for magnetic 14:2:1 phases', *AIP Advances*, vol. 13, no. 3, p. 035112, Mar. 2023, doi: 10.1063/5.0116650.

[179] A. Agrawal, P. D. Deshpande, A. Cecen, G. P. Basavarsu, A. N. Choudhary, and S. R. Kalidindi, 'Exploration of data science techniques to predict fatigue strength of steel from composition and processing parameters', *Integrating Materials and Manufacturing Innovation*, vol. 3, no. 1, pp. 90–108, Dec. 2014, doi: 10.1186/2193-9772-3-8.

[180] A. Vishina, O. Yu. Vekilova, T. Björkman, A. Bergman, H. C. Herper, and O. Eriksson, 'High-throughput and data-mining approach to predict new rare-earth free permanent magnets', *Physical Review B*, vol. 101, no. 9, p. 094407, Mar. 2020, doi: 10.1103/PhysRevB.101.094407.

[181] S. Giaremis, G. Katsikas, G. Sempros, M. Gjoka, C. Sarafidis, and J. Kioseoglou, 'Ab initio, artificial neural network predictions and experimental synthesis of mischmetal alloying in Sm–Co permanent magnets', *Nanoscale*, vol. 14, no. 15, pp. 5824–5839, Apr. 2022, doi: 10.1039/D2NR00364C.

[182] Y. Wang, Y. Tian, T. Kirk, O. Laris, J. H. Ross, R. D. Noebe, V. Keylin, and R. Arróyave, 'Accelerated design of Fe-based soft magnetic materials using machine learning and stochastic optimization', *Acta Materialia*, vol. 194, pp. 144–155, Aug. 2020, doi: 10.1016/j.actamat.2020.05.006.

[183] T. D. Rhone, W. Chen, S. Desai, S. B. Torrisi, D. T. Larson, A. Yacoby, and E. Kaxiras, 'Data-driven studies of magnetic two-dimensional materials', *Scientific Reports*, vol. 10, no. 1, Art. no. 1, Sep. 2020, doi: 10.1038/s41598-020-72811-z.

[184] S.-O. Kaba, B. Groleau-Paré, M.-A. Gauthier, A.-M. Tremblay, S. Verret, and C. Gauvin-Ndiaye, 'Prediction of Large Magnetic Moment Materials With Graph Neural Networks and Random Forests', Nov. 14, 2022, *arXiv*: arXiv:2111.14712. doi: 10.48550/arXiv.2111.14712.

[185] A. K. Choudhary, D. Hohs, A. Jansche, T. Bernthaler, D. Goll, and G. Schneider, 'A data-driven approach to predict the saturation magnetization for magnetic 14:2:1 phases from chemical composition', *AIP Advances*, vol. 14, no. 1, p. 015060, Jan. 2024, doi: 10.1063/5.0171922.

[186] V. Pusch, T. Bernthaler, F. Köppen, A. Wilde, A. Nagel, G. Schneider, 'Korrelative und quantitative Gefügeanalyse der Korn- und Domänenstruktur von Eisen-Neodym-Bor-Sintermagneten', *Praktischen Metallographie – Fortschritte in der Metallographie*, vol. 42, pp. 171–176, 2010.

[187] L. Exl, J. Fischbacher, A. Kovacs, H. Özelt, M. Gusenbauer, K. Yokota, T. Shoji, G. Hrkac, and T. Schrefl, 'Magnetic microstructure machine learning analysis', *Journal of Physics: Materials*, Nov. 2018, doi: 10.1088/2515-7639/aaf26d.

[188] M. Neumann, R. Cabiscol, M. Osenberg, H. Markötter, I. Manke, J. H. Finke, and V. Schmidt, 'Characterization of the 3D microstructure of Ibuprofen tablets by means of synchrotron tomography', *Journal of Microscopy*, vol. 274, no. 2, pp. 102–113, 2019, doi: 10.1111/jmi.12789.

[189] F. Jiang, Q. Gu, H. Hau, and N. Li, 'Grain segmentation of multi-angle petrographic thin section microscopic images', in *2017 IEEE International Conference on Image Processing (ICIP)*, Sep. 2017, pp. 3879–3883. doi: 10.1109/ICIP.2017.8297009.

[190] W. Shen, X. Wang, Y. Wang, X. Bai, and Z. Zhang, 'DeepContour: A Deep Convolutional Feature Learned by Positive-Sharing Loss for Contour Detection', presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 3982–3991. Accessed: Sep. 17, 2023. [Online]. Available: https://openaccess.thecvf.com/content_cvpr_2015/html/Shen_DeepContour_A_Deep_2015_CVPR_paper.html

[191] G. Bertasius, J. Shi, and L. Torresani, 'DeepEdge: A Multi-Scale Bifurcated Deep Network for Top-Down Contour Detection', presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 4380–4389. Accessed: Sep. 17, 2023. [Online]. Available: https://openaccess.thecvf.com/content_cvpr_2015/html/Bertasius_DeepEdge_A_Multi-Scale_2015_CVPR_paper.html

[192] J. J. Lim, C. L. Zitnick, and P. Dollar, 'Sketch Tokens: A Learned Mid-level Representation for Contour and Object Detection', presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2013, pp. 3158–3165. Accessed: Sep. 17, 2023. [Online]. Available: https://openaccess.thecvf.com/content_cvpr_2013/html/Lim_Sketch_Tokens_A_2013_CVPR_paper.html

[193] D. Martin, C. Fowlkes, D. Tal, and J. Malik, 'A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics', in *Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001*, Jul. 2001, pp. 416–423 vol.2. doi: 10.1109/ICCV.2001.937655.

[194] David Schuller, Dominic Hohs, Ralf Loeffler, Timo Bernthaler, Dagmar Goll, Gerhard Schneider, 'Analysis of soft magnetic materials by electron backscatter diffraction as a powerful tool | AIP Advances | AIP Publishing', *AIP Advances*, vol. 8, no. 4, p. 047612, 2018, doi: doi.org/10.1063/1.4994200.

[195] D. Goll, 'Quantitative Ermittlung von Gefügestrukturinformationen mittel korrelativer Mikroskopie und EBSD am weichmagnetischen Verbundwerkstoffen für elektrische Maschinen Tagungsband Prakt. Met. Sonderband 50; 229-234', 2016, Accessed: Sep. 17, 2023. [Online]. Available: https://www.hs-aalen.de/de/users/258/publications/10888

[196] D. Goll, ‘Analyse maßgeblicher Gefügecharakteristika von Elektroband zur Minimierung magnetischer Verluste von Elektromotoren Tagungsband Prakt. Met. Sonderband 51; 79-84’, 2017, Accessed: Sep. 17, 2023. [Online]. Available: https://www.hs-aalen.de/de/users/258/publications/10885

[197] D. Manara, A. Seibert, T. Gouder, O. Beneš, L. Martel, J.-Y. Colle, J.-C. Griveau, O. Walter, A. Cambriani, O. Dieste Blanco, D. Staicu, T. Wiss, and J.-F. Vigier, ‘2 - Experimental methods’, in *Advances in Nuclear Fuel Chemistry*, M. H. A. Piro, Ed., in Woodhead Publishing Series in Energy. , Woodhead Publishing, 2020, pp. 89–158. doi: 10.1016/B978-0-08-102571-0.00003-3.

[198] H. Tas, ‘Orientation determinations with polarized light’, *Metallography*, vol. 6, no. 1, pp. 1–8, Feb. 1973, doi: 10.1016/0026-0800(73)90012-8.

[199] L. Böhme, L. Morales-Rivas, S. Diederichs, and E. Kerscher, ‘Crystal c-axis mapping of hcp metals by conventional reflected polarized light microscopy: Application to untextured and textured cp-Titanium’, *Materials Characterization*, vol. 145, pp. 573–581, Nov. 2018, doi: 10.1016/j.matchar.2018.09.024.

[200] M. Peternell, F. Kohlmann, C. J. L. Wilson, C. Seiler, and A. J. W. Gleadow, ‘A new approach to crystallographic orientation measurement for apatite fission track analysis: Effects of crystal morphology and implications for automation’, *Chemical Geology*, vol. 265, no. 3, pp. 527–539, Jul. 2009, doi: 10.1016/j.chemgeo.2009.05.021.

[201] X. Wang, S. Yang, and M. Seita, ‘Combining polarized light microscopy with machine learning to map crystallographic textures on cubic metals’, *Materials Characterization*, vol. 190, p. 112082, Aug. 2022, doi: 10.1016/j.matchar.2022.112082.

[202] K.-W. Jin and M. De Graef, ‘c-Axis orientation determination of α-titanium using Computational Polarized Light Microscopy’, *Materials Characterization*, vol. 167, p. 110503, Sep. 2020, doi: 10.1016/j.matchar.2020.110503.

[203] B. Gaskey, L. Hendl, X. Wang, and M. Seita, ‘Optical characterization of grain orientation in crystalline materials’, *Acta Materialia*, vol. 194, pp. 558–564, Aug. 2020, doi: 10.1016/j.actamat.2020.05.027.

[204] M. Seita, M. M. Nimerfroh, and M. J. Demkowicz, ‘Acquisition of partial grain orientation information using optical microscopy’, *Acta Materialia*, vol. 123, pp. 70–81, Jan. 2017, doi: 10.1016/j.actamat.2016.10.021.

[205] K. Hara, T. Kojima, K. Kutsukake, H. Kudo, and N. Usami, ‘A machine learning-based prediction of crystal orientations for multicrystalline materials’, *APL Machine Learning*, vol. 1, no. 2, p. 026113, May 2023, doi: 10.1063/5.0138099.

[206] B. Lan, T. Ben Britton, T.-S. Jun, W. Gan, M. Hofmann, F. P. E. Dunne, and M. J. S. Lowe, ‘Direct volumetric measurement of crystallographic texture using acoustic waves’, *Acta Materialia*, vol. 159, pp. 384–394, Oct. 2018, doi: 10.1016/j.actamat.2018.08.037.

[207] O. Ilchenko, Y. Pilgun, A. Kutsyk, F. Bachmann, R. Slipets, M. Todeschini, P. O. Okeyo, H. F. Poulsen, and A. Boisen, ‘Fast and quantitative 2D and 3D orientation mapping using Raman microscopy’, *Nat Commun*, vol. 10, no. 1, Art. no. 1, Dec. 2019, doi: 10.1038/s41467-019-13504-8.

[208] D. H. Hurley, S. J. Reese, and F. Farzbod, ‘Application of laser-based resonant ultrasound spectroscopy to study texture in copper’, *Journal of Applied Physics*, vol. 111, no. 5, p. 053527, Mar. 2012, doi: 10.1063/1.3692386.

[209] R. Liu, A. Agrawal, W. Liao, A. Choudhary, and M. De Graef, ‘Materials discovery: Understanding polycrystals from large-scale electron patterns’, in *2016 IEEE International Conference on Big Data (Big Data)*, Dec. 2016, pp. 2261–2269. doi: 10.1109/BigData.2016.7840857.

[210] Z. Ding, E. Pascal, and M. D. Graef, 'Indexing of electron back-scatter diffraction patterns using a convolutional neural network', *Acta Materialia*, vol. 199, pp. 370–382, 2020, doi: 10.1016/j.actamat.2020.08.046.

[211] Z. Ding, C. Zhu, and M. D. Graef, 'Determining crystallographic orientation via hybrid convolutional neural network', *Materials Characterization*, vol. 178, p. 111213, 2021, doi: 10.1016/J.MATCHAR.2021.111213.

[212] Y.-F. Shen, R. Pokharel, T. J. Nizolek, A. Kumar, and T. Lookman, 'Convolutional neural network-based method for real-time orientation indexing of measured electron backscatter diffraction patterns', *Acta Materialia*, vol. 170, pp. 118–131, May 2019, doi: 10.1016/j.actamat.2019.03.026.

[213] H. M. Fitzek, R. Schmidt, M. Nachtnebel, J. Rattenberger, A. Zankel, F. Hofer, and H. Schröttner, 'Correlative microscopy and machine learning –new tools for material characterization', presented at the 20th Plansee Seminar, Reutte, Austria, 2022. [Online]. Available: https://pure.tugraz.at/ws/portalfiles/portal/46004234/Poster_Fitzek_Plansee.pdf

[214] M. Wittwer and M. Seita, 'A machine learning approach to map crystal orientation by optical microscopy', *npj Computational Materials*, vol. 8, no. 1, p. 8, Dec. 2022, doi: 10.1038/s41524-021-00688-1.

[215] O. E. Randle Valerie, *Introduction to Texture Analysis: Macrotexture, Microtexture, and Orientation Mapping, Second Edition*, 2nd ed. Boca Raton: CRC Press, 2009. doi: 10.1201/9781420063660.

[216] A. K. Choudhary, T. Bernthaler, T. Grubesa, A. Jansche, and G. Schneider, 'Machine Learning for Microstructures Classification in Functional Materials', *TechConnect Briefs*, pp. 114–117, Oct. 2021.

[217] F. Bolzoni, F. Leccabue, O. Moze, L. Pareti, M. Solzi, and A. Deriu, '3 *d* and 4 *f* magnetism in Nd2Fe14− *x* Co *x* B and Y2Fe14− *x* Co *x* B compounds', *Journal of Applied Physics*, vol. 61, no. 12, pp. 5369–5373, Jun. 1987, doi: 10.1063/1.338276.

[218] K. Shimizu, K. Ichinose, Y. Fukuda, and M. Shimotomai, '143Nd NMR study of 3d and 4f magnetism in Nd2(Fe1−xCox)14B and Nd2(Fe1−yNiy)14B', *Solid State Communications*, vol. 96, no. 9, pp. 671–674, Dec. 1995, doi: 10.1016/0038-1098(95)00527-7.

[219] Y. Yang, D. E. Tharp, G. J. Long, O. A. Pringle, and W. J. James, 'A Mössbauer effect study of the structural and magnetic properties of Y2(Fe1− *x* Al *x*) 14B', *Journal of Applied Physics*, vol. 61, no. 8, pp. 4343–4345, Apr. 1987, doi: 10.1063/1.338961.

[220] E. Burzo, N. Plugaru, V. Pop, and L. Stanciu, 'Bulk Magnetic Properties of The Y2TxFe14-xB Compounds, Where T = Al, Ni OR Co', *Solid State Communications*, vol. 58, no. 10, pp. 803–805, 1986.

[221] H. Fujii, W. E. Wallace, and E. B. Boltich, 'Concerning magnetic characteristics of (R2-xR′x)Fe12Co2B (R = Pr and Nd, R′ = Tb and Dy)', *Journal of Magnetism and Magnetic Materials*, vol. 61, no. 3, pp. 251–256, Oct. 1986, doi: 10.1016/0304-8853(86)90034-X.

[222] C. Chacon and O. Isnard, 'Determination of the substitution scheme of gallium and magnetic features of Nd2Fe14−xGaxB', *Journal of Applied Physics*, vol. 88, no. 6, pp. 3570–3577, Sep. 2000, doi: 10.1063/1.1286726.

[223] J. Q. Xie, C. H. Wu, Y. C. Chuang, and F. M. Yang, 'Effect of gallium on the magnetic properties of Nd2Fe11.5Co2.5B compounds', *Solid State Communications*, vol. 71, no. 5, pp. 329–332, Aug. 1989, doi: 10.1016/0038-1098(89)90764-3.

[224] M. Jurczyk, 'Effect of substitution of Al and Mo on the magnetic properties of R2Fe12-xTxCo2B (R=synthetic mischmetal, didymium and neodymium)', *Journal of Magnetism and Magnetic Materials*, vol. 73, no. 2, pp. 199–204, Jun. 1998, doi: https://doi.org/10.1016/0304-8853(88)90293-4.

[225] W. Zhang, S. Zhang, A. Yan, H. Zhang, and B. Shen, 'Effect of the substitution of Pr for Nd on microstructure and magnetic properties of nanocomposite Nd2Fe14B/α-Fe magnets', *Journal of Magnetism and Magnetic Materials*, 2001.

[226] Z. Chen, B. R. Smith, D. N. Brown, and B. M. Ma, 'Effect of Zr substitution for rare earth on microstructure and magnetic properties of melt-spun (Nd0.75Pr0.25)12.5-xZrxFe82B5.5 (x=0–3) ribbons', *Journal of Applied Physics*, vol. 91, no. 10, pp. 8168–8170, May 2002, doi: https://doi.org/10.1063/1.1446119.

[227] J. Chaboy, N. Plugaru, J. Bartolomé, and G. Subías, 'Enhancement of anisotropy in Nd2Fe14B driven by Eu substitution', *Physical Review B*, vol. 67, no. 1, p. 014415, Jan. 2003, doi: 10.1103/PhysRevB.67.014415.

[228] D. W. Lim, H. Kato, M. Yamada, G. Kido, and Y. Nakagawa, 'High-field magnetization process and spin reorientation in (Nd1−xDyx)2Fe14B single crystals', *Physical Review B*, vol. 44, no. 18, pp. 10014–10020, Nov. 1991, doi: 10.1103/PhysRevB.44.10014.

[229] Y. Yang and X. Zhang, 'Intrinsic magnetic hardness of R2(Fe1−xMnx)14B', *Journal of Applied Physics*, vol. 67, no. 9, pp. 4765–4767, May 1990, doi: 10.1063/1.344779.

[230] J. M. D. Coey, 'Intrinsic magnetic properties of compounds with the Nd2Fe14B structure', *Journal of the Less Common Metals*, vol. 126, pp. 21–34, 1986, doi: https://doi.org/10.1016/0022-5088(86)90245-6.

[231] A. Kowalczyk, 'Magnetic and Crystallographic Properties of Substituted Pr2Fel4-xMxB Compounds (M--Si, Ga, Cr and Cu)', *Journal of Magnetism and Magnetic Materials*, vol. 82, no. 01, pp. L1–L4, 1989.

[232] F. R. De Boer, H. Ying-kai, Z. Zhi-dong, D. B. De Mooij, and K. H. J. Buschow, 'Magnetic and crystallographic properties of ternary rare-earth compounds of the type R2Fe14C', *Journal of Magnetism and Magnetic Materials*, vol. 72, no. 2, pp. 167–173, Apr. 1988, doi: 10.1016/0304-8853(88)90184-9.

[233] C. Abache and J. Oesterreicher, 'Magnetic anisotropies and spin reorientations of R2Fe14B-type compounds', *Journal of Applied Physics*, vol. 60, no. 10, pp. 3671–3679, Nov. 1986, doi: 10.1063/1.337574.

[234] F. Compounds, X. C. Kou, X. K. Sun, and Y. C. Chuang, 'Magnetic Anisotropy and Exchange Interaction in Gd2Fe14C and Gd2Fe14B Compounds', *Solid State Communications*, vol. 73, no. 01, pp. 87–89, 1990.

[235] B. Cheng, S. Fu, Y. Yang, and W. J. James, 'Magnetic behavior and anomalous thermal expansion of Nd2(Fe1−x Mnx)14B', *Journal of Applied Physics*, vol. 64, no. 10, pp. 5543–5545, Nov. 1988, doi: 10.1063/1.342330.

[236] M. Jurczyk and W. E. Wallace, 'Magnetic behavior of R1.9Zr0.1Fe14B and R1.9Zr0.1Fe12Co2B compounds', *Journal of Magnetism and Magnetic Materials*, vol. 59, no. 3, pp. L182–L184, Jun. 1986, doi: 10.1016/0304-8853(86)90409-9.

[237] S. Y. Jiang, F. Pourarian, J. X. Yan, S. G. Sankar, and W. E. Wallace, 'Magnetic characteristics of NdTbFe14-xCoxB and Nd2-xTbxCo14 alloys', *Journal of Magnetism and Magnetic Materials*, vol. 74, no. 2, pp. 181–185, Sep. 1988, doi: 10.1016/0304-8853(88)90066-2.

[238] V. Villas-Boas, F. P. Missell, and S. F. Da Cunha, 'Magnetic properties of La2(Fe1− *x* Co *x*)14B and Nd2(Fe1− *x* Co *x*)14B', *Journal of Applied Physics*, vol. 64, no. 10, pp. 5549–5551, Nov. 1988, doi: 10.1063/1.342332.

[239] Y. Fukuda, A. Fujita, and M. Shimotomai, 'Magnetic properties of monocrystalline Nd2(Fe, Co, Ni)14B', *Journal of Alloys and Compounds*, vol. 193, pp. 256–258, 1993.

[240] S. Hirosawa, Y. Yamaguchi, K. Tokuhara, H. Yamamoto, S. Fujimura, and M. Sagawa, 'Magnetic properties of Nd2(Fe1-xMx)14B measured on single crystals (M =Al, Cr, Mn and Co)', *IEEE Transactions on Magnetics*, vol. 23, no. 5, pp. 2120–2122, Sep. 1987, doi: 10.1109/TMAG.1987.1065623.

[241] N. C. Koon, 'Magnetic properties of R2Fe14B single crystals', in *New Frontiers in Rare Earth Science and Applications*, China: Academic Press, 1985, p. 936. doi: https://doi.org/10.1016/B978-0-12-767662-3.50048-6.

[242] S. Bao-gen, D. Jun, B.-X. Gu, Z. Zhi-ying, H. Homburg, Z. Jian-gao, and S. Methfessel, 'Magnetic properties of rapidly quenched (Nd2Fe14B)1-x(Nd1.1Fe4B4)x alloys', *Journal of Magnetism and Magnetic Materials*, vol. 92, pp. 53–58, 1990.

[243] F. Bolzoni, F. Leccabue, O. Moze, L. Paret, and M. Solzi, 'Magnetocrystalline anisotropy of Ni and Mn substituted Nd2Fe14B compounds', *Journal of Magnetism and Magnetic Materials*, vol. 67, pp. 373–377, 1987, doi: https://doi.org/10.1016/0304-8853(87)90197-1.

[244] M. J. Hawton and W. D. Corner, 'Magnetocrystalline anisotropy of R2Fe14B, R= Dy, Ho, Gd, measured using high field torque magnetometry', *Journal of Magnetism and Magnetic Materials*, vol. 72, pp. 52–58, 1988, doi: https://doi.org/10.1016/0304-8853(88)90269-7.

[245] Y. Fu-Ming, H. Ying-Kait, and F. R. D. Boer, 'Magnetic Anisotropy in Pr2(Fe1-xCox)14B Compounds', *Journal of the Less Common Metals*, vol. 132, pp. 265–272, 1987.

[246] A. Kowalczyk, P. Stefański, A. Szlaferek, and A. Wrzeciono, 'Spin phase diagrams and magnetic anisotropy in Nd2Fe14−xMxB compounds (M = Si, Cr and Cu)', *Journal of Magnetism and Magnetic Materials*, vol. 83, no. 1–3, pp. 147–148, Jan. 1990, doi: 10.1016/0304-8853(90)90464-2.

[247] M. J. Kim, Y. B. Kim, C. S. Kim, and T. K. Kim, 'Spin reorientation and magnetocrystalline anisotropy of (Nd1-xDyx)Fe14B', *Journal of Magnetism and Magnetic Materials*, pp. 49–54, 2001.

[248] S. K. Chen, J. G. Duh, and H. C. Ku, 'Structural and magnetic properties of pseudoternary Nd2(Fe1− *x* Cu *x*)14B compounds', *Journal of Applied Physics*, vol. 63, no. 8, pp. 2739–2741, Apr. 1988, doi: 10.1063/1.340969.

[249] W. E. Wallace, S. G. Sankar, E. Oswald, and Y. Jiang, 'Studies of Substituted R2 T14 B and R2 Co17 Systems (T = Fe or Co)', *MRS Online Proceedings Library*, vol. 96, no. 17, 1987, doi: https://doi.org/10.1557/PROC-96-17.

[250] L. H. Lewis, V. Panchanathan, and J.-Y. Wang, 'Technical magnetic properties of melt-spun (Nd1−xPrx)2Fe14B at low temperature', *Journal of Magnetism and Magnetic Materials*, vol. 176, no. 2–3, pp. 288–296, Dec. 1997, doi: 10.1016/S0304-8853(97)00648-3.

[251] C. Quan, Y. Wang, L. Yin, and J. Zhao, 'The effects of Ga addition on the structure and agnetic properties of Nd2Fe14B compounds', *Solid State Communications*, vol. 72, no. 9, pp. 955–957, 1989, doi: https://doi.org/10.1016/0038-1098(89)90435-3 Get rights and content.

[252] K. Kobayashi, K. Urushibata, T. Matsushita, and T. Akiya, 'The origin of unique initial magnetization curves in Dy-substituted (Nd1− *x* , Dyx)2Fe14B (*x* = 0−1) sintered magnets', *Journal of Applied Physics*, vol. 111, no. 2, p. 023907, Jan. 2012, doi: 10.1063/1.3676209.

[253] M. A. Susner, B. S. Conner, B. I. Saparov, M. A. McGuire, E. J. Crumlin, G. M. Veith, H. Cao, K. V. Shanavas, D. S. Parker, B. C. Chakoumakos, and B. C. Sales, 'Flux growth and characterization of Ce-substituted Nd2Fe14B single crystals', *Journal of Magnetism and Magnetic Materials*, vol. 434, pp. 1–9, Jul. 2017, doi: 10.1016/j.jmmm.2016.10.127.

[254] S. Sinnema, R. J. Radwanski, J. J. M. Franse, D. B. de Mooij, and K. H. J. Buschow, 'Magnetic properties of ternary rare-earth compounds of the type R2Fe14B', *Journal of Magnetism and Magnetic Materials*, vol. 44, no. 3, pp. 333–341, Oct. 1984, doi: 10.1016/0304-8853(84)90261-0.

[255] C. Lin, Z.-X. Liu, Y.-X. Sun, C.-X. Bai, and T.-S. Zhao, 'Effect of exchange interaction on spin reorientation in the Nd2Fe14B system', *Physical Review B*, vol. 39, no. 10, pp. 7273–7276, Apr. 1989, doi: 10.1103/PhysRevB.39.7273.

[256] S. Feng, H. Zhou, and H. Dong, 'Using deep neural network with small dataset to predict material defects', *Materials & Design*, vol. 162, pp. 300–310, Jan. 2019, doi: 10.1016/j.matdes.2018.11.060.

[257] Y. Zhang and C. Ling, 'A strategy to apply machine learning to small datasets in materials science', *npj Computational Materials*, vol. 4, no. 1, Art. no. 1, May 2018, doi: 10.1038/s41524-018-0081-z.

[258] P. Xu, X. Ji, M. Li, and W. Lu, 'Small data machine learning in materials science', *npj Computational Materials*, vol. 9, no. 1, Art. no. 1, Mar. 2023, doi: 10.1038/s41524-023-01000-z.

[259] P. Raccuglia, K. C. Elbert, P. D. F. Adler, C. Falk, M. B. Wenny, A. Mollo, M. Zeller, S. A. Friedler, J. Schrier, and A. J. Norquist, 'Machine-learning-assisted materials discovery using failed experiments', *Nature*, vol. 533, no. 7601, Art. no. 7601, May 2016, doi: 10.1038/nature17439.

[260] Y. Sano, H. Onodera, H. Yamauchi, and H. Yamamoto, 'Magnetic properties of the 3d sublattice in pseudoternary compounds Y2Fe14-xMxB with M = Co and Mn', *Journal of Magnetism and Magnetic Materials*, vol. 79, no. 1, pp. 67–75, Apr. 1989, doi: 10.1016/0304-8853(89)90293-X.

[261] W. Cheng, 'Prediction of magnetic remanence of NdFeB magnets by using novel machine learning intelligence approach — Support vector regression', in *2014 IEEE 13th International Conference on Cognitive Informatics and Cognitive Computing*, Aug. 2014, pp. 431–435. doi: 10.1109/ICCI-CC.2014.6921494.

[262] K. Simonyan and A. Zisserman, 'Very Deep Convolutional Networks for Large-Scale Image Recognition', Apr. 10, 2015, *arXiv*: arXiv:1409.1556. doi: 10.48550/arXiv.1409.1556.

[263] D. Hasler and S. E. Suesstrunk, 'Measuring colorfulness in natural images', in *Human Vision and Electronic Imaging VIII*, SPIE, Jun. 2003, pp. 87–95. doi: 10.1117/12.477378.

[264] A. Buslaev, V. I. Iglovikov, E. Khvedchenya, A. Parinov, M. Druzhinin, and A. A. Kalinin, 'Albumentations: Fast and Flexible Image Augmentations', *Information*, vol. 11, no. 2, Art. no. 2, Feb. 2020, doi: 10.3390/info11020125.

[265] X. J. Yin, I. P. Jones, and I. R. Harris, 'The microstructural characterisation of Nd-Fe-B alloys. II: Microstructural investigation of cast Nd-Fe-B materials', *Journal of Magnetism and Magnetic Materials*, vol. 125, no. 1, pp. 91–102, Jul. 1993, doi: 10.1016/0304-8853(93)90822-J.

[266] S. I. Wright, M. M. Nowell, S. P. Lindeman, P. P. Camus, M. De Graef, and M. A. Jackson, 'Introduction and comparison of new EBSD post-processing methodologies', *Ultramicroscopy*, vol. 159, pp. 81–94, Dec. 2015, doi: 10.1016/j.ultramic.2015.08.001.

[267] R. van den Goorbergh, M. van Smeden, D. Timmerman, and B. Van Calster, 'The harm of class imbalance corrections for risk prediction models: illustration and simulation using logistic regression', *Journal of the American Medical Informatics Association*, vol. 29, no. 9, pp. 1525–1534, Sep. 2022, doi: 10.1093/jamia/ocac093.

[268] H. Rezatofighi, N. Tsoi, J. Gwak, A. Sadeghian, I. Reid, and S. Savarese, 'Generalized Intersection Over Union: A Metric and a Loss for Bounding Box Regression', presented at the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE Computer Society, Jun. 2019, pp. 658–666. doi: 10.1109/CVPR.2019.00075.

[269] H. Kervadec, J. Bouchtiba, C. Desrosiers, E. Granger, J. Dolz, and I. B. Ayed, 'Boundary loss for highly unbalanced segmentation', *Medical Image Analysis*, vol. 67, p. 101851, Jan. 2021, doi: 10.1016/j.media.2020.101851.

[270] D. Hu, H. Yang, and X. Hou, 'Distance Field-Based Convolutional Neural Network for Edge Detection', *Computational Intelligence and Neuroscience*, vol. 2022, p. e1712258, Mar. 2022, doi: 10.1155/2022/1712258.

[271] S. Dangi, C. A. Linte, and Z. Yaniv, 'A distance map regularized CNN for cardiac cine MR image segmentation', *Medical Physics*, vol. 46, no. 12, pp. 5637–5651, 2019, doi: 10.1002/mp.13853.

[272] P. Ribalta Lorenzo, M. Marcinkiewicz, and J. Nalepa, 'Multi-modal U-Nets with Boundary Loss and Pre-training for Brain Tumor Segmentation', in *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, A. Crimi and S. Bakas, Eds., in Lecture Notes in Computer Science. Cham: Springer International Publishing, 2020, pp. 135–147. doi: 10.1007/978-3-030-46643-5_13.

[273] H. Trevor, T. Robert, and F. Jerome, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed. New York: Springer, 2009.

[274] L. I.-K. Lin, 'A Concordance Correlation Coefficient to Evaluate Reproducibility', *Biometrics*, vol. 45, no. 1, pp. 255–268, 1989, doi: 10.2307/2532051.

[275] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and É. Duchesnay, 'Scikit-learn: Machine Learning in Python', *Journal of Machine Learning Research*, vol. 12, no. 85, pp. 2825–2830, 2011.

[276] F. Chollet and others, 'Keras: The Python Deep Learning library', *Astrophysics Source Code Library*, p. ascl:1806.022, Jun. 2018.

[277] P. J. Huber, 'Robust Estimation of a Location Parameter', *The Annals of Mathematical Statistics*, vol. 35, no. 1, pp. 73–101, Mar. 1964, doi: 10.1214/aoms/1177703732.

[278] K. Gauen, R. Dailey, J. Laiman, Y. Zi, N. Asokan, Y.-H. Lu, G. K. Thiruvathukal, M.-L. Shyu, and S.-C. Chen, 'Comparison of Visual Datasets for Machine Learning', in *2017 IEEE International Conference on Information Reuse and Integration (IRI)*, San Diego, CA: IEEE, Aug. 2017, pp. 346–355. doi: 10.1109/IRI.2017.59.

[279] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, 'Focal Loss for Dense Object Detection', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 02, pp. 318–327, 20AD, doi: 10.1109/TPAMI.2018.2858826.

[280] C. Nwankpa, W. Ijomah, A. Gachagan, and S. Marshall, 'Activation Functions: Comparison of trends in Practice and Research for Deep Learning', Nov. 08, 2018, *arXiv*: arXiv:1811.03378. doi: 10.48550/arXiv.1811.03378.

[281] A. Kirillov, E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, T. Xiao, S. Whitehead, A. C. Berg, W.-Y. Lo, P. Dollár, and R. Girshick, 'Segment Anything', Apr. 05, 2023, *arXiv*: arXiv:2304.02643. Accessed: Apr. 06, 2023. [Online]. Available: http://arxiv.org/abs/2304.02643

[282] X. Li, L. Cui, J. Li, Y. Chen, W. Han, S. Shonkwiler, and S. McMains, 'Automation of intercept method for grain size measurement: A topological skeleton approach', *Materials & Design*, vol. 224, p. 111358, Dec. 2022, doi: 10.1016/j.matdes.2022.111358.

[283] 'E1382 Standard Test Methods for Determining Average Grain Size Using Semiautomatic and Automatic Image Analysis'. Accessed: Oct. 08, 2023. [Online]. Available: https://www.astm.org/e1382-97r15.html

[284] M. Teague, 'Image analysis via the general theory of moments', *Journal of the Optical Society of America*, vol. 70, no. 8, pp. 920–930, 1980, doi: 10.1364/JOSA.70.000920.

[285] J. Nunez-Iglesias, A. J. Blanch, O. Looker, M. W. Dixon, and L. Tilley, 'A new Python library to analyse skeleton images confirms malaria parasite remodelling of the red blood cell membrane skeleton', *PeerJ*, vol. 6, p. e4312, Feb. 2018, doi: 10.7717/peerj.4312.

[286] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, 'ImageNet: A large-scale hierarchical image database', in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, Jun. 2009, pp. 248–255. doi: 10.1109/CVPR.2009.5206848.

[287] A. Krizhevsky, 'Learning Multiple Layers of Features from Tiny Images', Master's thesis, University of Toronto, 2009. [Online]. Available: https://www.cs.toronto.edu/~kriz/cifar.html

[288] K. He, X. Zhang, S. Ren, and J. Sun, 'Identity Mappings in Deep Residual Networks', Jul. 25, 2016, *arXiv*: arXiv:1603.05027. doi: 10.48550/arXiv.1603.05027.

[289] M. Tan and Q. V. Le, 'EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks', Sep. 11, 2020, *arXiv*: arXiv:1905.11946. doi: 10.48550/arXiv.1905.11946.

[290] X. Glorot and Y. Bengio, 'Understanding the difficulty of training deep feedforward neural networks', in *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, JMLR Workshop and Conference Proceedings, Mar. 2010, pp. 249–256. Accessed: Nov. 09, 2023. [Online]. Available: https://proceedings.mlr.press/v9/glorot10a.html

[291] David J. C. MacKay, 'A Practical Bayesian Framework for Backpropagation Networks | Neural Computation | MIT Press', *Neural Computation*, vol. 4, no. 3, pp. 448–472, 1992.

[292] K. Team, 'Keras documentation: Keras Applications'. Accessed: Jul. 26, 2022. [Online]. Available: https://keras.io/api/applications/

[293] W. W. Sun and K. P. Kodama, 'Magnetic anisotropy, scanning electron microscopy, and X ray pole figure goniometry study of inclination shallowing in a compacting clay-rich sediment', *Journal of Geophysical Research: Solid Earth*, vol. 97, no. B13, pp. 19599–19615, 1992, doi: 10.1029/92JB01589.

[294] M. Li, D. Chen, S. Liu, and F. Liu, 'Grain boundary detection and second phase segmentation based on multi-task learning and generative adversarial network', *Measurement*, vol. 162, p. 107857, Oct. 2020, doi: 10.1016/j.measurement.2020.107857.

[295] H. Abrams, 'Grain size measurement by the intercept method', *Metallography*, vol. 4, no. 1, pp. 59–78, Feb. 1971, doi: 10.1016/0026-0800(71)90005-X.

[296] G. Tvrtko, 'Skalenübergreifende Quantifizierung der Anisotropie von Sintermagneten auf Fe-Nd-B-Basis', Doctoral dissertation, Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany, 2025.

[297] A. Conneau, K. Khandelwal, N. Goyal, V. Chaudhary, G. Wenzek, F. Guzmán, E. Grave, M. Ott, L. Zettlemoyer, and V. Stoyanov, 'Unsupervised Cross-lingual Representation Learning at Scale', in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, Online: Association for Computational Linguistics, 2020, pp. 8440–8451. doi: 10.18653/v1/2020.acl-main.747.

[298] B. Bischl, G. Casalicchio, M. Feurer, P. Gijsbers, F. Hutter, M. Lang, R. G. Mantovani, J. N. van Rijn, and J. Vanschoren, 'OpenML Benchmarking Suites', Nov. 22, 2021, *arXiv*: arXiv:1708.03731. Accessed: Sep. 18, 2023. [Online]. Available: http://arxiv.org/abs/1708.03731

[299] P. Gijsbers, E. LeDell, J. Thomas, S. Poirier, B. Bischl, and J. Vanschoren, 'An Open Source AutoML Benchmark', Jul. 01, 2019, *arXiv*: arXiv:1907.00909. Accessed: Sep. 18, 2023. [Online]. Available: http://arxiv.org/abs/1907.00909

[300] A. Stoll and P. Benner, 'Machine learning for material characterization with an application for predicting mechanical properties', *GAMM-Mitteilungen*, vol. 44, no. 1, p. e202100003, 2021, doi: 10.1002/gamm.202100003.

[301] Chang, R., Wang, YX. & Ertekin, E, 'Towards overcoming data scarcity in materials science: unifying models and datasets with a mixture of experts framework', *npj Computational Materials*, vol. 8, p. 242, 2022, doi: doi.org/10.1038/s41524-022-00929-x.

[302] J. Nelson and S. Sanvito, 'Predicting the Curie temperature of ferromagnets using machine learning', *Physical Review Materials*, vol. 3, no. 10, p. 104405, Oct. 2019, doi: 10.1103/PhysRevMaterials.3.104405.

[303] J. M. Chambers, T.J. Hastle, *Statistical Models in S*, 1st ed. New York, US: Routledge, Boca Raton, 2017. Accessed: Sep. 18, 2023. [Online]. Available: https://www.routledge.com/Statistical-Models-in-S/Chambers-Hastie/p/book/9780412830402

[304] X. Li, Y. Wang, S. Basu, K. Kumbier, and B. Yu, 'A Debiased MDI Feature Importance Measure for Random Forests', in *Advances in Neural Information Processing Systems*, H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, Eds., Curran Associates, Inc., 2019. [Online]. Available: https://proceedings.neurips.cc/paper/2019/file/702cafa3bb4c9c86e4a3b6834b45aedd-Paper.pdf

[305] S. Nirmalraj, A. S. M. Antony, P. Srideviponmalar, A. S. Oliver, K. J. Velmurugan, V. Elanangai, and G. Nagarajan, 'Permutation feature importance-based fusion techniques for diabetes prediction', *Soft Computing*, Apr. 2023, doi: 10.1007/s00500-023-08041-y.

[306] S. M. Lundberg and S.-I. Lee, 'A Unified Approach to Interpreting Model Predictions', in *Advances in Neural Information Processing Systems*, Curran Associates, Inc., 2017. Accessed: Mar. 29, 2023. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html

[307] Y. V. Karpievitch, E. G. Hill, A. P. Leclerc, A. R. Dabney, and J. S. Almeida, 'An Introspective Comparison of Random Forest-Based Classifiers for the Analysis of Cluster-Correlated Data by Way of RF++', *PLoS One*, vol. 4, no. 9, p. e7087, Sep. 2009, doi: 10.1371/journal.pone.0007087.

[308] D. H. Wolpert, 'Stacked generalization', *Neural Networks*, vol. 5, no. 2, pp. 241–259, Jan. 1992, doi: 10.1016/S0893-6080(05)80023-1.

[309] M. Mukaka, 'A guide to appropriate use of Correlation coefficient in medical research', *Malawi Medical Journal*, vol. 24, no. 3, pp. 69–71, Sep. 2012.

[310] Z. Qiao, S. Dong, Q. Li, X. Lu, R. Chen, S. Guo, A. Yan, and W. Li, 'Performance prediction models for sintered NdFeB using machine learning methods and interpretable studies', *Journal of Alloys and Compounds*, vol. 963, p. 171250, Nov. 2023, doi: 10.1016/j.jallcom.2023.171250.

[311] G. Palmer, S. Du, A. Politowicz, J. P. Emory, X. Yang, A. Gautam, G. Gupta, Z. Li, R. Jacobs, and D. Morgan, 'Calibration after bootstrap for accurate uncertainty quantification in regression models', *npj Computational Materials*, vol. 8, no. 1, pp. 1–9, May 2022, doi: 10.1038/s41524-022-00794-8.

[312] M. Volpi and D. Tuia, 'Dense Semantic Labeling of Subdecimeter Resolution Images With Convolutional Neural Networks', *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 2, pp. 881–893, Feb. 2017, doi: 10.1109/TGRS.2016.2616585.

[313] W. Baccouch, S. Oueslati, B. Solaiman, and S. Labidi, 'A comparative study of CNN and U-Net performance for automatic segmentation of medical images: application to cardiac MRI', *Procedia Computer Science*, vol. 219, pp. 1089–1096, Jan. 2023, doi: 10.1016/j.procs.2023.01.388.

[314] K. P. Mingard, P. N. Quested, and M. S. Peck, 'Determination of grain size by EBSD – Report on a round robin measurement of equiaxed Titanium', NPL Report MAT 56, 2012. [Online]. Available: https://eprintspublications.npl.co.uk/5412/

[315] B. Anthony and V. M. Miller, 'A "step" in the right direction: Quantifying error in twinning analysis due to EBSD scan resolution', *Scripta Materialia*, vol. 242, p. 115929, Mar. 2024, doi: 10.1016/j.scriptamat.2023.115929.

[316] J. Schindelin, I. Arganda-Carreras, E. Frise, V. Kaying, M. Lingair, T. Pietzsch, S. Preibisch, C. Rueden, S. Saalfeld, B. Schmid, J. Tinevez, D. White, V. Hartenstein, K. Eliceiri, P. Tomancak, A. Cardona, 'Fiji: an open-source platform for biological-image analysis', *Nature Methods*, vol. 9, pp. 676–682, 2012.

[317] 'ZEISS ZEN analyzer', Zeiss Zen Core. Accessed: Sep. 17, 2023. [Online]. Available: https://www.zeiss.com/microscopy/en/products/software/zeiss-zen-analyzer.html

[318] M. T. Ribeiro, S. Singh, and C. Guestrin, 'Anchors: High-Precision Model-Agnostic Explanations', *AAAI*, vol. 32, no. 1, Apr. 2018, doi: 10.1609/aaai.v32i1.11491.

[319] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, 'SLIC Superpixels Compared to State-of-the-Art Superpixel Methods', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 11, pp. 2274–2282, Nov. 2012, doi: 10.1109/TPAMI.2012.120.

[320] Z. Chen, M. Li, C. Liu, Z. Ma, Y. Han, J. Gao, W. Wei, Z. Sheng, and H. Du, 'Magnetic Domain Structure in Ferromagnetic Kagome Metal DyMn6Sn6', *Frontiers in Physics*, vol. 9, 2021, Accessed: Jul. 19, 2022. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fphy.2021.685510

[321] A. Hubert and R. Schäfer, 'Domain Theory', in *Magnetic Domains: The Analysis of Magnetic Microstructures*, A. Hubert and R. Schäfer, Eds., Berlin, Heidelberg: Springer, 1998, pp. 99–335. doi: 10.1007/978-3-540-85054-0_3.

[322] R. Löffler, D. Goll, G. Guth, T. Bernthaler, V. Pusch, and G. Schneider, 'Lichtmikroskopische Analyse der intrinsischen Eigenschaften hart- magnetischer Phasen aus der Domänenstruktur', Carl Zeiss Microscopy GmbH, White paper, Apr. 2013.

[323] Q. Sun, M. Zhu, J. Bai, and Q. Wang, 'Relationship between Remanence and Micromorphology of Nd-Fe-B Permanent Magnets Revealed by Fractal Theory and EBSD Data', *Fractal and Fractional*, vol. 7, no. 5, Art. no. 5, May 2023, doi: 10.3390/fractalfract7050393.

List of publications

Peer-reviewed journal publications

A. K. Choudhary, D. Hohs, A. Jansche, T. Bernthaler, D. Goll, and G. Schneider, 'A data-driven approach to predict the saturation magnetization for magnetic 14:2:1 phases from chemical composition', AIP Advances, vol. 14, no. 1, p. 015060, Jan. 2024, doi: 10.1063/5.0171922.

A. K. Choudhary, T. Grubesa, A. Jansche, T. Bernthaler, D. Goll, and G. Schneider, 'Deep learning and correlative microscopy for quantification of grain orientation in sintered FeNdB-type permanent magnets by domain pattern analysis', Acta Materialia, vol. 264, p. 119563, Nov. 2023, doi: 10.1016/j.actamat.2023.119563.

A. K. Choudhary, A. Kini, D. Hohs, A. Jansche, T. Bernthaler, O. Csiszár, D. Goll, and G. Schneider, 'Machine learning-based Curie temperature prediction for magnetic 14:2:1 phases', AIP Advances, vol. 13, no. 3, p. 035112, Mar. 2023, doi: 10.1063/5.0116650.

A. K. Choudhary, A. Jansche, T. Grubesa, F. Trier, D. Goll, T. Bernthaler, and G. Schneider, 'Grain size analysis in permanent magnets from Kerr microscopy images using machine learning techniques', Materials Characterization, vol. 186, p. 111790, Apr. 2022, doi: 10.1016/j.matchar.2022.111790.

A. K. Choudhary, A. Jansche, T. Bernthaler, and G. Schneider, 'Machine Learning for Microstructure Quantification of Different Material Classes', Practical Metallography, vol. 57, no. 7, pp. 475–501, Jul. 2020, doi: 10.3139/147.110632.

Conference proceedings (non peer-reviewed publication)

A. K. Choudhary, T. Bernthaler, T. Grubesa, A. Jansche, and G. Schneider, 'Machine Learning for Microstructures Classification in Functional Materials', TechConnect Briefs, no. 2021, pp. 114–117, Oct. 2021.