Deep Learning-Driven Image Analysis for Studying Biomolecular Interactions

Zur Erlangung des akademischen Grades einer
DOKTORIN DER INGENIEURWISSENSCHAFTEN (DR.-ING.)

von der KIT-Fakultät für Chemieingenieurwesen und Verfahrenstechnik des
Karlsruher Instituts für Technologie (KIT)
genehmigte

DISSERTATION

von
M. Sc. Safoura Vaez
aus Esfahan, Iran

Tag der mündlichen Prüfung: 04.08.2025

Erstgutachter: Prof. Dr.-Ing Matthias Franzreb

Zweitgutachter: Prof. Dr. Jörg Lahann

Dedication

This dissertation is dedicated to my mum, dad, husband, and sister. Words cannot fully express my gratitude for your unwavering support and belief in me. I am deeply indebted to you all for standing by me and supporting me through every step of this journey.

This work is for those who strive for knowledge and innovation, hoping it contributes, even in a small way, to the progress of science and understanding.

Acknowledgements

I am deeply grateful to the many individuals whose support, guidance, and encouragement have been invaluable in bringing this PhD dissertation to completion.

First and foremost, I would like to express my deepest gratitude to my advisor, Prof. Dr. Jörg Lahann, for granting me the invaluable opportunity to join his lab and become part of his academic community. I am immensely thankful for the chance to work under his guidance, as he continually pushed me beyond my limits, inspiring me to exceed my own expectations. His encouragement to think creatively and approach projects with innovative perspectives has greatly shaped my growth as a scientist. His humble demeanor, sharp intellect, and vast scientific expertise have profoundly influenced my development.

I have been fortunate to have two exceptional advisors during my Ph.D. studies. I would also like to extend my sincere thanks to Prof. Dr. Matthias Franzreb for welcoming me into his academic group and providing unwavering support throughout my journey. His humility and accessibility made him an invaluable mentor, especially during the final stages of my Ph.D. I will always cherish the memories of the Palmspring seminars, which were filled not only with insightful scientific discussions but also with enjoyable moments.

A special thanks to my former supervisors, Dr. Azam Jiehanipour and Dr. Keikhosro Karimi, for shaping my scientific journey. Their guidance taught me to think critically, plan experiments, tackle challenges, and seek solutions with perseverance, lessons that have profoundly influenced my growth as a researcher.

I am truly grateful to my family, my parents and my husband, for their unconditional love and support, which have been my greatest source of strength. Your belief in me has carried me through every challenge. To my siblings, your encouragement during difficult times has meant the world to me, and I feel so lucky to have you in my life.

I would also like to express my sincere appreciation to my friends and colleagues, Bahar, Tahereh, Gözde, Martina, Meike and Haseeb, for making this journey unforgettable with their valuable feedback, insightful discussions, and unwavering support.

I am grateful to Bianca Posselt, Dr. Erik Strandberg, Dr. Stefan Heißler, and Dr. Alexander Welle for their invaluable training and assistance in using the CD spectroscopy, IRRAS, and TOF SIMS instruments. My thanks extend to my colleagues and lab members, both from Germany and the US, for making this journey unforgettable, and for their invaluable feedback and thoughtful discussions.

Lastly, I want to express my special thanks to Dr. Angela Weiss, Astrid Biedermann, and Stefanie Sellheim-Ret for their support with all administrative matters at the Institute of Functional Interfaces (IFG).

Abstract

In recent years, machine learning approaches have garnered significant interest across a wide range of scientific disciplines. By utilizing large datasets, machine learning (ML) algorithms are capable of identifying intricate patterns and correlations that conventional methods may overlook or take significantly longer to analyze. This integration of machine learning not only accelerates data processing but also enhances predictive accuracy and provides more refined control over experimental measurements and analysis. While experimental approaches remain essential for generating the data required to train machine learning models, ML offers a complementary tool that can accelerate analysis and guide future experiments. Once adequately trained, these models can provide rapid, scalable, and cost-effective insights, enhancing research efficiency in fields such as molecular biology, chemistry, and materials science.

This dissertation is motivated by the potential of ML techniques to increase our understanding of i) protein-DNA binding affinities and ii) surface chemistry using deep learning approaches, with fast and scalable methods that can uncover complex patterns and provide insights beyond traditional analytical techniques.

Developing simple, fast, scalable and precise predictive analytical techniques for the stratification of protein-DNA interaction is essential for enhancing the fundamental understanding of biological processes, disease mechanisms, and for the development of innovative biotechnological and medical applications. However, this remains an unresolved challenge to date. We discovered that valuable information about protein-DNA interactions can be derived from the stains left behind by drying droplets of mixtures of these biological macromolecules. To decipher the intricate stain patterns, a deep-learning neural network (InceptionV3) was applied to polarized light microscopy images obtained from drying droplet deposits of different histone-DNA mixtures. These stain patterns not only demonstrated reproducibility but also enable a comprehensive categorization of various DNA (dependent on

both DNA type and size) and their binding affinity with histone. Eukaryotic DNA binds with a higher affinity to histone than prokaryotic DNA, a trend that resulted in higher prediction accuracy. This is further corroborated by the fact that the average prediction accuracy is higher for longer DNA strands compared to shorter DNA strands. These findings suggest that a simple method like drying a droplet of a protein-DNA mixture solution onto a solid surface could serve as a reliable indicator for predicting protein-DNA bindings affinities. Moreover, following the neural network training on polarized light images of various DNA and their mixtures with histone, the pre-trained model accurately predicted both unknown DNA samples and the binding affinities of unknown histone-DNA samples which had not been included in the training image set. The convolutional neural network (CNN) successfully categorizes unknown histone-DNA samples into strong and medium binders, achieving prediction accuracy rates of 84.4% and 96.25%, respectively. This scalable approach offers the potential for rapid screening of new protein candidates capable of interacting with DNA, facilitating predictions regarding their binding affinity. By employing this advanced methodology, researchers can efficiently identify proteins that demonstrate desirable interactions with DNA, thus accelerating the discovery process in molecular biology and related fields.

Similar to the classification of histone-DNA interactions, the same approach was applied to characterizing material surfaces by classifying surface chemistry using deep learning algorithms. Categorizing surface chemistry through a simple, fast, precise, and low-tech method presents a significant challenge, yet holds considerable importance across various scientific disciplines, including medical implants, biosensors, and regenerative medicine. This study demonstrated that it is possible to differentiate surface chemistry by analyzing the stain patterns generated when protein solutions are deposited onto substrates with varying structural polymer coatings. These coatings share the same polymer backbone but have different functional groups, highlighting the effectiveness of this approach in discerning subtle

differences in surface chemistry. A deep learning neural network (InceptionV3) was employed to classify polarized light microscopy images of dried droplet deposits on different surfaces. These stain patterns exhibited high reproducibility across different surface chemistries, facilitating comprehensive surface classification with 96% accuracy. To demonstrate the generalizability of our approach, a pre-trained CNN was tested on images from copolymerized polymer surfaces not included in the training set, achieving a classification accuracy of 96%. These findings are significant because they demonstrate that suitably pre-trained CNNs can predict polymer surface chemistry beyond their original training set. This scalable approach can be used for rapid screening of new or unknown polymer surface chemistries.

Kurzzusammenfassung

In den letzten Jahren haben Ansätze des maschinellen Lernens in einer Vielzahl wissenschaftlicher Disziplinen großes Interesse geweckt. Durch die Nutzung großer Datensätze sind Algorithmen des maschinellen Lernens in der Lage, komplizierte Muster und Korrelationen zu erkennen, die bei herkömmlichen Methoden übersehen werden oder deren Analyse erheblich länger dauert. Diese Integration des maschinellen Lernens beschleunigt nicht nur die Datenverarbeitung, sondern verbessert auch die Vorhersagegenauigkeit und ermöglicht eine verfeinerte Kontrolle über experimentelle Messungen und Analysen. Obwohl experimentelle Ansätze nach wie vor unerlässlich sind, um die Daten für das Training von Machine-Learning-Modellen zu generieren, stellt maschinelles Lernen ein ergänzendes Werkzeug dar, das Analysen beschleunigen und zukünftige Experimente gezielt unterstützen kann. Nach ausreichendem Training können diese Modelle schnelle, skalierbare und kosteneffiziente Erkenntnisse liefern und so die Forschungseffizienz in Bereichen wie Molekularbiologie, Chemie und Materialwissenschaften steigern.

Diese Dissertation ist motiviert durch das Potenzial von ML-Techniken, unser Verständnis von i) Protein-DNA-Bindungsaffinitäten und ii) Oberflächenchemie mithilfe von Deep-Learning-Ansätzen zu erweitern, mit schnellen und skalierbaren Methoden, die komplexe Muster aufdecken und Einblicke über traditionelle analytische Techniken hinaus bieten können.

Die Entwicklung einfacher, schneller, skalierbarer, und präziser prädiktiver Analyseverfahren für die Stratifizierung von Protein-DNA-Interaktionen ist für die Verbesserung unseres grundlegenden Verständnisses biologischer Prozesse und Krankheitsmechanismen sowie für die Entwicklung innovativer biotechnologischer und medizinischer Anwendungen unerlässlich. Dies ist jedoch bis heute eine ungelöste Herausforderung. Wir haben entdeckt, dass wertvolle Informationen über Protein-DNA-

Wechselwirkungen aus den Flecken abgeleitet werden können, die beim Trocknen von Tröpfchenmischungen dieser biologischen Makromoleküle zurückbleiben. Um die komplizierten Fleckenmuster zu entschlüsseln, wurde ein neuronales Netzwerk mit tiefem Lernen auf polarisierte Lichtmikroskopie-Bilder angewandt, die beim Trocknen von Tropfenablagerungen verschiedener Histon-DNA-Mischungen entstanden. Diese Trocknungsmuster zeigten nicht nur Reproduzierbarkeit, sondern ermöglichen auch eine umfassende Kategorisierung verschiedener DNA (abhängig von DNA-Typ und -Größe) und ihrer Bindungsaffinität mit Histon. Eukaryotische DNA bindet mit einer höheren Affinität an Histon als prokaryotische DNA, ein Trend, der zu einer höheren Vorhersagegenauigkeit führte. Dies wird auch durch die Tatsache bestätigt, dass die durchschnittliche Vorhersagegenauigkeit für längere DNA-Stränge höher ist als für kürzere DNA-Stränge. Diese Ergebnisse deuten darauf hin, dass eine einfache Methode wie das Trocknen eines Tropfens einer Protein-DNA-Mischungslösung auf einer festen Oberfläche als zuverlässiger Indikator für die Vorhersage von Protein-DNA-Bindungsaffinitäten dienen könnte. Nach dem Training des neuronalen Netzes auf polarisierten Lichtbildern verschiedener DNA und deren Mischungen mit Histon konnte das vortrainierte Modell sowohl DNA-Proben als auch die Bindungsaffinitäten von Histon-DNA-Proben, die nicht im Trainingsbildsatz enthalten waren, genau vorhersagen. Das neuronale Faltungsnetzwerk (CNN) kategorisiert unbekannte Histon-DNA-Proben erfolgreich in starke und mittlere Bindungen und erreicht dabei Genauigkeitsraten von 84.4 % bzw. 96.25 %. Dieser skalierbare Ansatz bietet das Potenzial für ein schnelles Screening neuer Proteinkandidaten, die in der Lage sind, mit der DNA zu interagieren, und erleichtert Vorhersagen über ihre Bindungsaffinität. Durch den Einsatz dieser fortschrittlichen Methodik können Forscher effizient Proteine identifizieren, die wünschenswerte Wechselwirkungen mit der DNA aufweisen, und so den Entdeckungsprozess in der Molekularbiologie und verwandten Bereichen beschleunigen.

Ähnlich wie bei der Klassifikation von Histon-DNA-Interaktionen wurde derselbe Ansatz auf die Charakterisierung von Materialoberflächen angewendet, indem die Oberflächenchemie mit Deep-Learning-Algorithmen klassifiziert wurde. Die Klassifikation von Oberflächenchemie durch eine einfache, schnelle, präzise und kostengünstige Methode stellt eine erhebliche Herausforderung dar, hat jedoch eine große Bedeutung in verschiedenen wissenschaftlichen Disziplinen, einschließlich medizinischer Implantate, Biosensoren und regenerativer Medizin. In dieser Studie wurde gezeigt, dass es möglich ist, die Oberflächenchemie zu differenzieren, indem die Fleckenmuster analysiert werden, die entstehen, wenn Proteinlösungen auf Substrate mit unterschiedlichen strukturellen Polymerbeschichtungen aufgebracht werden. Diese Beschichtungen haben dasselbe Polymergerüst, weisen aber unterschiedliche funktionelle Gruppen auf, was die Wirksamkeit dieses Ansatzes bei der Unterscheidung feiner Unterschiede in der Oberflächenchemie unterstreicht. Ein neuronales Deep-Learning-Netzwerk (InceptionV3) wurde eingesetzt, um polarisierte Lichtmikroskopie-Bilder von getrockneten Tröpfchenablagerungen auf verschiedenen Oberflächen zu klassifizieren. Diese Trocknungsmuster zeigten eine hohe Reproduzierbarkeit über verschiedene Oberflächenchemien hinweg und ermöglichten eine umfassende Oberflächenklassifizierung mit 96 % Genauigkeit. Um die Verallgemeinerbarkeit unseres Ansatzes zu demonstrieren, wurde ein vortrainiertes CNN an Bildern von copolymerisierten Polymeroberflächen getestet, die nicht in der Trainingsgruppe enthalten waren, und erreichte eine Klassifizierungsgenauigkeit von 96 %. Diese Ergebnisse sind von großer Bedeutung, da sie zeigen, dass entsprechend vortrainierte CNNs die Chemie von Polymeroberflächen über ihren ursprünglichen Trainingssatz hinaus vorhersagen können. Dieser skalierbare Ansatz kann für ein schnelles Screening neuer oder unbekannter chemischer Eigenschaften von Polymeroberflächen verwendet werden.

List of Abbreviations

Alde (CHO)	Aldehyde
Alk	Alkyne
A	Amine
AFM	Atomic force microscopy
AM	Aminomethyl
ATR-FTIR	Attenuated total reflection- Fourier transform infrared spectroscopy
Ave pooling	Average pooling
BGD	Batch Gradient Descent
Bi	Bismuth
BSA	Bovine serum albumin
Br	Bromide
CVD	Chemical vapor deposition
Cl	Chloride
ChIP	Chromatin immunoprecipitation
CD	Circular dichroism
CCA	Constant contact angle
CCR	Constant contact radius
CNNs	Convolutional neural networks
DL	Deep learning
DNA	Deoxyribonucleic acid
2D	2 dimensional
F	Fluoride

FC layers	Fully-connected layers
Grad-CAM	Gradient-weighted class activation mapping
GPU	Graphics processing unit
Her	Herring DNA
HSA	Human serum albumin
H	Hydrogen
OH	Hydroxymethyl
IR	Infrared
IRRAS	Infrared reflection-absorption spectroscopy
IgG	Immunoglobulin G
LSD	Least significant difference
LSTM	Long short-term memory
MALDI	Matrix - assisted laser desorption ionization
Max pooling	Maximum pooling
MD	Molecular dynamics
ML	Machine-learning
MM	Mixed mode
Multi Image Alignment	MIA
NMR	Nuclear magnetic resonance
PDB	Protein Data Bank
PLM	Polarize light microscopy
PMT	Photomultiplier tube
PPX	Poly(p-xylylene)
ReLU	Rectified linear unit
RGB	Red Green Blue

RH	Relative humidity
RBMs	Restricted Boltzmann machines
RNA	Ribonucleic acid
Sal $_{20\ kbp}$	Salmon DNA
Sal $_{1\ kbp}$	Sheared Salmon DNA
SELEX	Systematic evolution of ligands by exponential enrichment
SEM	Scanning electron microscopy
SGD	Stochastic Gradient Descent
SPR	Surface plasmon resonance spectroscopy
t-SNE	t-distributed stochastic neighbor embedding
TOF-SIMS	Time-of-flight secondary ion mass spectrometry
TPCL	Triple phase contact line
XPS	X rays photoelectron spectroscopy

List of Symbols

A	Absorption/ Absorbance value
Å	Angstrom
$(NH_4)_2$ SO_4	Ammonium sulfate
cm	Centimeter
c	Concentration
Da	Dalton
°	Degree
°C	Degree Celsius
[θ]	Degree of ellipticity
EtBr	Ethidium bromide
ε_L	Extinction coefficients (left)
ε_R	Extinction coefficients (right)
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
tanh	Hyperbolic Tangent
kbp	Kilo base pairs
kcal mol^{-1}	kilocalories per mole
keV	Kiloelectron volts
ƛ	Lambda DNA
w	Learnable parameter
α	Learning rate
L	Loss function
µL	Microliter
µm	Micrometer

mdeg	Millidegrees
mg	Milligram
mg/mL	Milligrams per milliliter
mm	Millimetres
mM	Millimolar
nm	Nanometer
ns	Nanosecond
l	Path length
KCl	Potassium chloride
R	Ratio
σ	Sigmoid
$NaHCO_3$	Sodium bicarbonate
Na_2CO_3	Sodium carbonate

Table of Contents

1. Introduction

Over the last few decades, there has been significant attention directed toward studying the dried patterns of biologically relevant sessile droplets. [1] Experimentally, the drying process of particle droplets (e.g., DNA, protein, blood, etc.) adheres to a simple and fast procedure. In this process, particles are dispersed in a solvent, such as water or buffer, and a defined volume of the solution is deposited onto a substrate as a droplet. During the drying process, the solvent gradually evaporates, leading to an increase in the local concentration of particles, and results in the development of unique, characteristic, and reproducible patterns. [1, 2] These characteristic patterns arise due to the interplay of heat transfer, mass transport, and fluid dynamics within the liquid, with key factors such as contact-line dynamics, surface hydrophobicity, surface-tension-driven forces, Marangoni flow, and thermal instabilities contributing to the phenomenon, often resulting in the well-known coffee-ring effect. [3-6] Previous studies have demonstrated that these unique patterns serve as "fingerprints" for biomolecular screening and classification. [2, 7]

The interaction between protein and DNA plays a critical role in almost all of the biological processes, including mechanisms associated with health and disease. [8, 9] In recent years, numerous experiments have been conducted to investigate protein-DNA interactions, employing a wide range of techniques, both in vivo and in vitro, such as conventional chromatin immunoprecipitation (ChIP), electrophoretic mobility shift assay (EMSA), Systematic Evolution of Ligands by Exponential Enrichment (SELEX)-based methods, nuclear magnetic resonance (NMR), X-ray crystallography, fluorescence-based techniques, circular dichroism (CD) spectroscopy, atomic force microscopy (AFM), and surface plasmon resonance (SPR) spectroscopy. [10-13] These methods have provided insights into the nature of protein-DNA interactions, yet each comes with its own set of limitations and challenges that

must be carefully considered when choosing the most appropriate technique for a given study. ChIP requires specific-grade antibodies, which may not always be available or of sufficient quality. EMSA faces challenges due to the rapid dissociation of protein-DNA complexes during electrophoresis, which hinders detection, while slow dissociation may lead to underestimation of binding density. SELEX-based methods are labour-intensive and time-consuming, often requiring weeks or months to complete. NMR necessitates multimodal analytics, making it complex, resource-intensive, and technically demanding. X-ray crystallography requires specialized equipment and facilities, limiting accessibility. Fluorescence-based assays, such as those using ethidium bromide, pose safety risks due to mutagenic properties, while fluorescence dye displacement assays can suffer from low sensitivity. CD spectroscopy has limited sensitivity to small conformational changes and can experience signal overlap in large protein-DNA complexes. AFM and SPR depend on the proper immobilization of biomolecules; incorrect orientation affects accuracy, and SPR also faces mass-transfer limitations that distort binding kinetics. [10, 14-18] Given these challenges, there is a growing need for alternative approaches that are accurate, straightforward, rapid, cost-effective, and non-toxic. Recent advances in computational modelling, including molecular dynamics (MD) simulations, have been explored for predicting protein-DNA binding affinities. However, MD simulations are often constrained by high computational costs and limited applicability to large biomolecular complexes. Consequently, in this dissertation, machine learning techniques, with a particular emphasis on deep learning, are proposed as powerful tools for analyzing protein-DNA complex data and providing efficient alternatives to traditional experimental approaches. [19-24]

In addition to protein-DNA interactions, the interfacial characteristics of biomaterials play a crucial role in determining key performance attributes, including cell adhesion, [25, 26] biocompatibility, [27] and wettability. [28] These surface properties influence biomedical

applications, including tissue engineering, [29, 30] medical implants, [31] and drug delivery systems. [32] Accurate surface analysis is essential for selecting the appropriate polymer coating for specific biomedical applications, ensuring optimal coating performance. Surface characterization techniques are commonly used to identify potential surface contaminants or variations in surface chemistry that could impact key properties such as adhesion, [33] wettability, [34] biological integration, surface fouling, [35] or optical performance. [36, 37] The most commonly used techniques [38] for surface composition characterization are IR spectroscopy, [39] XPS, [40, 41] SIMS, [38, 42, 43] ellipsometry, [44] atomic force microscopy (AFM), [45] and contact angle measurement. [46] However, each of these methods has certain inherent limitations. For instance, IR spectroscopy may demonstrate limited sensitivity in analyzing samples with low thickness or weak absorption bands, and overlapping absorption bands further complicate accurate interpretation. [47, 48] XPS requires ultra-high vacuum conditions and involves complex data interpretation, limiting accessibility. TOF-SIMS is expensive, requires an ultra-high vacuum, and demands sophisticated data processing algorithms alongside expertise in surface chemistry and mass spectrometry. [36] Ellipsometry suffers from low optical contrast in transparent, low-polarizability films, particularly at solid/liquid interfaces, making it difficult to accurately characterize film properties. [49] Additionally, contact angle measurement is limited by time-dependent variations caused by evaporation or contamination, which hinder the achievement of stable and reproducible results. To address these challenges, an efficient and reproducible method is needed for analyzing surface properties in a straightforward and scalable manner. Similar to protein-DNA interaction study, surface characterization and recognition can also benefit from deep learning-based image analysis techniques, providing a novel approach for identifying surface patterns and correlating them with functional properties.

Deep learning, a subset of machine learning, has revolutionized data-driven analysis by automatically identifying complex patterns from large datasets without requiring manual feature extraction. [50-56] By leveraging deep learning algorithms, researchers can efficiently process vast amounts of image data, uncover intricate correlations, and enhance predictive accuracy. Additionally, deep learning enables more refined control over measurements and analyses, making it a promising tool for both biomolecular interaction studies and surface characterization. [57, 58] In a previous study conducted in Professor Lahann's lab, a deep learning-based method was developed to predict single amino acid mismatches in peptides by analyzing stain patterns left by drying droplets. Using polarized light microscopy images of dried amyloid-beta peptide deposits, deep learning models successfully identified structural variations with high accuracy. [2] Inspired by this approach, the application of deep learning has been extended to two major research areas: (i) studying protein-DNA interactions and (ii) investigating surface chemistries.

In this study, a conventional image-based neural network was employed to analyze extensive datasets generated from protein solutions prepared with varying salt conditions. A systematic workflow was established, in which precise amounts of human serum albumin (HSA) and immunoglobulin G (IgG) were dispensed, with controlled variations in salt types and concentrations. The resulting stain patterns were captured using an automated polarized light microscope (PLM). This approach enabled the efficient creation of a large dataset capturing subtle variations in solution composition. This simple, fast, and scalable method provides a powerful tool for studying protein behavior across diverse chemical environments.

Moreover, a pre-trained neural network was employed for image analysis to examine extensive sets of images derived from dried protein-DNA solution samples. An automated process was developed to efficiently deposit precise amounts of protein-DNA complexes and capture images of the resulting stain patterns using an automated polarized light microscope

(PLM). This streamlined approach generated thousands of images in just a few hours. This method, which is simple, rapid, and cost-efficient, has been applied successfully in studying protein-DNA interactions across both eukaryotic and prokaryotic DNA specimens. This innovative technique offers a powerful tool for researchers to explore different levels of protein-DNA interactions on a large scale, potentially accelerating discoveries in the field of molecular biology.

Similarly, a conventional image-based neural network was utilized to analyze large sets of images obtained from drying Bovine Serum Albumin (BSA) solutions on various functionalized surfaces. The developed automated workflow, which includes the deposition of defined volumes of BSA solution in a massively parallel manner, was applied, followed by the capture of images of the resulting stain patterns using an automated polarized light microscope. This method has been successfully applied to a wide range of polymer surfaces with different functional groups and properties, enabling the study of surface classification and recognition based on BSA-stained patterns. This method not only saves time and reduces cost but also ensures the continuous reproducibility of surface properties critical for the functionality of polymeric materials in various application.

2. Background

2.1. Pattern Formation on Solid State Substrates

Deposits formed by the evaporation of droplets containing non-volatile substances is a well-known natural phenomenon, attracting significant attention due to its fundamental aspects and practical applications. [59-62] This phenomenon has implications in various fields involving evaporation on the surface such as inject printing, [63] nanomaterials assembly, [64] fabrication process, [65] and colloidal crystals. [66] It also affects the performance of applications including, electronic devices, [67] matrix-assisted laser desorption ionization (MALDI) spectrometry, [68] surface-enhanced Raman spectroscopy, [69] fluorescence microarrays, DNA or RNA microarray, [70] and disease diagnosis. [71-73] One specific occurrence during this process which is called the “coffee ring effect”, is contact line pinning and forming the ring-like residues at the droplet's edges. [4, 60, 74, 75] The particles within the initial droplet are spread out, defining an initial equilibrium state. It was demonstrated that the evaporation rate is most pronounced near the periphery, primarily due to the curvature of the droplet. As the droplet dries, the system undergoes a phase transition away from its initial equilibrium state. This process leads to the development of concentration gradients, triggering the development of various flows within the droplet. [76] The flow caused by evaporation pulls particles or solute toward the contact line, to compensate the excessive loss of mass. [59] As the solvent evaporates and the droplet fractures to relieve stress, a macroscopic fingerprint pattern begins to emerge. Different stain patterns form when water droplets containing biomolecules like proteins and DNA evaporate. [1] The final dried pattern is influenced by several factors, including substrate properties, nature of solute components (size, chemical composition, and concentration), as well as environmental factors such as temperature and relative humidity. [1, 77, 78]

Lee et al. studied the effect of substrate wettability on the formation of ring-shaped patterns by nanofluids. They found that substrates with lower contact angles resulted in wider ring widths, as the smaller contact angle increased the droplet's surface area, leading to faster evaporation and higher flow velocity. [79] Their study showed that lower contact angle promotes the formation of more ring-shaped patterns, while higher contact angle substrates, such as stainless steel, exhibit fewer ring-shaped patterns than glass. Specifically, a larger contact angle results in greater height and smaller contact diameter for a fixed droplet volume. Consequently, as the height at the center is maximized during evaporation, more particles will likely remain at the center. [79] Similarly, Uno et al. examined the formation of patterns during the evaporation of polymer solutions on surfaces with varying hydrophobicity. Their research revealed that on hydrophilic surfaces, droplets maintained their initial contact area while their volume gradually decreased, eventually forming circular residue patterns. Microscopic analysis indicated that many particles accumulated at the contact boundary due to the presence of a thin water layer, which resulted from the surface's hydrophilic nature. In contrast, on hydrophobic surfaces, the contact area decreased as evaporation progressed, with no significant particle adsorption occurring in the early stages. Instead, particle aggregation took place once their concentration exceeded a critical threshold, leading to the formation of small deposits after complete evaporation. [80]

As mentioned, the second parameter that affects the final dried pattern is the nature of the solute components. The movement of particles within a droplet can be influenced by their size. Smaller particles typically migrate towards the three-phase contact line of the droplet, forming ring-shaped patterns, while larger particles tend to gather nearer the center in the residue. Therefore, particle size variation leads to distinct residue patterns.

Previous study showed that at low concentrations and small particle sizes (below 2% by vol and smaller than 13 nm), a ring-shaped pattern emerged. Conversely, at high

concentrations and larger particle sizes (above 3% by vol and larger than 20 nm), a uniform pattern was observed. [79] Consequently, various residue patterns depending on particle size, impact the functionality of multiple applications. [81, 82] Chen et al. investigated the impact of salt ions on protein pattern formation and observed that salt plays a crucial role in influencing protein aggregation. In the absence of salt, no protein patterns were observed. However, with the addition of salt, protein molecules undergo aggregation and self-assembly. At low salt concentrations (e.g., 0.1 × PBS buffer), dendritic-shaped aggregates and some scalloped microstructures form. As the salt concentration increases, rosette-shaped patterns emerge due to salt crystals promoting nucleation events. [83]

Finally, the third parameter influencing the final dried pattern is the set of environmental factors. Li et al. observed that during slow drying process (T = 25 °C), the majority of particles accumulate at the droplet periphery, creating a ring-like structure. Conversely, rapid drying at high temperatures (T = 75 °C) leads to uniform particle deposition across the droplet surface, with minimal accumulation at the periphery. They suggest that uniform deposition can be achieved through straightforward control of evaporation kinetics, without the need to alter the droplet composition or modify particles. [84, 85] Moreover, a previous study showed that relative humidity (RH) significantly influences the evaporation dynamics and pattern formation of drying droplets, including blood droplets. RH affects the contact angle of the droplet, thereby impacting the initial evaporation rate. Higher RH levels reduce the evaporation rate, allowing more time for internal fluid movement and particle redistribution. As a result, the width of mobile plaques in the corona and the fine peripheral region increases with increasing RH (**Figure 2.1**). These findings highlight the crucial role of environmental conditions in shaping the final deposition patterns of dried droplets. [86]

Figure 2.1. Effect of relative humidity on the residue left after the evaporation of a sessile blood drop. All experiments were conducted using the same droplet volume (V = 14.2 µl) across a range of relative humidity (RH) levels, with the droplets placed on a microscope-grade ultraclean glass substrate at room temperature (23.8 °C) and atmospheric pressure. Adapted from [86]

2.1.1. General Physics of Drying Droplets

As illustrated in **Figure 2.2**, two modes of profile evolution occur during the evaporation of a droplet: the constant contact radius (CCR) and the constant contact angle (CCA). In the CCR mode, the diameter of the droplet (or the contact radius) stays constant while the height of the droplet decreases during evaporation. In the CCA mode, the contact radius of the droplet gradually decreases over time, but the contact angle remains constant. The combination of both the contact angle and contact radius modes during the drying process is referred to as the mixed mode (MM). [71] As mentioned above, the evaporation mode relies on the solid substrate where the droplet is located. CCR is typical on hydrophilic substrates, while CCA is prevalent on hydrophobic surfaces. [87]

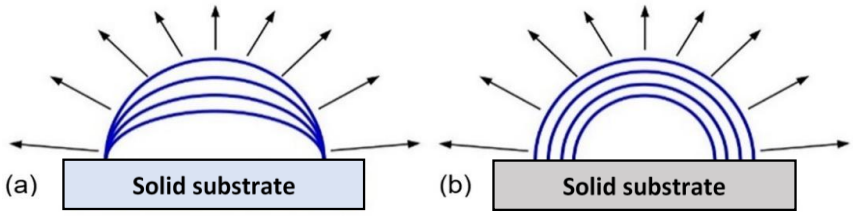


Figure 2.2. Schematic representation of two modes of drying process of droplets. (a) constant contact radius (CCR) and (b) constant contact angle (CCA). Adapted from [71]

The two primary flows within an evaporating droplet are Marangoni flow and capillary flow (**Figure 2.3**). [88] In capillary flow, the drying occurs in the CCR mode, where the droplet remains pinned to the substrate surface. Droplets with a contact angle below 90° exhibit the highest evaporative flux at the outer edge, known as the Triple Phase Contact Line (TPCL). [71] As a result, fluid is needed to replace the evaporated liquid at the TPCL. This causes fluid to flow radially outward from the center to the edge. [71]

Marangoni flow is induced by a gradient in surface tension at the droplet's external interface. [89] This tension gradient can originate from two factors. Firstly, a temperature gradient across the droplet surface (thermal Marangoni effect) is induced by variations in the evaporative flux across the droplet surface. Since evaporation is an endothermic process, the bulk liquid maintains a higher temperature compared to the liquid at the droplet's surface. [61] The second factor is a shift in local composition (solute Marangoni effect), where the concentration of dissolved solute varies across the droplet. Different solutes can have diverse impacts on surface tension. [90] At room temperature, capillary flow often dominates drying within a droplet, while Marangoni flow driven by temperature becomes more prominent on heated surfaces. [1]

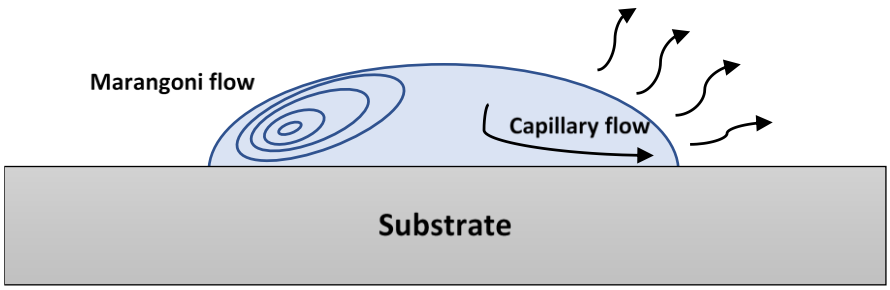


Figure 2.3. The flow patterns inside an evaporating droplet, illustrating both Capillary and Marangoni flows. The lines indicate the direction of the flow. Adapted from [91]

2.1.2. Biomolecules Within Drying Droplets

Biological fluids, or biofluids, are a type of complex fluid that typically contain salts. [92, 93] The drying behavior of droplets containing biologically relevant substances, such as DNA, proteins, plasma, blood, bacteria, and algal dispersions, has attracted considerable attention in recent decades. [94-97] These droplets form distinct patterns upon drying, which have promising applications in medical fields such as biosensors, diagnostics, drug delivery, and combating antimicrobial resistance. [71, 98, 99] Consequently, substantial progress has been made in understanding these patterns and developing advanced image-based analysis techniques for potential biomedical applications. [1] The drying process of drops of these fluids involves several physical and physiochemical steps depending on several parameters including nature of salute components, the substrate chemistry, and environmental conditions. [89, 100, 101]

Deoxyribonucleic acid (DNA)

Recent research focused on leveraging the drying droplet method to comprehend self-assembly, microarray techniques, and the patterning of DNA in diverse microenvironments. [102, 103] DNA is a double helix polymer consisting of polynucleotide chains, where each nucleotide comprises a 2-deoxyribose sugar, a phosphate group, and a nitrogenous base. [104] The nitrogenous bases, as depicted in **Figure 2.4**, consist of purines (adenine 'A' and guanine 'G') and pyrimidines (thymine 'T' and cytosine 'C'). [105] The deoxyribose sugar and phosphate group form a backbone to which the nitrogenous bases attach directly to the sugar unit. [106] Notably, DNA carries a negative charge, which is attributed to the phosphate groups in the DNA backbone, making both the phosphate backbone and overall DNA structure negatively charged. [107] Each nitrogenous base possesses a unique structure and can create particular hydrogen bonds (**Figure 2.4**, red boxes) because of its electron-accepting and donating

characteristics. Adenine (A) pairs with thymine (T), and guanine (G) pairs with cytosine (C), resulting in equal amounts of A and T, and G and C in a given sample of DNA. [108] The arrangement of base pairs creates major and minor grooves, offering unique chemical environments for interaction. [109] The major groove provides multiple interaction sites, resulting in strong binding to drugs or ligands compared to the minor groove. The major groove has a width of 11.6 Å and a depth of 8.5 Å, allowing larger molecules like proteins to easily fit in. In contrast, the minor groove, which is smaller and has fewer binding sites, is often unoccupied due to its 8.2 Å depth, making it accessible for smaller drug molecules to interact. Since many antibiotic and anticancer drugs consist of small molecules, the minor groove acts as their primary binding site. [105, 110]

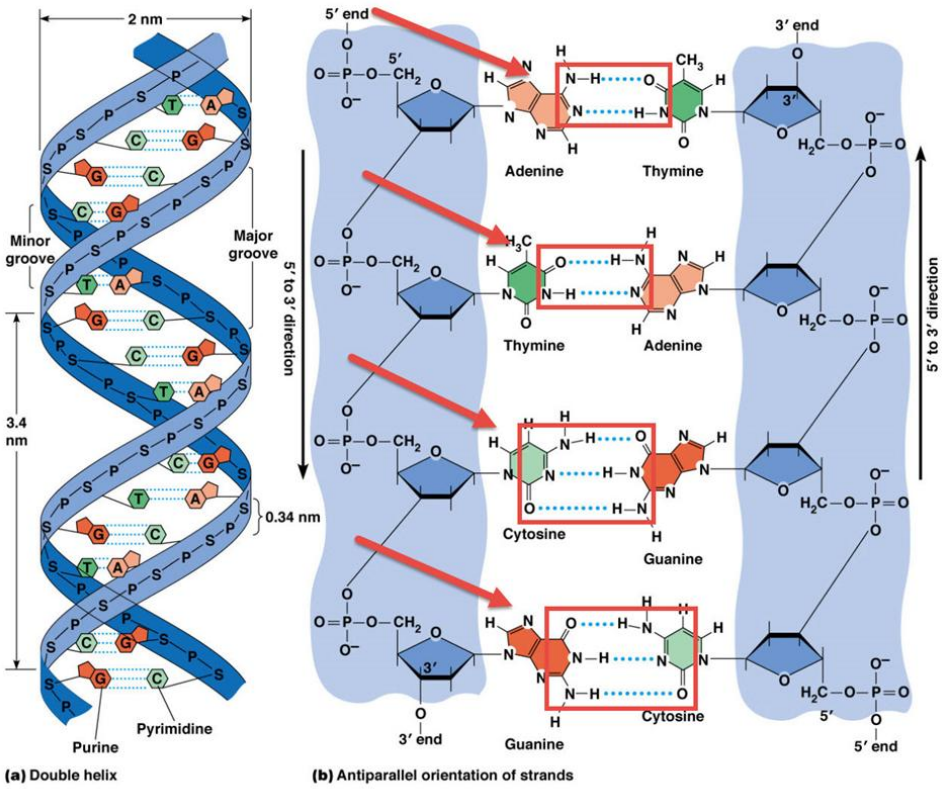


Figure 2.4. DNA molecule structure. (a) Double helix model showing two antiparallel strands twisted around each other. The strands consist of a sugar-phosphate backbone and nitrogenous base pairs: adenine (A) pairs with thymine (T), and guanine (G) pairs with cytosine (C), connected by hydrogen bonds. Major and minor grooves are visible along the helix. (b) Base pairing and strand orientation illustrating the antiparallel arrangement (5′ to 3′ and 3′ to 5′) and specific hydrogen bonding between complementary bases. 'A' pairs with 'T' through two hydrogen bonds, and 'G' pairs with 'C' through three hydrogen bonds. Adapted from [107]

Li et al. introduced a method for nucleic acid detection by applying the hybridization-induced suppression of the coffee ring effect. Typically, as a droplet evaporates on a solid surface, suspended spherical particles migrate outward, forming a ring-shaped pattern. However, non-spherical particles tend to adhere to each other at the air-water interface, resisting this outward flow and suppressing the coffee ring effect. In this method, suspended microspheres were functionalized with oligonucleotide probes complementary to target DNA. Upon hybridization, these probes connected multiple microspheres, forming non-spherical particle agglomerates that resist capillary flow, resulting in more uniform particle deposition (**Figure 2.5**). They exhibited high specificity and could even detect a single nucleotide mismatch. Due to the simplicity of its operation and the visual readout without requiring a special detector, the "coffee ring" approach demonstrates the immense potential for inexpensive and convenient nucleic acid detection in resource-limited settings. [111] Moreover, another study highlights the impact of DNA strand length, varying in the number of base pairs, on drying dynamics, the coffee-stain effect, nanoscale structure, and aggregation. Intriguingly, the study suggested a connection between DNA viscosity and drying behavior. Their findings indicated that lower viscosity, observed in solutions with shorter DNA chains, facilitates mobility within the droplet, promoting continuous deposit growth during drying. The crystallization mechanism likely involves DNA strands serving as nucleation sites at the solid-liquid interface, causing dendrite crystal formation through diffusion-limited growth. Conversely, for longer DNA chains, crystallization was attributed to "faceted growth," primarily a nucleation-limited process. [112]

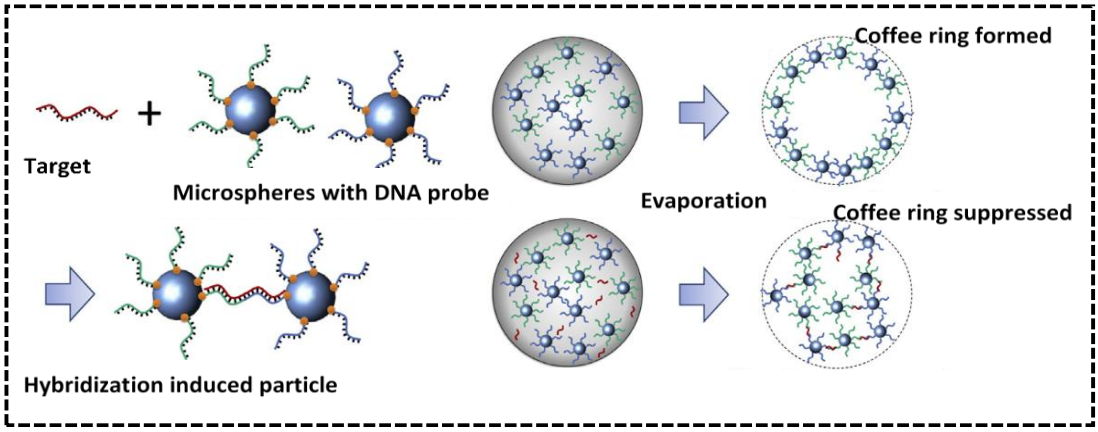


Figure 2.5. Detection of target DNA using the coffee ring effect. Microspheres functionalized with DNA probes hybridize with target DNA to form non-spherical aggregates, suppressing the coffee ring effect and resulting in uniform particle deposition. This method enables simple, highly specific, and low-cost nucleic acid detection without the need for specialized equipment. Adapted from [111]

Histones

Over two decades, there has been significant attention focused on the drying of droplets containing various types of proteins. Previous studies categorized the protein-drying droplets into three types: globular proteins, fibrous proteins, and composite proteins (a mix of globular and fibrous proteins). [1] Histones are essential globular proteins that organize DNA into chromatin, ensuring proper packaging while maintaining accessibility for vital processes such as replication, transcription, repair, and recombination. [113-115] These proteins prevent DNA tangling, protect it from damage, and serve as molecular spools around which DNA wraps to form nucleosomes, thereby regulating gene expression and replication. [116, 117] These nucleosomes, in turn, are intricately organized into tightly packed chromatin fibers. Without histones, the DNA strands within chromosomes would extend to considerable lengths. [118] For example, a human cell contains roughly 1.8 meters of DNA when fully extended. However, when wrapped around histones, this length is condensed to approximately 90 micrometers (0.09 mm) of chromatin fibers with a diameter of 30 nanometers. Histones are classified into five families: H1/H5, known as linker histones, and H2, H3, and H4, referred to as core

histones. The nucleosome core comprises two H2A-H2B dimers and a H3-H4 tetramer. The histone octamer complex consists of two copies each of H2A, H2B, H3, and H4 proteins. This complex forms the protein core around which approximately 146 or 147 base pairs of DNA are wrapped. The DNA wraps around the histone octamer in about 1.67 turns of a left-handed superhelix. This complex of DNA and histone proteins is called a nucleosome and forms the fundamental unit of eukaryotic chromatin. [119-122] Between each nucleosome, there is a segment of DNA called "linker" DNA, which can vary in length from 20 to 80 base pairs. The nucleosomes themselves form a chain about 10 nanometers wide. This chain is then folded into a denser fiber that is about 30 nanometers wide. This folded structure further combines into higher-order formations. [123, 124] A protein called linker histone H1 (or H5) helps organize this linker DNA between nucleosomes (**Figure 2.6**). When linker histones attach to a nucleosome, they shield additional 22 base pairs of DNA around the nucleosome structure, as shown by DNase I cleavage experiments. [125] This whole package, including the nucleosomes, linker histones, and the DNA they interact with, is called a chromatosome. [125, 126]

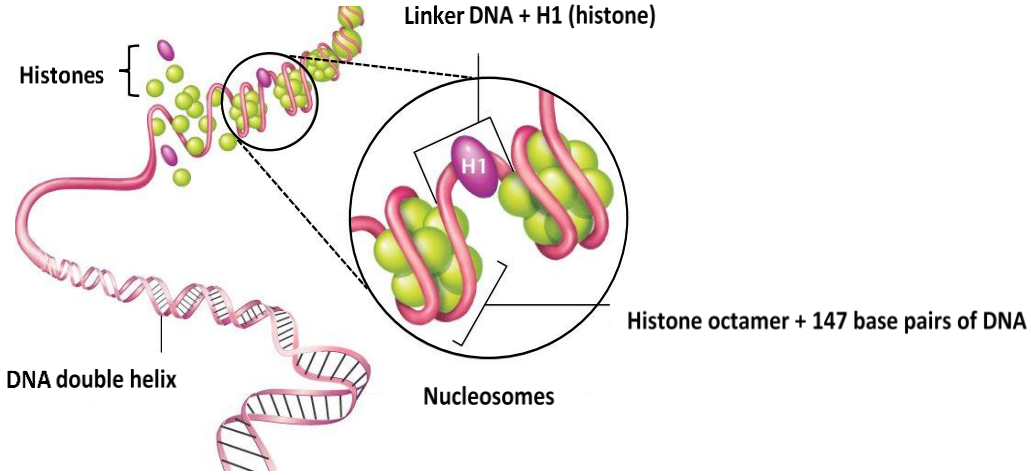


Figure 2.6. Nucleosomes structure. Nucleosomes consist of DNA wrapped around histone octamers, with linker DNA and histone H1 facilitating chromatin compaction and organization. Adapted from [127]

2.2. Protein-DNA Interaction Classification

Since the latter half of the previous century, it has been well-established that the interaction between a protein and a DNA molecule plays a pivotal role in the functioning of living cells and the overall sustainability of life. This interaction governs fundamental molecular and cellular processes, including transcription, transcriptional regulation, recombination, replication, DNA repair, DNA packaging, and DNA modifications. [128] Studies have often approached this subject from two primary perspectives: i) a transcriptomic level, investigating the binding of specific proteins to particular DNA sequences or genes and how this interaction influences gene expression, and ii) a chemical perspective, examining the structural aspects of the formed complexes. Notably, there is a significant overlap between these two approaches. [14, 129] In the 1960s, Leng and Felsenfeld [130] made significant strides in understanding how DNA molecules interact with proteins. They discovered patterns and preferences in the interactions between amino acids and DNA base sequences. Specifically, they found that polylysine polypeptides tend to interact preferentially with A-T-rich DNA, while polyarginine shows a preference for G-C-rich DNA.

A decade later, Seeman et al. [131] contributed more insights into the structure of these interactions. They utilized hydrogen-bonding atoms identified on the edges of DNA bases to propose that specific amino acid side chains have an affinity for particular nucleotides. Furthermore, they suggested that these interactions are more specific within the major groove of DNA rather than the minor groove. Subsequently, studies that involved model building, following McKay and Steitz's [132] pioneering use of X-ray crystallography to detect DNA-protein complexes, indicated that the Escherichia coli catabolite gene activator protein (CAP) binds to the major groove of DNA. Later, researchers like Pabo and Sauer, [133] along with Matthews, [134, 135] continued to delve into the connection between amino acids and DNA bases, incorporating electrostatic and van der Waals interactions into their models.

In summary, based on previous literatures, four forces govern protein-DNA interactions: (1) hydrogen bonding between amino acids and the grooves of the DNA helix, (2) electrostatic attraction between the negatively charged DNA backbone and positively charged amino acids, (3) hydrophobic interactions, including π-π stacking between aromatic amino acids and DNA bases, and (4) Van der Waals forces that stabilize interactions over large surface areas. [136, 137]

2.2.1. Histone-DNA Interactions

Histone-DNA interactions are crucial for chromosome structure and gene regulation. Acidic chromosomal proteins are prominent candidates for regulating specific genes, contrasting with the less specific DNA binding histones exhibited by core histones (basic proteins). Linker Histone (H1) found in most eukaryotic chromosomes, is known for its preferential binding to specific DNA regions and distinct folding patterns within chromosomes. [138-141] The investigations indicated that H1 induces compaction of the nucleosome chain, resulting in a more densely folded chromatin structure. H1, found on the outside of the nucleosome, keeps the higher order of structure by connecting the DNA between neighboring nucleosomes. Changes in how tightly chromatin is packed happen due to alterations in H1-DNA interactions. [142, 143] In previous studies, the preferential interaction of H1 with eukaryotic DNA in contrast to prokaryotic DNA has been demonstrated through filter binding assays utilizing nitrocellulose filters. In 2007, Al-Natour et al. showed the binding of highly lysine-rich H1 to superhelical DNA, favoring it over linear or nicked circular DNA forms as deduced from direct competition experiments. [144] Lymphocyte DNA fragments, weighing 2×10^6 Da exhibited a binding affinity with H1 at a magnitude at least 15 times greater than equivalent E. coli fragments of the same molecular weight (2×10^6 Da). Moreover, studies have shown that the distribution of preferential binding sites for histone I on fragmented DNA is strongly influenced

by the size of the DNA fragments. Larger DNA fragments are more likely to contain preferential binding sites for histone I, while smaller fragments tend to lack these sites. This finding highlights the critical role of DNA type and its fragment size in regulating the interaction between DNA and histones, which is important for understanding the dynamics of chromatin structure and the specificity of histone-DNA binding.[138]

A previous study has demonstrated that double-stranded DNA can induce the formation of secondary structures, such as helices and turns, in a peptide from the C-terminal (COOH-terminal) domain of histone H1, which is otherwise mostly unstructured in solution. This domain is located next to the globular region of histone H1 and may affect the shape of linker DNA where it enters and exits the nucleosome. As a result, it could play a role in gene regulation mediated by the histone tail domains. [145] Mello et al. (2012) showed the vibrational characteristics of DNA chemical groups, notably PO_2, were influenced in distinct ways by histone H1, protamine, and histone-mimicking macromolecules. Specifically, they reported that the shift of DNA PO_2^- antisymmetric stretching to a lower frequency accompanied by an enhanced intensity of this vibration is particularly influenced by lysine-rich histones. [146]

2.3. Surface Classification and Recognition of Functionalized Polymer Coatings

2.3.1. General Polymer Surface Properties and Classifications

The performance of biomaterials is significantly influenced by their interfacial properties, which affect key aspects like cell adhesion, [25] biocompatibility, [27] and wettability. [28] A common approach to enhance these surface properties is through the application of polymer coatings, which can impart functionalities such as antifouling, bactericidal, or improved biocompatibility. [147, 148] After modifying the polymer surfaces, characterization and control of the functional groups on material surfaces is essential for optimizing their performance for

various applications. [148] Moreover, surface characterization is frequently employed in industry to detect potential surface contaminants and to analyze variations in surface chemistry that may affect adhesion, wetting, biological integration, catalyst fouling, or other performance characteristics. [36]

Effective polymer surface classification provides a systematic framework to evaluate, compare, and predict the performance of polymer coatings under diverse functional requirements. This is especially critical in biomedical, environmental, and industrial applications, where surface-dependent behaviors such as protein binding significantly impact material performance.

Polymer surfaces can generally be divided into untreated and surface-modified categories. While untreated surfaces retain the native characteristics of the polymer, such as the inherent properties of natural polymers like chitosan, collagen, and cellulose, surface-modified variants are engineered through physical or chemical processes to enhance functionality for specific applications. Additionally, polymer film surfaces can be classified based on variations in molecular structure, composition and ratio of their constituent materials, as well as preprocessing methods and environmental conditions during their fabrication. [149, 150] A reliable classification system not only facilitates material selection and application-specific optimization but also supports the reproducibility of results across studies and industries.

2.3.2. Wettability-Based Surface Classification

Various classification methods have been developed to improve the design and optimization of surfaces, focusing on measurable properties such as wettability and hydration behavior. These parameters serve as critical indicators for evaluating and recognizing the functionality of polymer coatings. Previous studies have highlighted the importance of

wettability-driven classification in optimizing the design and performance of functionalized polymer coatings. Surfaces are commonly classified based on static contact angle measurements, distinguishing them as superhydrophilic (<10 °), hydrophilic (10 °–90 °), hydrophobic (90 °–150 °), or superhydrophobic (>150 °). These thresholds are commonly used to correlate wetting behavior with biological or industrial performance. Wettability is primarily governed by the balance of intermolecular forces at the solid–liquid interface and is affected by liquid properties, surface characteristics of the solid, and surrounding environmental conditions. This property can be deliberately tuned through physical modification of surface morphology or chemical alterations of the surface composition. Extensive research has revealed that modifying a material's surface texture and roughness can effectively influence its wetting behavior with different liquids. [151-153] In addition, surface functionalization techniques that introduce hydrophilic or hydrophobic chemical groups have been widely employed to tailor a surface's interaction with water, thereby achieving the desired wetting characteristics. [154, 155] As a key surface attribute, wettability significantly impacts the biocompatibility of polymer materials, particularly in biomedical contexts. It governs protein-surface interactions, where increased surface hydrophobicity generally leads to greater protein adsorption and more pronounced conformational changes in the adsorbed proteins. [156, 157] Conversely, highly hydrophilic surfaces form a strong hydration layer that acts as a protective barrier, preventing nonspecific molecular adhesion and enhancing antifouling performance. [158] Moreover, the ability of a surface to retain hydration is closely influenced by both the intrinsic properties of the material and its surface architecture. Key factors include the chemical composition, hydrophobicity, along with surface-specific features such as film thickness, and the packing density of surface molecules. These combined parameters play a critical role in determining the surface's interaction with its environment and thus contribute to the classification of polymer coatings based on wettability-related behavior.[158] Integrating multiple

physicochemical metrics, including contact angle, surface energy, and chemical composition, can lead to a more robust and comprehensive classification framework, enhancing the predictability of coating behavior in practical applications.

2.3.3. Thickness-Based Classification of Layer by Layer Films

In addition to wettability, film thickness represents a fundamental parameter for the classification and functional assessment of polymer-coated surfaces. This property is closely linked to key material behaviors, such as surface packing density and resistance to nonspecific protein adsorption, which are critical in both biomedical and industrial applications. [159] Thickness-dependent classification provides valuable insights into the performance of multilayer assemblies, particularly in layer-by-layer (LbL) fabricated coatings, where precise control over structural parameters is essential. Several challenges have been identified in the development of LbL films, particularly regarding the empirical selection of polycations and polyanions, as well as the inherently time-consuming nature of multilayer construction. The accurate control of film thickness remains difficult due to the sensitivity of the deposition process to environmental factors, including pH, temperature, and ionic strength. Moreover, the diversity in polymer, polymer interactions and growth kinetics, ranging from linear to exponential, complicates the formation of homogeneous and reproducible coatings. To assess and classify surface coatings based on thickness, various analytical techniques have been widely adopted. Among them, quartz crystal microbalance with dissipation monitoring (QCM-D), atomic force microscopy (AFM), and ellipsometry are frequently employed. However, inconsistencies among these methods have been reported, particularly in the quantification of ultrathin films, emphasizing the need for standardized and systematic approaches to thickness-based surface classification. Thickness-based classification serves as a critical parameter for the systematic characterization of polymer-coated surfaces. By correlating film thickness with

functional performance, this approach enhances the predictive capability of surface behavior under specific application conditions and supports the rational design of advanced coating systems. [160]

2.3.4. PPX Polymers and their Applications

Poly-p-xylylene polymers are highly regarded for their ability to modify surface properties by forming stable, conformal coatings, thereby enhancing the performance of a wide range of materials. Certain members of the poly-p-xylylene family, commercially known as Parylenes, are particularly valued for their exceptional solvent resistance at elevated temperatures, high melting points, low dielectric constants, and outstanding barrier properties. These characteristics make PPX highly suitable for applications requiring chemical stability, electrical insulation, and effective protection against environmental factors. There is a strong focus on understanding how these properties influence performance, particularly in electronic materials, biomaterials, and separation technologies. [161] Parylene coatings are found to be extensively used as barriers in implantable chemical sensors, stainless steel implants, pacemakers, stents, and catheters. Commercially available Parylenes include Parylene N derived from non-functionalized PCP, Parylene-C (produced from mono-chloro-PCP), and Parylene-D (derived from di-chloro-PCP). [162] Parylene C, a flexible dielectric polymer, is commonly utilized in electronic applications for its ability to enable the production of fully transparent and flexible devices. Serving as a substrate or encapsulation material, it offers a flexible medium conducive to high-frequency electronic signal operation due to its low dielectric loss properties. Additionally, its minimal moisture absorption significantly enhances the stability of compatible electronic technologies, such as oxides and organics. [163] Fluorinated variants such as Parylene HT or Parylene AF-4 offer enhanced thermal and UV stability, making them ideal for use in advanced thin-film transistors. [164] For instance, Kim et

al. demonstrated the potential of PPX derivatives, including PPX-AF4, as dielectrics for oxide-based semiconductors on flexible plastic substrates. In their study, a dielectric film of PPX-AF4, combined with other PPX variants, resulted in organic thin film transistors (TFTs) with significantly improved mobility, outperforming conventional SiO_2 dielectric top gates. Additionally, semiconductor-insulator TFT designs using a PPX-AF4 layer with barium zinc tin oxide (BZTO) as the semiconductor material exhibited promising mobility values, highlighting the potential of these fluorinated PPX materials in electronic applications. [165] Nonfunctional and halogen-containing PCPs are widely accessible and can be used as commercial precursors for barrier coatings like Parylene N, C, D, HT, and AF-4. These coatings exhibit high solvent, temperature, and chemical stability, along with low dielectric constants. [37]

With the growing focus on advanced medical treatments, such as artery stents, implants, bioadhesive sensors, and wound healing patches, there is an increasing demand for customized biointerfaces. PPX polymers are particularly promising in this field due to their high molecular weight, strong adhesive properties, and ability to provide uniform coverage. These characteristics make PPXs highly suitable for use as coatings in drug-eluting stents (DES). Significant advancements have been made in DES development, including the FDA-approved Taxus product by Boston Scientific. [166] Enhanced methods for synthesizing PCPs, featuring functional groups such as hydroxyl, amine, and aldehyde, have been developed. These precursors hold promise for producing bioactive coatings. [167, 168] For instance, using functionalized PPX coatings, poly-(hydroxymethyl-p-xylylene-co-p-xylylene) (PPX-HM), Vorwerk et al. designed biocompatible stents that did not rely on the release of therapeutic drugs. Their study demonstrated the in vitro biocompatibility of PPX-HM-coated stents, showing reduced platelet adhesion compared to uncoated controls. [161] In a recent study, researchers demonstrated the remarkable potential of PPX coatings by developing an alkyne-

functionalized variant that serves as an excellent platform for biomolecule attachment via the surface-oriented CuAAC click reaction. This functionalized PPX-Alkyne coating enabled the conjugation of poly (sulfobetaine methacrylate-co-Az), resulting in highly stable polymer layers that effectively inhibited cell adhesion and protein adsorption. [169] Its ability to support precise biomolecular modifications enhances its relevance in advanced medical technologies, further establishing PPX as a crucial material for next-generation biomedical coating.

2.4. Chemical Vapor Deposition Polymerization

Conventional wet chemical coating methods, such as spray, dip, and spin coating, can provide satisfactory results for many applications. However, they face significant challenges when precise thickness control, high-quality coatings, and uniform coverage of complex geometries are required. These limitations arise due to solvent evaporation during drying and baking, which can introduce defects like pores and cracks, compromising coating integrity. Additionally, the effectiveness of solvent-based coatings depends on the substrate's surface energy and chemical compatibility, making them unsuitable for certain materials. Furthermore, achieving uniform coatings on porous or curved surfaces often necessitates multiple processing steps. To overcome these challenges, chemical vapor deposition (CVD) polymerization has gained increasing attention, offering a solvent-free approach that ensures conformal, chemically pure, and defect-free coatings for biomedical and industrial applications. Based on the intended application area, various polymer layers with different functional groups can be derived using the CVD technique. [161, 170]

The production of poly(p-xylylene) (PPX) films and their functionalized variants, employ various polymerization techniques. One of these methods involves a reaction pathway utilizing paracyclophane (PCP) as a precursor molecule, a process first outlined by William Gorham at Union Carbide in 1966. [171] This Gorham technique utilizes heat to decompose

vapor-phase reactants associated with [2.2] paracyclophane (PCP). The resulting radical species polymerize on cooled surfaces through a chain growth mechanism. This cyclophane-based CVD polymerization process involves three distinct stages. Firstly, the PCP precursor undergoes sublimation at temperatures ranging from 100 to 200 °C under vacuum conditions of 0.1–0.3 Torr. Subsequently, the precursor gas is transported via a stream of inert gas, such as argon, to the pyrolysis zone, where the PCP precursor reacts at temperatures between 500 and 800 °C to form two 1,4-quinodimethane radicals while retaining the functional groups. Because of the minimal energy gap between these two states, typically ranging from 8 to 9 kcal mol^{-1}, the radical exhibits high reactivity and readily condenses at lower temperatures, forming PPX. For this, the pre-formed radicals are transported to a deposition stage maintained at lower temperatures (<30 °C). In this stage, the substrate is positioned, leading to the formation of parylene polymers or copolymers on the surface. In the deposition chamber, parameters including deposition temperature, monomer ratio, and deposition rate significantly influence the properties of the deposited polymer. For instance, the deposition temperature and rate directly impact the formation of the polymer, thereby affecting characteristics such as barrier properties, which in turn influences its suitability for various applications. Lower deposition temperatures are associated with higher growth rates and the deposition of polymers with higher molecular weights, resulting in superior thermal stability. PPX polymerized at a rapid deposition rate tends to exhibit a granular morphology with high surface roughness, whereas PPX polymerized at a slower deposition rate tends to have a smoother surface. The flexibility of the CVD method allows for the creation of functional parylene coatings with adjustable thickness, spatial and temporal compositions, and multi-layer arrangements. [172]

In conclusion, chemical vapor deposition (CVD) polymerization is a highly efficient technique for achieving high-quality surface functionalization and is applicable to a wide range of substrates, including metals, ceramics, glass, and synthetic materials. [173-177] Through CVD

polymerization, functionalized PCPs can introduce a variety of functional groups without altering the underlying backbone chemistry. [37, 178] Since CVD polymerization involves the direct deposition of polymer films from the gas phase, it eliminates the need for solvents, catalysts, or liquid phases in the process. [179] Moreover, CVD polymerization enables the integration of bio-based materials into thin film fabrication, offering a more sustainable approach to surface modification. Unlike conventional high-temperature deposition methods that may compromise temperature-sensitive substrates, CVD can be carried out at or even below room temperature, preserving the structural integrity of delicate materials. This process involves high-temperature monomer activation followed by low-temperature deposition, allowing the formation of high-quality polymer coatings without causing damage. As a result, CVD polymerization serves as a versatile and environmentally friendly technique suitable for a wide range of applications. [180, 181] This technique produces uniform, pinhole-free coatings with minimal impurities, which can be easily patterned. [180, 182] Chemical vapor deposition (CVD) polymerization offers a versatile engineering of surface properties, applicable to various substrate types. [162]

The experimental setup for chemical vapor deposition (CVD) polymerization is configured to regulate various factors, such as the temperature of pyrolysis, pressure, flow rate of the carrier gas, and the temperature of the substrate (**Figure 2.7**).

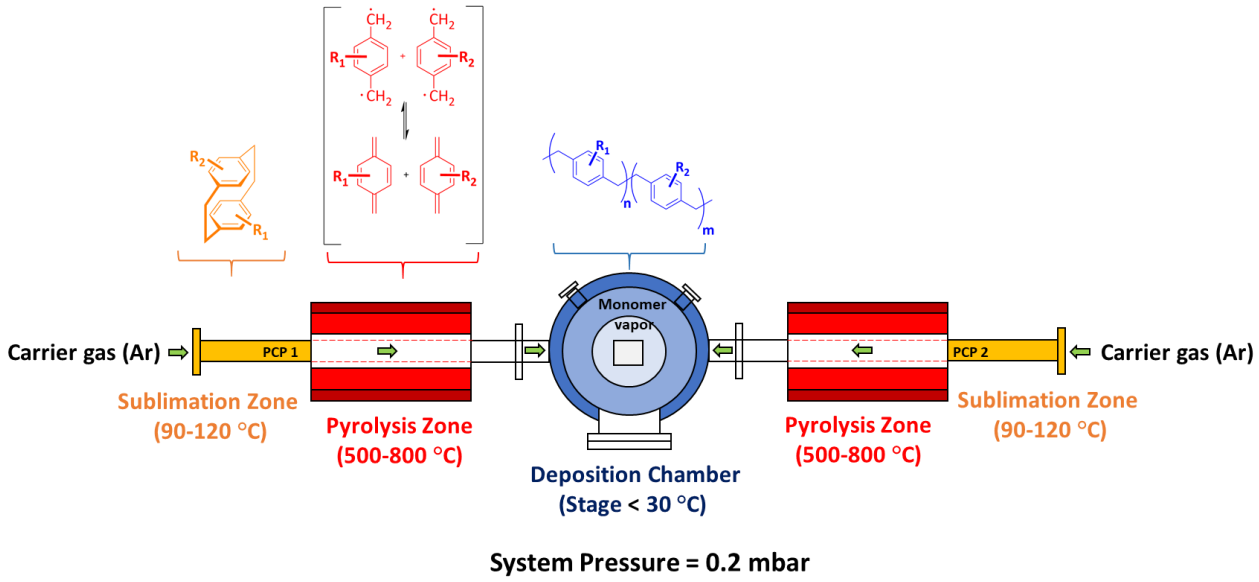


Figure 2.7. Conceptual overview and schematic representation of the CVD polymerization process. For copolymerization, both precursor components (PCP1 and PCP2) are sublimated and pyrolyzed simultaneously, enabling the spontaneous formation of copolymers in the deposition chamber. Adapted from [183]

In this dissertation, CVD polymerization is a sustainable method for depositing uniform, defect-free thin films with tunable properties. Its high-precision coating capability provides a reliable platform for studying droplet drying dynamics by minimizing surface defects and ensuring accurate observations, which is essential for optimizing applications related to droplet behavior. This approach also allows precise control over the thickness and composition of the surface, ensuring consistency across experiments. [184] In this study, CVD-coated surfaces play a key role in investigating protein/DNA interactions and surface chemistries recognition and classification. The uniformity and controllability of these coatings reduce surface irregularities, ensuring that droplet patterns, molecular interactions, and protein adsorption behaviors mainly depend on the intrinsic properties of the biomolecules rather than unintended surface variations. This level of surface consistency is particularly valuable for deep learning applications, as it provides high-quality, reliable data for training models to analyze droplet behavior, molecular interactions, and surface recognition with high accuracy.

2.5. Machine Learning

Machine learning is the study of algorithms that enable computers to automatically learn from data and enhance their performance based on experience, enabling autonomous decision-making without external aid. [185] These decisions are derived from identifying important patterns in complex datasets. There are several primary categories of machine learning algorithms, including supervised, unsupervised, and reinforcement learning, based on their learning approach and input and output data types, as well as the specific problems they target. [186, 187] Additionally, there are hybrid approaches and other common methods that provide natural extensions of machine learning problem formulations. Supervised learning is employed when data includes input variables along with corresponding output target values. [185] Supervised learning consist of two main categories: classification and regression. [188] Classification involves predicting an output variable that belongs to a finite set of known categories, such as distinguishing between "cat" or "dog" and determining "positive" or "negative" outcomes. In contrast, regression focuses on predicting an output variable that is a real or continuous value, such as estimating a "price". [189]

2.5.1. Deep Learning

As illustrated in **Figure 2.8**, deep learning, a subset of machine learning, plays a crucial role in various fields such as image recognition, natural language processing, and speech recognition. [190] Since 2006, deep learning has become a significant domain within machine learning, revolutionizing by promoting end-to-end learning, i.e., processing data from pixel level to real-world applications. [191] Unlike traditional machine learning models that rely on handcrafted features, deep learning models leverage multiple layers of artificial neurons to progressively

extract higher-level features from raw input data. At the heart of deep learning are artificial neural networks, which are inspired by the structure and functionality of the human brain.

In recent years, deep learning has enhanced the analysis of noisy biomarker data, improving assays for DNA sequences, gene expression, and other biological measurements. It also aids in phenotype prediction from genetic data, including traits and disease risks, by integrating additional data such as medical images and clinical histories. A promising approach involves predicting intermediate molecular phenotypes, such as gene expression or splicing, which are easier to forecast and provide valuable insights for disease prediction. Overall, deep learning's ability to process complex data and predict molecular states has made it a powerful tool in advancing genetic and medical research. [192]

The current surge in deep learning can be attributed primarily to three key factors: the remarkable enhancement in chip processing power, particularly the widespread use of GPUs, the substantial reduction in computing hardware costs, and the notable progress in machine learning algorithms. [193] The deep learning algorithms are divided into four categories based on the basic method that they are derived from: Convolutional Neural Networks (CNNs), Restricted Boltzmann Machines (RBMs), Sparse Autoencoder, and Long Short-Term Memory (LSTM). [194]

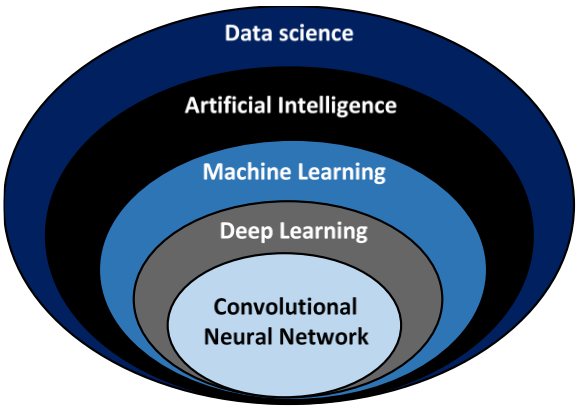


Figure 2.8. Data science methods include different approaches. Artificial intelligence (AI) is a part of data science and includes traditional programming as well as machine learning (ML). Machine learning involves several models and techniques, including deep learning (DL) and Convolutional neural networks (CNN). Adapted from [195]

The Multi-Layer Perceptron (MLP) is a fundamental feedforward neural network that serves as the basis for deep learning models. Conceptually inspired by the structure and function of the human brain, the MLP consists of three main layers: an input layer, an output layer, and typically one or more hidden layers. In the brain, the approximately 86 billion neurons work in parallel, receiving input through dendrites, processing it in the cell body, and transmitting output along axons. The synaptic connections between neurons are crucial for learning, as they strengthen or weaken over time based on experience, a process called synaptic plasticity. This adaptability forms the basis of memory and learning in the brain. Similarly, in an MLP, each neuron in a layer is connected to all neurons in the next layer, enabling efficient flow of information.

The input layer of the MLP functions like the brain's sensory cortex, receiving data from the external environment. This data is then passed through one or more hidden layers, where the network learns complex patterns and relationships through weighted connections. This process mirrors how the brain's association cortex processes and integrates sensory inputs, relating them to past experiences and knowledge. As in the brain, where synaptic connections adjust based on experience, the MLP adjusts the weights of its connections through a process known as backpropagation, fine-tuning these weights during training using gradient descent. Finally, the output layer generates a prediction based on the learned features, similar to how the brain responds to inputs and adapts to new information. **Figure 2.9** illustrates a biological neuron, versus single-neuron perceptron model where multiple input values ($x_1, x_2, \ldots, x_n$) and a bias term (represented as 1) are each assigned corresponding weights ($w_0, w_1, \ldots, w_n$). These weighted inputs are summed at the summation node ($\sum$) before being passed through an activation function (g), which determines the final output (y) (Eq 2.5.1). [196]

$$y = g\left(w_0 + \sum_{i=1}^{n} x_i \; w_i\right) \qquad \text{(Eq 2.5.1)}$$

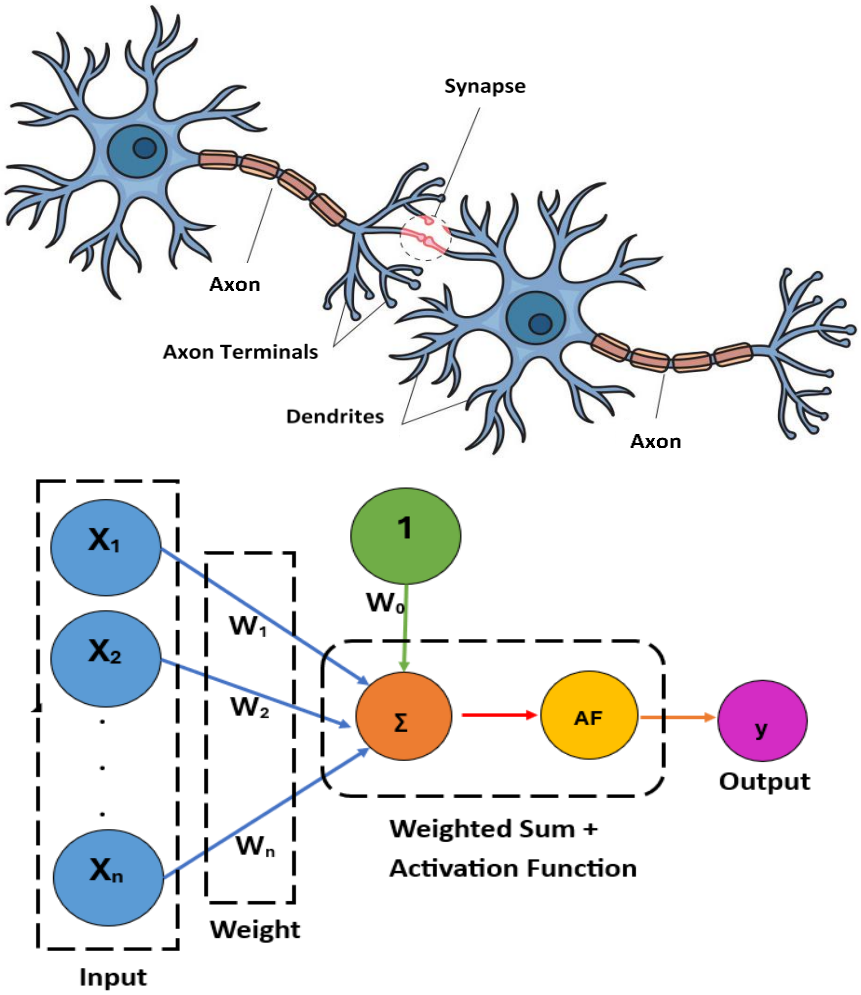


Figure 2.9. The schematic of biological neuron versus single-neuron perceptron. Adapted from [196]

2.5.2. Convolution Neural Networks (CNNs)

The Convolutional Neural Network (CNN) extends the concept of MLPs by incorporating specialized layers, such as convolutional and pooling layers, which enable automatic feature extraction from spatial data. Unlike MLPs, which treat all input features equally, CNNs preserve the spatial structure of images, allowing them to capture important local patterns such as edges, textures, and shapes. This makes CNNs particularly effective for image processing and classification, object detection, face recognition, automatic handwriting recognition, and other computer vision tasks. [197-204] The application of CNN has significantly expanded due to its ability to extract crucial patterns from images without human intervention. [205] For instance,

if a given dataset containing various images featuring both dogs and cats, a CNN that was trained with a large number of examples can automatically distinguish whether a particular image represents a dog or a cat by utilizing the key features it has acquired through learning. [206]

As shown in **Figure 2.10**, the image (which serves as input data for the Convolutional Neural Network) has three dimensions: width, height, and depth. Each layer of the CNN takes a 3D input and produces a 3D output of neuron activations. For RGB images, the depth is 3, representing the three-color channels, while the height and width correspond to the dimensions of the image. [203] CNN consists of three neural layers: convolutional, pooling, and fully connected. The initial two layers, convolution and pooling, primarily engage in feature extraction, while the third layer, a fully connected layer, translates these extracted features into the final output, such as classification. [207] The standard image classification CNN architecture is presented in **Figure 2.10**. [208] The training process of the network involves two stages: the forward stage (Forward propagation) and the backward stage (Backpropagation). [203, 209] In forward propagation, the network processes the input data through its layers to generate predictions. During this phase, the kernels in the convolutional layers and weights in the fully connected layers are applied to the input data. The predicted outputs are compared to the corresponding ground truth labels through a loss function, which quantifies the discrepancy between the predicted and actual values. In backpropagation, the loss value obtained from forward propagation is used to update the learnable parameters, such as kernels and weights. The backpropagation algorithm, combined with the gradient descent optimization method, adjusts these parameters to minimize the error, improving the model's performance in future iterations. Once enough iterations of the forward and backward stages have been completed, the network learning process can be stopped. [194]

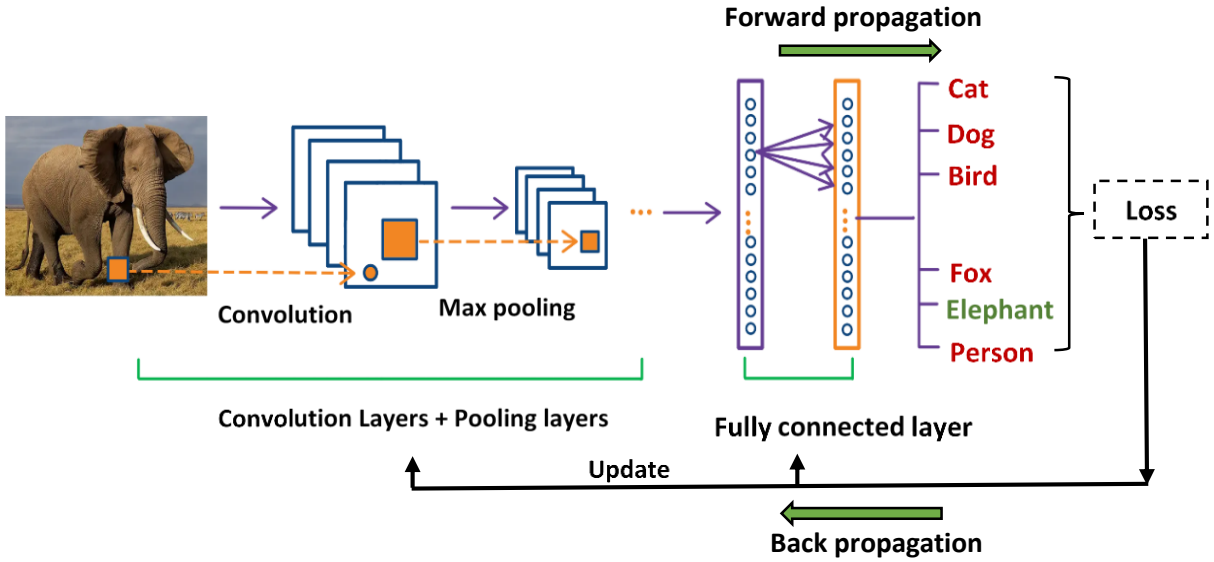


Figure 2.10. The pipeline of the general CNN architecture. A CNN consists of several parts, including: convolution layers, pooling layers, and fully connected layers. The model's performance is evaluated using a loss function during forward propagation on the training dataset. The learnable parameters, such as kernels and weights, are then updated based on the loss value through backpropagation using the gradient descent optimization algorithm. Adapted from [194]

Convolutional Layers

The convolutional layer acts as the fundamental building block of CNN. [210] Its primary role is feature extraction, which typically combines linear and nonlinear operations i.e., convolution operations and activation functions. [207] The convolution operation offers three primary benefits. [211] i) Efficiency through shared weights: Convolution operations contribute to increased model efficiency by reducing the number of parameters that need to be learned compared to fully connected neural networks. This efficiency arises from the shared weights across different positions, leading to a more compact representation of the learned features. ii) Translation invariance: It ensures that local feature patterns extracted by kernels remain invariant to translation as the kernels traverse through different positions within the image. This enables the detection of learned local patterns consistently across the image. iii) Spatial hierarchies and downsampling: By incorporating downsampling along with a pooling operation, convolution operations facilitate the learning of spatial hierarchies of feature

patterns. This process results in the capture of progressively larger fields of view, enhancing the network's ability to understand complex spatial relationships.

This layer takes input data along with a kernel (or filter), typically a 2D array of numbers, which contains learned weights, in the case of image input. The configuration of a convolutional layer is primarily determined by several hyperparameters, including the number of filters, their spatial dimensions (e.g.,3 × 3), stride, and padding. Stride indicates the degree of movement of the filter as it scans across the image. The output is determined by the kernel configuration. The kernel or filter moves horizontally across the input image with a specific stride value until it covers the entire width. It then shifts downward to the starting position on the left side of the image with the same stride value and repeats this process until the entire image has been traversed. This output then becomes the input for the next stage of processing (**Figure 2.11**). For instance, the convolution operation involves an input image of size 6 × 6 convolved with a 3 × 3 kernel to generate a 4 × 4 convoluted feature maps. At each position, a 3 × 3 region of the input is multiplied element-wise with the filter, and the results are summed to produce a single value in the output map. [203] The size of the resulting feature map is determined by the dimensions of the kernel in use. For RGB-colored images, the multiplication process occurs independently for each color channel (red, blue, and green) using the respective kernel. Subsequently, the outputs from these operations are combined to form the convolution output. [203]

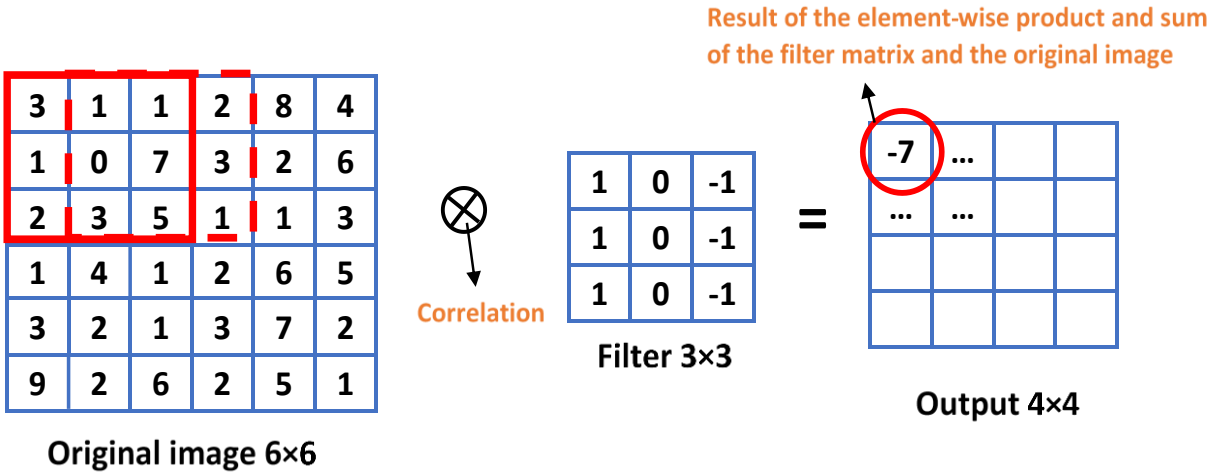


Figure 2.11. The operation of the convolutional layer. An illustration of the convolution operation is shown with a 3 × 3 kernel, and a stride of 1. The kernel is applied across the input tensor, performing element-wise multiplication at each location, followed by summing the results to produce the output value in the corresponding position of the output tensor, known as the feature map. Adapted from [212]

Padding is used to maintain the size of the output after applying convolution (**Figure 2.12**). By adding zeros around the borders of the image, the filter can slide across the entire image without reducing the output dimensions, thereby ensuring that the output size remains equal to the input size when using the same padding. This technique is particularly useful for preserving spatial information and preventing loss of data at the edges during the convolution operation. [213]

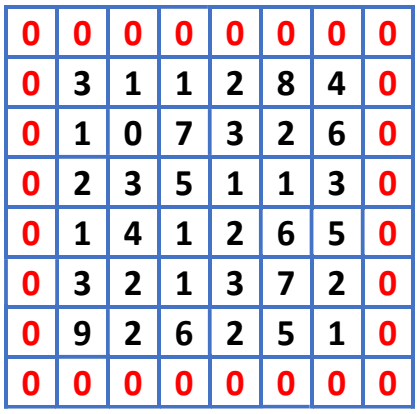


Figure 2.12. Illustration of zero-padding in image processing. The image is padded with zeros along its borders to enable the filter to slide over it while maintaining an output size that is equal to the input size. Adapted from [212]

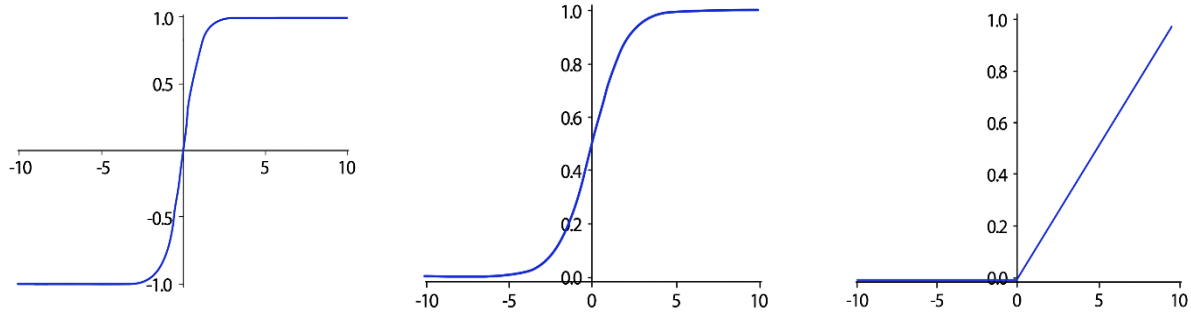
Size of the kernels, the number of kernels, padding, and stride are hyperparameters that must be defined before the training process begins. Kernel weights are considered parameters in a CNN, as they are learned and adjusted during the training process. [207]

Nonlinear Activation Function

The activation function selected for a neural network significantly influences its overall performance and the efficiency of its learning process. [214] Activation functions are essential because they introduce nonlinearity into the model, allowing neural networks to capture complex patterns in data. Effective activation functions typically exhibit properties such as differentiability to facilitate gradient-based optimization, simplicity for efficient computation, minimal parameters to reduce computational complexity, and robustness against saturation which helps prevent vanishing gradients that could obstruct parameter updates and hinder deep network training. [215]

The outputs of the convolutional layer pass through a nonlinear activation function (**Figure 2.13**), which is crucial for introducing nonlinearity into the network. However, when the input variable takes on either a large positive or small negative value, certain activation functions can saturate, becoming unresponsive to small changes in the input data. This saturation leads to gradients approaching zero during backpropagation, effectively stopping weight updates and obstructing the training of deep neural networks. To address this issue, Nair and Hinton proposed the Rectified Linear Unit (ReLU) activation function. ReLU eliminates the need for computationally expensive exponential calculations and determines activation values using a simple threshold rule. These attributes make ReLU computationally efficient and highly effective, establishing it as the default activation function in feedforward neural networks, where it has substantially improved training efficiency and overall performance. [216]

Other nonlinear functions, such as sigmoid and hyperbolic tangent (tanh), were historically favored for their mathematical resemblance to the behavior of biological neurons. [217]



a) $tanh\ (x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}}$ ***b)*** $\sigma\ (x) = \frac{1}{1+e^{-x}}$ ***c)*** $f\ (x) = max\ (0,x)$

Figure 2.13. Common activation functions used in neural networks. a) Hyperbolic Tangent (tanh), b) Sigmoid, and c) Rectified Linear Unit (ReLU). Adapted from [207]

Pooling Layers

Typically, after a convolutional layer, a pooling layer is employed to diminish the dimensions of feature maps and network parameters. This layer helps cut down on the parameters, resulting in faster training times. It downsamples each feature map in terms of height and width while maintaining the depth. Similar to convolutional layers, pooling layers exhibit translation invariance, as they consider neighboring pixels in their computations. The most frequently used strategies include average pooling and max pooling. **Figure 2.14** illustrates a max pooling operation as an example. For an initial set of 8×8 feature maps, this process reduces the output maps to dimensions of 4×4 using a max pooling operator with a size of 2×2 and a stride of 2. Max pooling selects the maximum value from the pooling window. Conversely, average pooling (Fp) can be calculated as shown in Eq. 2.5.2. In this equation, Fp represents the output value at position (*i, j*), obtained by applying the pooling operation (pool (*i, j*)). The operation is performed over a pooling window that contains M elements, where each element within the

window is represented by x_k, with k indicating the index of the input value in the pooling region. [203]

$$Fp = \text{pool}\ (i, j) = \frac{1}{M}\sum_{k=1}^{M} x_k \qquad \text{(Eq.2.5.2)}$$

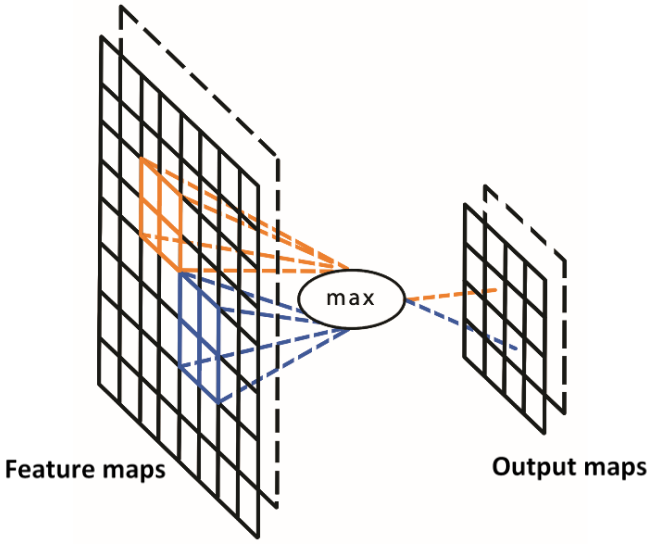


Figure 2.14. The operation of the max pooling layer. Adapted from [218]

Scherer et al. [219] carried out a comparative analysis of these two pooling operations and determined that max-pooling exhibits advantages, including quicker convergence, selection of superior invariant features, and enhanced generalization. In recent years, a range of efficient GPU implementations of CNN variations has emerged, with a majority of them favoring the max-pooling strategy.

Fully-Connected Layers

The fully connected layer represents the final component of a convolutional neural network (CNN). After the final pooling layer, as shown in **Figure 2.15**, the resulting two-dimensional feature maps are flattened into a one-dimensional vector. [207] This vector is then passed through one or more fully connected (dense) layers, where each node in the input is connected to every node in the output through trainable weights. These layers process the features previously

extracted by the convolutional operations and reduced in dimensionality by pooling, ultimately transforming them into the final outputs of the network, such as class scores in classification tasks. Typically, the final fully connected layer contains the same number of output units as there are target classes. [220] A nonlinear activation function, such as ReLU, is commonly applied after each fully connected layer, as previously discussed.

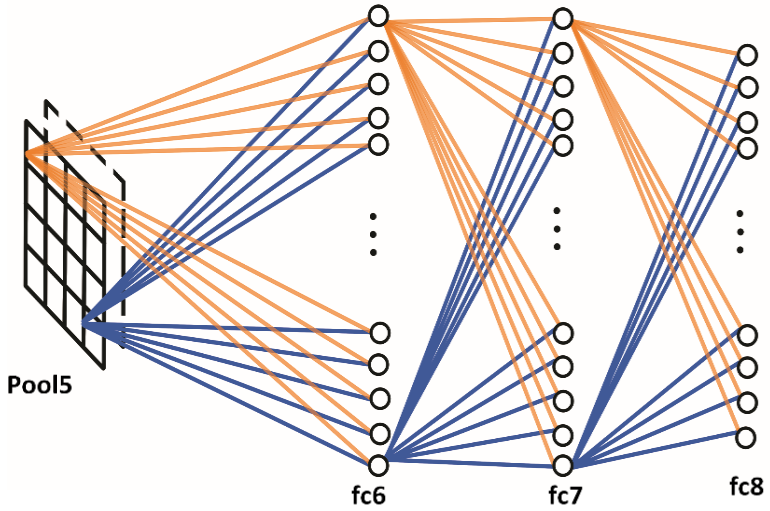


Figure 2.15. The operation of the fully-connected layer. Adapted from [218]

Final Layer Activation Function

The activation function applied to the last fully connected layer typically differs from the others. It is crucial to select an appropriate activation function based on the task at hand. For multiclass classification tasks, the commonly used activation function is Softmax. This function normalizes the output real values from the last fully connected layer into target class probabilities, ensuring that each value falls between 0 and 1, and the sum of all values equals 1. [217] The common choices for the activation function in the last layer for various types of tasks are summarized in **Table 2.1**.

Table 2.1. A list of frequently used activation functions in the final layer for different tasks. Adapted from [207]

Task	Last Layer Activation Function
Binary classification	Sigmoid
Multiclass single-class classification	Softmax
Multiclass multiclass-class classification	Sigmoid

2.5.3. Training Strategy of a Network

Training a network involves the iterative process of adjusting kernels in convolution layers and weights in fully connected layers to minimize the disparities between predicted outputs and provided ground truth labels in a training dataset. The backpropagation algorithm is widely utilized technique for training neural networks, where the loss function and gradient descent optimization algorithm are pivotal components in the process. [221]

Loss Function

A loss function, sometimes called a cost function, evaluates the agreement between the output predictions generated by the network through forward propagation and the provided ground truth labels. [222] For multiclass classification tasks, the commonly used loss function is cross-entropy, [223] while mean squared error is typically employed for regression tasks involving continuous values. The selection of the appropriate loss function is considered one of the hyperparameters and should be determined based on the specific requirements of the task at hand.

Gradient Descent

Gradient descent is a widely used optimization algorithm that iteratively adjusts the learnable parameters, kernels and weights, of the network to minimize the loss. The gradient of the loss

function indicates the direction in which the function exhibits the steepest rate of increase. Consequently, each learnable parameter is updated in the opposite direction of the gradient, with the step size determined by a hyperparameter known as the learning rate. The gradient represents the partial derivative of the loss concerning each learnable parameter mathematically (**Figure 2.16**). A single update of a parameter is formulated as follows (Eq. 2.5.3):

$$w := w - \alpha \times \frac{\partial L}{\partial w} \qquad \text{(Eq.2.5.3)}$$

Here, "w" represents each learnable parameter, "α" represents the learning rate, and "L" signifies the loss function. Noteworthy, setting an appropriate learning rate is crucial as it is one of the most significant hyperparameters to determine before training commences. In practice, due to constraints such as memory limitations, the gradients of the loss function regarding the parameters are calculated using a subset of the training dataset known as a mini-batch. These gradients are then utilized to update the parameters. [207] There are three main types of gradient descent algorithms: i) Batch Gradient Descent (BGD), which processes the entire dataset simultaneously during training and then updates the weights, making it a deterministic approach; ii) Stochastic Gradient Descent (SGD), which updates the weights after processing each individual data point one at a time; and iii) Mini-batch Gradient Descent, a hybrid approach that updates the weights using a small subset of data points at each step. These methods differ in the amount of data used to calculate the gradient, balancing the trade-off between the accuracy of parameter updates and the time required for each update. [224]

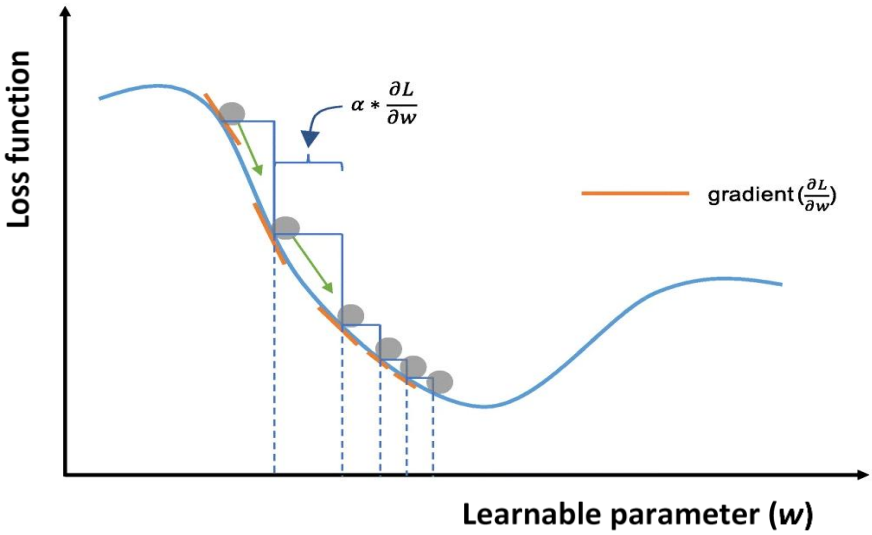


Figure 2.16. Schematic representation of the gradient descent concept. Learnable parameters are iteratively adjusted via gradient descent, minimizing loss (the difference between predicted and true values. Adapted from [207]

2.5.4. Challenges in Deep Learning Algorithms: Solutions and Strategies

In deep learning classification, data is systematically partitioned into training, validation, and test sets to facilitate efficient model development and evaluation. [225] The training set serves as the foundation for model learning. [226] During training, the model is exposed to training data, calculating errors through forward propagation and iteratively adjusting its learnable parameters via backpropagation to improve its ability to learn the underlying patterns. The validation set acts as a crucial checkpoint throughout this process. [227] By evaluating the model's performance on the validation set, researchers can fine-tune hyperparameters (such as learning rates and number of iterations), which are parameters that govern the training process, and ultimately select the optimal model configuration. The test set, ideally reserved for use at the conclusion of the project, provides an unbiased estimate of the model's generalizability to unseen data. This final evaluation step is crucial for determining how well the model would perform in real-world scenarios (**Figure 2.17**).

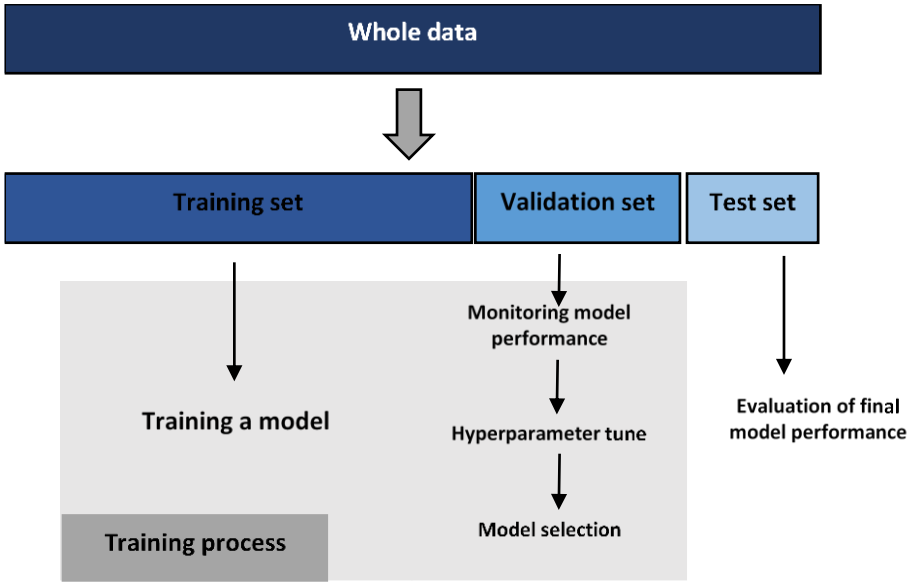


Figure 2.17. Typical data division for machine learning. The data is segregated into training (Training Set), validation (Validation Set), and test sets (Test Set). Training Set: Used to train the model. Forward propagation calculates the loss, guiding the update of learnable parameters through backpropagation. Validation Set: Monitors model performance during training, aiding hyperparameter tuning and model selection. Test Set: Evaluates the final model's generalizability on unseen data, ideally used only once at the project's conclusion. Adapted from [207]

It is important to have distinct validation and test sets because hyperparameter tuning and model selection take place during the training phase. This process relies on the model's performance on the validation set, which can result in some information from the validation set being unintentionally incorporated into the model, leading to overfitting, despite the model not being explicitly trained on the validation set for its learnable parameters. Consequently, the model will likely perform well on the validation set. To accurately assess the model's performance and generalizability on completely new data, an entirely separate test set is required.

Overfitting occurs when a model learns statistical patterns particular to the training set, resulting in memorization of irrelevant noise rather than learning the essential signal (**Figure 2.18**). In deep learning, overcoming overfitting and managing computation time are significant challenges due to the increased layers of abstraction. A standard method to detect overfitting to the training data involves monitoring the loss and accuracy on both the training and validation sets. To combat overfitting, several approaches have been suggested to reduce

overfitting such as using more training data, data augmentation, regularization (weight decay, dropout), batch normalization, and reduce architecture complexity. [203, 207] A model trained on a larger dataset often demonstrates better generalization, although achieving this may not always be feasible in many applications like medical imaging. [228]

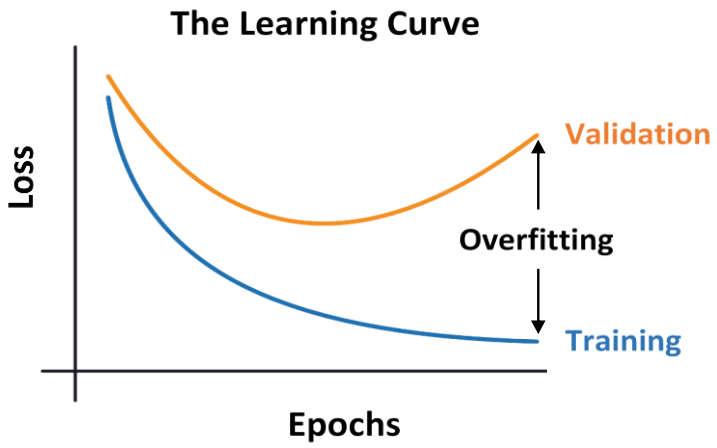


Figure 2.18. The loss curves for training and validation per epoch, illustrating overfitting. The overfitting phase is shown where the validation error (orange curve) begins to rise while the training error (blue curve) continues to decrease. This indicates that the model is learning the training data too well, capturing noise and specific details, which leads to a poor generalization on unseen data. Adapted from [229]

Data augmentation is primarily used to generate additional data from existing samples through various transformations, without incurring extra labeling costs. This is often achieved through random transformations such as flipping, cropping, and rotation. By applying such transformations, data augmentation can significantly expand the training dataset, making it especially valuable when working with limited data. [203, 230]

Besides data augmentation, transfer learning is another widely used and efficient technique for training a network with a small dataset (**Figure 2.19**). In this approach, a network is initially trained on a significantly large dataset, like ImageNet, which comprises 1.4 million images categorized into 1000 classes. Subsequently, this pre-trained network is repurposed and utilized for the specific given task. Transfer learning relies on the premise that common features

acquired from a sufficiently extensive dataset can be applied across diverse datasets. This ability to transfer learned features is a distinctive strength of deep learning, facilitating its utility across different domain tasks, even when datasets are limited in size. There are two methods for employing a pretrained network: fixed feature extraction and fine-tuning. Fine-tuning method involves not only replace the fully connected layers of the pre-trained model with new ones to train on a specific dataset, but also adjusting some or all of the kernels within the pre-trained convolutional base using backpropagation. This adjustment can involve fine-tuning all layers of the convolutional base or selectively freezing earlier layers while fine-tuning deeper ones. This approach is driven by the observation that early-layer features, such as edges, are more applicable across different datasets and tasks, while later features tend to become increasingly tailored to specific datasets or tasks. [207, 231]

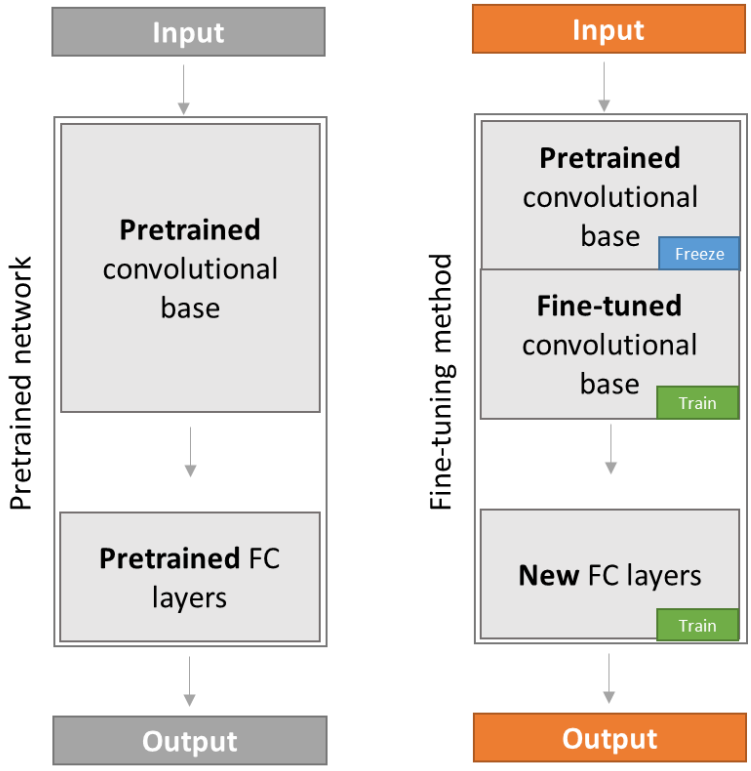


Figure 2.19. Schematic representation of the transfer learning technique. Utilizing transfer learning is a widespread and efficient approach for training a network with a limited dataset. Adapted from [207]

2.5.5. Gradient-Weighted Class Activation Map (Grad-CAM)

Gradient-weighted class activation map, developed by Selvaraju and colleagues, [232] is a powerful visualization technique that enhances the interpretability of convolutional neural networks (CNNs) in image classification tasks. It works by leveraging the gradient of the classification score with respect to the convolutional features from the final convolutional layer to identify the regions of an image most influential for the model's decision. While deep learning models have achieved significant success in computer vision tasks, their complexity often creates a challenge in understanding their decision-making processes. Grad-CAM addresses this issue by providing visual explanations that highlight the specific regions of an image most responsible for the model's predictions, thereby enhancing transparency and interpretability. The key innovation of Grad-CAM lies in its ability to generate class-specific localization maps, allowing it to highlight different areas of an image based on the model's classification. This technique generates heatmaps to visualize the regions of an image most relevant to a deep network's predictions, providing valuable insights into the model's decision-making process. This capability is particularly valuable in domains such as medical imaging, autonomous driving, and security, where model transparency is crucial. Moreover, Grad-CAM is an effective method for providing visual insights into deep networks, demonstrating its utility across various complex tasks, including object recognition, classification, and action recognition. For instance, it can accurately identify attributes like the breed of a dog, or the musical instrument being played. Beyond improving model interpretation and debugging, Grad-CAM plays a key role in advancing visual recognition algorithms and refining existing models. This versatile tool can be applied across various neural network architectures and tasks beyond image classification. By visualizing how CNNs make decisions, it helps users gain deeper insight into model behavior, identify potential issues, and understand which features

influence predictions. In conclusion, Grad-CAM enhances the interpretability and reliability of deep networks, making them more transparent and trustworthy. [232]

2.5.6. t-Distributed Stochastic Neighbor Embedding (t-SNE)

t-Distributed Stochastic Neighbor Embedding (t-SNE) is a nonlinear machine learning technique for reducing the dimensionality of high-dimensional data, enabling its representation in two- or three-dimensional space. Proposed by Maaten and Hinton as an enhancement of Stochastic Neighbor Embedding, this nonlinear technique focuses on maintaining the local relationships within the data. By positioning similar points close to each other and separating dissimilar ones, t-SNE provides an effective way to visualize and interpret complex datasets in reduced dimensions. [233, 234] Although t-SNE visualizations often suggest the presence of clusters, these apparent groupings can be significantly influenced by the choice of parameters, particularly perplexity, as different perplexities often yield different visualizations. Perplexity can be viewed as a trade-off between preserving the global and local structures of data. Specifically, a higher perplexity value considers a broader range of neighboring points, while a lower value focuses on a smaller set of neighbors, making the algorithm more sensitive to local variations in the data. Maaten and Hinton recommended typical perplexity values in the range of 5 to 50. [234] The optimal value of perplexity depends on the data density, with larger and denser datasets typically requiring a higher perplexity value.[235] In most implementations, the default value of perplexity is set to 30. [2] Furthermore, the size and distance between clusters in t-SNE plots should not be interpreted as an indication of true global distances or cluster separability. Studies have demonstrated that t-SNE can effectively identify well-separated clusters and, with specific parameter settings, can approximate a simplified form of spectral clustering.

2.6. Characterization Methods-Cross Validation with Experimental Data

2.6.1. Fluorescence Spectroscopy

Fluorescence spectroscopy is a powerful technique for studying protein-ligand and protein-DNA interactions. [236] Fluorescence is a type of luminescence that occurs when a molecule absorbs photons, exciting its electrons to a higher energy level, typically from a singlet ground state to a singlet-excited state. As the molecule returns to its ground state, it emits a photon of lower energy and longer wavelength than the absorbed one. Fluorescence spectroscopy analyses this emitted fluorescence to study the molecular properties and environment. (**Figure 2.20**). [237]

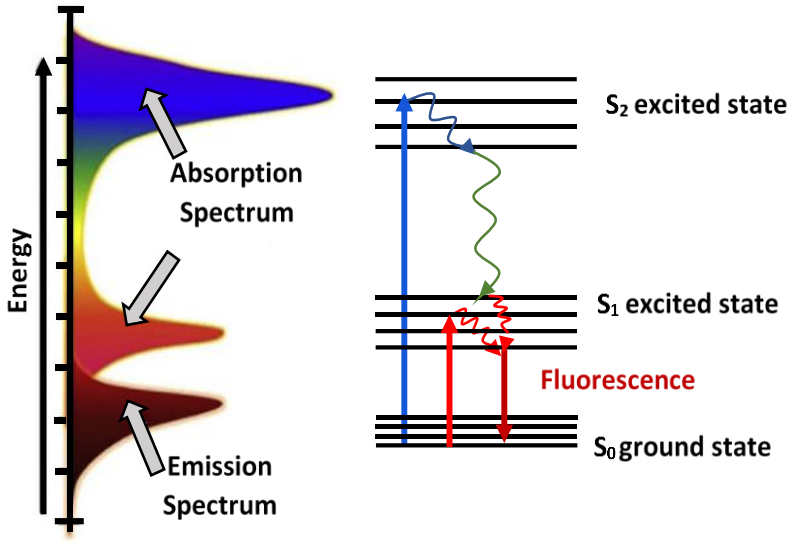


Figure 2.20. Jablonski diagram. This diagram illustrating various transitions between a molecule's energy states. Adapted from [237]

A fluorescence spectrometer is an analytical instrument primarily designed to measure the intensity of fluorescent light emitted by a sample, while also determining the wavelengths at which this fluorescence occurs by recording the emission spectrum. It consists of a light source, which provides the excitation light, and monochromators that select specific

wavelengths for excitation and emission (**Figure 2.21**). The excitation monochromator ensures the sample is exposed to the desired wavelength, while the emission monochromator, typically placed at a 90° angle to the excitation light to minimize background interference from scattered light, isolates the emitted fluorescence. The emitted light is detected by a photomultiplier tube (PMT), which converts it into an electrical signal for processing and display. This setup allows for fluorescence measurements, making fluorescence spectrometers beneficial tools in valuable scientific and analytical applications. [238]

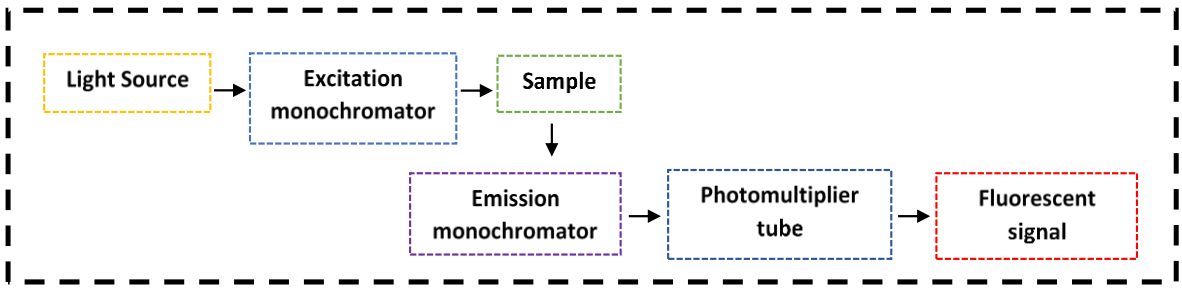


Figure 2.21. Diagram of simplified fluorescence spectroscopy setup. Adapted from [239]

Fluorescence spectroscopy assays present several limitations, particularly in terms of sensitivity when analyzing low-concentration samples or molecules with weak fluorescence. Additionally, fluorescence signals are prone to interference from quenching and background noise, which can hinder accurate data interpretation. [240-243] The requirement for specific fluorescent labels further complicates sample preparation. [243] Moreover, handling dyes such as ethidium bromide necessitates skilled personnel and stringent safety precautions and protocols to minimize risks to health and the environment. Consequently, there is a growing demand for more robust, interference-resistant, and highly sensitive alternatives to fluorescence-based assays.

2.6.2. Time of Flight Secondary Ion Mass Spectrometry (TOF-SIMS)

Time-of-Flight Secondary Ion Mass Spectrometry (ToF-SIMS) is an analytical method that prioritizes surface analysis. [244] [245] It involves bombarding the specimen's surface with a focused primary ion beam, which induces sputtering and the emission of secondary ions. These ions, representing the surface composition, are then accelerated into a "flight tube" where their mass is determined based on their flight time to a detector (**Figure 2.22**). By measuring the mass-to-charge ratios of these ions, SIMS provides insights into the elemental, isotopic, or molecular composition of the surface with a depth resolution typically ranging from 1 to 2 nm. Due to significant variations in ionization probabilities among elements sputtered from different materials, precise calibration against standards is essential for achieving accurate quantitative results with SIMS. A primary ion beam impacts the surface, resulting in the emission of various secondary particles. These include secondary electrons, photons, neutrals, as well as positive and negative secondary ions from the sample. The sputtering yield is influenced by factors such as the energy of the primary ion, the nature of the atoms or molecules, experimental conditions, and the surrounding atmosphere. Secondary ions emitted from organic samples are subsequently analyzed using a mass spectrometer, generating positive or negative mass spectra. These spectra typically include peaks corresponding to precursor or molecular ions, as well as fragment ions characteristic of the sample surface. ToF-SIMS operates in three distinct modes: surface spectroscopy, surface imaging, and depth profiling, making it a versatile tool for detailed surface characterization in various scientific and industrial applications.

TOF-SIMS has some limitations that must be considered despite its high sensitivity and surface specificity. One of the primary constraints is its requirement for ultrahigh vacuum conditions, which increases operational complexity and necessitates the use of advanced

instrumentation and maintenance protocols. This not only elevates the overall cost of the analysis but also restricts the range of compatible sample types. In particular, volatile, hydrated, or biologically sensitive materials may be unsuitable for analysis under such vacuum environments, limiting the applicability of TOF-SIMS in certain fields. These factors collectively reduce the technique's flexibility, especially in studies involving delicate or non-vacuum-stable specimens. [43]

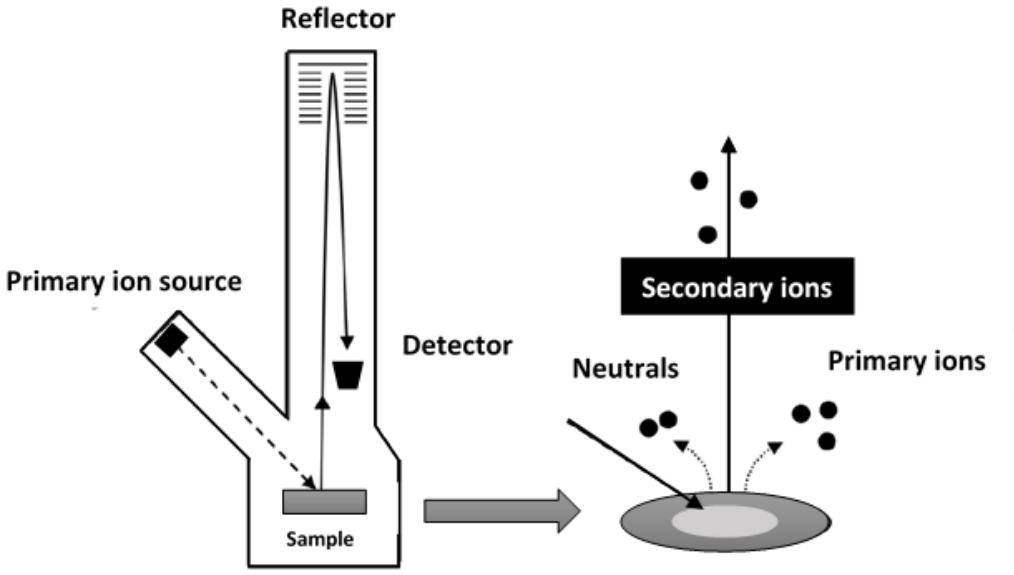


Figure 2.22. Schematic representation of the TOF-SIMS instrument. Adapted from [246]

2.6.3. Infrared (IR) Spectroscopy

Infrared (IR) spectroscopy is a popular tool utilized by both organic and inorganic chemists to analyze a molecule's structure by examining its vibrational and rotational energy level changes when it absorbs infrared radiation. [247, 248] The frequencies absorbed are specific to the molecule's functional groups and atomic masses. Importantly, for a vibration to be IR-active, it must induce a change in the molecule's dipole moment. This capability enables IR spectroscopy to provide insights into the specific functional groups and overall structure of molecules. The electromagnetic spectrum categorizes infrared radiation into three main regions

of the near-infrared, mid-infrared, and far-infrared based on their relationship to the visible spectrum (**Figure 2.23**). [248] [249] The near-infrared region with higher energy, ranging from approximately 14000 to 4000 cm^{-1} (0.8-2.5 µm wavelength), can stimulate overtone or harmonic vibrations. Mid-infrared wavelengths ranging from 4000 to 400 cm^{-1} (2.5-25 µm) are suitable for examining fundamental vibrations and the associated rotational-vibrational structure. Far-infrared radiation, adjacent to the microwave region with wavelengths ranging from 400 to 10 cm^{-1} (25-1000 µm), possesses lower energy and is employed in rotational spectroscopy.

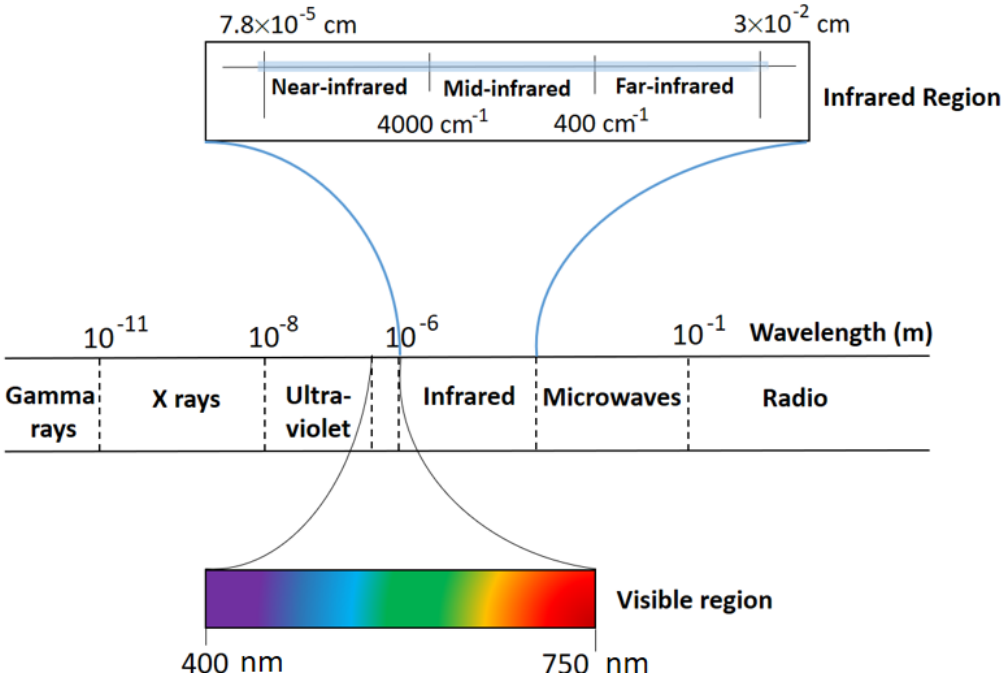


Figure 2.23. Illustration of the electromagnetic spectrum. It highlights the division of the infrared regions into three categories: Near-infrared (12820-4000 cm^{-1}), Mid-infrared (4000-400 cm^{-1}), and Far-infrared (400-33 cm^{-1}). Adapted from [250]

Infrared spectroscopy (**Figure 2.24**) is a widely used method for analyzing surfaces and adsorbate films. Infrared reflection-absorption spectroscopy (IRRAS), also known as reflection-absorption, provides detailed insights into the chemical composition and structure of thin surface layers and adsorbed molecules. [251, 252] IRRAS is a vibrational spectroscopy technique where infrared light is reflected at a grazing incidence angle from a reflective surface.

Most studies have concentrated on metal substrates due to their perfect reflectivity, which ensures maximum sensitivity and comparable detection limits to other surface science techniques. The first efforts in the mid-1980s to apply external reflection measurements to nonmetal substrates like carbon, [253] silicon, [254] or oxide surfaces led to very complex and significantly distorted spectra.

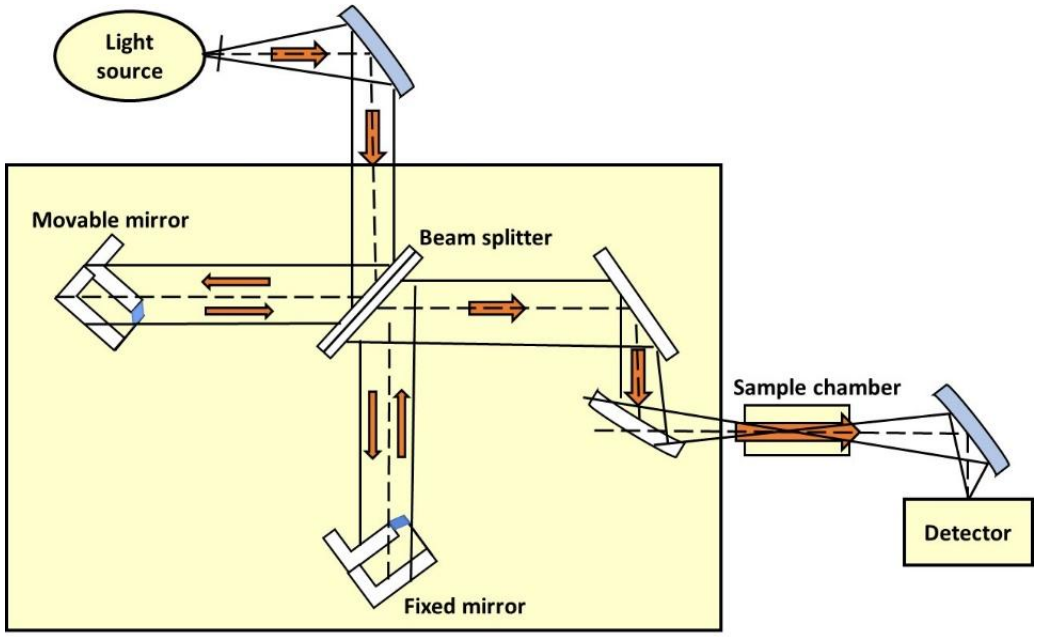


Figure 2.24. Schematic diagram for infrared spectroscopy. Adapted from [255]

When molecules are bound to a conductive surface, an additional selection rule applies: the induced dipole moment must be perpendicular to the plane of the surface. If the induced dipole moment is perpendicular to the surface plane (**Figure 2.25 A**), the image dipole will align in the same direction, enhancing the intensity of the reflected light. Conversely, if the dipole moment is parallel to the surface plane, the image dipole will have an equal magnitude but opposite direction to the induced dipole, causing the vectors to cancel each other out (**Figure 2.25 C**) and resulting in no observable frequency. If the induced dipole is at an intermediate angle, only the perpendicular component of the dipole moment will contribute to the observed signal (**Figure 2.25 B**).

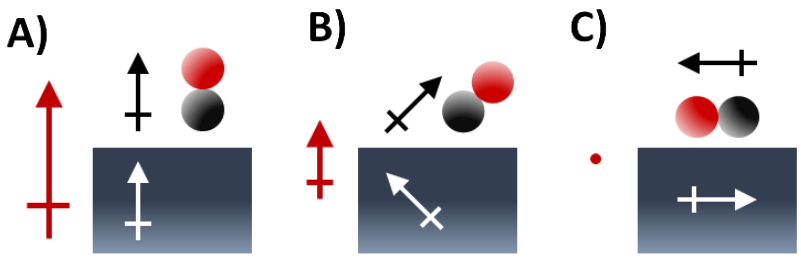


Figure 2.25. Illustration of image dipole theory. The black, white, and red vectors represent the dipole moment of the molecule, the image dipole induced within a reflective substrate, and the resultant vector obtained by summing the black and white vectors, respectively. A) The induced dipole moment is perpendicular to the surface. B) The induced dipole moment is at an intermediate angle to the surface. C) The induced dipole moment is parallel to the surface plane. Adapted from [256]

The infrared (IR) technique exhibits certain limitations, particularly in sensitivity when evaluating samples with minimal thickness or functional groups characterized by weak absorption bands [47]. Furthermore, overlapping absorption bands can complicate the interpretation of IR spectra. Besides, this characterization method requires advanced and costly instrumentation, as well as the expertise of skilled operators for effective operation and maintenance.

2.6.4. Circular Dichroism (CD) Spectroscopy

Circular dichroism (CD) spectroscopy is a widely utilized technique for investigating the secondary structures of optically active molecules, such as proteins. [257] In a CD spectrometer, the sample is placed in a cuvette, and a beam of light is directed through the sample. The light, referring to all electromagnetic waves, undergoes circular polarization, where its plane of polarization rotates either clockwise (right circular polarization) or counterclockwise (left circular polarization) over time as it propagates **(Figure 2.26)**. Clockwise and counterclockwise are sometimes used, but their interpretation depends on the perspective, whether from the source of light or the target sample.

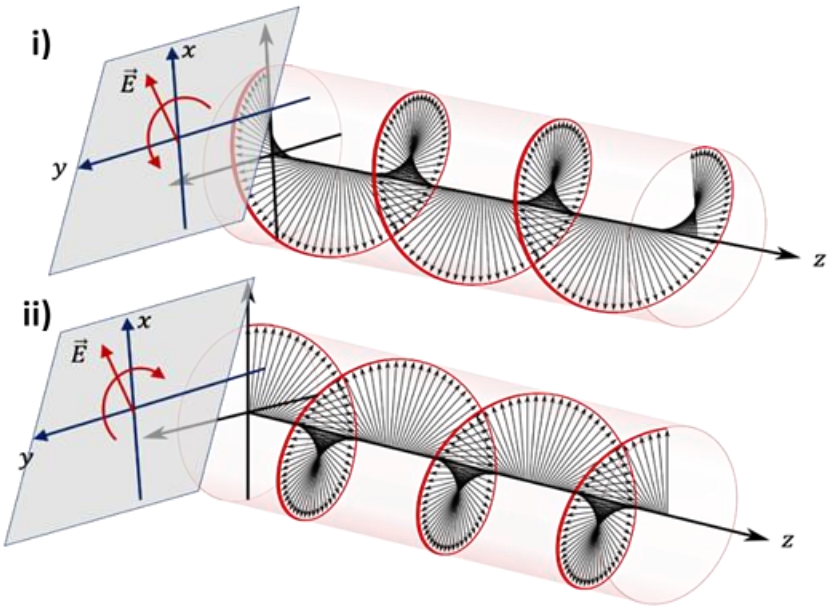


Figure 2.26. Schematic representation of light circular polarization. i) right circularly polarized ii) left circularly polarized light. Adapted from [258]

The sample is typically sequentially irradiated with left- and right-circularly polarized light, and the absorption is measured. Chiral molecules interact differently with circularly polarized light depending on the direction of rotation due to their intrinsic asymmetry, typically absorbing more light in one direction than the other (Eq.2.6.1). The difference in absorption between left and right circularly polarized light is described by the equation below (Eq.2.6.2), where ε_L and ε_R represent the molar extinction coefficients for left and right circularly polarized light, respectively. The variables c and l denote the molar concentration and path length (cuvette width in cm). The difference in absorption can be related to the difference in extinction coefficients ($\Delta\varepsilon$) (Eq.2.6.3).

$$A = \varepsilon c l \qquad \text{(Eq.2.6.1)}$$

$$\Delta A = A_L - A_R = (\varepsilon_L - \varepsilon_R) c l \qquad \text{(Eq.2.6.2)}$$

$$\Delta\varepsilon = \varepsilon_L - \varepsilon_R \qquad \text{(Eq.2.6.3)}$$

CD is reported not only as the difference in absorption or extinction coefficients but also as the degree of ellipticity, [θ]. The relationship between [θ] and Δε is described by the equation in bellow (Eq.2.6.4):

$$[\theta] = 3298\,\Delta\varepsilon \qquad \text{(Eq.2.6.4)}$$

Since the CD spectra of proteins uniquely represents their conformation, CD can be utilized to monitor structural changes due to complex formation like protein-DNA complex formation, folding/unfolding, denaturation caused by temperature increase, amino acid sequence changes, or mutations. This makes CD an effective tool for studying the kinetics of proteins and conducting stability investigations and interaction modeling in dynamic systems. [259]

While CD spectroscopy is a useful technique, it does have certain drawbacks in protein-DNA interaction studies. One key limitation is its lack of specificity, which can make it difficult to distinguish between different molecular components. When studying large complexes, the signals from various structural elements may overlap, complicating the interpretation of conformational changes. This overlap can reduce the clarity of the CD signal or distort the resulting spectra. Additionally, CD spectroscopy is not very sensitive to small or subtle changes in structure, particularly during protein-DNA interactions, making it less effective at detecting minor conformational shifts. [15, 16]

3. Material and Methods

3.1. Chemicals

All chemicals used in the experiments (**Table 3.1**) were of analytical grade and applied without additional purification. Milli-Q water, obtained through a MilliQ-Plus purification system (Merck Millipore), was utilized in all procedures.

Table 3.1. List of Chemicals and materials

Chemicals and Materials	**Company**
Human serum albumin (HSA)	Sigma Aldrich
Immunoglobulin G from human serum	Sigma Aldrich
Histone (Type II-A, lyophilized powder)	Sigma Aldrich
Histone H1 Protein, 20 mg	Sigma Aldrich
Bovine serum albumin	Sigma Aldrich
Lambda DNA (lambda phage), 500 µg/tube	Thermo Fisher Scientific
Deoxyribonucleic acid, low molecular weight from salmon sperm	Sigma Aldrich
Ethidium bromide stock solution, biotechnological quality	VWR Chemicals
Histone from calf thymus	Sigma Aldrich
hsDNA, Deoxyribonucleic acid, partially degraded from herring sperm	Sigma Aldrich
Deoxyribonucleic acid sodium salt from salmon testes	Sigma Aldrich
Ammonium sulfate, for molecular biology, ≥99.0%	Sigma Aldrich
HEPES (biological buffer) ≥99.5% (titration)	Sigma Aldrich
Potassium chloride	Sigma Aldrich
Sodium hydroxide	Merck KGaA

Sodium bicarbonate	Merck KGaA
Sodium carbonate (anhydrous)	Merck KGaA
PCP-H	Curtiss-Wright Surface Technologies
PCP-Cl	SCS GmbH, Surrey, UK
PCP-F	TCI Deutschland GmbH
Glass wafer	Optrovision
Silicon wafer	Si-mat
Gold coated Si wafer	Georg Albert
Microtube	Sigma Aldrich
Syringe filter 0.2 µm	Sartorius Stedim Biotech
Cuvette 1 mm thickness	Helma

3.2. Instrumentation

Table 3-2 lists all the instruments used in this study, including those for cleaning, the CVD process, drop dispersion via a pipetting system, microscopy, and chemical analysis of the respective samples.

Table 3.2. List of instruments

Instruments	**Company**
NanoDrop One	Thermo Scientific
SB3 tube rotator	Stuart
Plasma cleaner	PIE Scientific
Microplate pipetting robot (epMotion 5070)	Eppendorf
1-channel dispenser (TS10)	Eppendorf
Climate chamber (ICH750)	Memmert
Optical microscope (BX-53F)	Olympus
Scanning electron microscopy (SEM)	TESCAN
Time-of -flight secondary-ion mass spectroscopy (TOF-SIMS)	ION-TOF GmbH
Fluorescence spectroscopy	TECAN

Infrared reflection-absorption spectroscopy (IRRAS) Bruker VERTEX 80 FTIR	Bruker Optik GmbH
CD spectroscopy (J-1500 spectropolarimeter)	JASCO

3.3. Software

Table 3.3 provides an overview of the software used for the deep learning approach, imaging, generating input data for deep learning, and performing statistical analysis.

Table 3.3. List of software

Software	**Company**
CellSens	Olympus
MATLAB R2022a	MathWorks
Origin 2022b	OriginLab Corporation
Opus	Bruker

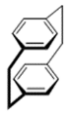
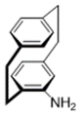
3.4. Chemical Vapor Deposition (CVD) Polymerization Coating

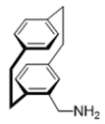
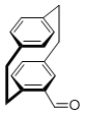
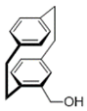
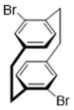
Chemical vapor deposition (CVD) polymerization was employed to apply a layer of poly (p-xylylene) on glass surface. In this procedure, the monomer is initially vaporized and then pyrolyzed, forming reactive intermediates that deposit and polymerize on the cooled clean glass surface. [171] Three distinct zones in CVD machine (sublimation zone, the pyrolysis zone, and the deposition chamber) maintain different temperatures but share the same pressure conditions, which are below 0.2 mbar. A weighted amount of the precursor is introduced into the sublimation zone, located 3-5 cm away from the oven (depends on the precursor). In this zone, PCP sublimates at approximately 100-110 °C and is transported within a stream of argon carrier gas (20 sccm) into the pyrolysis zone. Pyrolysis occurs at 660 °C for PCP-H, PCP-Cl_2,

PCP-Br_2, PCP-F_8, PCP-Aldehyde, PCP-Amine, PCP-Methyl amine, and PCP-Alkyne (**Table 3.4**). For PCP-Methyl Hydroxyl the pyrolysis temperature is 540 °C. The vaporized cyclic dimer is gradually disassembled into reactive radicals by breaking the ethylene bond. Subsequently, these vaporized monomers make their way to the deposition chamber, where the substrates (glass surfaces) are positioned on a stage that is cooled to 25 °C. The concentration of monomers increases upon condensation on the cooled surface until an exothermic reaction occurs. Two radicals combine to form an uncoupled biradical dimer, serving as the initial molecule for the polymerization process. [260] As additional monomers attach, the polymer chain extends, creating a transparent, and uniform film on the substrate. The use of a rotating sample stage guarantees a uniform thickness of the polymer film.

The glass wafers, sized at 120 mm × 80 mm and with a thickness of 0.1 ± 0.05 mm (Optrovision, München, Germany), were subjected to cleansing using a plasma cleaner (Tergeo, Union City, CA, USA). In each deposition batch, two slides of glass substrate were put in the deposition chamber. Adjacent to it, two silicon (Si) wafers were placed to gather data about the thickness of the coating. The deposition speed was determined and could be adjusted to a rate of 0.3-0.5 Å/s by altering the distance between the precursor and the oven, using a quartz crystal detector situated within the deposition chamber.

Table 3.4. Polymer structure of different precursors.

Name of precursor	Polymer structure
PCP-H (PCP-n)	
PCP-NH_2	

PCP-CH_2NH_2	
PCP- Alkyne	
PCP-CHO	
PCP- CH_2OH	
PCP-Cl_2	
PCP-Br_2	
PCP-F_8	

3.5. Histone-DNA Solutions

The Deoxyribonucleic acid sodium salt from salmon tests, Deoxyribonucleic acid from herring sperm, Lambda DNA, and Histone H1 protein were acquired. Salmon (Sal $_{1\ kbp}$) DNA was generated by sonicating Sal $_{20\ kbp}$ DNA, yielding fragments with an average length of 1000 base pairs (**Figure S1**). The DNA was sonicated on ice at 20 kHz for a duration of 5 minutes. To form H1-DNA complexes, their previously equilibrated solutions were directly mixed in a binding buffer. This buffer consisted of 100 mM HEPES (pH 7.8), along with 150 mM potassium chloride and 50 mM ammonium sulfate. Histone was gradually added to the DNA at five ratios: 0.5 (R1), 1 (R2), 1.4 (R3), 3.4 (R4), and 6.8 (R5) moles of histone (H1) per 168

base pairs of DNAs. The overall mass concentration of H1 and DNA remained unchanged across all ratios. The solutions were gently stirred using a tube rotator (Stuart, Stone, UK) at 10 rpm for 60 minutes at room temperature (25 °C) and then stored at -20 °C.

3.6. Bovine Serum Albumin (BSA) Solutions

Bovine Serum Albumin (BSA) was dissolved into a 100 mM carbonate-bicarbonate buffer to a final concentration of 0.1 mg/mL. The buffer, adjusted to a pH of 9.2, was prepared using ultrapure water from a Milli-Q Plus system and contained 91 mM $NaHCO_3$ and 9 mM Na_2CO_3. The BSA solution was mixed for 30 minutes at room temperature using a tube rotator (Stuart, Stone, UK) set at 10 rpm. Upon thorough mixing, aliquots of the solution were stored at -20 °C until further use.

3.7. Droplet Deposition

A microplate pipetting robot with 96 wells (epMotion 5070, Eppendorf AG, Hamburg, Germany) was employed to dispense small droplets (2 μL) of the solutions onto the coated surface using a single-channel dispenser (TS10, Eppendorf AG, Hamburg, Germany). The pipetting setup was configured to distribute 96 droplets on each glass plate, forming a grid of 12 columns and 8 rows. In each experimental set, four different liquids were randomly positioned on the slide, with 24 droplets of each liquid distributed together (H1-DNA interaction section). To regulate the evaporation rate, the robot was placed inside a climate chamber (ICH 750, Memmert GmbH + Co. KG, Schwabach, Germany) maintained at a controlled temperature of 23 ± 0.5 °C and humidity of 40 ± 3%. Following the drying of the droplets (approximately 50 minutes), images of the deposited patterns were captured using a polarized light microscope (Olympus polarizing optical microscope BX-53F, Tokyo, Japan) equipped with an automated stage. All images were captured under uniform microscope

settings, with consistent resolution, 10× magnification, and exposure time, and were combined using the Multi Image Alignment (MIA) algorithm from CellSens software (Olympus, Tokyo, Japan) to ensure data comparability and interpretability. Each image of a dried droplet had rectangular dimensions of 2344 × 1878 pixels in JPG format. To enhance processing speed without sacrificing image quality, they were reduced to 75% of their original size, allowing for faster import into the network for training. Following this resizing, the images were then preprocessed into square dimensions.

3.8. Convolutional Neural Network of Training and Testing Set Images

The training and subsequent processing of PLM images using MATLAB software (R2022a, MathWorks Inc.) involved the selection of the InceptionV3, a pre-trained CNN network, due to its swift response and satisfactory accuracy. InceptionV3, developed by Google, is a deep convolutional neural network architecture specifically designed for image classification tasks. It features "Inception modules" to efficiently capture multi-scale features in an image and comprises 315 layers, making it relatively deep (**Figure 3.1**). The network requires input images to be resized to 299 × 299 pixels, necessitating preprocessing before feeding them into the network for training. Maintaining a uniform format across all groups facilitated consistent comparisons and preserved the reliability of the analysis. Standardizing all images to a resolution of 299 × 299 pixels ensured consistent input dimensions and reduced the influence of original image formats on the outcomes. Employing a transfer learning approach, the network, initially pre-trained on a substantial dataset of image features, underwent fine-tuning with a relatively small new set of images. In this process, the final classification layer was excluded and retrained with the new dataset. Fine-tuning involved adjusting parameters across all layers with a consistent global learning rate of 0.001, a minimum batch size of 32 images.

For each training set, approximately 75-80% of the images per class were used for training, with 10% of them randomly selected for validation during the process. After the network was trained, a separate set of images, accounting for about 15-20% of the total, was reserved for testing to assess its performance. To maintain data integrity, the training, validation, and testing datasets were kept completely separate, with no overlap between them. To enhance the generalization of the network, minimize the risk of overfitting, and prevent it from simply memorizing the training data, images were augmented by applying random horizontal and vertical reflections, each with a 50% probability, across all trained networks.

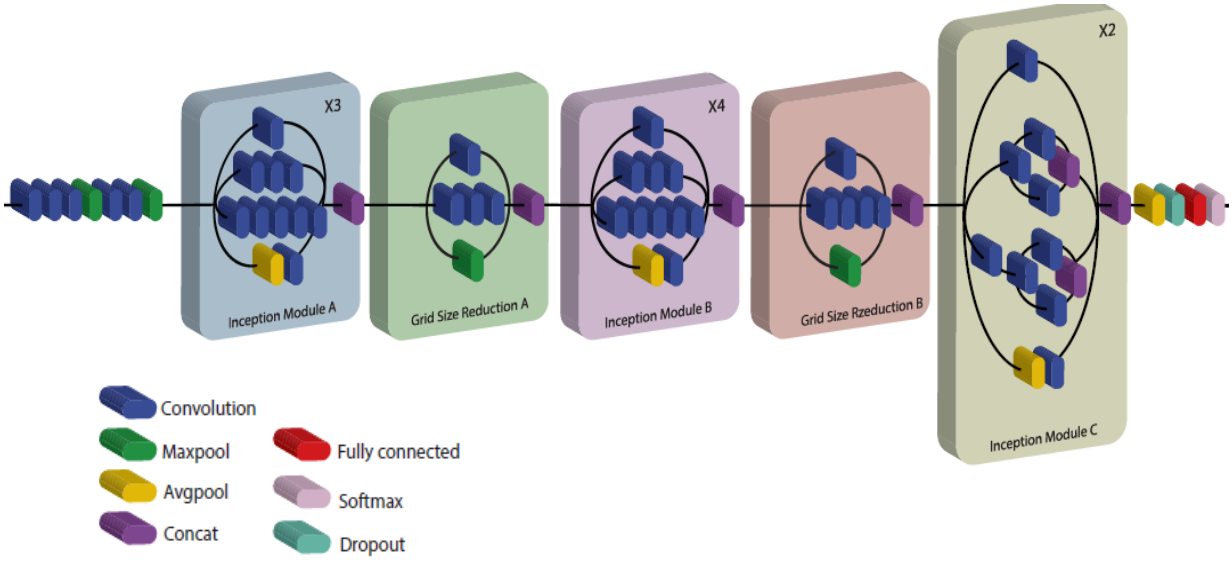


Figure 3.1. Architecture of the pre-trained InceptionV3 model. The model comprises multiple Inception modules that apply convolutional filters of various sizes in parallel to extract multi-scale features. It includes convolutional and max-pooling layers, followed by fully connected layers and a final softmax layer for classification. Adapted from [261]

The network's performance was evaluated by analyzing overall accuracy and confusion matrices for the test datasets. Additionally, to determine how well the model could generalize to new data, a completely unseen/unknown set of images was introduced for classification. This approach allowed for a comparison between these new images and those previously learned during training. By conducting this thorough evaluation, CNN's reliability and effectiveness in classifying diverse datasets were validated.

To identify the key regions within images that significantly influence the classification decisions of the convolutional neural network (CNN), the Gradient-weighted Class Activation Mapping (Grad-CAM) algorithm was applied as a visualization tool. By utilizing Grad-CAM, the most influential areas contributing to the network's decision-making process were highlighted, providing valuable insights into its interpretability. This technique was implemented using MATLAB (Release 2023a, MathWorks Inc.), ensuring an effective analysis of the model's focus during classification.

To demonstrate the network's ability to effectively cluster data, the t-distributed Stochastic Neighbor Embedding (t-SNE) algorithm, known for its strength in visualizing high-dimensional data, was applied to the Softmax layer of the trained convolutional neural network (CNN). This layer, typically used for classification tasks, converts the network's raw output (logits) into a probability distribution. The t-SNE technique was implemented using the MATLAB Machine Learning Package, with a learning rate of 500 and a perplexity value set to 30, to reduce the complex high-dimensional data to a lower-dimensional space while maintaining the relationships between data points. By mapping the Softmax layer outputs to a two-dimensional representation, t-SNE allowed for clear visualization of distinct clusters corresponding to different classes, providing valuable insights into the network's performance and the underlying patterns within the data.

3.9. Scanning Electron Microscopy (SEM)

Analyzing the structural features of the H1/DNA complex and salt of stain droplets was carried out using scanning electron microscopy (SEM) (TESCAN VEGA3). To address potential surface charging effects, a fine layer of gold was sputtered onto the samples before performing SEM imaging. The SEM images were obtained at an electron accelerating voltage of 15 kV, maintaining a working distance of 7.2 mm.

3.10. Time of Flight Secondary Ion Mass Spectrometry (TOF-SIMS)

Time-of-Flight Secondary Ion Mass Spectrometry (ToF-SIMS) experiments were performed using an ION-TOF instrument (ION-TOF GmbH, Münster, Germany). The instrument featured a liquid metal primary-ion source utilizing a Bi cluster and a non-linear time-of-flight analyzer. The Bi source, operating in the "bunched" mode for spectrometry, generated short primary-ion pulses (<1 ns) providing Bi^{1+} or Bi^{3+} ion pulses at 25 keV energy with a lateral resolution of approximately 4 μm. As the droplets' size exceeded the maximum deflection range of the primary-ion gun (500 × 500 μm^2), images were acquired using the manipulator stage scan mode. Negative polarity spectra were calibrated based on the C^-, CH^-, and CH_2^- peaks. Spectrometry was performed in static SIMS mode, ensuring the primary-ion dose remained below $<10^{11}$ ions cm^{-2}. High lateral-resolution images were obtained in a primary-ion source mode, offering a lateral resolution of around 200 nm with nominal mass resolution in "burst alignment" mode. Charge compensation was essential due to the glass substrate, necessitating the use of an electron flood gun delivering electrons of 21 eV. The secondary-ion reflectron was tuned accordingly to ensure precise measurements.

For the surface recognition section, the main chamber pressure was 5×10^{-9} mbar. For high mass resolution, the Bi source was operated in the “high current bunched” mode providing short Bi_3^+ primary ion pulses at 25 keV energy and a lateral resolution of approximately 5 μm. The short pulse length of 1.1 ns allowed for high mass resolution. Primary ion doses were kept below 2×10^{11} ions cm^{-2} (static SIMS limit) for all measurements. Spectra were calibrated on the omnipresent CH^-, C_2^-, C_2H^-, OH^-; or on the CH^+, CH_2^+, CH_3^+, and $C_2H_3^+$ peaks. Spectra were normalized by the total ion dose.

3.11. Infrared Reflection-Absorption Spectroscopy (IRRAS)

Infrared spectral analysis of the polymer films was conducted using a Bruker VERTEX 80 FTIR (Bruker Optik GmbH, Ettlingen, Germany). The spectra were scanned with a resolution of 2 cm^{-1} across the range of 500 to 4000 cm^{-1}. The device featured a horizontal reflection unit for measurements in grazing incidence reflection mode with an 80° incident angle to the surface normal. Background correction was carried out using the onboard Bruker OPUS software.

3.12. Circular Dichroism (CD) Spectroscopy

The far-UV CD spectra of the protein solutions (BSA) were captured using a J-1500spectropolarimeter (JASCO, Germany) at a temperature of 20 °C. For the solution samples it was conducted in quartz glass cuvettes with a 1 mm optical path length within the wavelength range of 260 to 190 nm, with measurements taken at 0.5 nm intervals. Each sample underwent two repeated scans at a scan rate of 100 nm min^{-1}, an 8 s response time, and 8 nm bandwidth. The obtained data were averaged for each sample, along with its respective baseline obtained from the protein-free sample (buffer solution). For protein samples on solid surface, the quartz glass was used. The protein concentration used was 0.1 mg mL^{-1} in a 20 mM carbonate-bicarbonate buffer, with and without 50 mM at pH 9.2.

3.13. Ethidium Bromide Displacement Assay

The ethidium bromide displacement assays, following the method by Geall et al. (2000), were conducted in 10 mM HEPES buffer (pH 7.8) with 15 mM KCl and 5 mM $(NH_4)_2 SO_4$. [262] Steady-state fluorescence measurements were performed using a Tecan Spark multimode reader. In a 96-well black plate maintained at a temperature of 293.15 K, a working volume of 200 µL solution was prepared, containing 0.1 M EtBr for every 168 base pairs of DNAs. Ethidium bromide solution was introduced into the stirring solution and allowed to equilibrate

for 15 minutes. Subsequently, aliquots of histone were added to the stirring solution, and fluorescence was measured after 30 minutes of equilibration.

3.14. Statistical Analysis

Statistical analysis of the results was performed using the analysis of variance, employing both the least significant difference (LSD) and Tukey methods. For least significant difference (LSD), the analysis utilized SAS 9.1.3 software (SAS Institute, Inc., 1999, Cary, NC, USA). Tukey analysis was done with Origin software (2022b). The LSD and Tukey methods were specifically employed to detect significant differences, with a predetermined significance level set at $p < 0.05$, as outlined by Montgomery and Runger (2011). [263]

4. Results and Discussion

4.1. Effect of Various Salts on Stain Patterns

To evaluate the effect of various salts on dried droplet patterns, a library comprising twelve different salts and concentrations was prepared (**Table 4.1**). To achieve uniform hydrophobic substrates with consistent droplet deposition (water contact angle: $80 \pm 1°$) over large areas, chemical vapor deposition (CVD) polymerization of (PPX-H) was employed. The process enabled precise control over the coating thickness, which was maintained in the range of 50-55 nm to ensure consistent surface characteristics and reproducibility. Human serum albumin (HSA) and Immunoglobulin G (IgG) were dissolved separately in each salt solution at a concentration of 0.1 mg/mL (as indicated in the table below). A defined 2 µL volume of the solution was dispensed onto the PPX-H-coated surface and left to dry for 45 minutes under controlled conditions (23 °C, 40% humidity). A total of 7400 images were collected, with 6200 allocated for the training and validation sets, and 1200 images for the test set.

Table 4.1. A library of buffer solutions with various salt types and concentrations.

Nr	Salts	Protein
1	NH_4HCO_3 (0.1 M) + KH_2PO_4 (0.1 M)	HSA
2	NH_4HCO_3 (0.1 M) + KH_2PO_4 (0.1 M)	IgG
3	NH_4HCO_3 (0.1 M) + KH_2PO_4 (0.05 M)	HSA
4	NH_4HCO_3 (0.1 M) + KH_2PO_4 (0.05 M)	IgG
5	NH_4HCO_3 (0.1 M) + $NaHCO_3$ (0.1 M)	HSA
6	NH_4HCO_3 (0.1 M) + $NaHCO_3$ (0.1 M)	IgG
7	NH_4HCO_3 (0.1 M) + $NaHCO_3$ (0.05 M)	HSA

8	NH_4HCO_3 (0.1 M) + $NaHCO_3$ (0.05 M)	IgG
9	NH_4HCO_3 (0.1 M) + Tris (0.1 M)	HSA
10	NH_4HCO_3 (0.1 M) + Tris (0.1 M)	IgG
11	NH_4HCO_3 (0.05 M) + $NaHCO_3$ (0.1 M)	HSA
12	NH_4HCO_3 (0.05 M) + $NaHCO_3$ (0.1 M)	IgG
13	$(NH_4)_2SO_4$ (0.05 M) + $NaHCO_3$ (0.1 M)	HSA
14	$(NH_4)_2SO_4$ (0.05 M) + $NaHCO_3$ (0.1 M)	IgG
15	HEPES (0.05 M) + $NaHCO_3$ (0.1 M)	HSA
16	HEPES (0.05 M) + $NaHCO_3$ (0.1 M)	IgG
17	KH_2PO_4 (0.1 M) + HEPES (0.1 M)	HSA
18	KH_2PO_4 (0.1 M) + Tricin (0.1 M)	HSA
19	Tricin (0.05 M) + $NaHCO_3$ (0.1 M)	HSA
20	Tricin (0.05 M) + $NaHCO_3$ (0.1 M)	IgG
21	Tris (0.1 M) + $NaHCO_3$ (0.1 M)	HSA
22	Tris (0.1 M) + $NaHCO_3$ (0.1 M)	IgG
23	Tris (0.05 M) + $NaHCO_3$ (0.1 M)	HSA
24	Tris (0.05 M) + $NaHCO_3$ (0.1 M)	IgG

Figure 4.1 illustrates the distinct patterns emerging from each solution. Generally, patterns from the same salt solution (with identical salt types and concentrations) exhibit noticeable similarities. In most cases, the dissolved Human Serum Albumin (HSA) and Immunoglobulin G (IgG) do not show significant differences in their resulting patterns (e.g., samples 23 and 24). However, in some instances, notable variations can be observed between HSA and IgG, even within the same salt conditions (e.g., samples 5 and 6).

Inception V3 was employed to classify various patterns, demonstrating exceptional performance. As depicted in **Figure 4.2**, the Convolutional Neural Network (CNN) achieved an impressive average prediction accuracy of 99%. The trained network effectively distinguished between Human Serum Albumin (HSA) and IgG within identical salt solutions, those containing the same salt components at equal concentrations (e.g., samples 3 and 4). Notably, it was capable of differentiating between samples with varying salt concentrations while maintaining the same protein type and salt components (e.g., samples 1 and 3). Furthermore, the model also successfully identified differences in samples where the protein and salt concentration remained unchanged, but the salt type varied (e.g., samples 17 and 18).

The highest misclassification rate was observed in sample 10, where 6% of the samples were incorrectly classified as belonging to sample 9. Despite sharing similar salt components at comparable concentrations, these samples differ in protein content, contributing to the misclassification. The second-highest misclassification rate (4%) was recorded for samples 2 and 7.

Further analysis revealed that misclassifications in sample 2 primarily involved samples 1 and 4, with 2% of the errors linked to their resemblance to sample 1 in terms of salt composition and concentration, and another 2% attributed to its similarity to sample 4 in protein type. Similarly, sample 7 exhibited misclassification errors with sample 11, as these samples share the same protein type and salt components but differ in salt concentrations.

The high classification accuracy indicates that the model effectively captures subtle differences in pattern features, even when those differences arise from minimal changes in salt composition or protein type. This robustness suggests that deep learning approaches like Inception V3 can be valuable tools for analyzing complex biochemical data. Additionally, these results emphasize the potential of CNNs in applications beyond basic classification, such as predictive modeling for understanding how proteins behave under different environmental conditions.

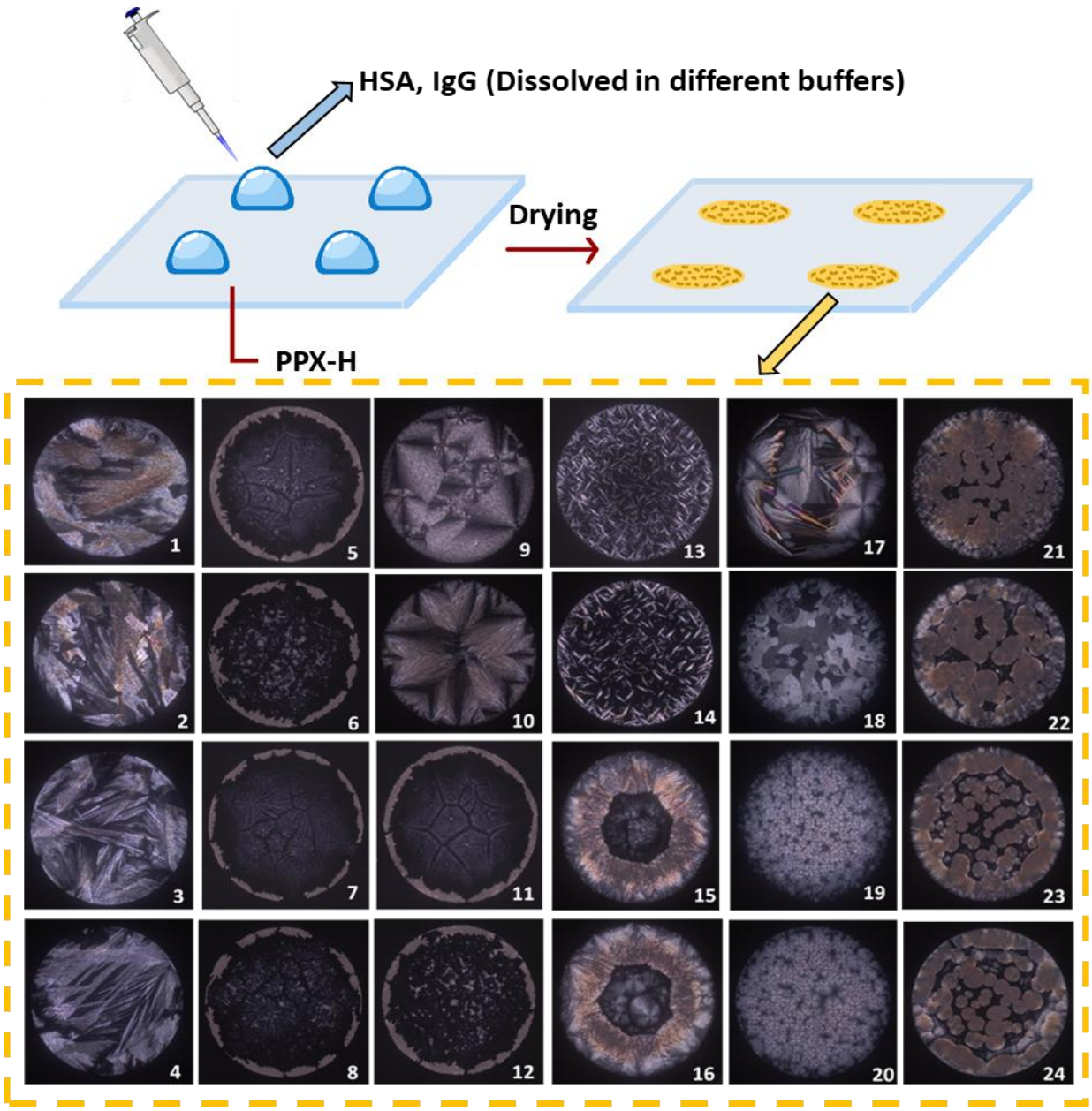


Figure 4.1. Representative PLM images showing the patterns formed by two proteins (HSA and IgG), dissolved in different buffers, and deposited as 2 µL droplets of each solution onto a PPX-H-coated glass wafer. The corresponding buffer compositions are detailed in Table 4.1. Odd-numbered samples contain HSA, while even-numbered samples contain IgG, except for sample 18, which also contains HSA.

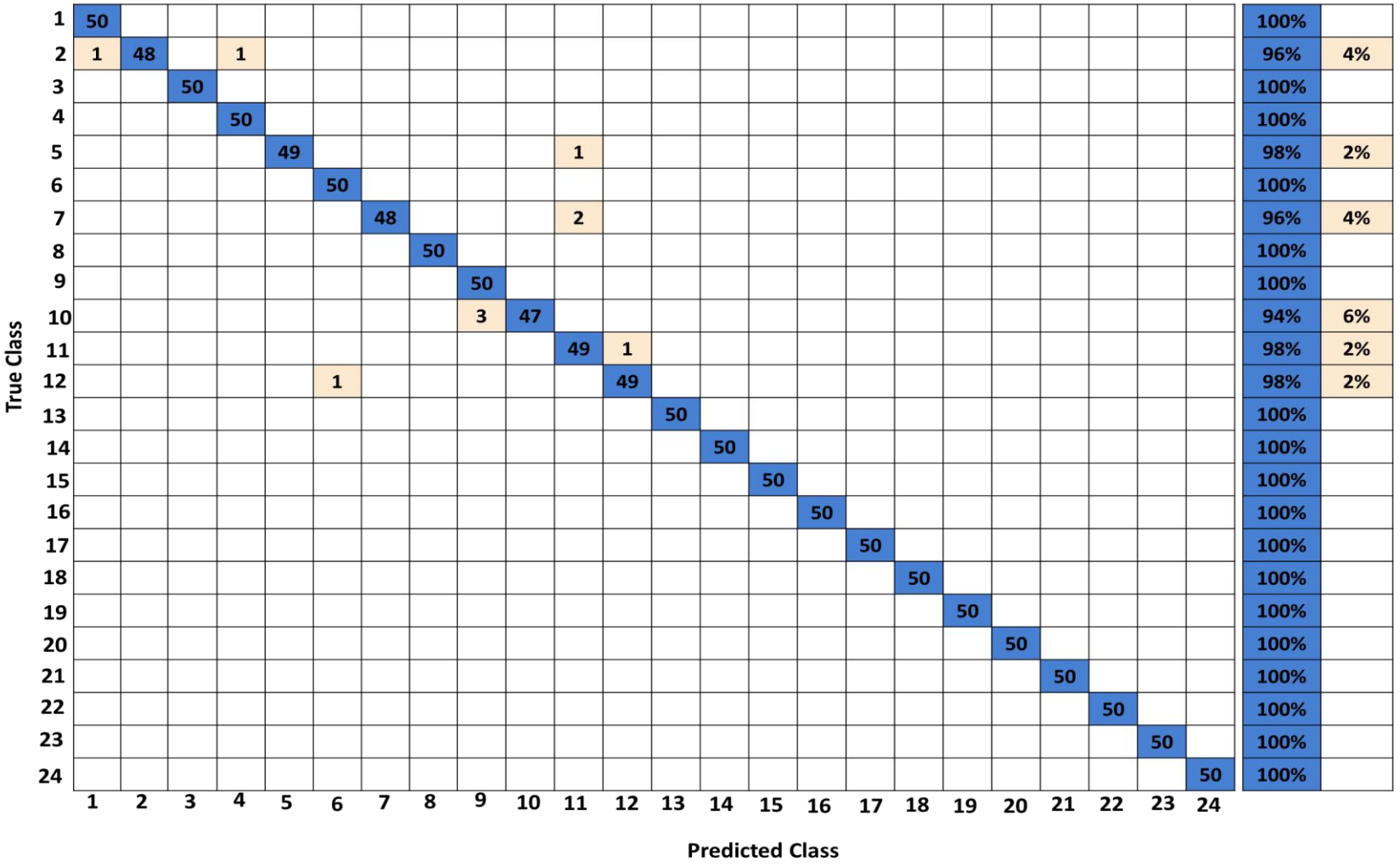


Figure 4.2. Confusion matrix obtained from the stains of various salts containing two different proteins (HSA and IgG). The stains were prepared by depositing 2 μL droplets of each solution onto hydrophobic glass wafers coated with PPX-H. The model parameters were optimized with a global learning rate of 0.001, a minimum batch size of 32 images, and up to 20 epochs.

4.2. Screening and Classification of Protein-DNA Ratios with Various Total Mass Concentrations

In the previous study, a single protein dissolved in buffer solutions was analyzed and classified. In the present study, the approach was expanded by introducing an additional biomolecule, DNA, to explore more complex classification scenarios. By incorporating DNA into the system, the potential of a deep learning model to effectively differentiate mixtures containing multiple biomolecular components was investigated. To evaluate the ability of Convolutional Neural Networks (CNN) in classifying mixtures of two components, specifically protein and DNA, various ratios of histone to Calf thymus DNA and Salmon DNA were prepared (**Tables 4.2, 4.3**). The significance of this study lies in assessing whether the CNN could discern patterns or classifications based on not only the ratios of the components, the relative amounts of histone and DNA, but also the total mass concentration of the mixtures. By focusing on both the component ratios and the total mass concentrations, the research aimed to provide insights into the application of CNNs for analyzing complex biological mixtures and identifying key parameters that should be taken into account when studying interactions between proteins and nucleic acids.

To ensure a systematic evaluation, two specific total mass concentrations, 0.1 mg/mL and 0.3 mg/mL, were selected. Uniform hydrophobic surfaces with consistent droplet behavior across large areas were obtained using chemical vapor deposition (CVD) polymerization of PPX-H. This technique enabled precise control of the coating thickness, which was maintained at 50 ± 5 nm to ensure stable surface properties and experimental reproducibility. A defined 2 µL volume of the solution was dispensed onto the coated surface and left to dry for 50 minutes under controlled conditions (23 °C, 40% humidity). This enabled an evaluation of the CNN's ability to classify different mixture ratios according to their total mass concentration, regardless of variations in the proportions of individual components. By comparing results across different

total mass concentration levels, it was assessed whether high classification accuracy could be maintained by the network despite changes in overall sample composition.

Table 4.2. Average prediction accuracies of two various total mass concentrations of histone-Calf thymus DNA mixture.

Scenario	Group	Prediction Accuracy	Histone/Calf DNA (mg/mg)	Total Mass Concentration
I (Low-Ratios)	A	90%	1/5	0.3 mg/ml
	B		1/10	
	C		1/15	
II (High-Ratios)	D	95%	5/1	
	E		15/1	
	F		30/1	
I (Low-Ratios)	G	93%	1/5	0.1 mg/ml
	H		1/10	
	I		1/15	
II (High-Ratios)	J	100%	5/1	
	K		15/1	
	L		30/1	

Table 4.3. Average prediction accuracies of histone-Salmon DNA mixture.

Scenario	Group	Prediction Accuracy	Histone/ Salmon DNA (mg/mg)	Total Mass Concentration
I (Proof of concept)	A′′	87%	1/5	0.1 mg/ml
	B′′		1/10	
	C′′		1/15	
II (Proof of concept)	D′′	97%	5/1	
	E′′		15/1	
	F′′		30/1	

In total, 7,600 images of deposition stains from histone-Calf and -Salmon DNA mixtures were prepared. Among these, 6,650 images were allocated for training and validation, while an additional 900 images comprised the test set that the network had not encountered during training. This division ensured a robust evaluation of the model's performance. For the histone-Calf thymus DNA mixtures, ratios of 1/15, 1/10, 1/5, 5/1, 15/1, and 30/1 were

examined at two total mass concentrations of 0.1 mg/mL and 0.3 mg/mL. These specific ratios were selected to evaluate how varying proportions of histone to DNA influence the CNN's ability to classify the mixtures accurately. **Figure 4.3** presents the CNN-derived confusion matrix, offering a clear representation of the network's performance across different scenarios. Overall, InceptionV3 achieved an impressive average prediction accuracy of 94%.

For the lower ratios, which contained a higher amount of Calf DNA, the average prediction accuracies were 90% for a total mass concentration of 0.3 mg/mL and 93% for 0.1 mg/mL. This suggests that the network performed slightly better at the lower concentration, likely due to the agglomeration of the complex at the 0.3 mg/mL total mass concentration. Furthermore, the average prediction accuracy for the lower histone-to-DNA ratios (1/15, 1/10, 1/5) was lower than that for the higher ratios (5/1, 15/1, and 30/1).

Notably, the results indicated that, regardless of the total mass concentration (either 0.1 or 0.3 mg/mL), the trend in prediction accuracies remained consistent across two categories of low and high ratios. Additionally, the CNN demonstrated the ability to distinguish similar ratios with varying total mass concentrations. This ability to differentiate between mixtures with identical histone-to-DNA ratios but differing total mass concentrations highlights the network's sensitivity to changes in mass, which could reflect subtle variations in sample composition. The histone/Salmon DNA mixtures were evaluated at a total mass concentration of 0.1 mg/mL, and consistent with the observations for Calf thymus DNA, the average prediction accuracy was higher for high histone-to-DNA ratios (97%) compared to low ratios (87%). These results prompted further investigation into the protein-DNA binding affinity. This suggests that CNN is capable of detecting changes in histone/DNA ratios in the same total mass concentrations, which reflect the outcomes of physiochemical interactions.

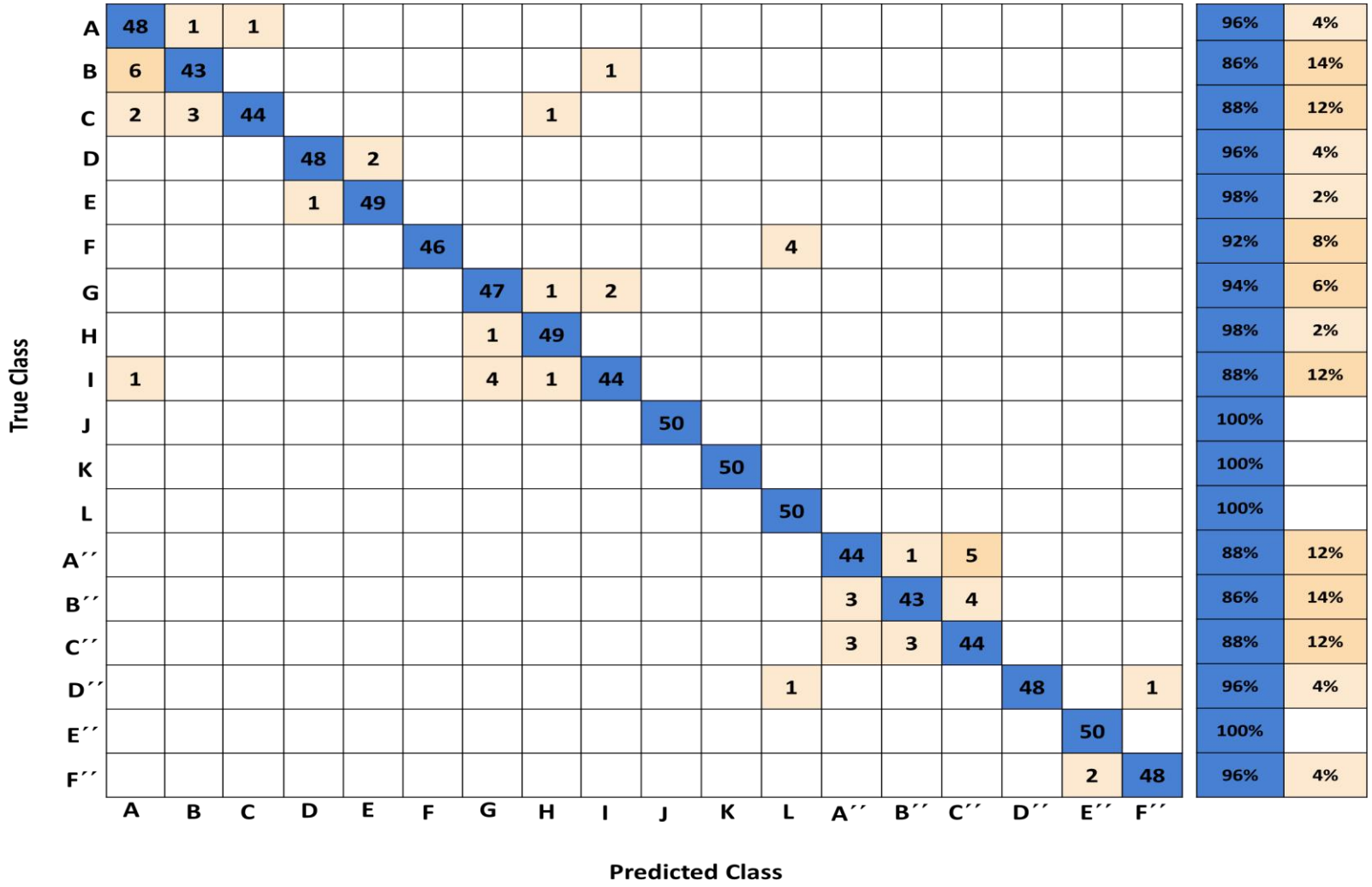


Figure 4.3. Confusion matrix obtained from the stains of various histone/DNA ratios (Calf and Salmon) at two different total mass concentrations. The stains were prepared by depositing 2 μL droplets of histone/DNA complexes, dissolved in an aqueous HEPES buffer solution, onto hydrophobic glass wafers coated with PPX-H. The model parameters were optimized with a global learning rate of 0.001, a minimum batch size of 32 images, and up to 40 epochs.

4.3. Deep Learning-Based Classification of Linker Histone (H1)-DNA Interactions

Developing a simple technique to classify protein-DNA complexes according to their binding affinities would be highly interesting for many applications. The findings presented in this section have been previously published in *Small Science* and can be accessed via https://doi.org/10.1002/smsc.202400252. [261] To assess the ability of deep learning approaches to predict the relative binding affinity of protein-DNA interactions, the composition of stains formed after drying droplets of H1-DNA mixtures was investigated, and their deposition patterns were analyzed using polarized light microscopy (PLM). **Figure 4.4.A** displays a PLM image depicting a typical drying pattern of the histone-Salmon $_{20\ kbp}$ DNA mixture dissolved in a HEPES buffer. Chemical vapor deposition (CVD) polymerization was employed to coat glass surfaces with PPX-H, resulting in uniform hydrophobic substrates with a coating thickness of 50 ± 5 nm. This approach ensured consistent droplet deposition over extensive regions of the surface. [264] Subsequently, a defined volume of 2 µL of the histone-DNA mixture solution was dispensed onto the coated surfaces and left to dry for 50 min under controlled conditions of 40% humidity and 23 °C temperature. In this investigation, solutions containing different ratios of histone (H1) and DNA, dissolved in the same binding buffer consisting of 100 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) buffer (pH 7.8), 150 mM potassium chloride, and 50 mM ammonium sulfate, were employed. [265] To incorporate kosmotropic ions into the buffer solution, potassium chloride and ammonium sulfate were used, which help reduce protein denaturation and increase the salting-out effect, promoting protein-DNA interactions. [266, 267] The deposition patterns from protein-DNA mixtures are compositionally simple, but structurally complex supramolecular systems. These systems are influenced by multiscale processes that are interconnected both locally and temporally. [60] [268] The manipulation of parameters such as the substrate on which the droplet is placed, or the

environmental conditions directly impacts the resulting stain patterns. As the protein-DNA solution evaporates, it becomes saturated, leading to the precipitation of its components alongside the crystallization of salts in the buffer. The deposition of salts and biomolecules begins at the edges of the stain and gradually moves toward the center. To gain deeper insights into the stain patterns, scanning electron microscopy (SEM) was employed. The SEM images revealed high-aspect structures distributed throughout the stain, providing a detailed view that closely matched the features seen in the PLM image, thus improving our understanding of the pattern's characteristics (**Figure 4.4.A**).

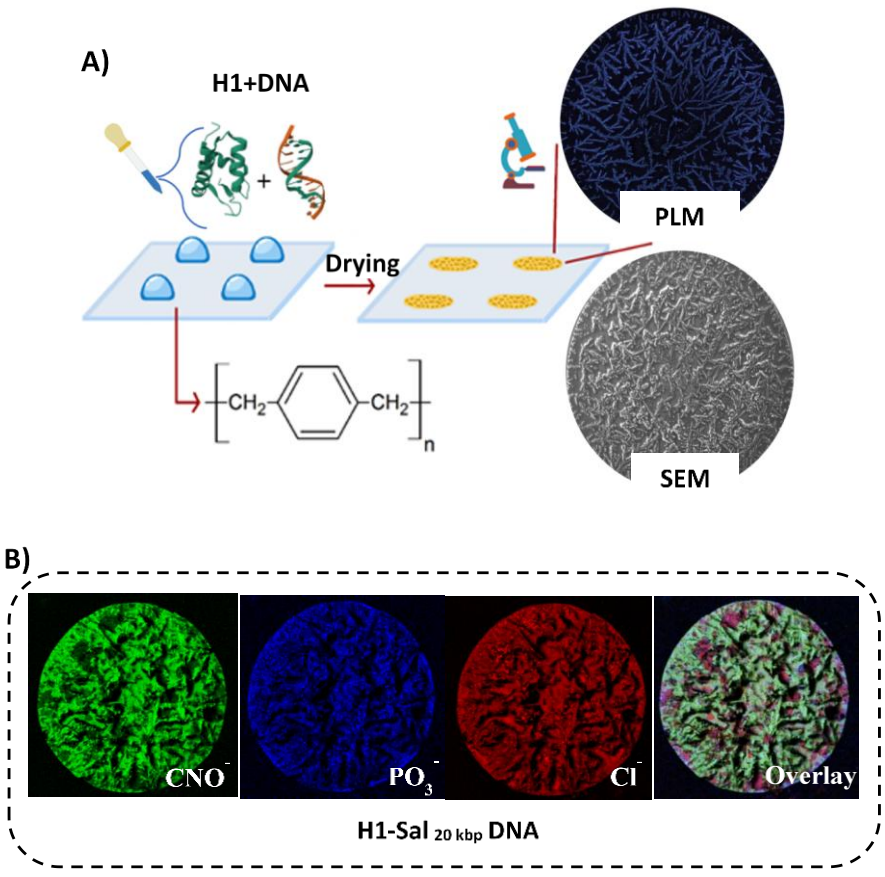


Figure 4.4. The deposition patterns of H1-DNA droplets provide detailed insights into protein-DNA interactions. These stains were created by placing 2 µL droplets of an aqueous HEPES buffer solution onto hydrophobic glass wafers coated with PPX-H A) A schematic of dispensing H1/DNA complex droplets, their drying process, and subsequent imaging. Representative images from PLM and SEM of a dried stain derived from a H1-Sal $_{20\ kbp}$ DNA mixture showcase complex deposition patterns. B) Analysis of the H1-Sal $_{20\ kbp}$ DNA mixture stain using TOF-SIMS imaging indicates the presence of PO_3^- (intensity color scale 0-2 counts), shown in blue. CNO^- fragments, marking the amino acids of histone and the nucleotides of DNA, are displayed in green (intensity color scale 0-30 counts). The distribution of chloride ions from the buffer solution is represented in red (intensity color scale 0-30 counts). The first row shows the RGB channels and their combined overlay. Adapted from [261]

To gain a deeper understanding of the variations in chemical composition within the pattern on the solid surface, time-of-flight secondary-ion mass spectrometry (TOF SIMS) was utilized, as illustrated in **Figure 4.4.B**, which shows the H1-Sal $_{20\ kbp}$ DNA complex stains. The analysis of the separate Sal $_{20\ kbp}$ DNA and H1 patterns can be found in supporting information (**Figure S2-3**). This imaging method is surface-sensitive with a sampling depth of approximately 2 nanometers. The signal from the PO_3^- ion is attributed to the phosphate backbone of the DNA crystal, whereas the CNO^- ion signal is indicative of the presence of both protein (amino acids) and DNA (nucleotides). The Cl^- signal was utilized to track the presence of salt originating from the buffer. The signals of CNO^- and PO_3^-, as well as the one of Cl^- reveal the homogeneous distribution over the entire deposited pattern and thus confirm co-deposition of the protein and DNA.

Based on a previous study, [2] it was hypothesized that the deposited patterns could be utilized to classify diverse types of DNA by discerning their differences in type and size, as well as distinguishing various levels of protein-DNA interactions (**Figure 4.5.A**). To investigate this hypothesis, a range of drying droplet stains generated from four distinct DNA samples, as well as their mixtures with H1, was analyzed (**Figure 4.5.B**), and approximately 400-500 PLM images were collected for each group. It is observed that approximately 100 images can be captured within a 60-minute timeframe. A transfer learning approach utilizing a commercially available deep learning (DL) neural network was employed for pattern analysis, thereby reducing the need for extensive training datasets. [2, 269]

The InceptionV3 network was chosen for its combination of high accuracy on the given dataset and shorter training time compared to other high-performance pretrained CNNs, such as NasNetLarge. [2]

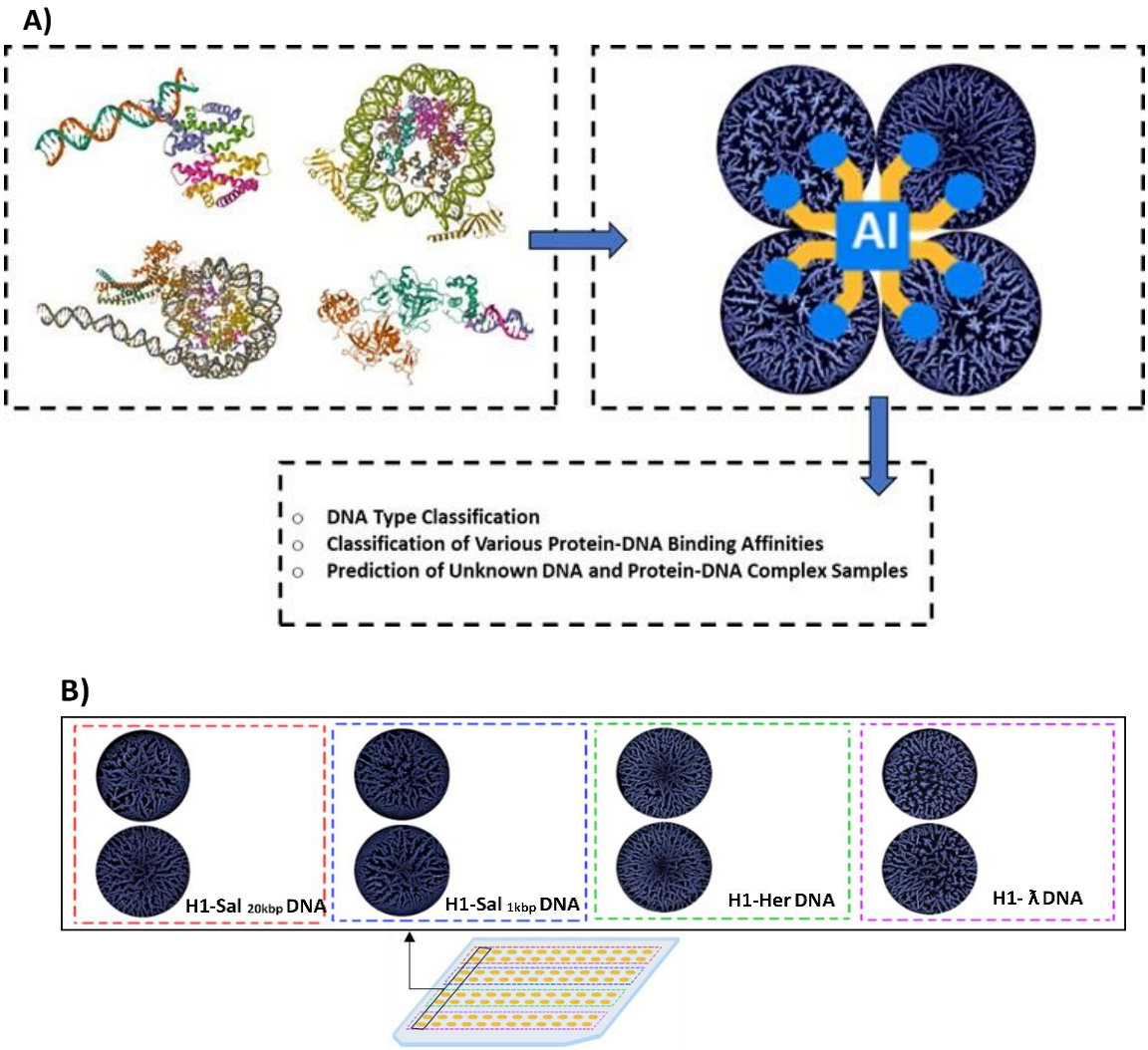


Figure 4.5. Schematic representation of the protein-DNA interaction study using deep learning approaches. A) The study aimed to classify a diverse level of histone-DNA interactions and the prediction of histone binding affinities to unknown DNA based on dried droplet patterns. B) Representative PLM images of H1-DNA stains from four distinct DNA types dissolved in HEPES buffer, deposited onto a CVD-coated glass slide, highlighting the variability in deposition patterns. From left to right: H1-Sal $_{20\ kbp}$ DNA, H1-Sal $_{1\ kbp}$ DNA, H1-Her DNA, and H1-λ DNA. Adapted from [261]

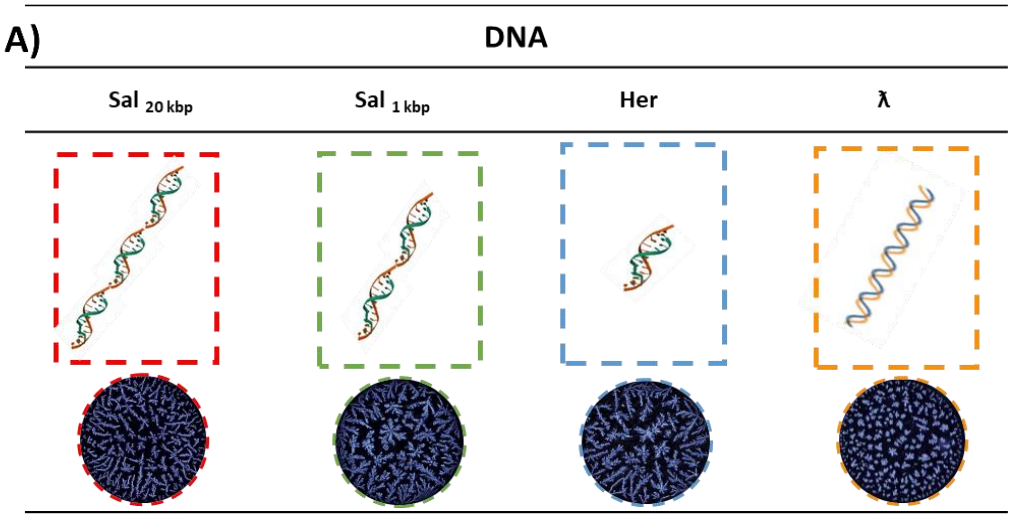
4.3.1 Classification of Various DNA

Figure 4.6.A presents PLM images of four distinct DNA types: Salmon DNA (Sal 20 kbp), Sheared Salmon DNA (Sal 1 kbp), Herring DNA (Her), and Lambda DNA (λ). These DNA types differ in both their genetic composition and molecular weight, influencing their structural and optical properties. Sal $_{20\ kbp}$ DNA, Sal $_{1\ kbp}$ DNA, and Her DNA originate from eukaryotic sources, sharing similarities in genetic content, while λ DNA is of prokaryotic origin,

representing a fundamentally different genomic structure. The variations in molecular weight also contribute to differences in DNA structure, which are evident in the PLM images.

To classify these DNA types based on their PLM images, the InceptionV3 deep learning (DL) network, a convolutional neural network known for its high efficiency in image classification tasks, was employed. The trained model achieved an outstanding accuracy rate of 100% in distinguishing between the four DNA categories, as demonstrated by the confusion matrix in **Figure 4.6.B**. The evaluation was conducted using a well-structured dataset: a combined training and validation set consisting of approximately 1,600 images (400 images per DNA type) and an independent test set of 320 images (80 per category). The test set comprised randomly selected images that were entirely unseen by the network during training, ensuring that the model's performance was evaluated on truly novel data.

The flawless classification performance underscores the robustness of the InceptionV3 model in identifying subtle, yet distinct optical patterns associated with different DNA types. This result highlights the potential of deep learning-based image analysis for DNA characterization and opens avenues for automated, high-throughput screening applications in molecular biology and biomedical research.



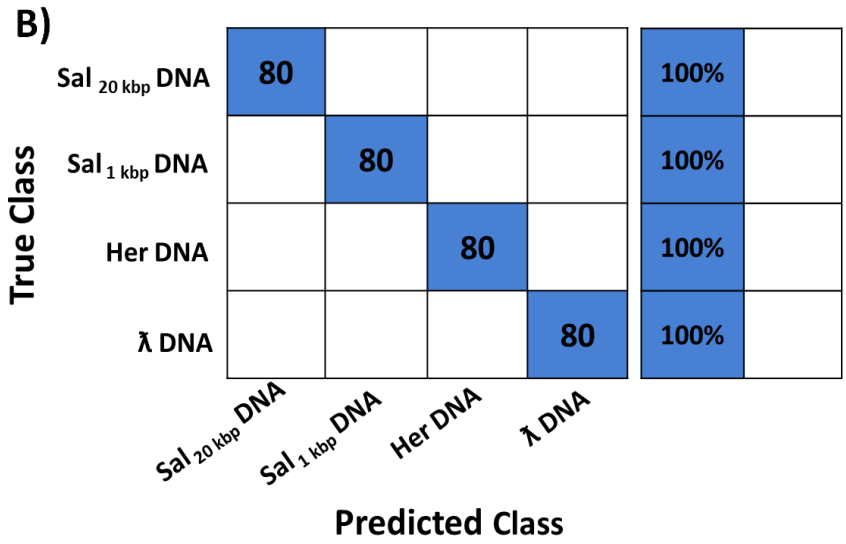


Figure 4.6. DL-based classification of different DNA types. A) PLM images of Sal $_{20\ kbp}$ DNA (outlined in red dashed line), Sal $_{1\ kbp}$ DNA (outlined in green dashed line), Her DNA (outlined in blue dashed line), and λ DNA (outlined in orange dashed line). B) Confusion matrix based on the deposition patterns of four distinct DNA types, varying in both size and type. The model parameters were optimized with a global learning rate of 0.001, a minimum batch size of 32 images, and up to 40 epochs. Adapted from [261]

Gradient-weighted class activation mapping (Grad-CAM) was applied to produce activation maps, highlighting the key features learned in the most informative regions of the PLM images. [270] **Figure 4.7** presents heat map layers of the PLM images, illustrating the deep learning network's focus on the crystalline areas of the stain patterns, rather than the surrounding regions (image background). This selective attention enables the network to differentiate between various DNA types based on their unique deposition patterns, size, and structural features. The Grad-CAM results highlight the network's capability to accurately classify DNA samples, offering valuable insights into the model's decision-making process. These visualizations not only enhance the interpretability of the deep learning model but also provide a more efficient and precise approach to analyzing and distinguishing different DNA types.

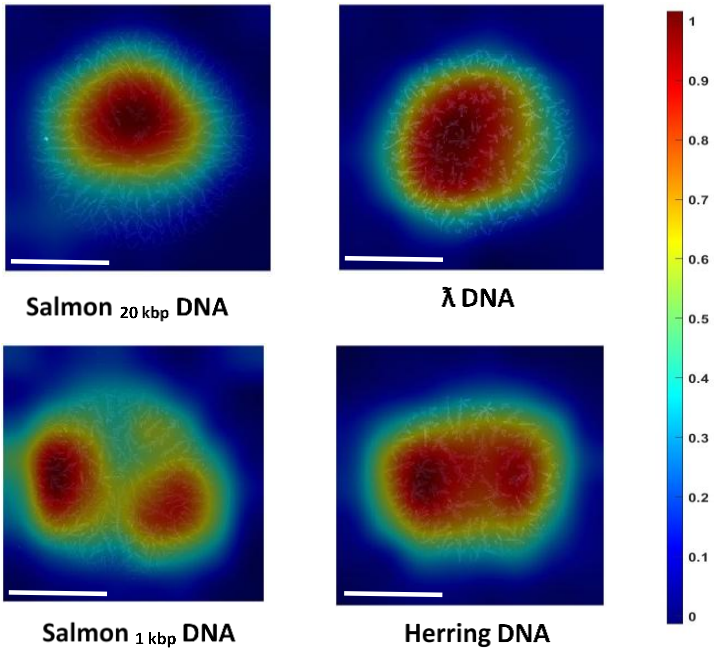


Figure 4.7. Grad-CAM activation maps. Heatmap overlays on the PLM images highlight regions most influential in the model's classification decisions, providing insight into how CNN interprets DNA patterns. Scale bars indicate 1 mm. Adapted from [261]

The t-distributed stochastic neighbor embedding (t-SNE) algorithm [234] (**Figure 4.8**) was applied to the "Softmax" layer of InceptionV3, resulting in clear clustering of four different DNA types. The output of this layer is a 4-dimensional array representing the spatial dimensions (x,y) of the images, along with the image channels and the batch dimension. [271] This observation further confirms that the stain patterns are distinctly separable and indicative, demonstrating their ability to reliably predict deposition patterns of particular DNA type generated under controlled conditions. The t-SNE results further highlight this by clearly clustering the DNA types, showcasing the model's capability to differentiate between them based on their unique deposition characteristics.

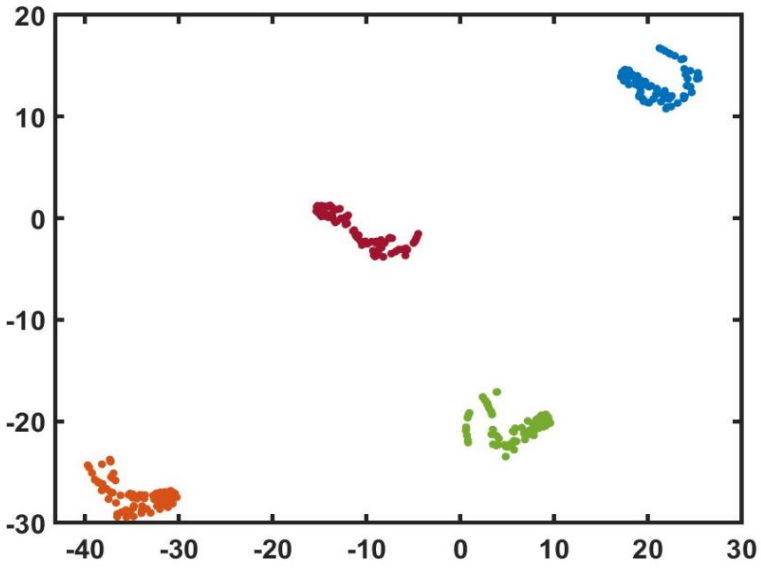


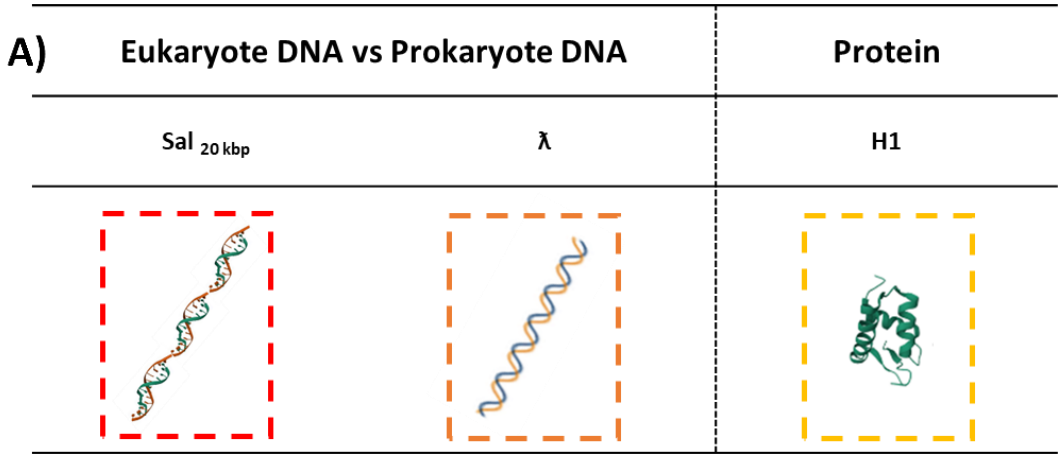
Figure 4.8. The t-SNE plot shows the results from the "Softmax Activation" layer of the trained CNN model. Sal $_{20\ kbp}$ DNA, Sal $_{1\ kbp}$ DNA, Her DNA, and ƛ DNA were represented by the colors red, green, blue, and orange, respectively. Adapted from [261]

4.3.2 Relative Affinity of H1-DNA Interaction-Based on DNA Type (Eukaryotic and Prokaryotic DNA)

The application of DL has emerged as a pivotal tool for comprehensive classification of protein-DNA complexes based on their binding affinity. Histone (H1) is commonly used as a DNA-binding candidate to study interactions with both prokaryotic and eukaryotic DNA. As previously mentioned, experimental studies have shown that H1 binds more strongly to eukaryotic DNA than to prokaryotic DNA. For example, lymphocyte DNA fragments were found to exhibit a binding affinity for H1 that is at least 15 times greater than that of E. coli DNA fragments of the same size, highlighting the selective interaction of H1 with eukaryotic DNA in comparison to prokaryotic DNA. [138, 144] Building on the previously described experimental methods, the InceptionV3 model was applied to classify different protein-DNA mixtures, which differed in DNA type and exhibited distinct binding affinities. [272, 273] Five various H1/DNA ratios (0.5-6.8 mole H1/168 base pairs DNA) were prepared for each DNA type. To achieve a more thorough comprehension, eukaryotic DNA and prokaryotic DNA with

similar number of base pairs, chosen from commercially available DNA samples, were gathered for comparative analysis.

As depicted in **Figure 4.9.A**, Sal $_{20\ kbp}$ DNA (sourced from eukaryotes), forms a double-helix molecule comprising approximately 20 K base pairs. Conversely, Lambda DNA is a linear, double-stranded molecule with around 48 K base pairs, from Escherichia coli bacteriophage. About 480 PLM images were collected for each ratio (400 images for the training and validation sets and 80 images for the testing set). The polarized light images (**Figure 4.9.B**) display distinct patterns in association with eukaryotic and prokaryotic DNA. The patterns that emerged from eukaryotic DNA before and after adding the H1 appear almost similar to the histone patterns. However, the deposited patterns for H1-prokaryotic DNA mixtures show distinct visual characteristics compared to those observed in the patterns of histone alone.



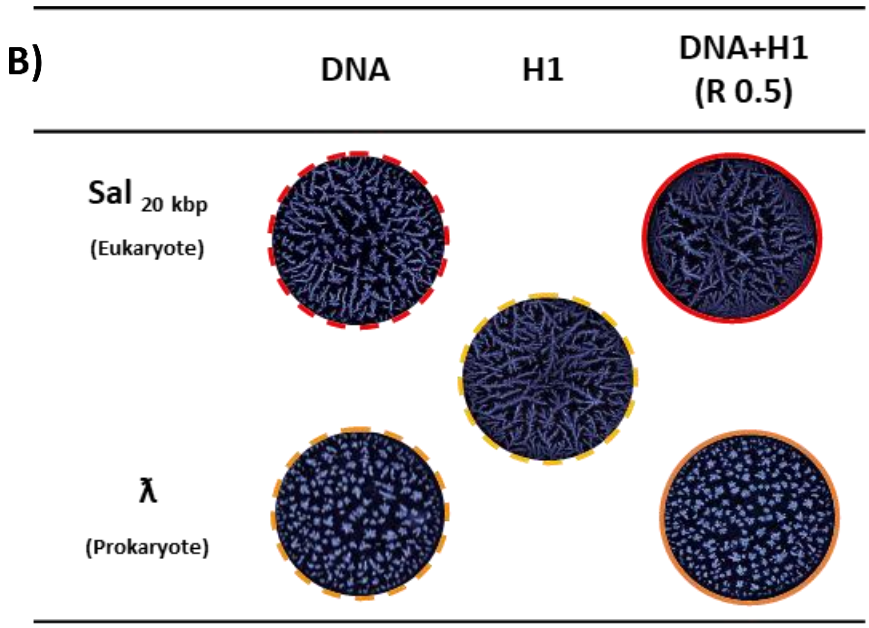


Figure 4.9. Effect of DNA type on H1-DNA binding affinity. A) Schematic representation of Sal $_{20\ kbp}$ (eukaryote) DNA, λ (prokaryote) DNA, and histone (H1). B) PLM images of Sal $_{20\ kbp}$ DNA (without histone), λ DNA (without histone), histone (without any DNA), and each H1-DNA mixture deposition patterns. Adapted from [261]

In **Figure 4.10**, the Inception V3 network classifies similar ratios of H1 with two different DNA types (eukaryotic and prokaryotic). The samples containing Sal $_{20\ kbp}$ DNA, which show a higher binding affinity for H1, reveal that even small quantities of histone can form affinity complexes, as detected by CNN. These changes are not visible to the naked eye but are detectable with CNN. As structural changes become more pronounced, the CNN's ability to distinguish between the deposition patterns of different H1/DNA ratios improves, reflected in the higher prediction accuracy obtained. [2] As depicted in **Figure 4.10**, the average prediction accuracy for various ratios of H1-Sal $_{20\ kbp}$ DNA (eukaryotic) was 99%, surpassing the 93% accuracy observed for H1-λ DNA (prokaryotic).

The CNN-derived confusion matrix indicated that prediction accuracies declined as the histone-to-base pair ratio decreased, reaching their lowest values for R0.5 and R1, which correspond to 0.5 and 1 mol of H1 per 168 base pairs of Sal $_{20\ kbp}$ DNA. According to the results, 2.5% of the stain images from the R0.5 group were incorrectly classified as belonging to the R1 group, while 5% of the R1 group were misclassified as R0.5. Increasing the quantity

of histone in Sal $_{20\ kbp}$ DNA (e.g., R1.4, R3.4, and R6.8) resulted in prediction accuracy rates reaching 100%. In contrast, the mixtures of H1-ƛ DNA (prokaryotic) showed a lower average prediction accuracy of 93%. The prediction accuracies for the R0.5, R1, R1.4, and R3.4 groups were 90%, 85%, 97.5%, and 80%, respectively, reflecting an average dissimilarity of approximately 88% among these groups. A complete prediction accuracy of 100% was only observed at the R6.8 ratio of H1-ƛ DNA. In comparison, the H1- Sal $_{20\ kbp}$ DNA groups already achieved this 100% accuracy at a lower histone concentration with the R1.4 ratio (**Figure 4.10**).

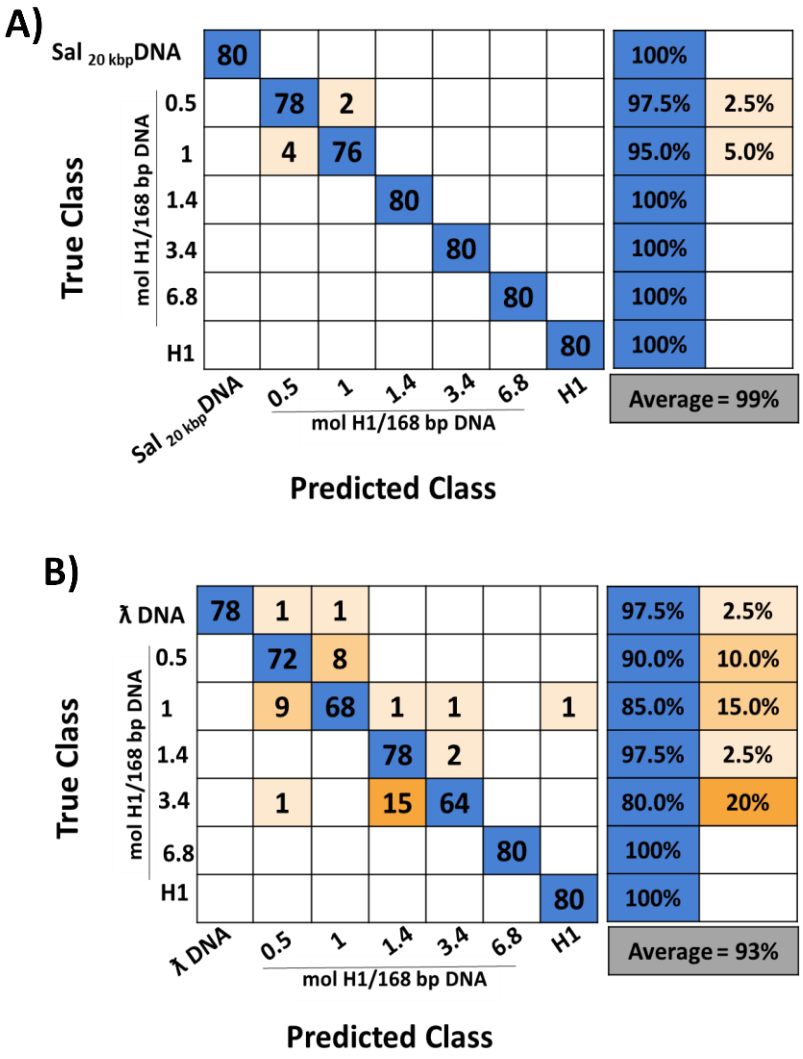
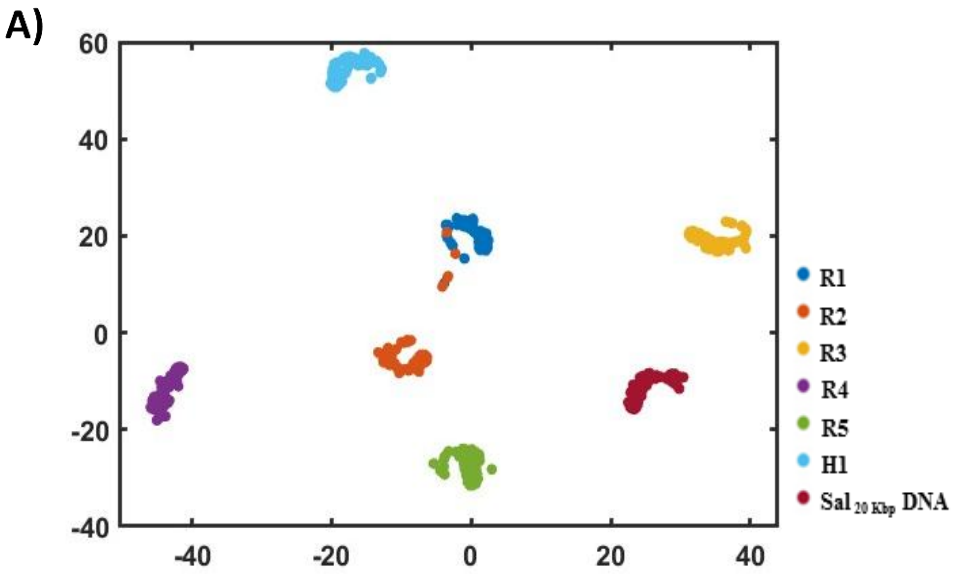


Figure 4.10. Confusion matrices obtained from stain patterns of H1-DNA mixtures, comparing the binding affinity of H1 for eukaryotic versus prokaryotic DNA. Two DNA types were used: A) Sal $_{20\ kbp}$ DNA and B) ƛ DNA, both combined with histone H1. The InceptionV3 model was trained with a global learning rate of 0.001, a minimum batch size of 32 images, and up to 40 epochs. Adapted from [261]

The t-SNE visualization (**Figure 4.11**) revealed distinct clustering among the various groups in H1-Sal $_{20\ kbp}$ DNA, indicating that the samples within this category exhibit well-defined and separable features. In contrast, no such clustering was observed among the different groups of H1-λ DNA, implying a higher degree of overlap or similarity in their features. This clear clustering in H1-Sal $_{20\ kbp}$ DNA suggests a stronger binding affinity between H1 and Sal $_{20\ kbp}$ DNA, as better-defined clusters indicate more consistent and distinguishable interactions at different ratios. The higher the binding affinity, the more pronounced the clustering between different H1-DNA ratios, further supporting the notion that binding affinity-driven structural variations contribute to the classification performance.



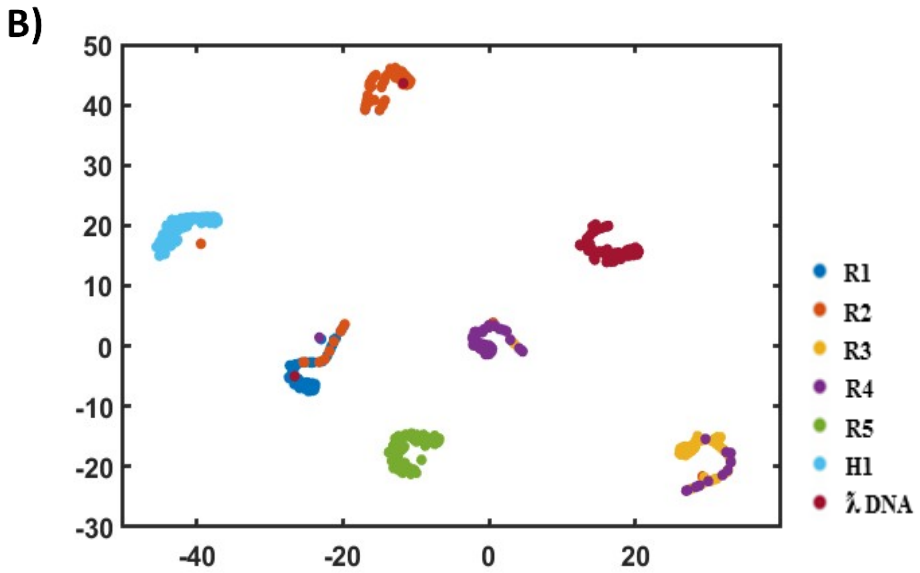


Figure 4.11. The t-SNE plots show the clustering of H1-eukaryote DNA vs H1-prokaryote DNA. The visualization of the 'Softmax' layer in the trained CNN model reveals distinct clusters associated with A) H1- Sal $_{20\ kbp}$ DNA and B) H1- λ DNA. Adapted from [261]

To validate the results of the CNN analysis, an ethidium bromide displacement experiment was performed. The results of this experiment confirm the higher binding affinity of H1-Sal $_{20\ kbp}$ DNA compared to H1-λ DNA, as indicated by fluorescence quenching of EtBr. Ethidium bromide typically interacts with DNA via a molecular intercalation mechanism. When DNA is associated with a strong binder, it causes the displacement of ethidium bromide from the DNA. This displacement leads to a decrease in fluorescence, due to deactivation of free EtBr by proton transfer from the excited singlet to water. [262] The findings from this assay, illustrated in **Figure 4.12**, indicate that the addition of H1 to a pre-incubated solution of Sal $_{20\ kbp}$ DNA-EtBr led to a more significant displacement of ethidium bromide from the complex than observed with the λ DNA-EtBr solution. This suggests that H1 exhibits a higher binding affinity for Sal $_{20\ kbp}$ DNA compared to λ DNA, which aligns with the conclusions drawn from the CNN analysis. The higher the average prediction accuracy for a DNA type, the greater the EtBr fluorescence quenching, suggesting a relatively higher binding affinity of that DNA type with H1.

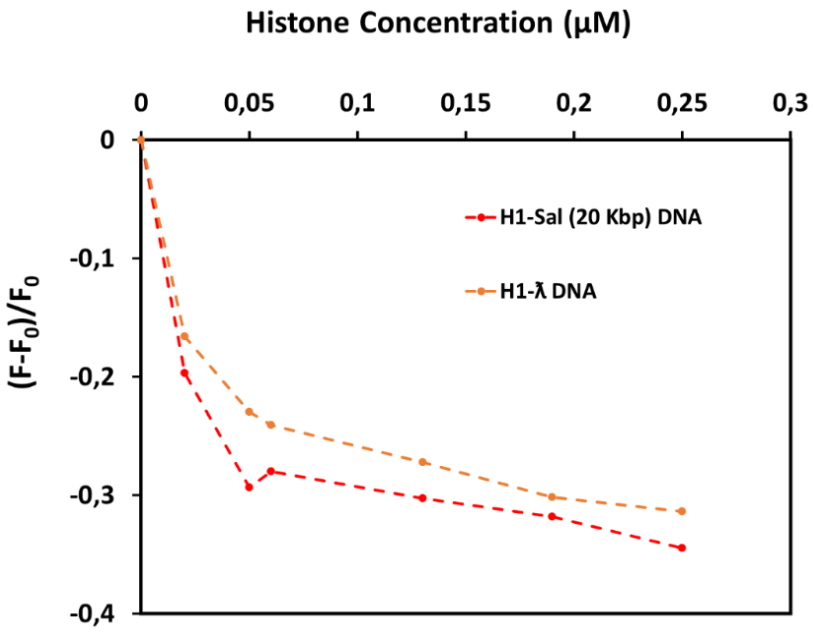


Figure 4.12. Ethidium bromide-DNA complex displacement assay. Each data point represents the average of samples obtained from two distinct experiments. The decrease in relative fluorescence intensity of the EtBr-Sal $_{20\ kbp}$ DNA (red dash-line), and EtBR- λ DNA (orange dash-line) complexes is a result of the interaction between H1 and each DNA. Adapted from [261]

4.3.3 Relative Affinity of H1-DNA Interaction-Based on DNA Size

DNA molecules with longer base pairs present more binding sites for H1, which raises the chances of H1-DNA interactions, and results in a tighter chromatin compaction. In fact, the length of DNA can impact on the distribution and density of H1 binding.

The study of Renz [138] revealed that the binding affinity of H1 differs among DNA fragments of the same origin but varying lengths. According to this study, H1 showed a higher affinity for longer eukaryotic DNA fragments compared to shorter fragments, as they compete more effectively for their binding. Moreover, Aviles et al. [274] showed that H1 displays a stronger binding affinity for high molecular weight Calf thymus DNA than for sonicated Calf thymus DNA, which has been sheared into shorter fragments. Alterations in the molecular weight of DNA can affect the binding affinity between DNA and proteins, potentially leading to modifications in the physical and chemical characteristics of the resulting protein-DNA complex. [141]

In this study, the Inception V3 model was utilized to classify the binding affinity between H1 and DNA, focusing on variations in DNA size. To have a better comparison, three different sizes of eukaryotic DNA, including Sal $_{20\ kbp}$ DNA, Sal $_{1\ kbp}$ DNA, and Herring (Her) DNA (50 bp) were gathered. **Figure 4.13.A** presents a schematic overview of the three different DNA lengths, while **Figure 4.13.B** provides PLM images of the droplet stain patterns formed by each DNA sample, H1 protein, and their respective complexes.

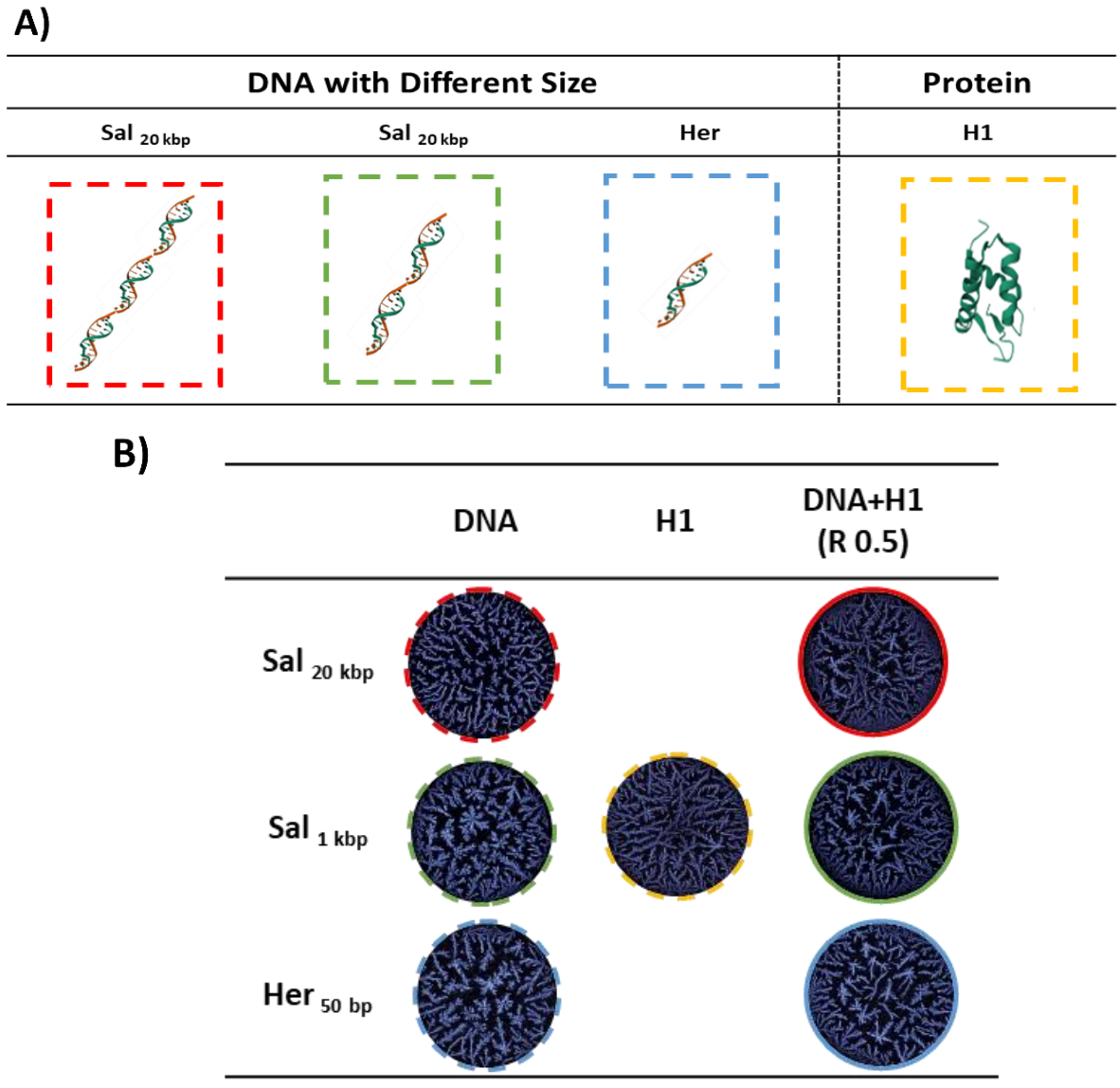
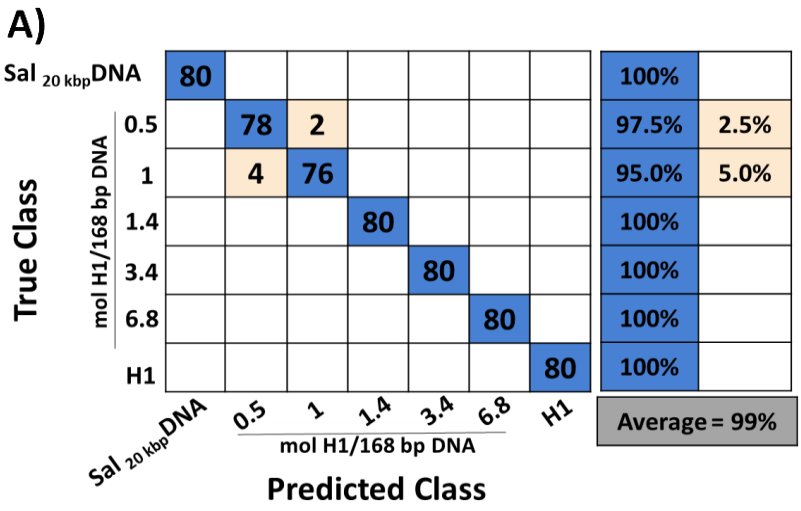


Figure 4.13. Effect of DNA fragments size on H1-DNA binding affinity. A) Schematic representation of Sal $_{20\ kbp}$ DNA, Sal $_{1\ kbp}$ DNA, Her DNA, and H1. B) PLM images of each DNA (without histone), histone (without DNA), and each H1-DNA mixture deposition patterns. Adapted from [261]

Using the training dataset of approximately 360 PLM images, a validation set of 40 images, and a test set of 80 images per group (**Figure 4.14**), we observed that the average classification accuracy from different ratios of H1-Sal $_{20\ kbp}$ DNA was notably higher compared to the prediction accuracy for H1-Sal $_{1\ kbp}$ DNA and H1-Her DNA complexes. This trend persisted even when the histone-to-DNA ratios were comparable across the groups, suggesting that the longer DNA fragment size (Sal $_{20\ kbp}$ DNA) enhances the model's ability to differentiate the complexes more effectively than the shorter DNA fragments. With the image set that had not been previously encountered by the DL network, InceptionV3 recorded 1%, 9% and 15% of total misclassification for H1-Sal $_{20\ kbp}$ DNA, H1-Sal $_{1\ kbp}$ DNA, and H1-Her DNA, respectively. At H1- Sal $_{20\ kbp}$ DNA ratios of 1.4 and above, the resulting stain patterns were highly distinguishable. In contrast, for shorter DNA fragments such as Sal $_{1\ kbp}$ and Her DNA, the stain patterns displayed a greater level of similarity across different ratios, resulting in lower average prediction accuracies. These similarities imply that for Sal $_{1\ kbp}$ DNA and Her DNA, a high concentration of histone is required to induce significant pattern alterations (leading to increased interactions and tighter DNA compaction), consequently leading to maximum classifications.



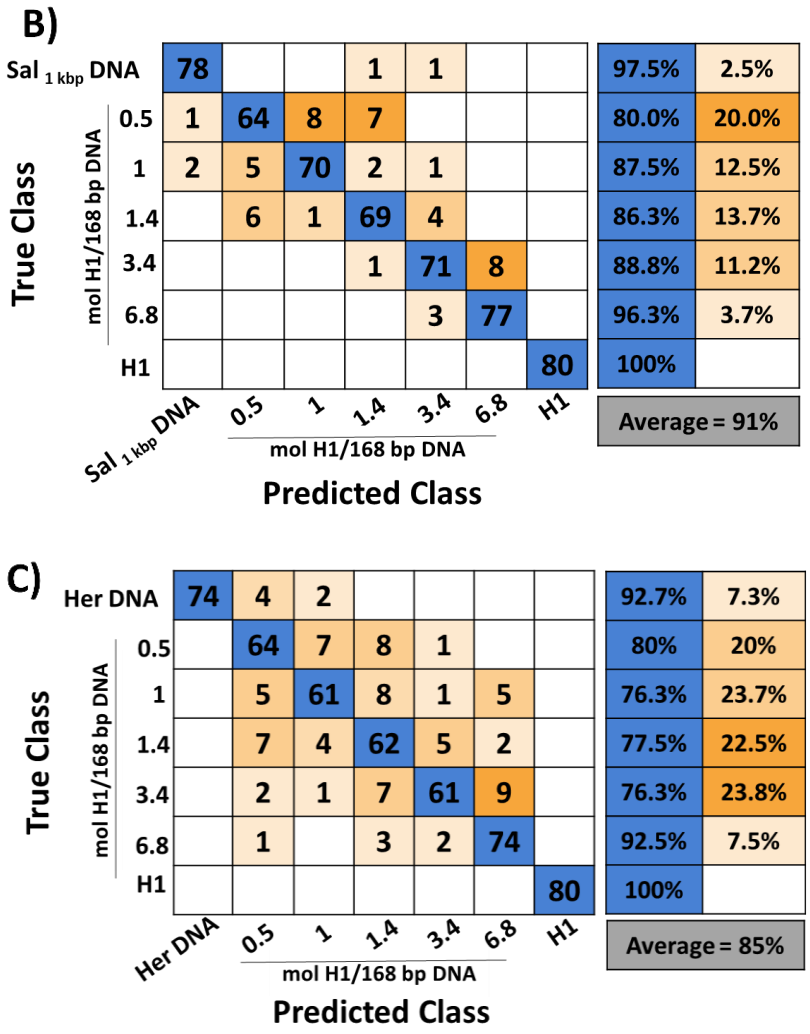


Figure 4.14. Confusion matrices derived from stain patterns of H1-DNA mixtures, comparing the binding affinity of H1 with eukaryotic DNAs differing in fragment sizes. Three eukaryotic DNA lengths were tested: A) 20 kbp, B) 1 kbp, and C) 50 bp, all combined with histone H1. The InceptionV3 model was trained with a global learning rate of 0.001, a minimum batch size of 32 images, and up to 40 epochs. Adapted from [261]

The t-SNE algorithm revealed clear clustering for H1-Sal $_{20\ kbp}$ DNA, moderate clustering for H1-Sal $_{1\ kbp}$ DNA, and the least clustering for H1-Her DNA (**Figure 4.15**). This pattern indicates that higher binding affinity corresponds to more distinct clustering, as observed for H1-Sal $_{20\ kbp}$ DNA, while lower binding affinity, as seen in H1-Her DNA, results in greater overlap and less-defined clusters.

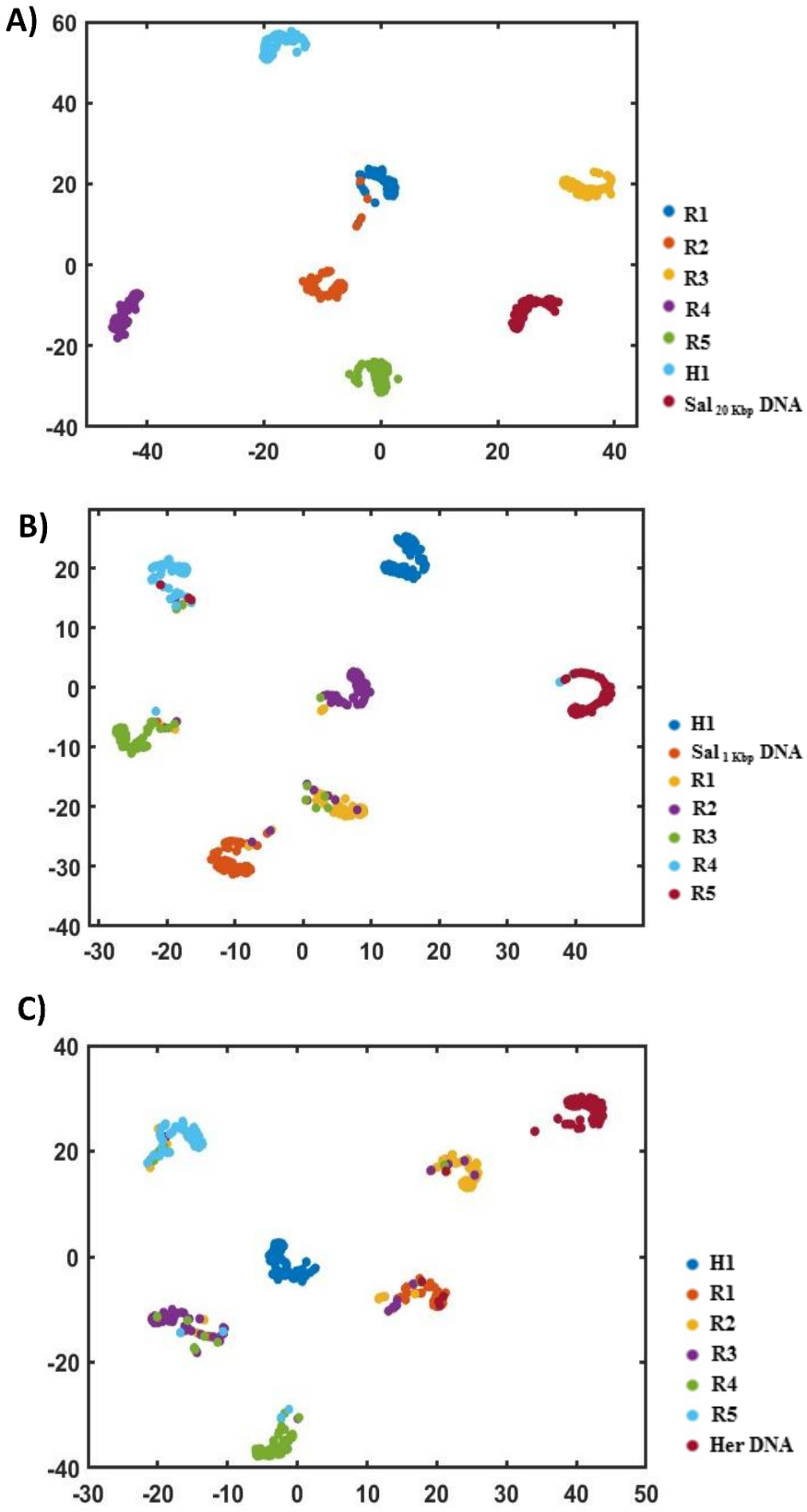


Figure 4.15. The t-SNE plots show the clustering of H1 interactions with eukaryotic DNA of varying fragment sizes. Visualization of the 'Softmax' layer in the trained CNN model highlights separate groupings corresponding to each DNA size: A) H1-Sal $_{20\ kbp}$ DNA, B) H1-Sal $_{1\ kbp}$ DNA, and C) H1-Her DNA, indicating a distinct separation between these clusters. Adapted from [261]

The binding affinity trend was confirmed through the ethidium bromide displacement experiment previously discussed. (**Figure 4.16).** In this experiment, H1-Sal $_{20\ kbp}$ DNA exhibited the highest binding affinity, as shown by the most significant displacement of ethidium bromide from the DNA complex and corresponding fluorescence quenching. This was followed by H1-Sal $_{1\ kbp}$ DNA, which demonstrated a moderate displacement of ethidium bromide, reflecting a moderate binding affinity. Lastly, H1-Her DNA showed the weakest interaction with histone I, as evidenced by the least displacement of ethidium bromide and minimal fluorescence quenching. These results further support the notion that DNA fragment size plays a crucial role in determining histone I binding affinity.

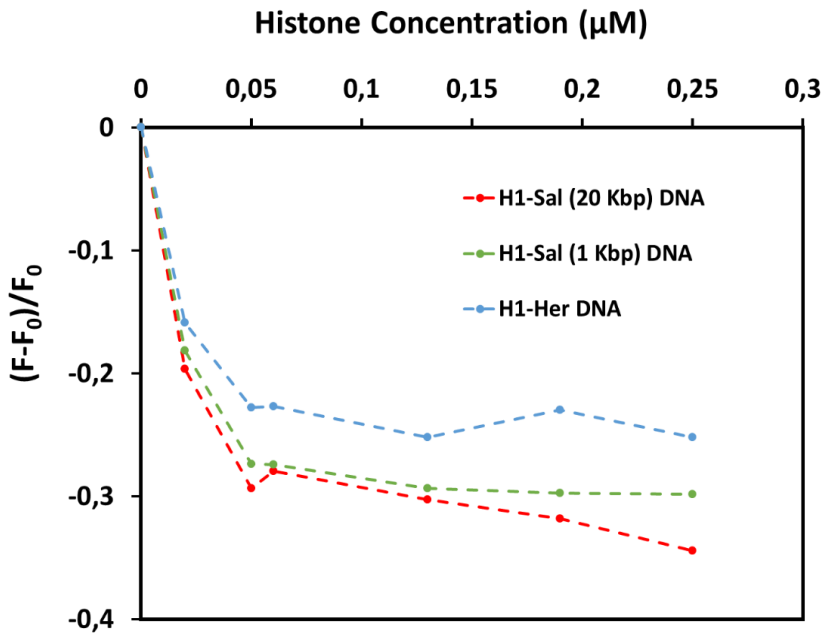


Figure 4.16. Ethidium bromide-DNA complex displacement assay. Each data point represents the average of samples obtained from two distinct experiments. The decrease in relative fluorescence intensity of the EtBr-Sal $_{20\ kbp}$ DNA (red dash-line), and EtBr-Sal $_{1\ kbp}$ DNA (green dash-line), and EtBr-Her DNA (blue dash-line) complexes is a result of the interaction between H1 and each DNA. Adapted from [261]

4.3.4 Stratification of Unknown Histone-DNA Interactions

To evaluate the performance of the trained neural network, we prepared images of unknown DNA, as well as H1-unseen/unknown DNA mixture samples that were not included in the

original network's training set. To achieve this purpose, the neural network was trained using 25 unique (approximately ten thousand PLM images) sample sets, each representing different H1-DNA ratios. These samples included varying proportions of H1 paired with four distinct DNA types: Sal $_{20\ kbp}$ DNA, Sal $_{1\ kbp}$ DNA, Herring (Her) DNA, and Lambda (ƛ) DNA. Additionally, control samples consisting of pure DNA without H1, as well as H1 alone without any DNA, were incorporated into the original training set. This diverse dataset provided the neural network with a comprehensive range of interaction patterns, enabling it to better recognize and classify unknown histone-DNA binding affinity across different DNA types and sizes. The pre-trained convolutional neural network (CNN) was utilized to predict the binding affinities between H1 and a range of unseen/unknown DNA samples based on their unique and unseen patterns. **Figure 4.17.A** depicts the workflow for the classification of unknown samples, which involves several steps, including collection of data, feature extraction and training and evaluation of the model.

Figure 4.17.B illustrates that the pre-trained network successfully predicted the binding affinities of DNA from the species that were part of the training set. A prediction accuracy of 100% was achieved, even though the CNN had not seen the images, as they were not part of the training set.

Moreover, the binding affinity of H1-Salmon Sperm $_{200\ bp}$ DNA (commercially available) was examined. The trained network accurately classified and predicted the binding affinity of unknown samples with 100% accuracy in eukaryotic groups. The DNA samples in both the unknown and predicted groups were derived from the same source, differing only in the length of their fragments. Consequently, the unknown samples were predominantly grouped into the closest related category, specifically identified as H1-Sal $_{1\ kbp}$ DNA groups. As shown in **Figure 4.17.B** for the H1-Salmon Sperm $_{200\ bp}$ DNA at ratio R3.4, 92.5% of images were classified into the H1- Sal $_{1\ kbp}$ DNA group (medium binding affinity), 5% of stain

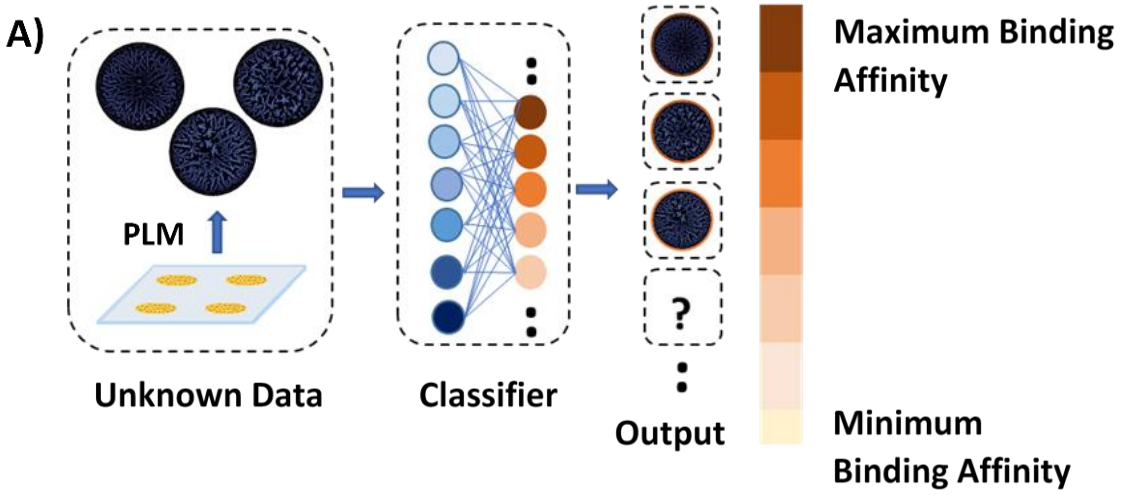
patterns were categorized as H1- Sal $_{20\,kbp}$ DNA (strong binding affinity), and 2.5% of stain patterns were classified into the H1-Her DNA group (weak binding affinity). Furthermore, the staining patterns of another ratio (R6.8) of H1-Salmon Sperm $_{200\,bp}$ DNA were examined using the trained network. The trained network successfully classified 100% of unknown images into the most similar group based binding affinity (medium binding affinity).

The classification results for images of H1-Calf DNA showed that the majority of the samples (84.4%) were identified as strong binders, while another 5.6% fell into the medium binder category. Notably, despite the CNN never having been trained on stain patterns containing Calf DNA, it accurately categorized 96% of these patterns as eukaryotic DNA, signifying specific H1 binding. Conversely, only 4% were misclassified as prokaryotic DNA (non-specific binding). This high classification accuracy is a promising indicator of the robustness of the deep learning model. The ability to classify binding affinity of unknown samples with such precision demonstrates the power of the DL approach, offering a significant advantage over traditional methods, which often rely on manual interpretation and are resource-intensive. This approach not only provides a simple and fast solution for classifying histone-DNA interactions but also delivers results with a high level of accuracy, reducing the time and effort typically required for extensive experimental analyses. Furthermore, the model can be easily adapted to classify new, previously unseen data, making it a highly flexible tool for rapid and scalable analysis of protein-DNA interactions.

Moreover, the impact of the test set sample size on prediction accuracy was investigated (**see Figure S4**). The CNN achieved a classification rate of 90% for Calf DNA samples as strong binders when using just ten images. As more images were added (specifically 50, 80, and 100), the average prediction accuracy improved slightly, reaching only 91%. These results suggest that reliable and accurate outcomes can be obtained with as few as ten images, highlighting the effectiveness and feasibility of this machine learning technique for rapid

analysis. This finding is particularly significant in situations where sample availability is limited or when there is a need for fast, initial predictions. Furthermore, the minimal increase in accuracy with the addition of more samples suggests that the CNN has successfully learned the key features of histone-DNA interactions in the training process, emphasizing the potential for cost-effective and time-saving applications in high-throughput analysis. This underscores the practical advantages of applying deep learning models in various experimental settings, where quick and reliable predictions can significantly reduce the need for extensive data collection and processing.

A)



B)

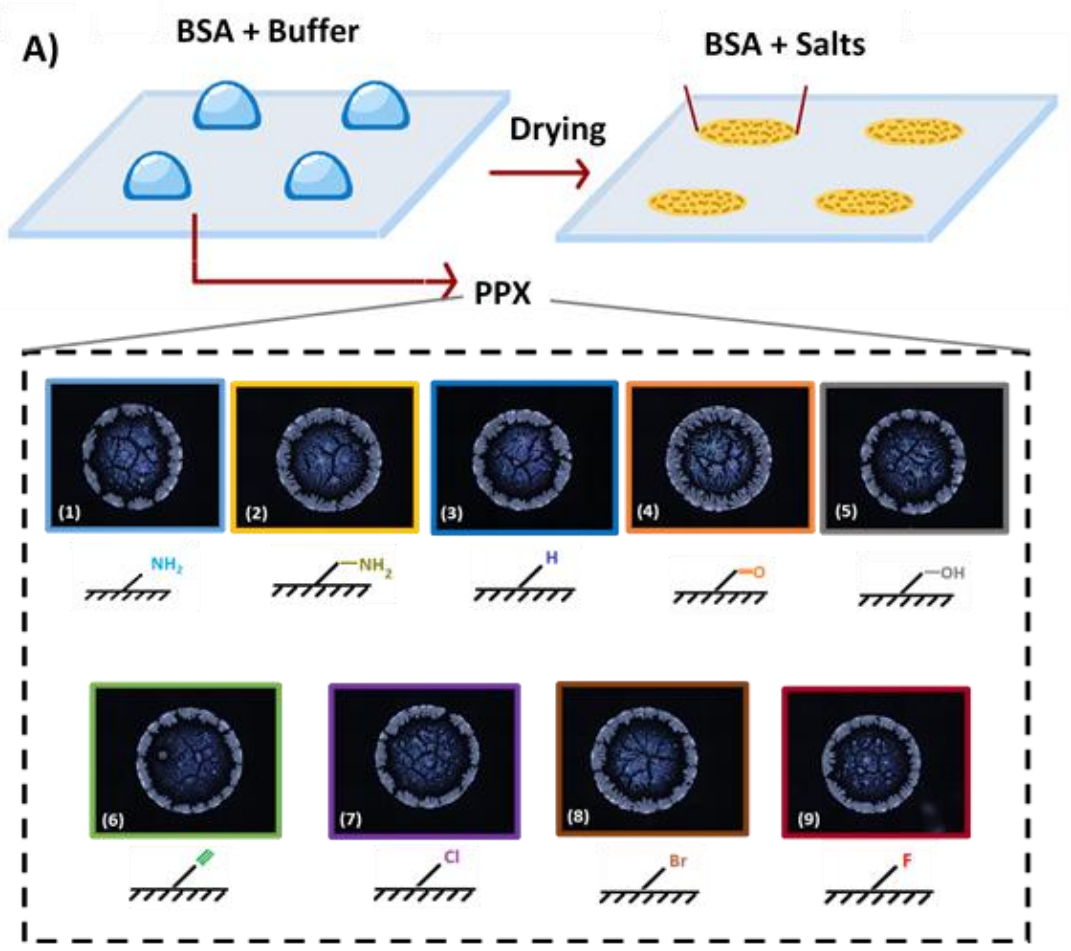
	Group	True Classes			
		Strong (Sal $_{20\ kbp}$ DNA)	Medium (Sal $_{1\ kbp}$ DNA)	Weak (Her DNA)	Non-Specific (λ DNA)
Unseen Samples	H1/Sal $_{20\ kbp}$ DNA = 3.4	100%	0%	0%	0%
	H1/Sal $_{1\ kbp}$ DNA = 3.4	0%	100%	0%	0%
	H1/Her DNA = 3.4	0%	0%	100%	0%
	H1/λ DNA = 3.4	0%	0%	0%	100%
Unknown Samples	H1/Sal $_{200\ bp}$ DNA = 3.4	5%	92.5%	2.5%	0%
	H1/Sal $_{200\ bp}$ DNA = 6.8	0%	100%	0%	0%
Unknown Samples and Ratios	H1/Calf DNA = 0.2	94%	2%	4%	0%
	H1/Calf DNA = 0.3	86%	6%	0%	8%
	H1/Calf DNA = 0.6	66%	12%	22%	0%
	H1/Calf DNA = 16	82%	6%	4%	8%
	H1/Calf DNA = 47	94%	2%	0%	4%

Figure 4.17. Evaluation of the pre-trained network's performance in classifying H1-DNA binding affinities using new species. A) Overview of the approach used to classify unknown samples. B) Prediction results for unseen or unknown H1-DNA samples. Adapted from [261]

4.4. Deep Learning-Based Surface Classification of Functional Polymer Coatings

The results presented in this chapter have been published in *Langmuir,* and can be accessed via https://doi.org/10.1021/acs.langmuir.4c03971. [183] To achieve uniform and functionalized substrates, chemical vapor deposition (CVD) polymerization of poly(p-xylylene) (PPX) was employed. In this study, the scientific scope was intentionally limited to poly(para-xylylene) (PPX) surfaces. The reason is that these polymers share identical polymer backbones but differ in their functional side groups. As the polymer surfaces become more diverse, the patterns should be even more characteristic. Additional aspects or selecting PPX substrates for this study include properties such as low roughness, homogeneous coating, transparency, and stability against water, ensuring consistent and reliable data acquisition. Ten different PPX films were deposited onto glass wafers, as shown in **Figure 4.18**, with coating thicknesses ranging from 50 to 60 nm. A defined volume of 2 µL of an aqueous bovine serum albumin (BSA) solution in buffer were deposited onto each coated surface and left to dry under controlled environmental conditions (40% humidity, 23 °C) for 45 ± 5 minutes. During the drying process of the BSA solution, as saturation is reached, the protein components start to precipitate, beginning at the droplet's periphery and gradually advancing toward the center. This process leads to the development of characteristic stain patterns. **Figure 4.18.A** presents polarized light microscopy (PLM) images illustrating typical drying patterns of BSA solution on nine different PPX-coated surfaces: PPX-Amine (A), PPX-Aminomethyl (AM), PPX-Hydrogen (H), PPX-Aldehyde (CHO), PPX-Hydroxymethyl (CH_2OH), PPX-Alkyne (Alk), PPX-Chloride (Cl), PPX-Bromide (Br), and PPX-Fluoride (F). As it is clear from **Figure 4.18.A**, dried BSA patterns on each surface were indistinguishable to the naked eye. To ensure an unbiased classification of the polymer surfaces, the InceptionV3 network was pre-trained using 400-500 images per group. InceptionV3 was selected for its superior accuracy and faster

training time compared to other advanced convolutional neural networks (CNNs), such as NasNetLarge [58, 275]. The training and validation sets comprised 85% of the total images, while 100 images (15% of the total) were reserved for the test set, which remained unseen by the trained network. The trained network was employed to accomplish two main goals: i) stratification of subtle variations in surface polymer chemistries, specifically polymers with identical backbone structures but differing functional groups, by analyzing dried BSA patterns, and ii) predict and categorize unknown surfaces using the trained network (**Figure 4.18.B**).



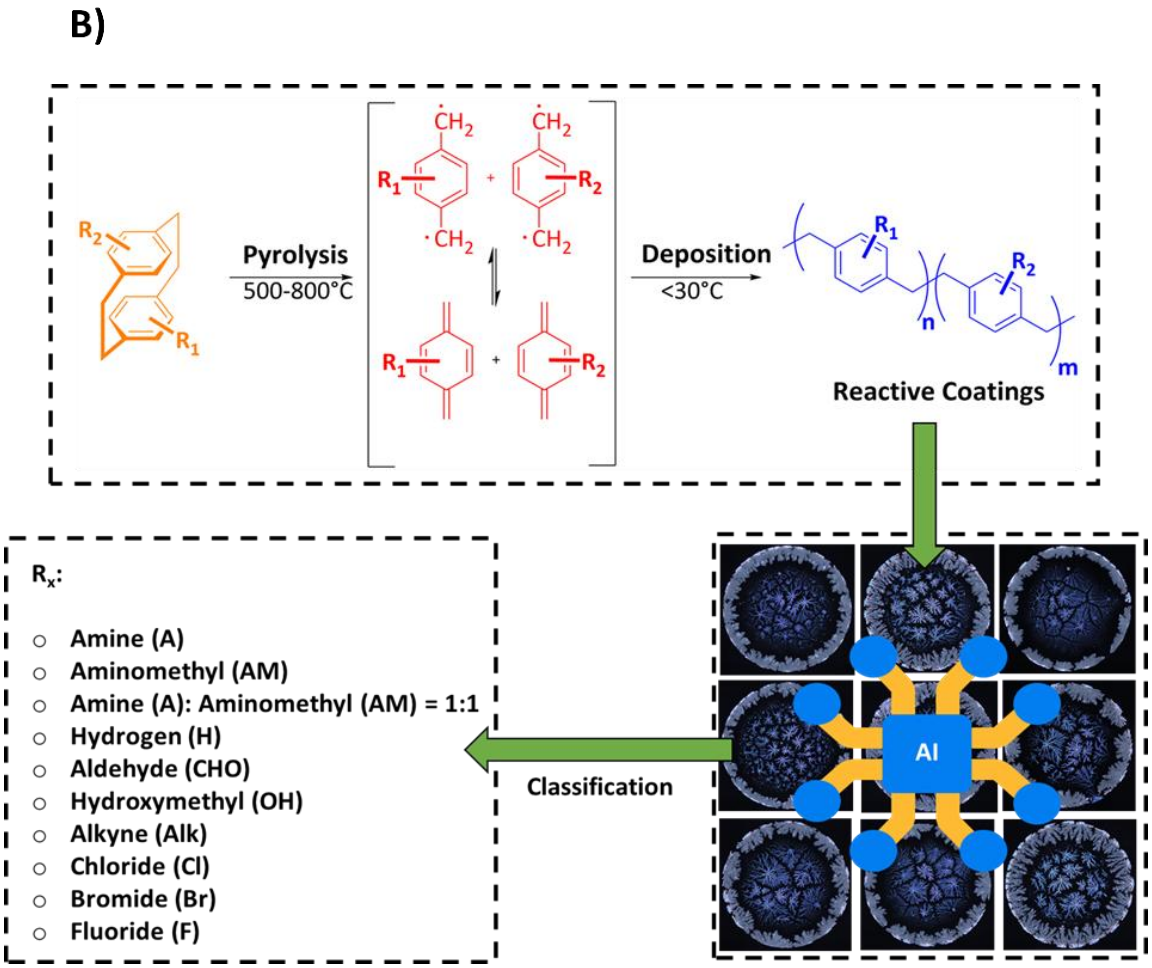


Figure 4.18. Classification of surface polymer chemistries through image analysis of deposition droplets using a deep learning (DL) approach. A) Representative polarized light microscopy (PLM) images showing deposition patterns of BSA on nine PPX-coated glass wafers. The stains were created by depositing 2 µL droplets of 0.1 mg/mL BSA dissolved in an aqueous 100 mM carbonate-bicarbonate buffer with 50 mM sodium chloride (pH=9.2). B) A schematic illustration of the deep learning-based recognition of polymer surfaces, alongside the chemical reaction scheme depicting the CVD polymerization of ten different PCPs utilized in this study. The aim of the study was to classify a wide range of functionalized CVD coatings and to predict unknown surface chemistries based on patterns formed by dried droplets. Adapted from [183]

4.4.1 Surface Characterization

To thoroughly investigate the physicochemical properties of the ten polymer-coated surfaces, a multi-technique approach was employed, utilizing Time-of-Flight Secondary Ion Mass Spectrometry (ToF-SIMS), Infrared Reflection-Absorption Spectroscopy (IRRAS) (**Figure S5-S6**), and contact angle measurements. ToF-SIMS provided detailed molecular and elemental surface composition data, enabling the identification of characteristic functional

groups and potential contaminants. Contact angle measurements were conducted to assess the wettability and hydrophobicity of the coatings, offering insights into their surface energy and interactions with aqueous environments. Additionally, IRRAS was used to analyze the vibrational signatures of functional groups, allowing for the confirmation of chemical modifications and variations among the polymer coatings. The combined results from these techniques provided a comprehensive understanding of the surface properties and confirmed that the CVD polymerization process was successfully performed, ensuring the presence of the corresponding functional groups on each surface.

Surface Chemistry Analysis by ToF-SIMS

Figure 4.19 shows the ToF-SIMS analysis of PPX coatings with different functionalization. All functional groups, except PPX-Alk, displayed distinct heteroatoms detectable by this method. The halogenated polymers, PPX-F, PPX-Cl, and PPX-Br were each identified by their unique F^-, Cl^-, and Br^- fragments, respectively. PPX-CHO and PPX-CH_2OH were distinguished by the presence of oxygen-containing fragments, with the most pronounced O^- signal observed for PPX-CHO and the strongest CH_2OH^+ signal for PPX-CH_2OH. It is important to note that low-level oxygen signals can be ambiguous, as they may also appear in other PPX films due to the interaction of molecular oxygen with free radicals during post-polymerization. PPX-A and PPX-AM were identified through nitrogen-containing fragments appearing as CN^+ with similar intensities, although PPX-AM showed a higher intensity of the $CH_2NH_2^+$ signal due to its additional methylene group. The full spectra across the 10-85 m/z range are available in supplementary information (**Figure S5**). Overall, the ToF-SIMS analysis confirmed the presence of the intended functional groups across the modified PPX surfaces. Each polymer exhibited characteristic ion fragments correlating with its chemical composition, enabling differentiation between halogenated, oxygenated, and aminated variants. These findings reflect the successful chemical modification of the PPX films. Similarly, Infrared

Reflection Absorption Spectroscopy (IRRAS) was utilized to confirm the existence and characteristics of the functional groups corresponding to each coated surface, as shown in **Figure S6**. The IRRAS spectra exhibited distinctive vibrational bands corresponding to the chemical functionalities introduced during surface modification. These results, alongside the ToF-SIMS findings, validated the successful incorporation of the desired functional groups, confirming the chemical identity of the modified PPX films.

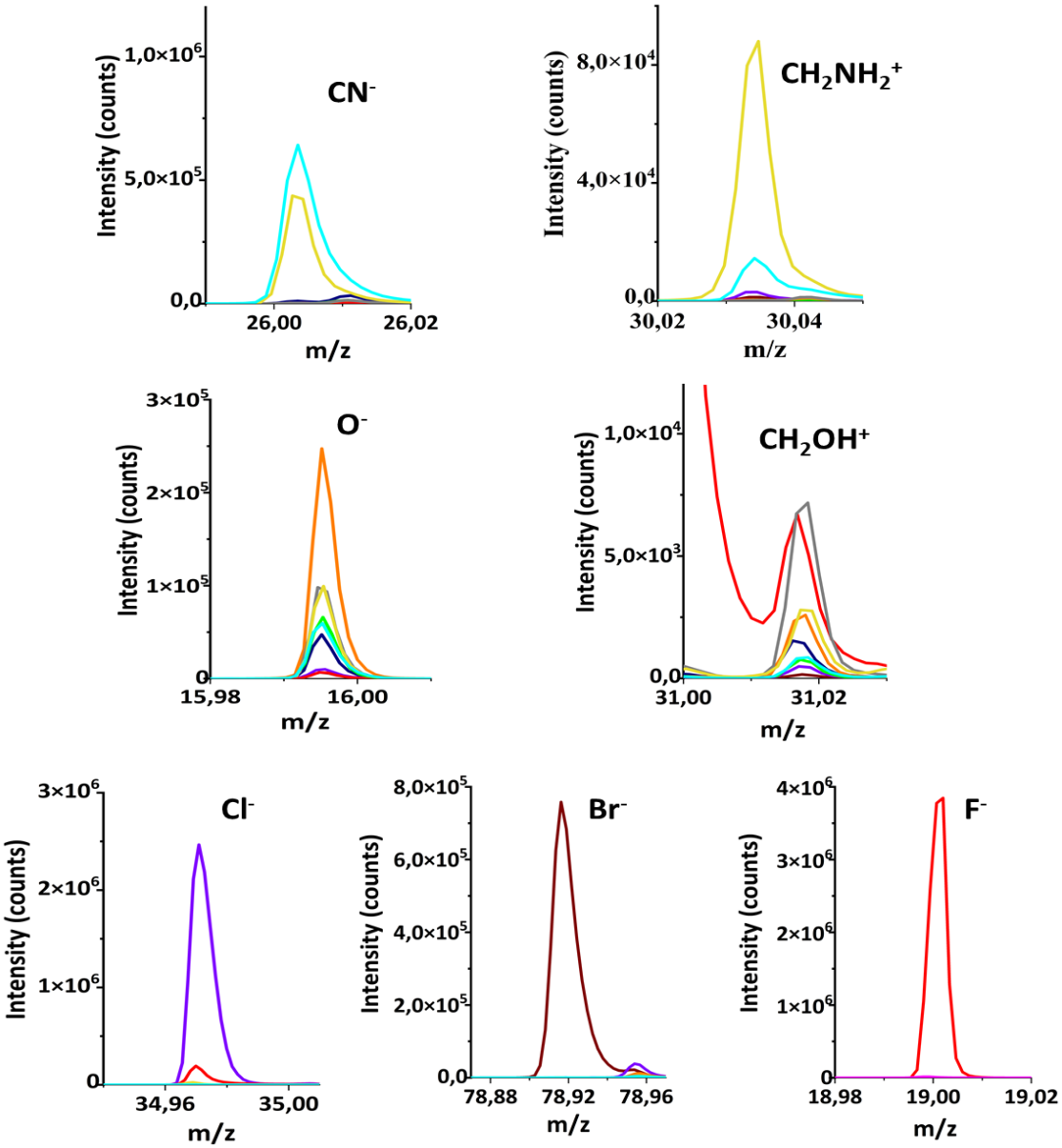


Figure 4.19. Characteristic TOF-SIMS spectral areas of PPX-coatings with varying functionalization. Light blue: PPX-NH_2, Yellow: PPX-AM, Pink: PPX-NH_2-co-AM, Orange: PPX-CHO, Gray: PPX-HM, Purple: PPX-Cl, Brown: PPX-Br, Red: PPX-F_8, and Green: PPX-Alkyne. Adapted from [183]

Surface Hydrophobicity

To assess the relative hydrophobicity (wettability) of different PPX surfaces, the diameter of dried droplets on each coating was measured, with an average of 150 measurements per surface to ensure statistical robustness and reproducibility (**Figure 4.20**). These measurements provided a quantitative comparison of the wettability of the various PPX coatings. The statistical analysis ranked the hydrophobicity of the coatings in a clear trend: PPX-CH_2OH exhibited the highest wettability and was determined to be the least hydrophobic, followed by PPX-CHO. The next group, consisting of PPX-Alk, PPX-A, PPX-Br, PPX-Cl, PPX-H, and PPX-AM, displayed similar hydrophobicity levels, indicating that their surface properties influenced droplet spreading in comparable ways. Finally, PPX-F emerged as the most hydrophobic among all tested surfaces, with the smallest droplet diameters, suggesting that its surface chemistry significantly reduced water affinity. These results highlight the role of functional groups in modifying the surface energy of PPX coatings and provide insights into their potential applications in areas where controlled wettability is critical.

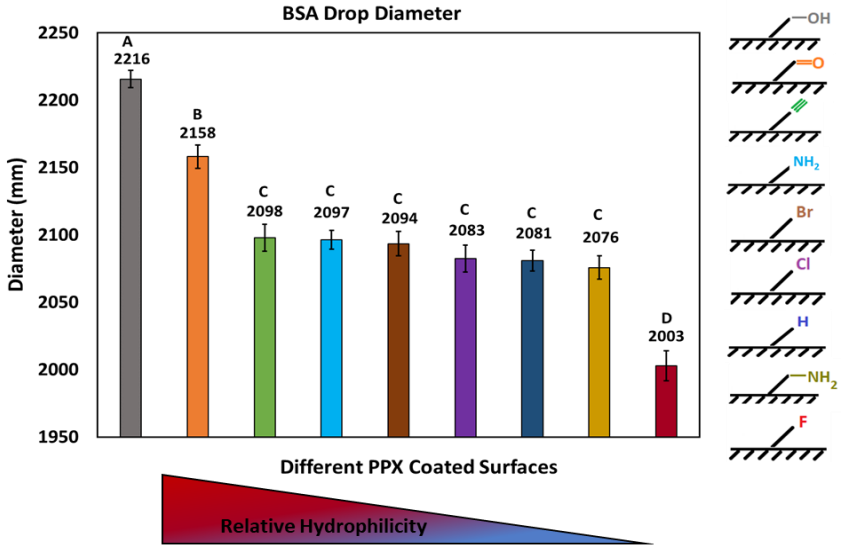


Figure 4.20. Average stain diameters of various CVD coatings were determined through image analysis (mean values, N=150, with error bars representing the standard error). Coating groups labeled with the same letters indicate no statistically significant differences, as determined by Tukey's test. Adapted from [183]

4.4.2. Classification and Identification of Surface Chemical Functionalities

To classify the BSA stains dried on various polymer surfaces, a total of 5,063 PLM images were utilized with the pre-trained Inception V3 network. All images were analyzed without any form of selection. Overall, the CNN achieved an average prediction accuracy of 96% across the nine different PPX-coated surfaces, indicating that the pre-trained network could accurately differentiate between polymer surfaces.

BSA adsorption onto different functional groups, each with unique physical (free surface energy) and chemical profiles (molecular moieties, hydrogen bonding, Lewis acid/base, polarizability, etc.), governs the interactions of the protein with the substrate, which, in turn, influences the resulting stain patterns. Previous studies have shown that surface characteristics play a crucial role in interfacial protein assembly. For example, Sarkar et al. observed that bovine serum albumin formed either multilayer or monolayer adsorption depending on substrate hydrophobicity. In this study, even subtle variations in surface chemistry led to distinct protein stain patterns, which the CNN effectively recognized and categorized. [276]

Figure 4.21.i shows that the CNN identified PPX-CH_2OH with 100% accuracy. This polymer coating exhibited the highest relative hydrophilicity, a property known to affect BSA adsorption on the surface [276]. Likewise, the PPX-CHO coated surface was also classified with 100% accuracy (refer to **Figure 4.20**).

As previously described, PPX-Alk, PPX-A, PPX-Br, PPX-Cl, PPX-H, and PPX-AM demonstrated similar relative hydrophobic properties (**Figure 4.20**). Misclassifications were most prevalent among halogenated coatings, with PPX-Br having the lowest prediction accuracy (83.8%). Notably, 13.7% of PPX-Br samples were misidentified by the CNN as either PPX-Cl or PPX-Alk, which share similar hydrophobic characteristics. While PPX-F

demonstrated greater hydrophobicity than other functionalized coatings, the CNN misclassified PPX-F coatings as either other halogenated PPX types or PPX-Alk. Grad-CAM analysis (**Figure S7**) revealed that most of these misclassifications were influenced by background interference and contrast from stain-free regions, which affected the CNN's decision-making process.

PPX-Alk had the second lowest prediction accuracy (91.3%), with around 7.5% of images misclassify as halogenated PPX surfaces. The CNN results indicate that most of misclassifications between alkyne and halogen groups can be attributed to their comparable hydrophobic properties. PPX-A achieved a prediction accuracy of 92.5%, with 2.5% of images misclassified as groups with similar hydrophobicity (PPX-AM and PPX-Alk), and 5% misclassified as PPX-CH_2OH coatings (**Figure S7**).

Subsequently, the model's performance in predicting more complex surface chemistries, i.e., unknown PPX-A/AM copolymer surfaces that were excluded from the training dataset, was evaluated. The pre-trained network effectively classified the unknown samples, with a majority being predicted as containing amine groups (56%) and aminomethyl groups (40%) (**Figure 4.21.i**).

In the following, as shown in **Figure 4.21.ii**, the training set was refined by excluding PPX-Br, allowing the model to focus and retrain on the remaining eight PPX-coated surfaces. This adjustment led to an improvement in overall prediction accuracy across the groups. Additionally, the misclassification pattern revealed a similar trend (**Figure 4.21.i**), with PPX-Alk and halogenated PPXs exhibiting the highest misclassification rate among all groups. Subsequently, images from the previously unknown PPX-A/AM copolymer and PPX-Br were tested by trained network. The results showed that all of the unknown PPX-A/AM binary copolymer samples were accurately classified as either PPX-A or PPX-AM. For the unknown PPX-Br coated surfaces, 56% of the images were categorized as halogenated PPX, while 95%

were grouped with polymer films that exhibiting hydrophobicity levels comparable to PPX-Br (**Figure 4.21.ii**).

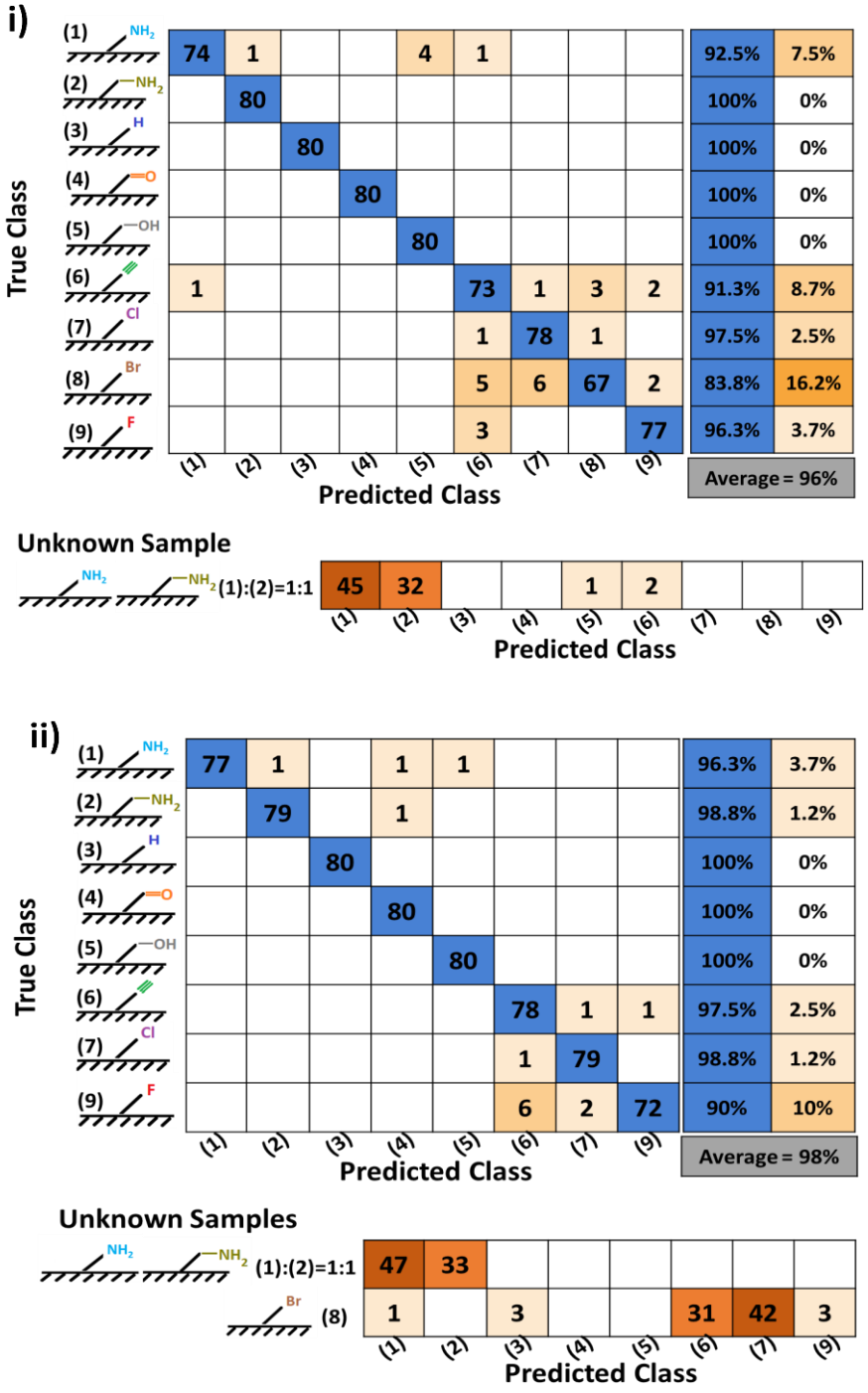


Figure 4.21. Confusion matrices derived from BSA stain pattern images of surfaces with various functionalized polymer CVD coatings. Deep learning-based recognition of PPX-coated glass wafers, with pre-training conducted i) including images of PPX-Br and ii) excluding them. BSA deposition stains were created by applying 2 µL droplets of a 100 mM carbonate-bicarbonate buffer (pH 9.2) containing 50 mM sodium chloride. The model parameters were optimized with a global learning rate of 0.001, a minimum batch size of 32 images, and up to 60 epochs. Adapted from [183]

Moreover, to assess the reproducibility of the experiment and the generalization capability of the trained network, we tested a small set of unknown PPX-Br samples collected from coated surfaces that were not included in the training, validation, or test sets. The results demonstrated high prediction accuracy in this "Few-Shot" study, aligning well with the findings of the original study (**Figure 4.22**).

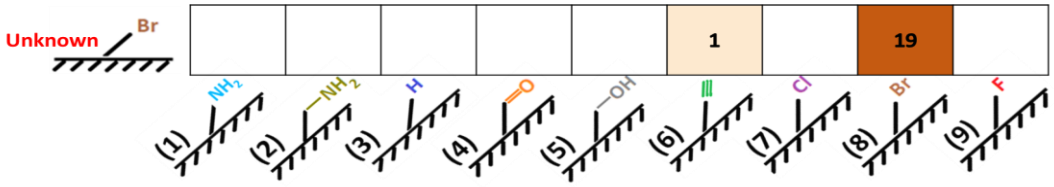


Figure 4.22. Evaluation of the trained network's reproducibility and generalization using unknown PPX-Br samples. High prediction accuracy was achieved in the Few-Shot study, consistent with the original findings. Adapted from [183]

Additionally, the Grad-CAM analysis revealed that the CNN primarily focused on central stain patterns during training, rather than edge patterns, indicating that stain size had a minimal impact on surface recognition (**Figure 4.23**). This suggests that the network learned to extract meaningful features from the internal structural characteristics of the stain patterns. Consequently, the classification was primarily driven by chemically relevant information encoded within the central morphology of the stains, which facilitated more robust and generalizable surface recognition across different PPX-coated surfaces.

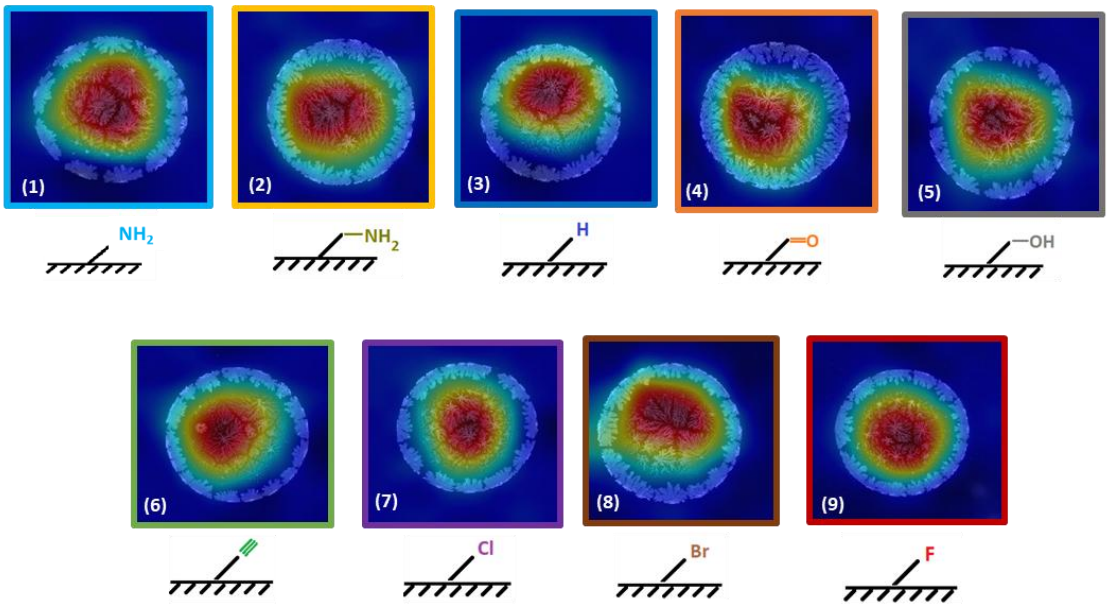


Figure 4.23. Grad-CAM activation maps. The heatmap overlay highlights regions that contributed most to the model's decision, offering insights into how the CNN interprets BSA patterns on various functionalized polymer surfaces. Adapted from [183]

The t-SNE analysis demonstrated clear clustering of BSA stain images from all distinct polymer surfaces, highlighting the reproducibility and uniqueness of the stain patterns for each PPX-coated surfaces (**Figure 4.24.i, 4.24.ii**). Moreover, this analysis performed with perplexities of 10, 50, and 100. The results consistently revealed well-defined clusters, indicating strong feature discrimination. This suggested that the network effectively identifies patterns and captures meaningful relationships at both global and local levels (**Figure S8**).

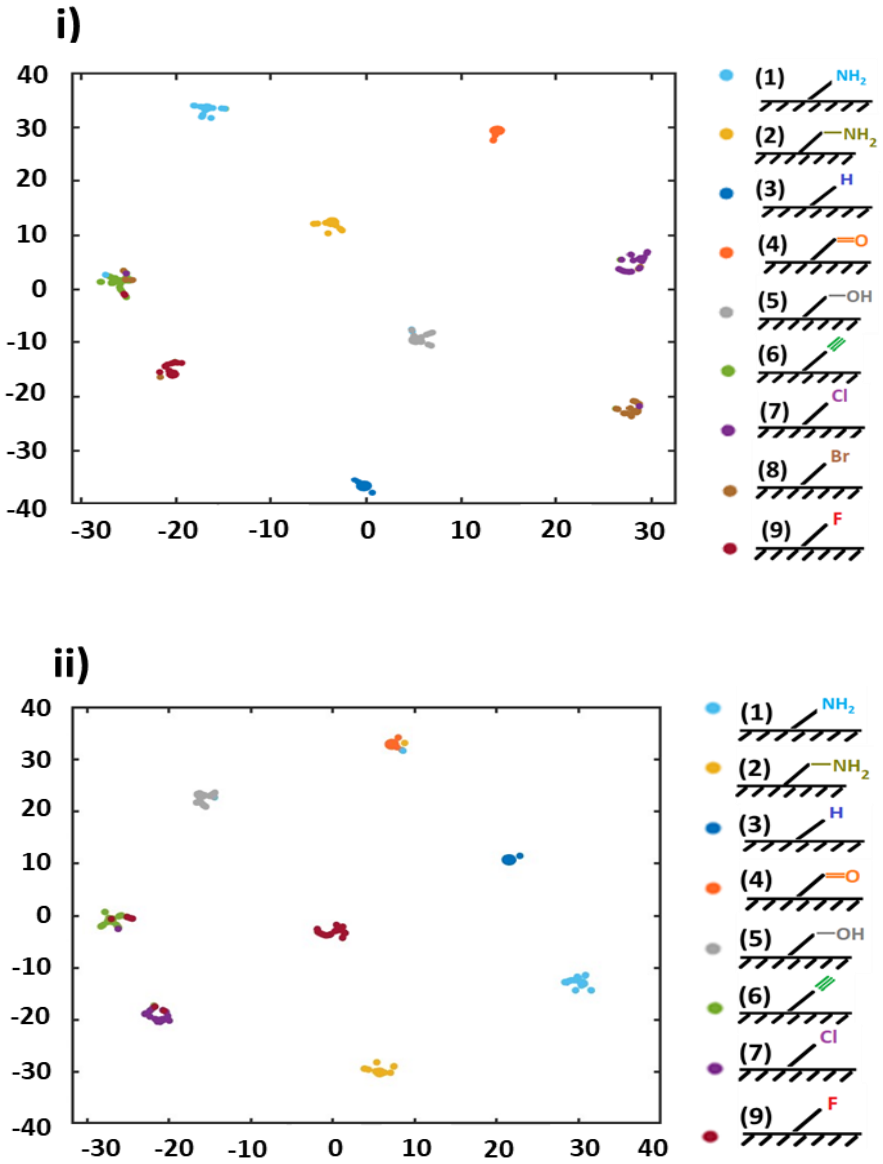


Figure 4.24. The t-SNE plots show the clustering of the various functionalized polymer surfaces. Visualization of the 'Softmax' layer in the trained CNN reveals distinct groupings corresponding to i) Nine PPX-coated glass wafers, including images of PPX-Br, and ii) excluding them. Adapted from [183]

4.4.3. Effect of Ionic Strength on CNN Classification

In this section, the effect of ionic strength on secondary structure changes and its influence on classification performance was investigated. As previously observed by Sarkar et al., [276] BSA adsorption was significantly influenced by ionic strength. It was suggested that ions in the protein solution interact with both proteins and surfaces, facilitating stronger binding of protein layers through electrostatic and hydrophobic interactions. These ionic interactions were found

to modify the structural organization, density, and orientation of the adsorbed protein layer, leading to distinct surface patterns. Additionally, it was reported that in the absence of ions, globular BSA molecules form a tilted monolayer on both hydrophilic and hydrophobic surfaces, with greater tilting occurring on hydrophobic surfaces, leading to a denser protein layer. [276, 277]

To investigate the influence of ionic strength on surface classification by CNN, PPX-A and PPX-AM surfaces were analysed and compared in both the presence and absence of sodium chloride. Notably, the secondary structure of BSA, dissolved in a dilute carbonate-bicarbonate buffer, remained unchanged at an ionic strength of 50 mM NaCl, as confirmed by CD spectroscopy analysis (**Figure 4.25.A**). However, when BSA was deposited onto the surfaces, the addition of 50 mM sodium chloride resulted in a noticeable increase in the prevalence of random coil structures in the CD spectrum (**Figure 4.25.B**). This observation suggests that the presence of NaCl promoted stronger BSA-surface binding through ionic interactions, which, in turn, led to a shift in the protein's secondary structure. The increased random coil formation implies that NaCl enhanced protein adsorption by altering molecular flexibility and interfacial interactions, ultimately resulting in a conformationally modified protein layer. These structural changes in BSA altered the organization of the adsorbed protein layer, affecting surface properties and crystallization behavior. Consequently, this reorganization influenced CNN classification outcomes, with a higher number of misclassifications observed in the absence of NaCl compared to samples containing NaCl.

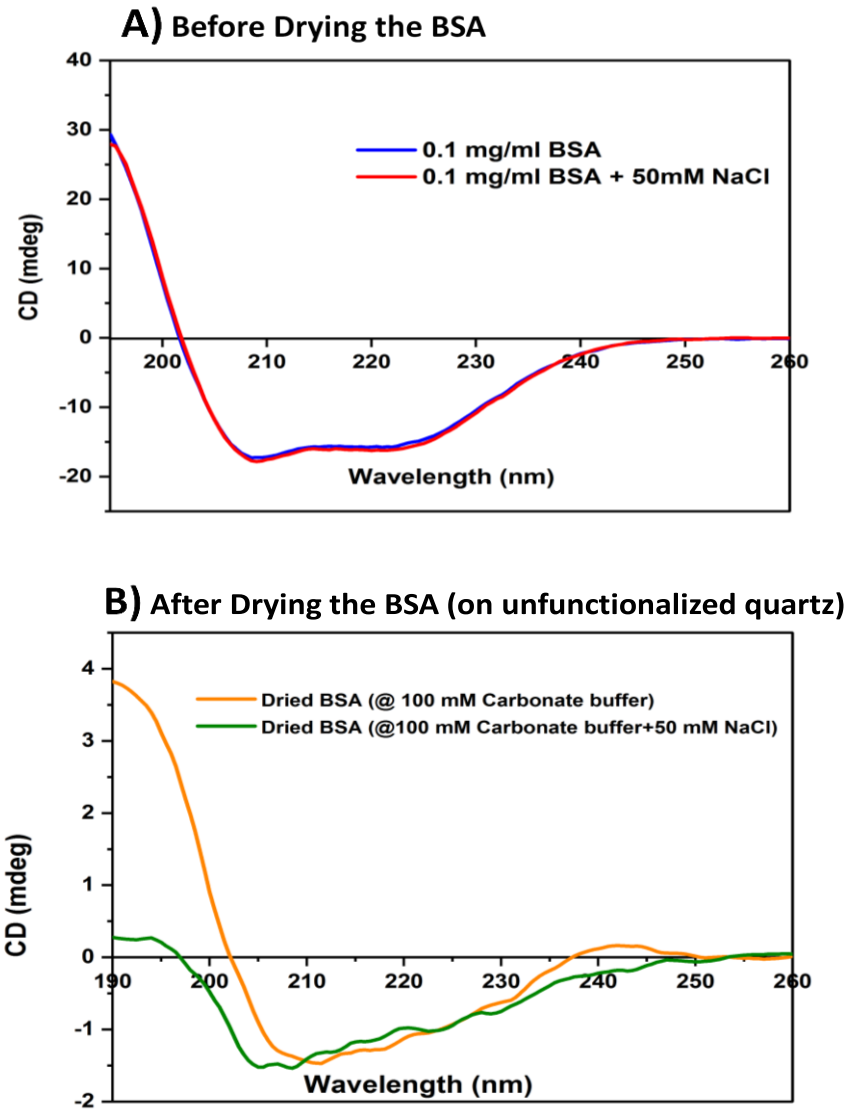


Figure 4.25. Influence of ionic strength on BSA secondary structure and its significance in classifying PLM images of BSA deposition patterns on amine and aminomethyl surfaces. A) Influence of ionic strength on BSA secondary structure when dissolved in a diluted carbonate-bicarbonate buffer with and without sodium chloride, and B) Impact of the drying process on BSA secondary structure, deposited on a quartz substrate in the absence and presence of sodium chloride, analyzed using CD spectroscopy. Adapted from [183]

Subsequently, the 50 mM sodium chloride was removed from the primary buffer solution, which contained 100 mM carbonate-bicarbonate buffer at pH 9.2. The resulting patterns, with and without sodium chloride, were distinctly different and easily distinguishable to the naked eye (see Figures **4.26.i and 4.26.ii**). For each group (amine and aminomethyl), around 80% of the total images were separated for the training and validation sets, while the remaining 20% were reserved for the testing set.

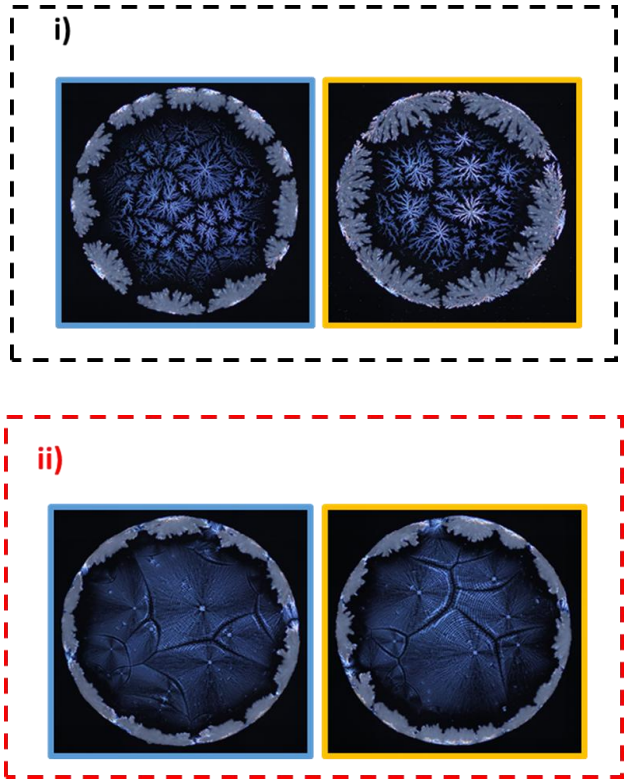


Figure 4.26. Effect of ionic strength on PLM images of deposition patterns of BSA on PPX- amine (blue boxes) and PPX-aminomethyl (yellow boxes) coated glass wafers. Stains were obtained by depositing 2 µL droplets of BSA solution in an aqueous i) 100 mM carbonate-bicarbonate buffer (pH=9.2) with 50 mM sodium chloride solution, and ii) 100 mM carbonate-bicarbonate buffer (pH=9.2) without sodium chloride. Adapted from [183]

Figure 4.27 presents the confusion matrices for each category, with panel i) showing the results with sodium chloride and panel ii) showing the results without sodium chloride. The CNN results indicate that as ionic strength decreased, the images from each category became harder to distinguish, leading to a drop in overall prediction accuracy from 99% with NaCl to 93% without NaCl.

To assess the performance of the network, we tested it on unknown samples, specifically BSA dissolved in carbonate-bicarbonate buffer both i) with and ii) without NaCl, dried on substrates coated with different PPX A/MA ratios (0.2/0.8, 0.5/0.5, and 0.8/0.2). For samples containing NaCl and dried on the PPX A/MA = 0.5/0.5 surface, the network classified 56% of the images as PPX-MA and 44% as PPX-A. Notably, it achieved 100% accuracy in distinguishing these samples from those with lower ionic strength. Similarly, when evaluating

samples ii) without NaCl on three copolymer surfaces, the model classified 54% of the binary copolymer (0.5/0.5) samples as PPX-MA, and 46% as PPX-A. For non-binary copolymers, the classification followed the dominant composition: 69% of PPX A/MA = 0.2/0.8 samples were identified as PPX-MA, while 31% were labeled as PPX-A, reflecting the major composition of the surface. Furthermore, the network demonstrated its ability to differentiate all samples with lower ionic strength from those with higher ionic strength with 100% accuracy.

To assess the relative hydrophobicity of each functionalized surface, 150 images of surfaces coated with amine and methyl amine groups were randomly collected, and the diameters of BSA droplets dissolved in buffers with and without sodium chloride were measured (**Figure S9**). Based on the results, for both buffer solutions (pH 9.2, with and without sodium chloride), the PPX-A and PPX-AM showed similar hydrophobicity.

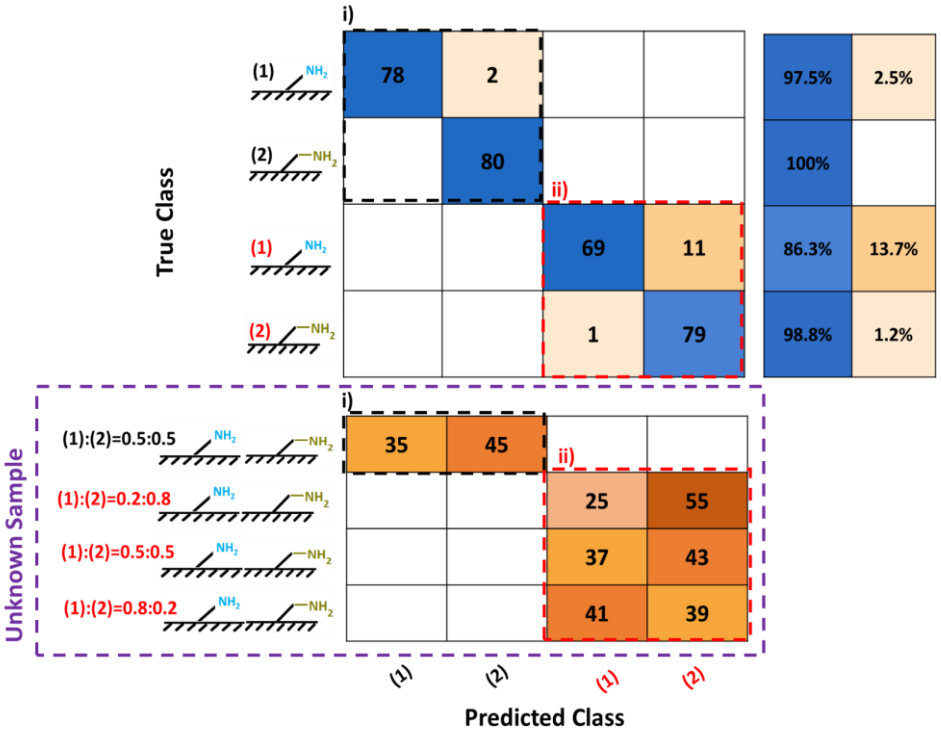


Figure 4.27. Confusion matrix derived from BSA stain pattern images, illustrating the model's ability to analyze the impact of ionic strength on surface recognition. BSA droplets, dissolved in an aqueous 100 mM carbonate-bicarbonate buffer (pH 9.2), were applied to glass wafers coated with PPX-amine and PPX-aminomethyl surfaces, (i) with 50 mM sodium chloride and (ii) without sodium chloride. The model parameters were fine-tuned across all layers, using a global learning rate of 0.001, a minimum batch size of 32 images, and a maximum of 60 epochs. Adapted from [183]

The t-SNE algorithm revealed more distinct clusters in BSA-stained images (i) with sodium chloride than (ii) without it (**Figure 4.28**). This indicates that the presence of NaCl enhanced the separation and organization of the data points, leading to clearer differentiation between the surface patterns. These findings further support the idea that ionic strength plays a critical role in modifying protein structure and its interaction with surfaces, ultimately impacting the classification performance and allowing CNN to more effectively distinguish between different surface chemistries.

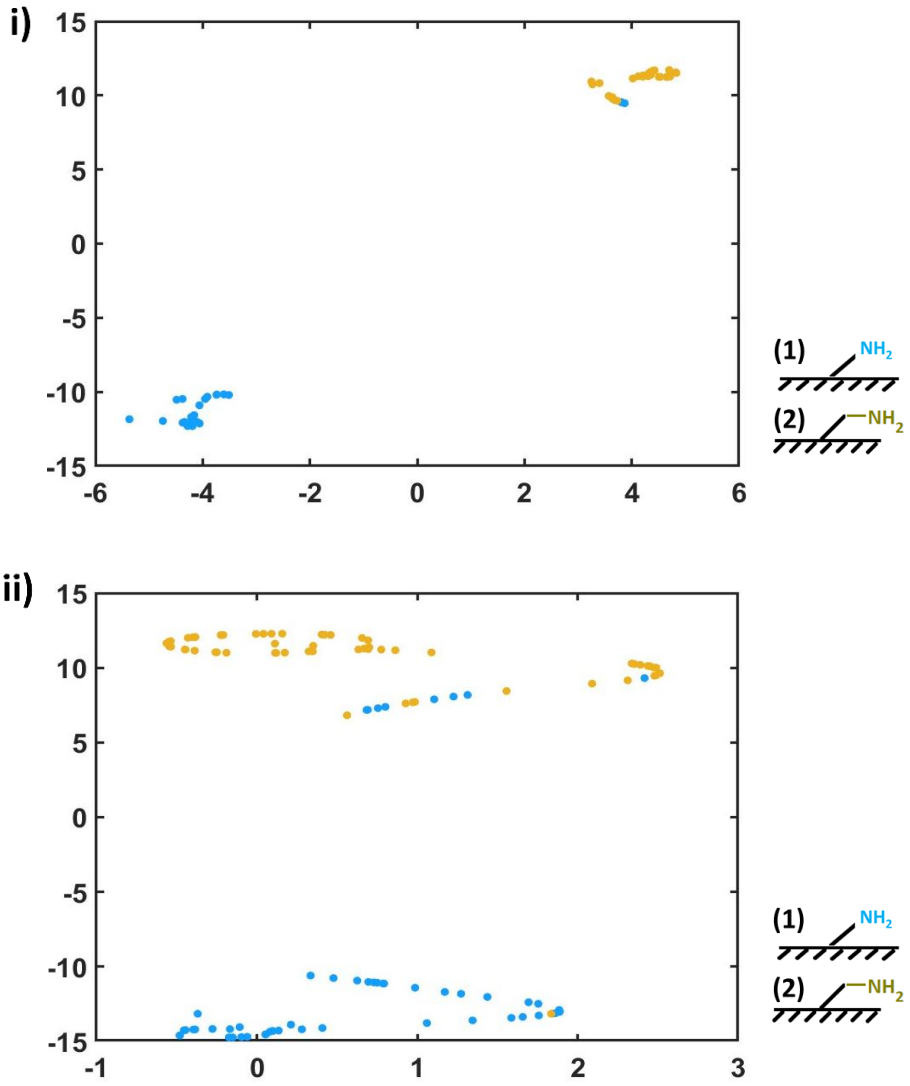
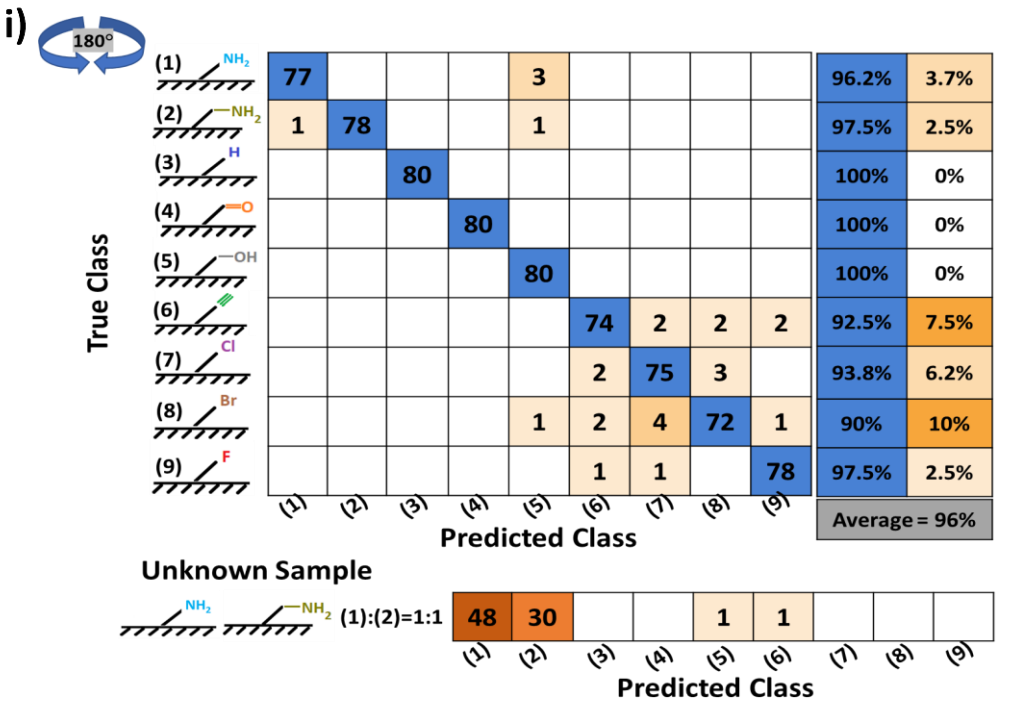


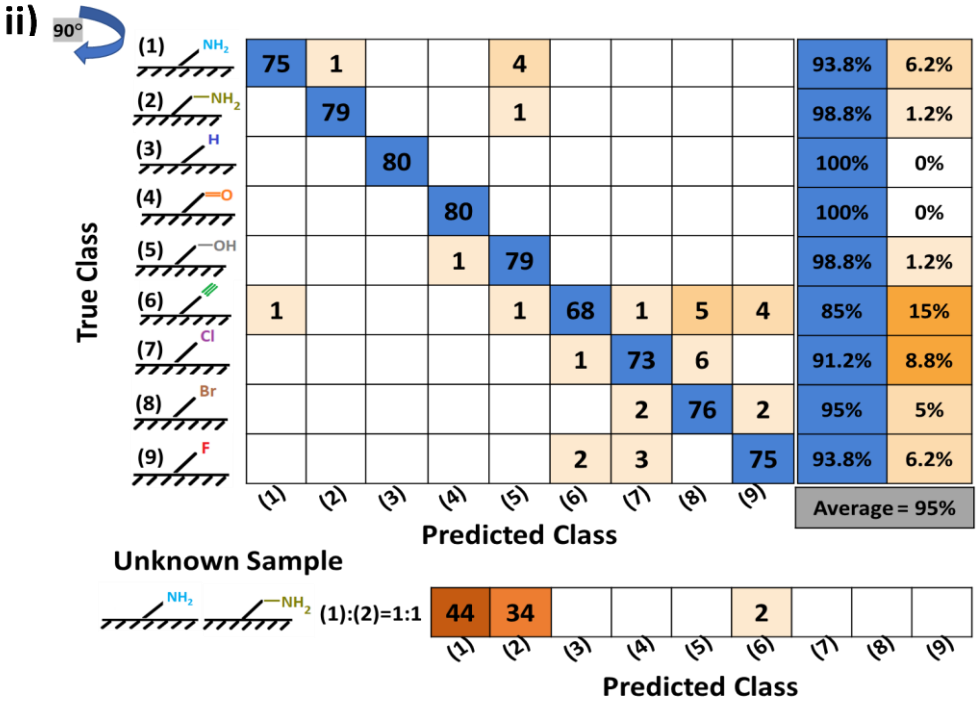
Figure 4.28. The t-SNE plots of the "Softmax" layer from the trained CNN demonstrate the clustering of two functionalized surfaces. Both in the i) presence and ii) absence of sodium chloride. Adapted from [183]

4.4.4. Effect of Image Rotation on Surface Classification

Addressing bias is a critical factor in the interpretation of deep learning analysis methods. [278] In this section, particular focus was placed on how the angle of observation might impact the results of the CNN analysis. To examine the presence of geometric training bias, the test set images were rotated at two different angles and in both directions, and the classification accuracy of surface chemistries was subsequently assessed using the pre-trained network.

The trained model (with the unrotated images) was evaluated by systematically rotating the images from each surface category. As previously mentioned, the CNN was tested on the original test dataset, achieving an impressive overall accuracy of 96%, as illustrated in the confusion matrix in **Figure 4.21.i**. The test images underwent (i) a 180-degree rotation and (ii, iii) a 90-degree rotation in both clockwise and counterclockwise directions (**Figure 4.29**). The results showed that the 180-degree rotations in both directions resulted in identical confusion matrices, with the average prediction accuracy remaining unchanged at 96% compared to the unrotated images (96%). For the 90-degree rotations, both clockwise and counterclockwise, there was a slight decrease in average prediction accuracy of 1%, lowering it to 95%, which is considered negligible. Furthermore, the trend of misclassification remained consistent across all three rotation scenarios when compared to the unrotated images.


i)
180°
(1) NH2
(2) NH2
(3) H
(4) O
(5) OH
(6)
(7) Cl
(8) Br
(9) F
True Class
77 3 96.2% 3.7%
1 78 1 97.5% 2.5%
80 100% 0%
80 100% 0%
80 100% 0%
74 2 2 2 92.5% 7.5%
2 75 3 93.8% 6.2%
1 2 4 72 1 90% 10%
1 1 78 97.5% 2.5%
(1) (2) (3) (4) (5) (6) (7) (8) (9)
Predicted Class
Average = 96%
Unknown Sample
NH2 NH2 (1):(2)=1:1
48 30 1 1
(1) (2) (3) (4) (5) (6) (7) (8) (9)
Predicted Class


ii)
90°
(1) NH2
(2) NH2
(3) H
(4) O
(5) OH
(6)
(7) Cl
(8) Br
(9) F
True Class
75 1 4 93.8% 6.2%
79 1 98.8% 1.2%
80 100% 0%
80 100% 0%
1 79 98.8% 1.2%
1 1 68 1 5 4 85% 15%
1 73 6 91.2% 8.8%
2 76 2 95% 5%
2 3 75 93.8% 6.2%
(1) (2) (3) (4) (5) (6) (7) (8) (9)
Predicted Class
Average = 95%
Unknown Sample
NH2 NH2 (1):(2)=1:1
44 34 2
(1) (2) (3) (4) (5) (6) (7) (8) (9)
Predicted Class

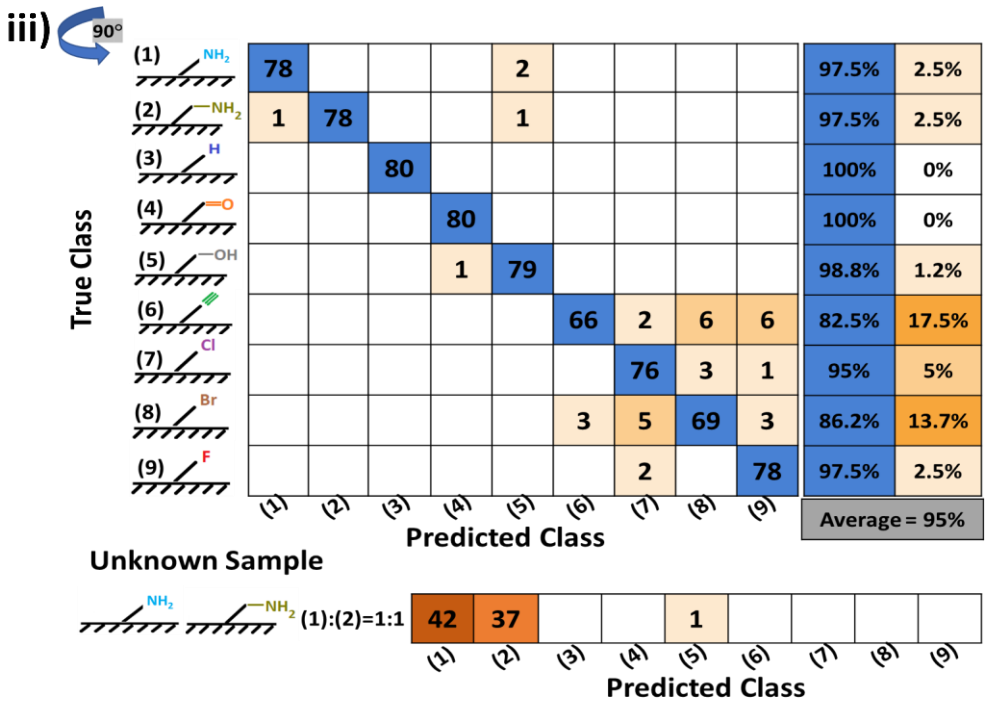


Figure 4.29. Assessment of geometric bias in the CNN analysis. (i) rotation of 180 degrees in both clockwise and counterclockwise directions, (ii) clockwise rotation of 90 degrees, and (iii) counterclockwise rotation of 90 degrees. Adapted from [183]

Following a 180-degree rotation in both directions, the unknown sample achieved a prediction accuracy of 97.5%, with classifications distributed as 60% PPX-A and 37.5% PPX-AM. For 90-degree rotations, accuracy remained high, with 97.5% accuracy for clockwise rotation (55% PPX-A and 42.5% PPX-AM) and 98.8% for counterclockwise rotation (52.5% PPX-A and 46.3% PPX-AM). These results indicate that classification accuracy remained consistent regardless of the rotation angle (90 or 180 degrees). Overall, the findings indicated that classification accuracy was unaffected by the rotation angle (90 and 180 degrees), demonstrating the network's capability to generalize and accurately classify surfaces regardless of image orientation.

5. Conclusion and Outlook

In this dissertation, neural networks were successfully trained, significantly aiding experimental efforts in the laboratory by often reducing the time and limitations associated with traditional methods. The chemical and physical characteristics of a biomacromolecule influenced the deposition patterns formed as a droplet of its solution dried on a material's surface. These deposition patterns were not only complex and characteristic but also highly reproducible. Consequently, they could serve as "fingerprints" for the screening and categorizing biomacromolecular interactions and various polymer surface chemistries. The main goal of this dissertation was to develop novel deep learning-based approaches for i) rating histone binding affinity to different DNA types and fragment sizes, and ii) classifying and identifying functional polymer-coated surfaces. By utilizing large datasets, machine learning algorithms can detect intricate patterns and correlations, significantly accelerating traditional analysis methods. In this study, the extensive dataset was generated using a method as simple as drying a droplet on a surface, allowing thousands of images to be produced within just a few hours. It enables low-cost, feasible experimentation while offering precise control over measurements and analyses.

In chapter 4.1, A large dataset consisting of two proteins (HSA and IgG) dissolved in different salt solutions was prepared. The results demonstrated that the CNN was capable of distinguishing not only between the proteins but also between different types of salts and even subtle variations in salt concentration. For example, in the same salt composition and concentration (NH_4HCO_3 (0.1 M) + KH_2PO_4 (0.05 M)), both proteins (HSA and IgG) were classified with 100% accuracy. Furthermore, when the same salt components and protein (HSA) were used, the trained network precisely distinguished between different salt concentrations (Tris at 0.05 M and 0.1 M) with 100% accuracy. Additionally, for constant salt

concentrations, the network accurately classified the type of components with high prediction accuracy. Overall, the trained network achieved an average prediction accuracy of 99%, demonstrating the CNN's ability to precisely detect even subtle changes in solution composition. This highlights the model's robustness and reliability in distinguishing small variations in experimental conditions.

In chapter 4.2, various mass ratios of histone/DNA mixture solutions were prepared, and two experimental scenarios were designed to evaluate the CNN's ability to distinguish not only between different protein-to-DNA ratios, but also between varying total mass concentrations of the mixtures. Scenario I included histone-Calf DNA mixtures with low histone-to-DNA mass ratios (1/5, 1/10, and 1/15), while Scenario II focused on histone-Calf DNA mixtures with higher protein-to-DNA ratios (5/1, 15/1, and 30/1). In both scenarios, the total mass concentrations of DNA and histone were maintained at two 0.1 and 0.3 mg/mL. In Scenario I, a higher rate of misclassification was observed compared to Scenario II at both total mass concentration levels, suggesting that an increase in histone content induces greater structural changes in the histone and/or DNA. These alterations in secondary structure resulted in more distinct dried stain patterns, which the CNN classified with higher accuracy. This trend was further corroborated using Salmon DNA as a proof of concept, where a similar pattern emerged, consistent with the results obtained for Calf DNA. The CNN also showed the capability to differentiate similar ratios with differing total mass concentrations. These findings prompted further exploration of protein-DNA binding affinity, revealing that the CNN can detect variations in histone/DNA ratios even within the same total mass concentrations, effectively capturing the effects of physicochemical interactions.

In chapter 4.3, this study utilized an advanced deep-learning method to investigate the relative binding affinity between DNA and H1. Specifically, using the InceptionV3 model to analyze stain images of histone-DNA complexes enabled accurate and predictive classification

of their binding affinities. The network indicated that H1 exhibits a stronger interaction with eukaryotic DNA compared to prokaryotic DNA. Moreover, due to longer DNA strands exhibiting a stronger binding affinity to histone, a higher prediction accuracy can be achieved for these in comparison to shorter DNA strands. This suggests a robust and highly discernible interaction pattern that the DL model readily identifies and classifies. These findings provide valuable insights into the intricate dynamics of protein-DNA interactions, highlighting the role of different parameters including DNA type and DNA size in binding affinity. Besides, all unseen/unknown images were precisely categorized into the most relevant groups, indicating the model's ability to assess unknown data effectively. Notably, the size of the test set had minimal impact on the results, indicating that just 10 images can yield an accuracy of over 90%. Although small sample sizes can be effective for testing, the model's robustness and reliability still require a large training dataset. In this research, approximately twelve thousand images were generated to pre-train the network and evaluate its performance using the test set, followed by evaluation on an unknown/unseen sample. Additionally, this study incorporated four distinct types and sizes of DNA, showcasing the model's ability to generalize effectively beyond the initial straightforward cases. The approach established in this research could enable the swift screening of candidates that bind to DNA or interactions between proteins and DNA, offering wide-ranging applications in biotechnology and molecular biology. To refine the training of the network further, additional image screening involving a varied array of proteins and DNA is essential, which would enhance the accuracy of binding affinity predictions and improve generalization.

In chapter 4.4, this study applied an advanced deep learning technique to identify surface chemical functionalization with different CVD coatings. By examining BSA stain patterns using the InceptionV3 model, the classification and identification of functional groups were highly accurate and predictive. Increasing ionic strength improved prediction accuracy,

which corroborates with the slight destabilizing effect of ionic strength on secondary structure of BSA. Future research should explore the impact of varying salt concentrations and different protein-surface interactions to further refine machine learning-based classification approaches in biointerface studies. Furthermore, the findings indicated that image rotation had no impact on prediction accuracy, eliminating concerns about geometric bias during the network training. This approach facilitated a rapid, cost-efficient, and straightforward assessment of surface chemistry candidates, showing great potential for broader applications in surface engineering. These findings show that the foundational methodology is not restricted to CVD-based polymer surfaces. In general, if there is a sufficiently diverse and representative dataset available for training, this work can easily be extended to functionalized polymer surfaces. For the process to be extended to a greater diversity of polymer coatings, the selection should be based on low roughness, homogeneity in coating, transparency, and stability against water. Given that material performance in various settings is closely tied to surface properties, this method was useful for analyzing protein-material interactions. The identification of functional groups as unique markers on material surfaces may allow researchers to predict and tailor surface characteristics for specific biomedical applications. This research highlights the flexibility and effectiveness of transfer learning and fine-tuning, demonstrating strong performance even in complex tasks, such as distinguishing chemically similar samples. Future studies could explore different network architectures to further assess and enhance the generalizability of these findings.

List of Figures

List of Tables

Supporting Information

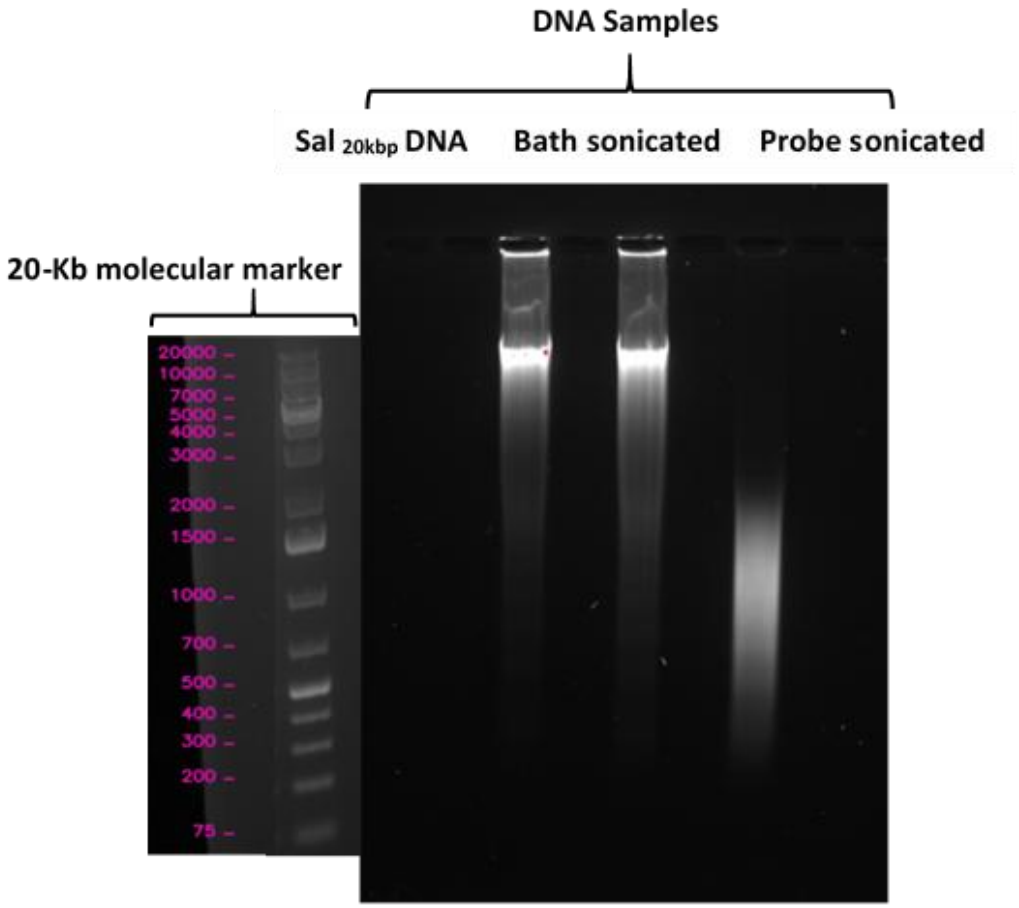


Figure S1) Agarose gel electrophoresis
Agarose gel electrophoresis (1%) for Sal $_{20\ kbp}$ DNA, sonicated Sal $_{20\ kbp}$ DNA (bath sonication for 5 second, and direct probe sonication for 5 min with 20 kHz. The determination of fragment sizes is based on their positioning on the gel compared to the molecular marker.

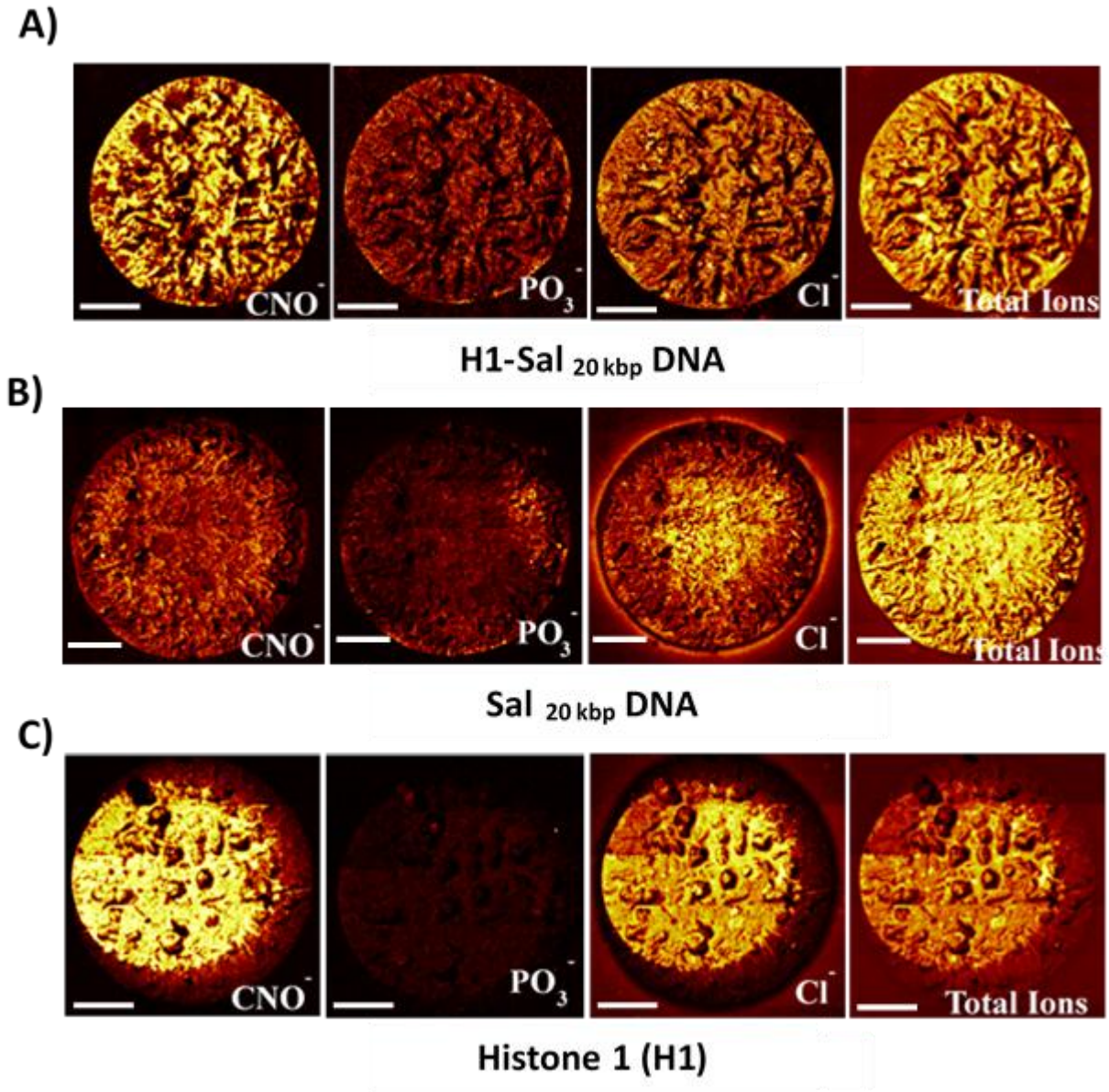


Figure S2) TOF-SIMS imaging.
Brown scale images of A) H1-Sal $_{20\ kbp}$ DNA sample: Amino acids of protein and base pairs of DNA were identified by CNO^- (scale 0-30), the backbone of DNA was identified by PO_3^- (scale 0-2), and buffer crystals were identified by Cl^- (scale 0-30); B) Sal $_{20\ kbp}$ DNA sample: CNO^- (scale 0-20), PO_3^- (scale 0-10), Cl^- (scale 0-30) and C) H1 (histone1) sample: CNO^- (scale 0-30), PO_3^- (scale 0-2), Cl^- (scale 0-30). All total ion images (scale 0-1300) represent the combined signals from all detected ions, displaying the variation in detected ionization over the droplet area. All scale bars represent 500 µm.

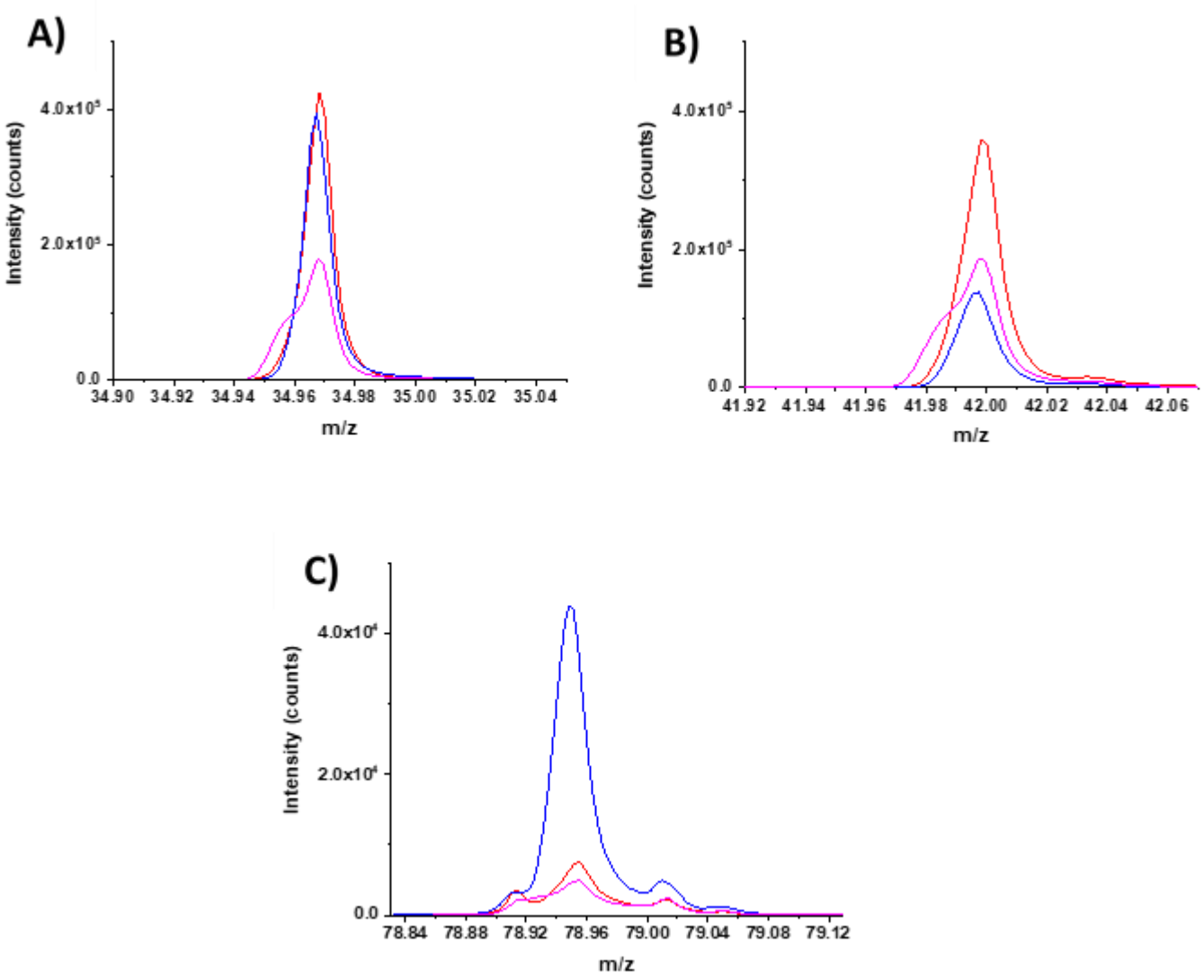


Figure S3) TOF-SIMS spectra.
Characteristic spectral areas, corresponding to the images in Figure S2, displaying the total areal intensity of A) Cl^-, B) CNO^- and C) $PO3^-$. The red, blue, and pink spectra represent histone (H1), Sal $_{20\ kbp}$ DNA, and H1+ Sal $_{20\ kbp}$ DNA mixture samples, respectively.

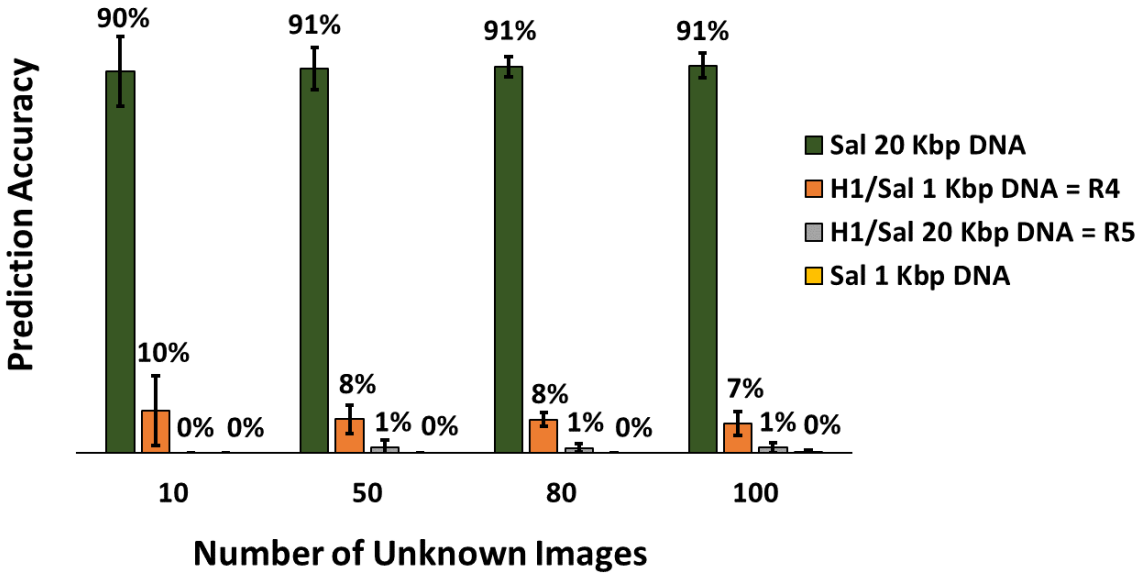


Figure S4) The impact of varying the quantity of unseen test images on the prediction accuracy of the trained neural network. The error bars in the graphical representation indicate the standard deviation (N=3). Groups sharing identical letters indicate no statistically significant differences based on least significant difference (LSD)-adjusted comparisons.

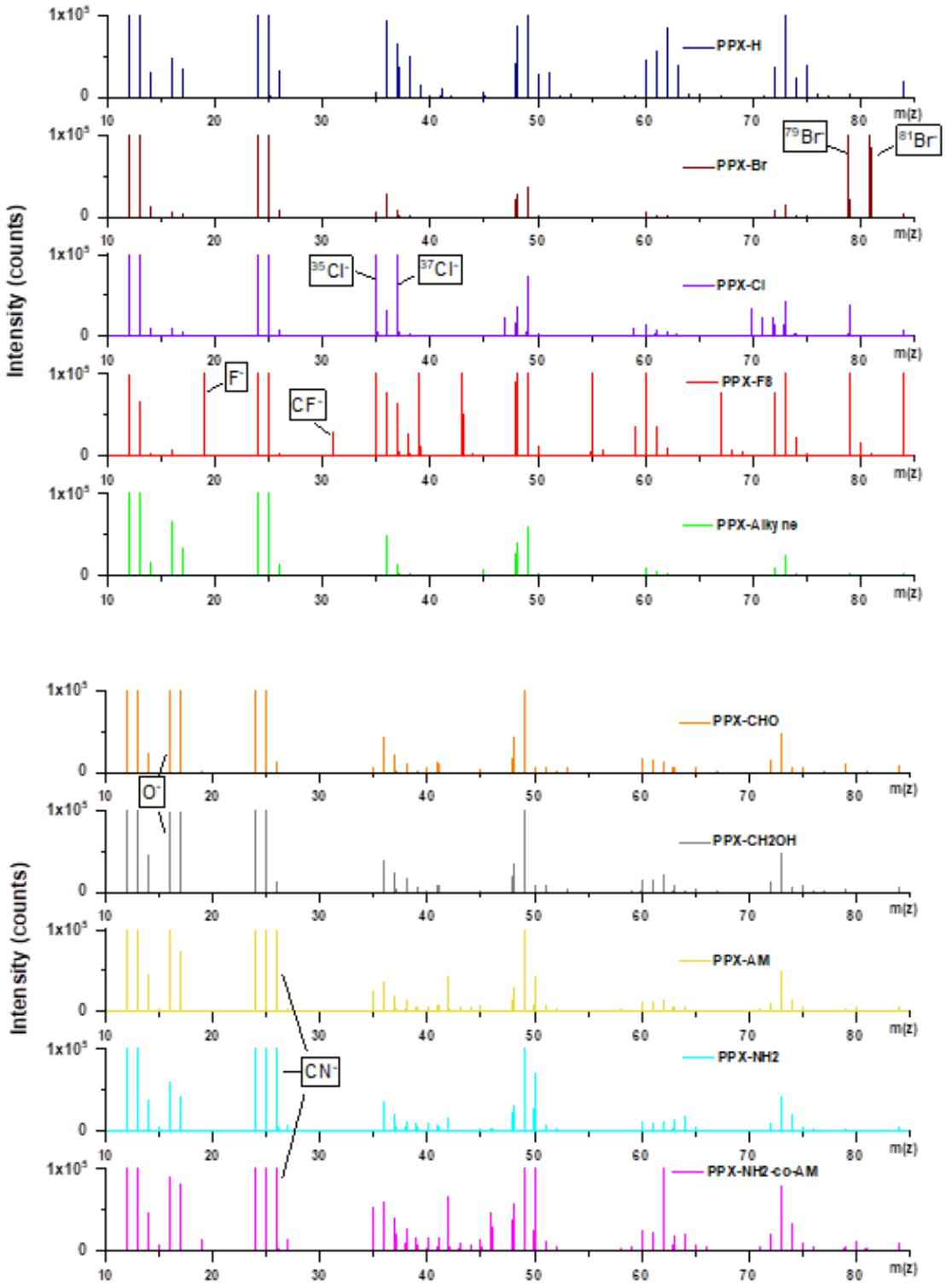


Figure S5) ToF-SIMS. Survey spectra of **negative polarity** measurements. All spectra are cut off at an upper intensity of x10^5 counts, to ensure the visibility of fragment peaks with medium intensity.

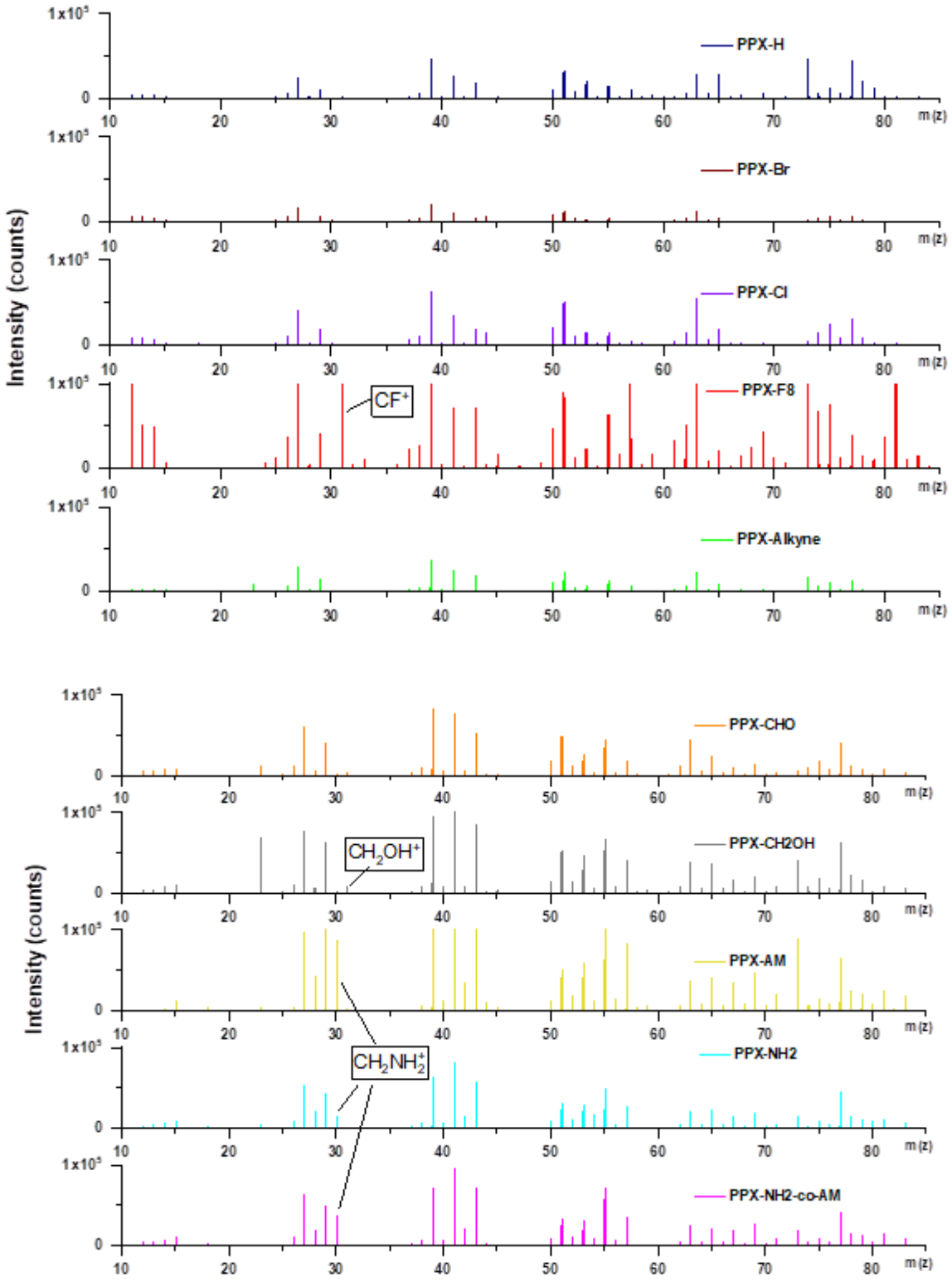
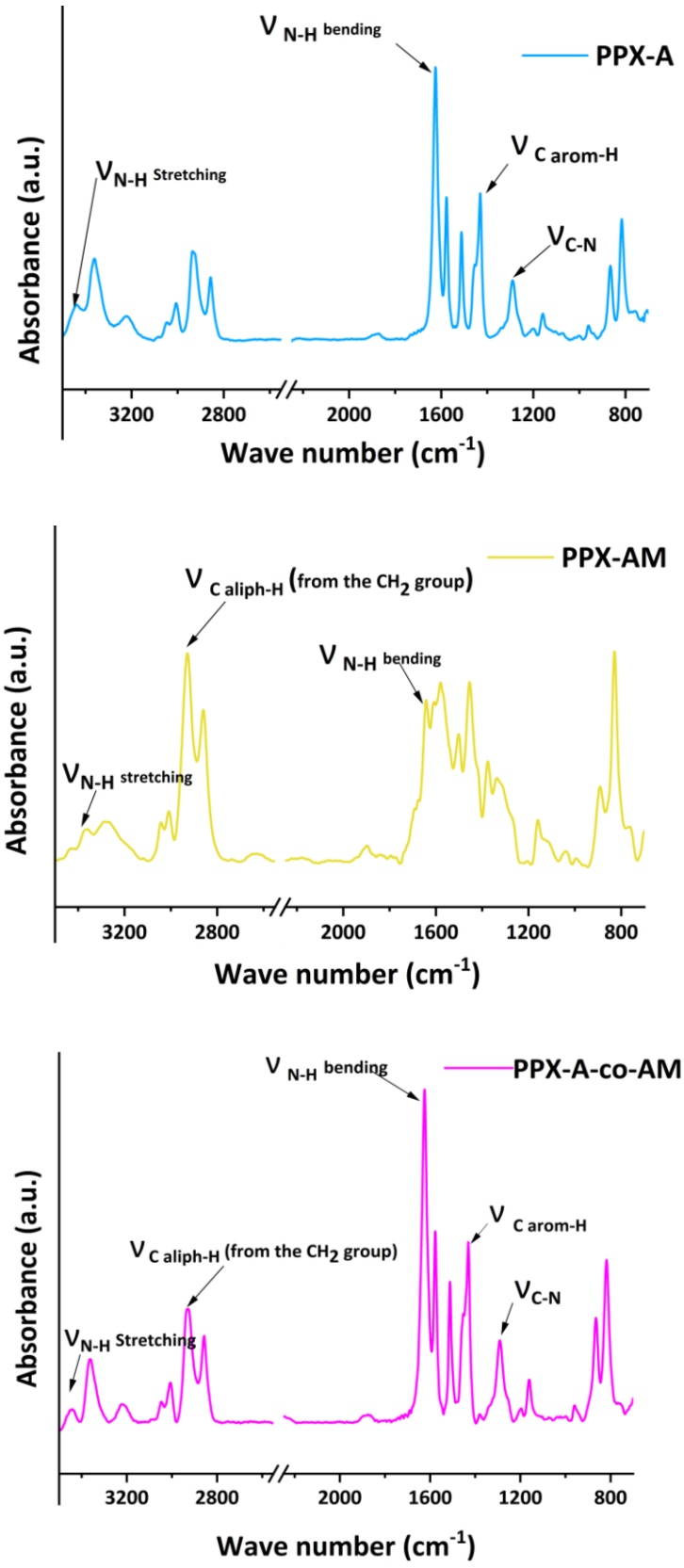
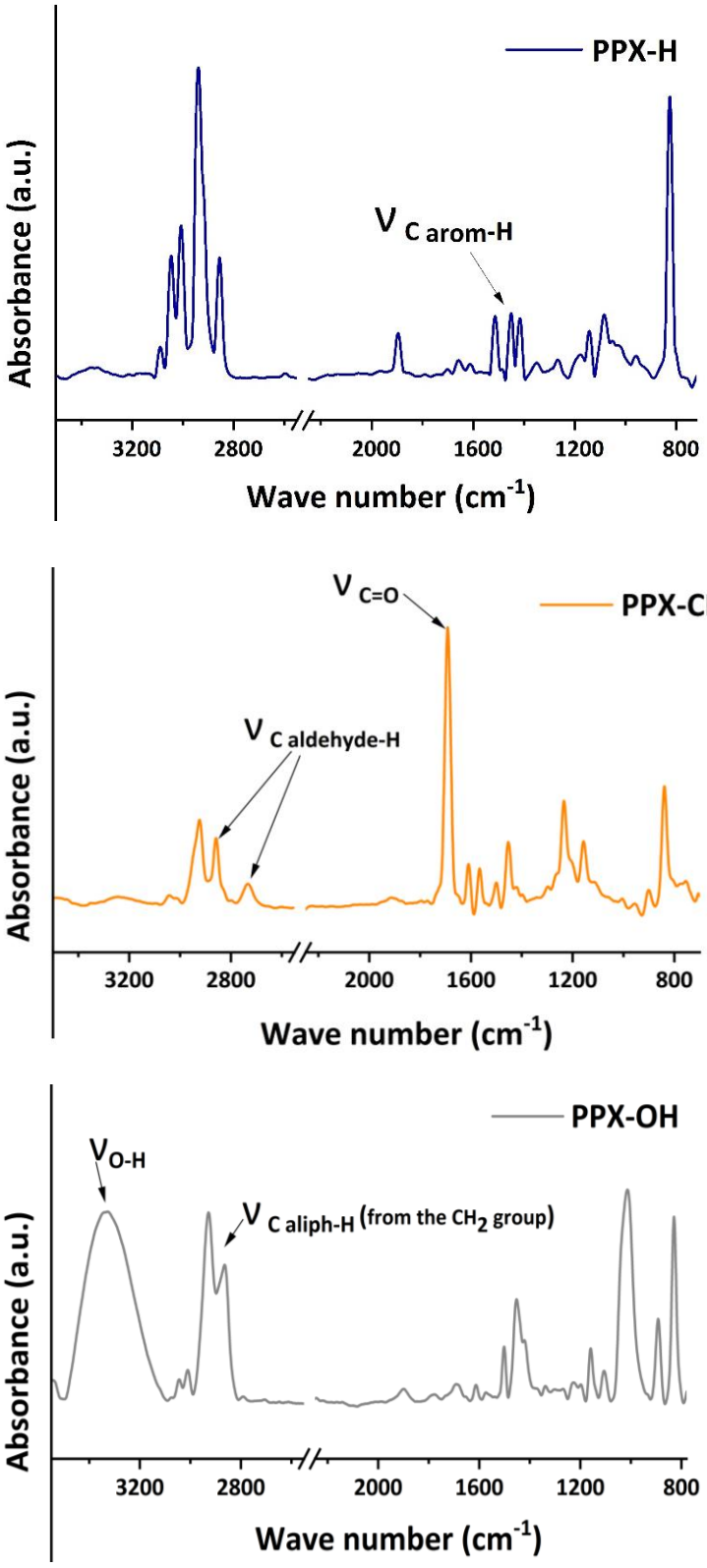


Figure S5) ToF-SIMS: Survey spectra of **positive polarity** measurements. All spectra are cut off at an upper intensity of x10^5 counts, to ensure the visibility of fragment peaks with medium intensity.


PPX-A
V N-H bending
V C arom-H
V C-N
V N-H Stretching
Absorbance (a.u.)
3200
2800
2000
1600
1200
800
Wave number (cm-1)
PPX-AM
V C aliph-H (from the CH2 group)
V N-H bending
V N-H stretching
Absorbance (a.u.)
3200
2800
2000
1600
1200
800
Wave number (cm-1)
PPX-A-co-AM
V N-H bending
V C arom-H
V C aliph-H (from the CH2 group)
V C-N
V N-H Stretching
Absorbance (a.u.)
3200
2800
2000
1600
1200
800
Wave number (cm-1)


PPX-H
V C arom-H
Absorbance (a.u.)
3200
2800
2000
1600
1200
800
Wave number (cm-1)
V C=O
PPX-CHO
V C aldehyde-H
Absorbance (a.u.)
3200
2800
2000
1600
1200
800
Wave number (cm-1)
PPX-OH
VO-H
V C aliph-H (from the CH2 group)
Absorbance (a.u.)
3200
2800
2000
1600
1200
800
Wave number (cm-1)

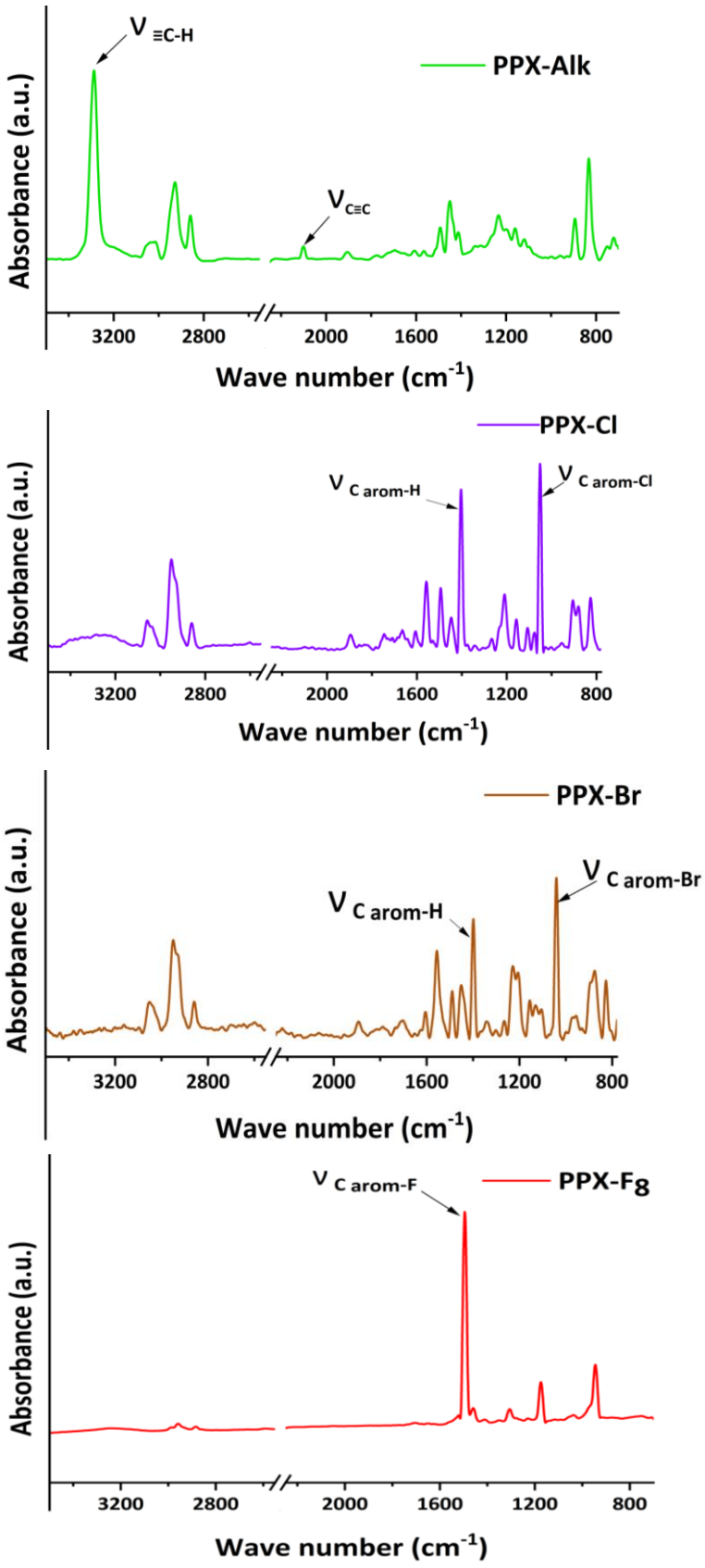


Figure S6) IRRAS of various PPXs

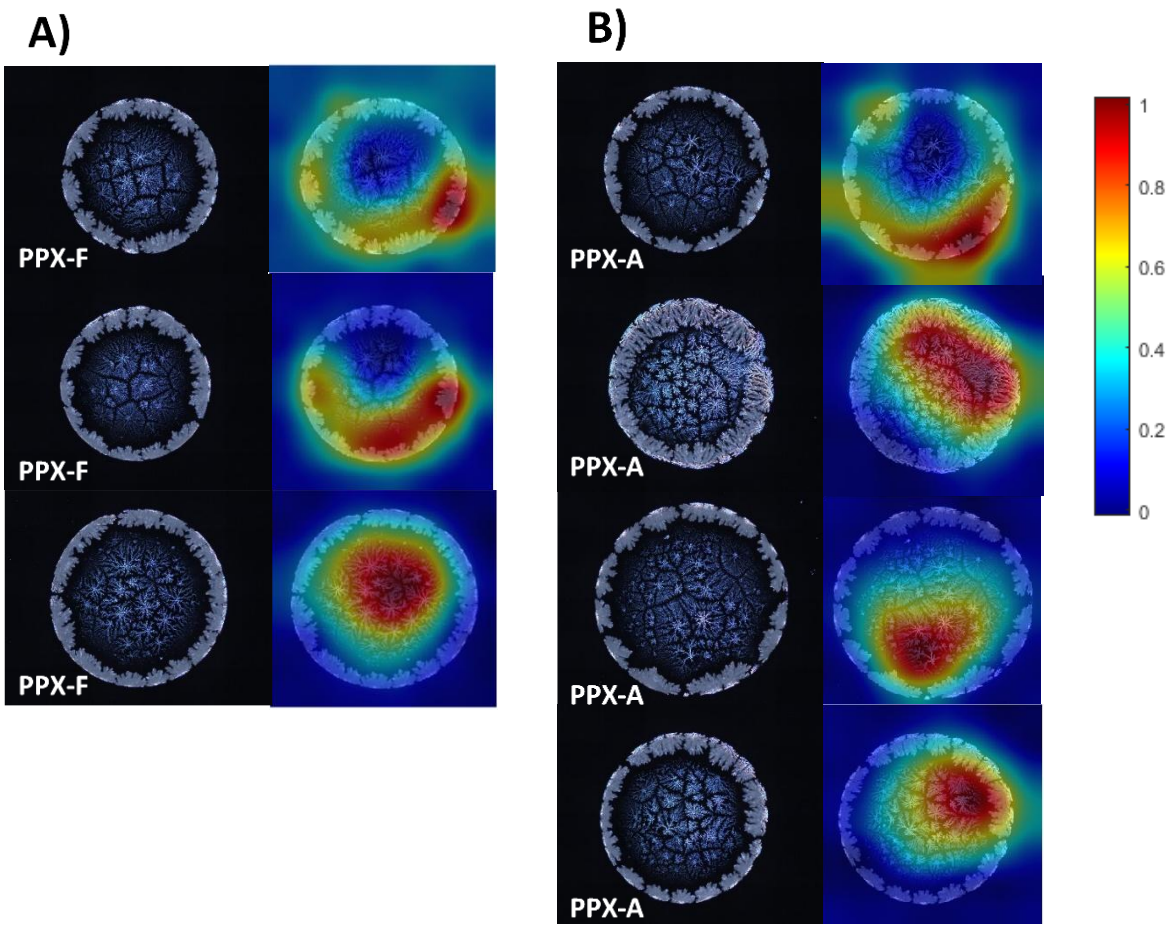


Figure S7) Gradient-weighted Class Activation Mapping (Grad-CAM). Grad-CAM illustrating misclassified PLM images: A) PPX-F as PPX-Alkyne; Three misclassifications occurred out of 80 unseen PLM images of BSA deposition patterns on PPX-F. B) PPX-A as PPX-CH_2OH; Four misclassifications occurred out of 80 unseen PLM images of BSA deposition patterns on PPX-A. Background interference, pattern-free spaces, and improper patterns collected from the edges of coated substrates (where the coating may have been insufficient) were the most causes of this misclassification.

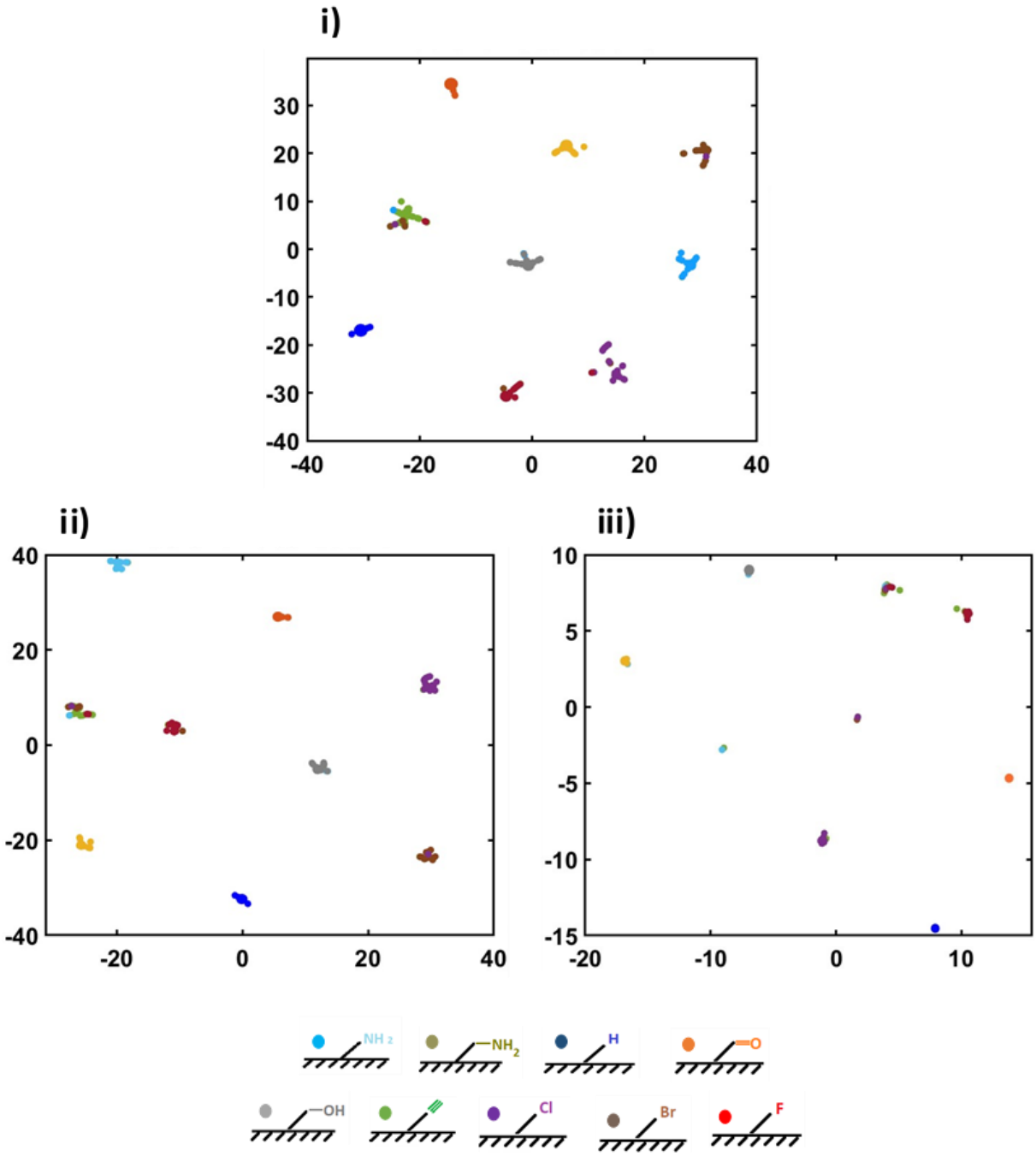


Figure S8) t-SNE visualizations of the feature space with perplexity values of i) 10, ii) 50, and iii) 100. These values were chosen to examine the robustness of clustering across both local (low perplexity) and global (high perplexity) relationships. The consistent clustering observed across all perplexity values indicates robust feature representations by the network.

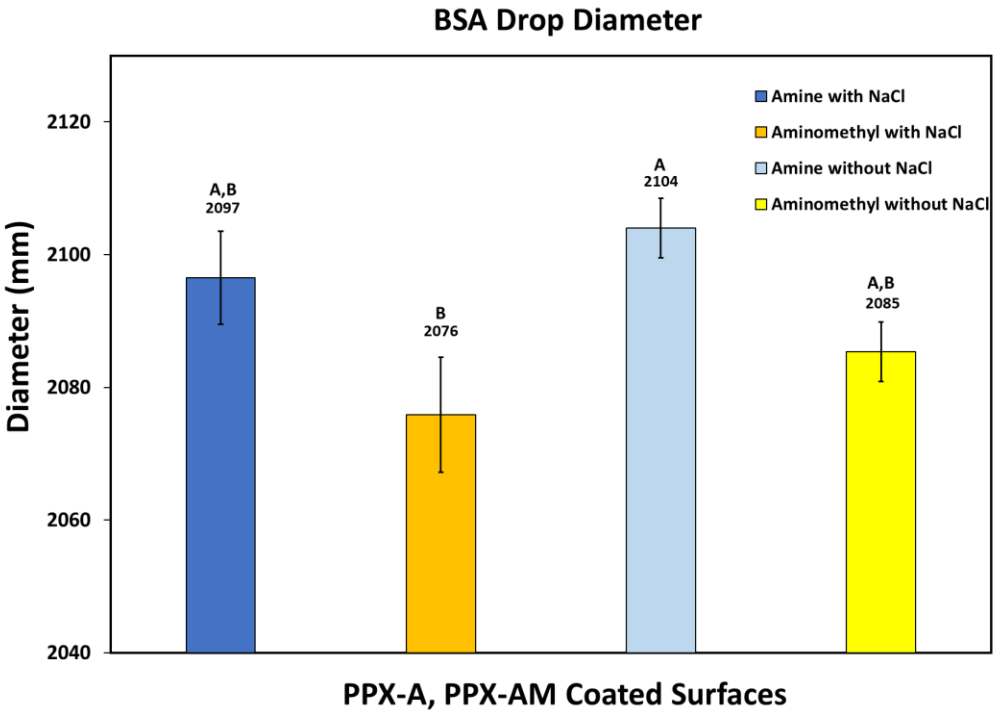


Figure S9) Hydrophobicity of PPX-A and PPX-AM functional surfaces, measured by the diameters size of dried BSA stains dissolved in buffer solutions with and without sodium chloride. A total of 150 randomly selected stains were analyzed per group. The error bars denote the standard error (SE). Groups sharing the same letters indicate no significant differences, as determined by Tukey-adjusted comparisons. Significant differences were identified using the Tukey method, with $p < 0.05$.

References

[1] A. Pal, A. Gope, A. Sengupta, Drying of bio-colloidal sessile droplets: Advances, applications, and perspectives, Advances in Colloid and Interface Science (2023) 102870.

[2] A. Jeihanipour, J. Lahann, Deep-Learning-Assisted Stratification of Amyloid Beta Mutants Using Drying Droplet Patterns, Advanced Materials 34(24) (2022) 2110404.

[3] D. Brutin, V. Starov, Recent advances in droplet wetting and evaporation, Chemical Society Reviews 47(2) (2018) 558-585.

[4] R.D. Deegan, O. Bakajin, T.F. Dupont, G. Huber, S.R. Nagel, T.A. Witten, Capillary flow as the cause of ring stains from dried liquid drops, Nature 389(6653) (1997) 827-829.

[5] H. Hu, R.G. Larson, Marangoni effect reverses coffee-ring depositions, The Journal of Physical Chemistry B 110(14) (2006) 7090-7094.

[6] R.G.G. Larson, M.A. López, D.W. Lim, J. Lahann, Complex protein patterns in drying droplets, MRS Online Proceedings Library (OPL) 1273 (2010) 1273-MM03-01.

[7] R.D. Deegan, Pattern formation in drying drops, Physical review E 61(1) (2000) 475.

[8] W. Yang, L. Deng, PreDBA: a heterogeneous ensemble approach for predicting protein-DNA binding affinity, Scientific Reports 10(1) (2020) 1-11.

[9] R. Chen, S. Subramanyam, A.H. Elcock, M. Spies, M.S. Wold, Dynamic binding of replication protein a is required for DNA repair, Nucleic acids research 44(12) (2016) 5758-5772.

[10] B. Dey, S. Thukral, S. Krishnan, M. Chakrobarty, S. Gupta, C. Manghani, V. Rani, DNA–protein interactions: methods for detection and analysis, Molecular and cellular biochemistry 365 (2012) 279-299.

[11] J.D. Nelson, O. Denisenko, K. Bomsztyk, Protocol for the fast chromatin immunoprecipitation (ChIP) method, Nature protocols 1(1) (2006) 179-185.

[12] L.M. Hellman, M.G. Fried, Electrophoretic mobility shift assay (EMSA) for detecting protein–nucleic acid interactions, Nature protocols 2(8) (2007) 1849-1861.

[13] J. Liu, G.D. Stormo, Combining SELEX with quantitative assays to rapidly obtain accurate models of protein–DNA interactions, Nucleic acids research 33(17) (2005) e141-e141.

[14] M.J. Guille, G.G. Kneale, Methods for the analysis of DNA-protein interactions, Molecular biotechnology 8 (1997) 35-52.

[15] R.W. Woody, [4] Circular dichroism, Methods in enzymology 246 (1995) 34-71.

[16] N.J. Greenfield, Using circular dichroism spectra to estimate protein secondary structure, Nature protocols 1(6) (2006) 2876-2890.

[17] L. Savinkova, E. Sharypova, N. Kolchanov, In vitro methods used to study DNA–protein interactions, Biology Bulletin Reviews 11 (2021) 344-357.

[18] Y.L. Lyubchenko, L.S. Shlyakhtenko, AFM for analysis of structure and dynamics of DNA and protein–DNA complexes, Methods 47(3) (2009) 206-213.

[19] A.U. Rehman, B. Khurshid, Y. Ali, S. Rasheed, A. Wadood, H.-L. Ng, H.-F. Chen, Z. Wei, R. Luo, J. Zhang, Computational approaches for the design of modulators targeting protein-protein interactions, Expert opinion on drug discovery 18(3) (2023) 315-333.

[20] P.C. Souza, V. Limongelli, S. Wu, S.J. Marrink, L. Monticelli, Perspectives on high-throughput ligand/protein docking with Martini MD simulations, Frontiers in Molecular Biosciences 8 (2021) 199.

[21] T.E. Cheatham III, M.A. Young, Molecular dynamics simulation of nucleic acids: successes, limitations, and promise, Biopolymers: Original Research on Biomolecules 56(4) (2000) 232-256.

[22] N. Bouatta, M. AlQuraishi, Structural biology at the scale of proteomes, Nature Structural & Molecular Biology 30(2) (2023) 129-130.

[23] M. Lee, Recent Advances in Deep Learning for Protein-Protein Interaction Analysis: A Comprehensive Review, Molecules 28(13) (2023) 5169.

[24] J. Si, R. Zhao, R. Wu, An overview of the prediction of protein DNA-binding sites, International journal of molecular sciences 16(3) (2015) 5194-5215.

[25] T.Y. Chang, V.G. Yadav, S. De Leo, A. Mohedas, B. Rajalingam, C.-L. Chen, S. Selvarasah, M.R. Dokmeci, A. Khademhosseini, Cell and protein compatibility of parylene-C surfaces, Langmuir 23(23) (2007) 11718-11725.

[26] E.M. Harnett, J. Alderman, T. Wood, The surface energy of various biomaterials coated with adhesion molecules used in cell culture, Colloids and surfaces B: Biointerfaces 55(1) (2007) 90-97.

[27] M. Golda-Cepa, W. Kulig, L. Cwiklik, A. Kotarba, Molecular Dynamics Insights into Water–Parylene C Interface: Relevance of Oxygen Plasma Treatment for Biocompatibility, ACS Applied Materials & Interfaces 9(19) (2017) 16685-16693.

[28] X.-M. Li, D. Reinhoudt, M. Crego-Calama, What do we need for a superhydrophobic surface? A review on the recent progress in the preparation of superhydrophobic surfaces, Chemical Society Reviews 36(8) (2007) 1350-1368.

[29] H. Cui, W. Wang, L. Shi, W. Song, S. Wang, Superwettable surface engineering in controlling cell adhesion for emerging bioapplications, Small Methods 4(12) (2020) 2000573.

[30] K.L. Menzies, L. Jones, The impact of contact angle on the biocompatibility of biomaterials, Optometry and vision science 87(6) (2010) 387-399.

[31] L. Tang, P. Thevenot, W. Hu, Surface chemistry influences implant biocompatibility, Current topics in medicinal chemistry 8(4) (2008) 270-280.

[32] D.S. Kohane, R. Langer, Biocompatibility and drug delivery systems, Chemical Science 1(4) (2010) 441-446.

[33] K. Anselme, L. Ploux, A. Ponche, Cell/material interfaces: influence of surface chemistry and surface topography on cell adhesion, Journal of Adhesion Science and Technology 24(5) (2010) 831-852.

[34] R. Tzoneva, N. Faucheux, T. Groth, Wettability of substrata controls cell–substrate and cell–cell adhesions, Biochimica et Biophysica Acta (BBA)-General Subjects 1770(11) (2007) 1538-1547.

[35] H.M. Lappin-Scott, J.W. Costerton, Bacterial biofilms and surface fouling, Biofouling 1(4) (1989) 323-342.

[36] H. Mei, T.S. Laws, T. Terlier, R. Verduzco, G.E. Stein, Characterization of polymeric surfaces and interfaces using time-of-flight secondary ion mass spectrometry, Journal of Polymer Science 60(7) (2022) 1174-1198.

[37] M. Plank, A. Berardi, A. Welle, E. Sauter, P. Krolla, C. Haret, M. Koenig, M. Stahlberger, Z. Hassan, S. Oßwald, Photo-Arbuzov Reactions as a Broadly Applicable Surface Modification Strategy, Advanced Functional Materials (2024) 2403408.

[38] Z. Ma, Z. Mao, C. Gao, Surface modification and property analysis of biomedical polymers used for tissue engineering, Colloids and Surfaces B: Biointerfaces 60(2) (2007) 137-157.

[39] F. Mangolini, A. Rossi, 5 Advances in attenuated total reflection (ATR) infrared spectroscopy: a powerful tool for investigating polymer surfaces and interfaces, Polymer Surface Characterization (2022) 173.

[40] E. De Giglio, N. Ditaranto, L. Sabbatini, Polymer surface chemistry: characterization by XPS, Polymer surface characterization (2014) 73-112.

[41] J. Kovac, Surface characterization of polymers by XPS and SIMS techniques, Mater. Technol 45(3) (2011) 191-197.

[42] B. Hagenhoff, High resolution surface analysis by TOF-SIMS, Microchimica Acta 132(2) (2000) 259-271.

[43] A.M. Belu, D.J. Graham, D.G. Castner, Time-of-flight secondary ion mass spectrometry: techniques and applications for the characterization of biomaterial surfaces, Biomaterials 24(21) (2003) 3635-3653.

[44] M. Koenig, R. Kumar, C. Hussal, V. Trouillet, L. Barner, J. Lahann, pH-Responsive Aminomethyl Functionalized Poly (p-xylylene) Coatings by Chemical Vapor Deposition Polymerization, Macromolecular Chemistry and Physics 218(9) (2017) 1600521.

[45] D. Wang, T.P. Russell, Advances in atomic force microscopy for probing polymer structure and properties, Macromolecules 51(1) (2018) 3-24.

[46] Y. Ko, B. Ratner, A. Hoffman, Characterization of hydrophilic-hydrophobic polymeric surfaces by contact angle measurements, Journal of Colloid and Interface Science 82(1) (1981) 25-37.

[47] D. Cozzolino, Benefits and limitations of infrared technologies in omics research and development of natural drugs and pharmaceutical products, Drug Development Research 73(8) (2012) 504-512.

[48] J. Kattner, H. Hoffmann, External reflection spectroscopy of thin films on dielectric substrates, Handbook of vibrational spectroscopy (2006).

[49] H. Arwin, Spectroscopic ellipsometry and biology: recent developments and challenges, Thin solid films 313 (1998) 764-774.

[50] F. Cui, Z. Zhang, C. Cao, Q. Zou, D. Chen, X. Su, Protein–DNA/RNA interactions: Machine intelligence tools and approaches in the era of artificial intelligence and big data, Proteomics 22(8) (2022) 2100197.

[51] C. Kauffman, G. Karypis, Computational tools for protein–DNA interactions, Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 2(1) (2012) 14-28.

[52] S. Patiyal, A. Dhall, G.P. Raghava, A deep learning-based method for the prediction of DNA interacting residues in a protein, Briefings in Bioinformatics 23(5) (2022) bbac322.

[53] K. Qu, L. Wei, Q. Zou, A review of DNA-binding proteins prediction methods, Current Bioinformatics 14(3) (2019) 246-254.

[54] H. Yang, H. Zheng, T. Zhang, A review of artificial intelligent methods for machined surface roughness prediction, Tribology International (2024) 109935.

[55] D.Y. Pimenov, A. Bustillo, T. Mikolajczyk, Artificial intelligence for automatic prediction of required surface roughness by monitoring wear on face mill teeth, Journal of Intelligent Manufacturing 29(5) (2018) 1045-1061.

[56] F. Lussier, V. Thibault, B. Charron, G.Q. Wallace, J.-F. Masson, Deep learning and artificial intelligence methods for Raman and surface-enhanced Raman scattering, TrAC Trends in Analytical Chemistry 124 (2020) 115796.

[57] P.M. Bhatt, R.K. Malhan, P. Rajendran, B.C. Shah, S. Thakar, Y.J. Yoon, S.K. Gupta, Image-based surface defect detection using deep learning: A review, Journal of Computing and Information Science in Engineering 21(4) (2021) 040801.

[58] S.A. Singh, K.A. Desai, Automated surface defect detection framework using machine vision and convolutional neural networks, Journal of Intelligent Manufacturing 34(4) (2023) 1995-2011.

[59] R.G. Larson, Re-shaping the coffee ring, Angewandte Chemie International Edition 51(11) (2012) 2546-2548.

[60] R.G. Larson, Transport and deposition patterns in drying sessile droplets, AIChE Journal 60(5) (2014) 1538-1571.

[61] D. Mampallil, H.B. Eral, A review on suppression and utilization of the coffee-ring effect, Advances in colloid and interface science 252 (2018) 38-54.

[62] S.S. Sazhin, Advanced models of fuel droplet heating and evaporation, Progress in energy and combustion science 32(2) (2006) 162-214.

[63] A. Friederich, J.R. Binder, W. Bauer, Rheological control of the coffee stain effect for inkjet printing of ceramics, Journal of the American Ceramic Society 96(7) (2013) 2093-2099.

[64] J. Zou, F. Kim, Diffusion driven layer-by-layer assembly of graphene oxide nanosheets into porous three-dimensional macrostructures, Nature communications 5(1) (2014) 5254.

[65] W. Han, Z. Lin, Learning from “coffee rings”: ordered structures enabled by controlled evaporative self-assembly, Angewandte Chemie International Edition 51(7) (2012) 1534-1546.

[66] L. Cui, Y. Li, J. Wang, E. Tian, X. Zhang, Y. Zhang, Y. Song, L. Jiang, Fabrication of large-area patterned photonic crystals by ink-jet printing, Journal of Materials Chemistry 19(31) (2009) 5499-5502.

[67] B.-J. de Gans, U.S. Schubert, Inkjet printing of well-defined polymer dots and arrays, Langmuir 20(18) (2004) 7789-7793.

[68] D. Mampallil, H. Eral, D. Van Den Ende, F. Mugele, Control of evaporating complex fluids through electrowetting, Soft Matter 8(41) (2012) 10614-10617.

[69] W. Wang, Y. Yin, Z. Tan, J. Liu, Coffee-ring effect-based simultaneous SERS substrate fabrication and analyte enrichment for trace analysis, Nanoscale 6(16) (2014) 9588-9593.

[70] R. Blossey, A. Bosio, Contact line deposits on cDNA microarrays: a “twin-spot effect”, Langmuir 18(7) (2002) 2952-2954.

[71] K. Sefiane, G. Duursma, A. Arif, Patterns from dried drops as a characterisation and healthcare diagnosis technique, potential and challenges: A review, Advances in Colloid and Interface Science 298 (2021) 102546.

[72] M.O. Kokornaczyk, N.B. Bodrova, S. Baumgartner, Diagnostic tests based on pattern formation in drying body fluids–A mapping review, Colloids and Surfaces B: Biointerfaces 208 (2021) 112092.

[73] Y. Choi, J. Han, C. Kim, Pattern formation in drying of particle-laden sessile drops of polymer solutions on solid substrates, Korean Journal of Chemical Engineering 28 (2011) 2130-2136.

[74] H.Y. Erbil, Evaporation of pure liquid sessile and spherical suspended drops: A review, Advances in colloid and interface science 170(1-2) (2012) 67-86.

[75] R. Picknett, R. Bexon, The evaporation of sessile or pendant drops in still air, Journal of colloid and Interface Science 61(2) (1977) 336-350.

[76] S.A. Okaiyeto, H.-W. Xiao, A.S. Mujumdar, Understanding the coffee ring effect: how it has led to advanced applications, Taylor & Francis (2023) pp. 1083-1084.

[77] D. Kaya, V. Belyi, M. Muthukumar, Pattern formation in drying droplets of polyelectrolyte and salt, The Journal of chemical physics (2010) 133(11).

[78] Z. Chen, K. Peng, B. Xu, Material assembly by droplet drying: From mechanics theories to applications, Droplet 2(4) (2023) e76.

[79] H.H. Lee, S.C. Fu, C.Y. Tso, C.Y. Chao, Study of residue patterns of aqueous nanofluid droplets with different particle sizes and concentrations on different substrates, International Journal of Heat and Mass Transfer 105 (2017) 230-236.

[80] K. Uno, K. Hayashi, T. Hayashi, K. Ito, H. Kitano, Particle adsorption in evaporating droplets of polymer latex dispersions on hydrophilic and hydrophobic surfaces, Colloid and polymer science 276 (1998) 810-815.

[81] B.M. Weon, J.H. Je, Capillary force repels coffee-ring effect, Physical Review E 82(1) (2010) 015305.

[82] A.D. Bermel, D. Bugner, Particle Size Effects in Pigmented Inkjet Inks, Journal of imaging science and technology 43(4) (1999) 320-324.

[83] G. Chen, G. J. Mohamed, Complex protein patterns formation via salt-induced self-assembly and droplet evaporation, The European Physical Journal E 33 (2010) 19-26.

[84] Y. Li, Q. Yang, M. Li, Y. Song, Rate-dependent interface capture beyond the coffee-ring effect, Scientific reports 6(1) (2016) 24628.

[85] N. Kim, Z. Li, C. Hurth, F. Zenhausern, S.-F. Chang, D. Attinger, Identification of fluid and substrate chemistry based on automatic pattern recognition of stains, Analytical Methods 4(1) (2012) 50-57.

[86] W.B. Zeid, D. Brutin, Influence of relative humidity on spreading, pattern formation and adhesion of a drying drop of whole blood, Colloids and Surfaces A: Physicochemical and Engineering Aspects 430 (2013) 1-7.

[87] N.D. Patil, R. Bhardwaj, Recent developments on colloidal deposits obtained by evaporation of sessile droplets on a solid surface, Journal of the Indian Institute of Science 99 (2019) 143-156.

[88] M. Parsa, S. Harmand, K. Sefiane, Mechanisms of pattern formation from dried sessile drops, Advances in colloid and interface science 254 (2018) 22-47.

[89] S. Tarafdar, Y.Y. Tarasevich, M. Dutta Choudhury, T. Dutta, D. Zang, Droplet drying patterns on solid substrates: from hydrophilic to superhydrophobic contact to levitating drops, Advances in Condensed Matter Physics 2018 (2018) 1-24.

[90] D. Zang, S. Tarafdar, Y.Y. Tarasevich, M.D. Choudhury, T. Dutta, Evaporation of a Droplet: From physics to applications, Physics Reports 804 (2019) 1-56.

[91] F. Giorgiutti-Dauphiné, L. Pauchard, Drying drops: Drying drops containing solutes: From hydrodynamical to mechanical instabilities, The European Physical Journal E 41 (2018) 1-15.

[92] Y.J. Carreón, M. Ríos-Ramírez, P. Vázquez-Vergara, S. Salinas-Almaguer, I. Cipriano-Urbano, A. Briones-Aranda, O. Díaz-Hernández, G.J.E. Santos, J. González-Gutiérrez, Effects of substrate temperature on patterns produced by dried droplets of proteins, Colloids and Surfaces B: Biointerfaces 203 (2021) 111763.

[93] J.M. Cameron, H.J. Butler, D.S. Palmer, M.J. Baker, Biofluid spectroscopic disease diagnostics: A review on the processes and spectral impact of drying, Journal of biophotonics 11(4) (2018) e201700299.

[94] A. Pal, A. Gope, G. Iannacchione, Temperature and concentration dependence of human whole blood and protein drying droplets, Biomolecules 11(2) (2021) 231.

[95] J.T. Wen, C.-M. Ho, P.B. Lillehoj, Coffee ring aptasensor for rapid protein detection, Langmuir 29(26) (2013) 8440-8446.

[96] M. Yang, D. Chen, J. Hu, X. Zheng, Z.-J. Lin, H. Zhu, The application of coffee-ring effect in analytical chemistry, TrAC Trends in Analytical Chemistry 157 (2022) 116752.

[97] W. Sempels, R. De Dier, H. Mizuno, J. Hofkens, J. Vermant, Auto-production of biosurfactants reverses the coffee ring effect in a bacterial system, Nature communications 4(1) (2013) 1757.

[98] Y. Zeng, D. Kai, Z. Niu, Z. Nie, Y. Wang, Y. Shao, L. Ma, F. Zhang, G. Liu, J. Chen, Coffee Ring Effect Enhanced Surface Plasmon Resonance Imaging Biosensor via 2-λ Fitting Detection Method, Biosensors 14(4) (2024) 195.

[99] J. Wilkinson, C. Tam, A. Askounis, S. Qi, Suppression of the coffee-ring effect by tailoring the viscosity of pharmaceutical sessile drops, Colloids and Surfaces A: Physicochemical and Engineering Aspects 614 (2021) 126144.

[100] D. Lohani, M.G. Basavaraj, D.K. Satapathy, S. Sarkar, Coupled effect of concentration, particle size and substrate morphology on the formation of coffee rings, Colloids and Surfaces A: Physicochemical and Engineering Aspects 589 (2020) 124387.

[101] Y.J. Carreón, M.L. Gómez-López, O. Díaz-Hernández, P. Vazquez-Vergara, R.E. Moctezuma, J.M. Saniger, J. González-Gutiérrez, Patterns in dried droplets to detect unfolded BSA, Sensors 22(3) (2022) 1156.

[102] Y.J. Cha, S.M. Park, R. You, H. Kim, D.K. Yoon, Microstructure arrays of DNA using topographic control, Nature communications 10(1) (2019) 2512.

[103] X. Fang, B. Li, E. Petersen, Y.-S. Seo, V.A. Samuilov, Y. Chen, J.C. Sokolov, C.-Y. Shew, M.H. Rafailovich, Drying of DNA droplets, Langmuir 22(14) (2006) 6308-6312.

[104] J.D. Watson, F. Crick, A structure for deoxyribose nucleic acid, (1953).

[105] G.S. Khan, A. Shah, D. Barker, Chemistry of DNA minor groove binding agents, Journal of photochemistry and photobiology B: Biology 115 (2012) 105-118.

[106] A. Travers, G. Muskhelishvili, DNA structure and function, The FEBS journal 282(12) (2015) 2279-2295.

[107] T. Roy, K. Szuttor, J. Smiatek, C. Holm, S. Hardt, Conformation and dynamics of long-chain end-tethered polymers in microchannels, Polymers 11(3) (2019) 488.

[108] C. Fonseca Guerra, F.M. Bickelhaupt, J.G. Snijders, E.J. Baerends, The nature of the hydrogen bond in DNA base pairs: the role of charge transfer and resonance assistance, Chemistry–A European Journal 5(12) (1999) 3581-3594.

[109] C.L. Kielkopf, S. White, J.W. Szewczyk, J.M. Turner, E.E. Baird, P.B. Dervan, D.C. Rees, A structural basis for recognition of A· T and T· A base pairs in the minor groove of B-DNA, Science 282(5386) (1998) 111-115.

[110] X. Gao, P. Mirau, D.J. Patel, Structure refinement of the chromomycin dimer-DNA oligomer complex in solution, Journal of molecular biology 223(1) (1992) 259-279.

[111] Y. Li, Z. Zhao, M.L. Lam, W. Liu, P.P. Yeung, C.-C. Chieng, T.-H. Chen, Hybridization-induced suppression of coffee ring effect for nucleic acid detection, Sensors and Actuators B: Chemical 206 (2015) 56-64.

[112] A. Askounis, Y. Takata, K. Sefiane, V. Koutsos, M.E. Shanahan, "Biodrop" evaporation and ring-stain deposits: The significance of DNA length, Langmuir 32(17) (2016) 4361-4369.

[113] P. Tropberger, R. Schneider, Going global: novel histone modifications in the globular domain of H3, Epigenetics 5(2) (2010) 112-117.

[114] M. Grunstein, Histones as regulators of genes, Scientific American 267(4) (1992) 68-75.

[115] C.L. Peterson, M.-A. Laniel, Histones and histone modifications, Current Biology 14(14) (2004) R546-R551.

[116] A.J. Bannister, T. Kouzarides, Regulation of chromatin by histone modifications, Cell research 21(3) (2011) 381-395.

[117] M. Ljungman, P.C. Hanawalt, Efficient protection against oxidative DNA damage in chromatin, Molecular carcinogenesis 5(4) (1992) 264-269.

[118] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, P. Walter, Chromosomal DNA and its packaging in the chromatin fiber, Molecular Biology of the Cell. 4th edition, Garland Science (2002).

[119] G. Arents, R.W. Burlingame, B.-C. Wang, W.E. Love, E.N. Moudrianakis, The nucleosomal core histone octamer at 3.1 A resolution: a tripartite protein assembly and a left-handed superhelix, Proceedings of the National Academy of Sciences 88(22) (1991) 10148-10152.

[120] C.A. Davey, D.F. Sargent, K. Luger, A.W. Maeder, T.J. Richmond, Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 Å resolution, Journal of molecular biology 319(5) (2002) 1097-1113.

[121] C.M. Wood, J.M. Nicholson, S.J. Lambert, L. Chantalat, C.D. Reynolds, J.P. Baldwin, High-resolution structure of the native histone octamer, Acta Crystallographica Section F: Structural Biology and Crystallization Communications 61(6) (2005) 541-545.

[122] R.K. McGinty, S. Tan, Histone, nucleosome, and chromatin structure, Fundamentals of chromatin, Springer (2013) pp. 1-28.

[123] V. Graziano, S. Gerchman, D. Schneider, V. Ramakrishnan, Histone H1 is located in the interior of the chromatin 30-nm filament, Nature 368(6469) (1994) 351-354.

[124] V. Ramakrishnan, Histone H1 and chromatin higher-order structure, Critical Reviews™ in Eukaryotic Gene Expression 7(3) (1997).

[125] K. Rippe, J. Mazurkiewicz, N. Kepper, Interactions of histones with DNA: nucleosome assembly, stability, dynamics, and higher order structure, DNA Interactions with Polymers and Surfactants. John Wiley & Sons, Isnc (2008) 135-72.

[126] M. Noll, R.D. Kornberg, Action of micrococcal nuclease on chromatin and the location of histone H1, Journal of molecular biology 109(3) (1977) 393-404.

[127] M. Klug, Dynamics of DNA methylation in differentiating hematopoietic cells, Doctoral dissertation (2010).

[128] Y. Ofran, V. Mysore, B. Rost, Prediction of DNA-binding residues from sequence, Bioinformatics 23(13) (2007) i347-i353.

[129] R.A.C. Ferraz, A.L.G. Lopes, J.A.F. da Silva, D.F.V. Moreira, M.J.N. Ferreira, S.V. de Almeida Coimbra, DNA–protein interaction studies: a historical and comparative analysis, Plant Methods 17(1) (2021) 1-21.

[130] M. Leng, G. Felsenfeld, The preferential interactions of polylysine and polyarginine with specific base sequences in DNA, Proceedings of the National Academy of Sciences 56(4) (1966) 1325-1332.

[131] N.C. Seeman, J.M. Rosenberg, A. Rich, Sequence-specific recognition of double helical nucleic acids by proteins, Proceedings of the National Academy of Sciences 73(3) (1976) 804-808.

[132] D.B. McKay, T.A. Steitz, Structure of catabolite gene activator protein at 2.9 Å resolution suggests binding to left-handed B-DNA, Nature 290(5809) (1981) 744-749.

[133] C.O. Pabo, R.T. Sauer, Protein-DNA recognition, Annual review of biochemistry 53(1) (1984) 293-321.

[134] B.W. Matthews, No code for recognition, Nature 335(6188) (1988) 294-295.

[135] N.M. Luscombe, R.A. Laskowski, J.M. Thornton, Amino acid–base interactions: a three-dimensional analysis of protein–DNA interactions at an atomic level, Nucleic acids research 29(13) (2001) 2860-2874.

[136] L. Jen-Jacobson, L.E. Engler, L.A. Jacobson, Structural and thermodynamic strategies for site-specific DNA binding proteins, Structure 8(10) (2000) 1015-1023.

[137] E.C. Beckwitt, M. Kong, B. Van Houten, Studying protein-DNA interactions using atomic force microscopy, Seminars in cell & developmental biology, Elsevier (2018) pp. 220-230.

[138] M. Renz, Preferential and cooperative binding of histone I to chromosomal mammalian DNA, Proceedings of the National Academy of Sciences 72(2) (1975) 733-736.

[139] J. Zlatanova, J. Yaneva, DNA sequence specific interactions of histone H1, Molecular biology reports 15 (1991) 53-56.

[140] A.T. Rodríguez, L. Pérez, F. Morán, F. Montero, P. Suau, Cooperative interaction of the C-terminal domain of histone H1 with DNA, Biophysical chemistry 39(2) (1991) 145-152.

[141] T. Diez-Caballero, F.X. Aviles, A. Albert, Specific interaction of histone H1 with eukaryotic DNA, Nucleic Acids Research 9(6) (1981) 1383-1394.

[142] L. Liao, R. Cole, Differences among subfractions of H1 histone in their interactions with linear and superhelical DNA. Circular dichroism, Journal of Biological Chemistry 256(13) (1981) 6751-6755.

[143] A.L. Turner, M. Watson, O.G. Wilkins, L. Cato, A. Travers, J.O. Thomas, K. Stott, Highly disordered histone H1− DNA model complexes and their condensates, Proceedings of the National Academy of Sciences 115(47) (2018) 11964-11969.

[144] Z. Al-Natour, A.H. Hassan, Effect of salt on the binding of the linker histone H1 to DNA and nucleosomes, DNA and cell biology 26(6) (2007) 445-452.

[145] R. Vila, I. Ponte, M. Collado, J.L.R. Arrondo, P. Suau, Induction of secondary structure in a COOH-terminal peptide of histone H1 by interaction with the DNA: an infrared spectroscopy study, Journal of Biological Chemistry 276(33) (2001) 30898-30903.

[146] M.L.S. Mello, B. Vidal, Changes in the infrared microspectroscopic characteristics of DNA caused by cationic elements, different base richness and single-stranded form, (2012) e43169.

[147] W. Sun, W. Liu, Z. Wu, H. Chen, Chemical surface modification of polymeric biomaterials for biomedical applications, Macromolecular Rapid Communications 41(8) (2020) 1900430.

[148] D.N.G. Krishna, J. Philip, Review on surface-characterization applications of X-ray photoelectron spectroscopy (XPS): Recent developments and challenges, Applied Surface Science Advances 12 (2022) 100332.

[149] T. Šušteršič, V. Gribova, M. Nikolic, P. Lavalle, N. Filipovic, N.E. Vrana, The effect of machine learning algorithms on the prediction of layer-by-layer coating properties, ACS omega 8(5) (2023) 4677-4686.

[150] S.K. Nemani, R.K. Annavarapu, B. Mohammadian, A. Raiyan, J. Heil, M.A. Haque, A. Abdelaal, H. Sojoudi, Surface modification of polymers: methods and applications, Advanced Materials Interfaces 5(24) (2018) 1801247.

[151] W. Huang, A. Samanta, Y. Chen, S. Baek, S.K. Shaw, H. Ding, Machine learning model for understanding laser superhydrophobic surface functionalization, Journal of manufacturing processes 69 (2021) 491-502.

[152] H. Busscher, A. Van Pelt, P. De Boer, H. De Jong, J. Arends, The effect of surface roughening of polymers on measured contact angles of liquids, Colloids and surfaces 9(4) (1984) 319-331.

[153] V. Belaud, S. Valette, G. Stremsdoerfer, M. Bigerelle, S. Benayoun, Wettability versus roughness: Multi-scales approach, Tribology International 82 (2015) 343-349.

[154] C.-J. Liang, J.-D. Liao, A.-J. Li, C. Chen, H.-Y. Lin, X.-J. Wang, Y.-H. Xu, Relationship between wettabilities and chemical compositions of candle soots, Fuel 128 (2014) 422-427.

[155] J. Zhang, Y. Han, A topography/chemical composition gradient polystyrene surface: toward the investigation of the relationship between surface wettability and surface structure and chemical composition, Langmuir 24(3) (2008) 796-801.

[156] M.S. Kim, G. Khang, H.B. Lee, Gradient polymer surfaces for biomedical applications, Progress in polymer science 33(1) (2008) 138-164.

[157] H. Uyen, J. Schakenraad, J. Sjollema, J. Noordmans, W. Jongebloed, I. Stokroos, H. Busscher, Amount and surface structure of albumin adsorbed to solid substrata with different wettabilities in a parallel plate flow cell, Journal of biomedical materials research 24(12) (1990) 1599-1614.

[158] Y. Tang, J. Wei, Y. Liu, Y. Chang, J. Zheng, Machine Learning Aided Design and Optimization of Antifouling Surfaces, Langmuir 40(43) (2024) 22504-22515.

[159] W. Yang, S. Chen, G. Cheng, H. Vaisocherova, H. Xue, W. Li, J. Zhang, S. Jiang, Film thickness dependence of protein adsorption from blood serum and plasma onto poly (sulfobetaine)-grafted surfaces, Langmuir 24(17) (2008) 9211-9214.

[160] V. Gribova, A. Navalikhina, O. Lysenko, C. Calligaro, E. Lebaudy, L. Deiber, B. Senger, P. Lavalle, N.E. Vrana, Prediction of coating thickness for polyelectrolyte multilayers via machine learning, Scientific Reports 11(1) (2021) 18702.

[161] J. Kim, A. Patra, S. Pal, N.L. Abbott, J. Lahann, Emergent Properties, Functions, and Applications of Phane-Based Polymers, Advanced Functional Materials 34(47) (2024) 2315891.

[162] Z. Hassan, D. Varadharajan, C. Zippel, S. Begum, J. Lahann, S. Bräse, Design Strategies for Structurally Controlled Polymer Surfaces via Cyclophane-Based CVD Polymerization and Post-CVD Fabrication, Advanced Materials 34(37) (2022) 2201761.

[163] B.J. Coelho, J.V. Pinto, J. Martins, A. Rovisco, P. Barquinha, E. Fortunato, P.V. Baptista, R. Martins, R. Igreja, Parylene C as a multipurpose material for electronics and microfluidics, Polymers 15(10) (2023) 2277.

[164] T. Moss, A. Greiner, Functionalization of Poly (para-xylylene) s-Opportunities and Challenges as Coating Material, Advanced Materials Interfaces 7(11) (2020) 1901858.

[165] J. Kim, S.C. Jang, K. Bae, J. Park, H.-D. Kim, J. Lahann, H.-S. Kim, K.J. Lee, Chemically tunable organic dielectric layer on an oxide TFT: poly (p-xylylene) derivatives, ACS Applied Materials & Interfaces 13(36) (2021) 43123-43133.

[166] J.E. Puskas, L.G. Muñoz-Robledo, R.A. Hoerr, J. Foley, S.P. Schmidt, M. Evancho-Chapman, J. Dong, C. Frethem, G. Haugstad, Drug-eluting stent coatings, Wiley Interdisciplinary Reviews: Nanomedicine and Nanobiotechnology 1(4) (2009) 451-462.

[167] C.J. Friedmann, S. Ay, S. Bräse, Improved synthesis of enantiopure 4-hydroxy [2.2] paracyclophane, The Journal of Organic Chemistry 75(13) (2010) 4612-4614.

[168] C. Zippel, T. Bartholomeyzik, C. Friedmann, M. Nieger, Z. Hassan, S. Bräse, Regioselective ortho-Palladation of [2.2] Paracyclophane Scaffolds: Accessing Planar and Central Chiral N, C-Palladacycles, European Journal of Organic Chemistry 2021(36) (2021) 5090-5093.

[169] P.-J. Chen, H.-Y. Chen, W.-B. Tsai, Fabrication of Low-Fouling Surfaces on Alkyne-Functionalized Poly-(p-xylylenes) Using Click Chemistry, Polymers 14(2) (2022) 225.

[170] X. Deng, J. Lahann, Orthogonal surface functionalization through bioactive vapor-based polymer coatings, Journal of Applied Polymer Science (2014) 131(14).

[171] W.F. Gorham, A new, general synthetic method for the preparation of linear poly-p-xylylenes, Journal of Polymer Science Part A-1: Polymer Chemistry 4(12) (1966) 3027-3039.

[172] Y. Elkasabi, M. Yoshida, H. Nandivada, H.Y. Chen, J. Lahann, Towards Multipotent Coatings: Chemical Vapor Deposition and Biofunctionalization of Carbonyl-Substituted Copolymers, Macromolecular Rapid Communications 29(11) (2008) 855-870.

[173] H.-Y. Chen, M. Hirtz, X. Deng, T. Laue, H. Fuchs, J. Lahann, Substrate-independent dip-pen nanolithography based on reactive coatings, Journal of the American Chemical Society 132(51) (2010) 18023-18025.

[174] X. Jiang, H.-Y. Chen, G. Galvan, M. Yoshida, J. Lahann, Vapor-based initiator coatings for atom transfer radical polymerization, Advanced Functional Materials (2008) 18 (1), 27-35.

[175] H. Nandivada, H.Y. Chen, J. Lahann, Vapor-based synthesis of poly [(4-formyl-p-xylylene)-co-(p-xylylene)] and its use for biomimetic surface modifications, Macromolecular rapid communications 26(22) (2005) 1794-1799.

[176] A. Kausar, Polymer coating technology for high performance applications: Fundamentals and advances, Journal of Macromolecular Science, Part A 55(5) (2018) 440-448.

[177] T.M. Hafshejani, X. Zhong, J. Kim, B. Dadfar, J. Lahann, Chemical and Topological Control of Surfaces Using Functional Parylene Coatings, Organic Materials 5(02) (2023) 98-111.

[178] A.M. Ross, D. Zhang, X. Deng, S.L. Chang, J. Lahann, Chemical-vapor-deposition-based polymer substrates for spatially resolved analysis of protein binding by imaging ellipsometry, Analytical chemistry 83(3) (2011) 874-880.

[179] J. Lahann, Reactive polymer coatings for biomimetic surface engineering, Chemical Engineering Communications 193(11) (2006) 1457-1468.

[180] J. Lahann, Vapor-based polymer coatings for potential biomedical applications, Polymer international 55(12) (2006) 1361-1370.

[181] D. Klemm, B. Heublein, H.P. Fink, A. Bohn, Cellulose: fascinating biopolymer and sustainable raw material, Angewandte chemie international edition 44(22) (2005) 3358-3393.

[182] K.Y. Suh, R. Langer, J. Lahann, A novel photoderinable reactive polymer coating and its use for microfabrication of hydrogel elements, (2004) 1401-1405.

[183] S. Vaez, D. Shahbazi, M. Koenig, M. Franzreb, J. Lahann, Deep Learning Based Surface Classification of Functionalized Polymer Coatings, Langmuir (2025) in press.

[184] M.M. Byranvand, F. Behboodi-Sadabad, A.A. Eliwi, V. Trouillet, A. Welle, S. Ternes, I.M. Hossain, M.R. Khan, J.A. Schwenzer, A. Farooq, Chemical vapor deposited polymer layer for efficient passivation of planar perovskite solar cells, Journal of Materials Chemistry A 8(38) (2020) 20122-20132.

[185] S. Sah, Machine learning: a review of learning types, (2020).

[186] P. Chaovalit, L. Zhou, Movie review mining: A comparison between supervised and unsupervised classification approaches, Proceedings of the 38th annual Hawaii international conference on system sciences, IEEE (2005) pp. 112c-112c.

[187] E.F. Morales, H.J. Escalante, A brief introduction to supervised, unsupervised, and reinforcement learning, Biosignal processing and classification using computational learning and intelligence, Elsevier (2022) pp. 111-129.

[188] V. Nasteski, An overview of the supervised machine learning methods, Horizons. b 4(51-62) (2017) 56.

[189] S.B. Kotsiantis, I. Zaharakis, P. Pintelas, Supervised machine learning: A review of classification techniques, Emerging artificial intelligence applications in computer engineering 160(1) (2007) 3-24.

[190] P.P. Shinde, S. Shah, A review of machine learning and deep learning applications, 2018 Fourth international conference on computing communication control and automation (ICCUBEA), IEEE (2018) pp. 1-6.

[191] G.E. Hinton, S. Osindero, Y.-W. Teh, A fast learning algorithm for deep belief nets, Neural computation 18(7) (2006) 1527-1554.

[192] A. Esteva, A. Robicquet, B. Ramsundar, V. Kuleshov, M. DePristo, K. Chou, C. Cui, G. Corrado, S. Thrun, J. Dean, A guide to deep learning in healthcare, Nature medicine 25(1) (2019) 24-29.

[193] L. Deng, A tutorial survey of architectures, algorithms, and applications for deep learning, APSIPA transactions on Signal and Information Processing 3 (2014) e2.

[194] Y. Guo, Y. Liu, A. Oerlemans, S. Lao, S. Wu, M.S. Lew, Deep learning for visual understanding: A review, Neurocomputing 187 (2016) 27-48.

[195] Y.-D. Zhang, X. Jiang, S.-H. Wang, Fingerspelling recognition by 12-layer CNN with stochastic pooling, Mobile Networks and Applications (2022) 1-13.

[196] F.M. Shiri, T. Perumal, N. Mustapha, R. Mohamed, A Comprehensive Overview and Comparative Analysis on Deep Learning Models, CNN, RNN, LSTM, GRU, arXiv preprint arXiv (2023) 2305.17473.

[197] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition, Proceedings of the IEEE 86(11) (1998) 2278-2324.

[198] Z. Li, F. Liu, W. Yang, S. Peng, J. Zhou, A survey of convolutional neural networks: analysis, applications, and prospects, IEEE transactions on neural networks and learning systems 33(12) (2021) 6999-7019.

[199] P. Arena, A. Basile, M. Bucolo, L. Fortuna, Image processing for medical diagnosis using CNN, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment 497(1) (2003) 174-178.

[200] N. Sharma, V. Jain, A. Mishra, An analysis of convolutional neural networks for image classification, Procedia computer science 132 (2018) 377-384.

[201] J. Du, Understanding of object detection based on CNN family and YOLO, Journal of Physics: Conference Series, IOP Publishing, (2018) p. 012029.

[202] M. Coşkun, A. Uçar, Ö. Yildirim, Y. Demir, Face recognition based on convolutional neural network, 2017 international conference on modern electrical and energy systems (MEES), IEEE (2017) pp. 376-379.

[203] M. Sahu, R. Dash, A survey on deep learning: convolution neural network (CNN), Intelligent and Cloud Computing: Proceedings of ICICC 2019, Volume 2, Springer (2021) pp. 317-325.

[204] M. Han, J. Chen, L. Li, Y. Chang, Visual hand gesture recognition with convolution neural network, 2016 17th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD), IEEE (2016) pp. 287-291.

[205] Y. Liu, H. Pu, D.-W. Sun, Efficient extraction of deep image features using convolutional neural network (CNN) for applications in detecting and analysing complex food matrices, Trends in Food Science & Technology 113 (2021) 193-204.

[206] T. Adriyanto, R.A. Ramadhani, R. Helilintar, A. Ristyawan, Classification of dog and cat images using the CNN method, Ilk. J. Ilm 14(3) (2022) 203-208.

[207] R. Yamashita, M. Nishio, R.K.G. Do, K. Togashi, Convolutional neural networks: an overview and application in radiology, Insights into imaging 9 (2018) 611-629.

[208] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, Advances in neural information processing systems (2012) 25.

[209] X. Zhou, Understanding the convolutional neural networks with gradient descent and backpropagation, Journal of Physics: Conference Series, IOP Publishing (2018) p. 012028.

[210] M.M. Taye, Theoretical understanding of convolutional neural network: Concepts, architectures, applications, future directions, Computation 11(3) (2023) 52.

[211] M.D. Zeiler, Hierarchical convolutional deep learning in computer vision, Doctoral dissertation (2013).

[212] K. Dutta, R. Lenka, M.S. Sarowar, Improvement of Denoising in Images Using Generic Image Denoising Network (GID Net), 2021 IEEE 2nd International Conference on Applied Electromagnetics, Signal Processing, & Communication (AESPC), IEEE (2021) pp. 1-6.

[213] F. Alrasheedi, X. Zhong, P.-C. Huang, Padding module: Learning the padding in deep neural networks, IEEE Access 11 (2023) 7348-7357.

[214] M. Varshney, P. Singh, Optimizing nonlinear activation function for convolutional neural networks, Signal, Image and Video Processing 15(6) (2021) 1323-1330.

[215] W. Hao, W. Yizhou, L. Yaqin, S. Zhili, The role of activation function in CNN, 2020 2nd International Conference on Information Technology and Computer Application (ITCA), IEEE (2020) pp. 429-432.

[216] Y. Wang, Y. Li, Y. Song, X. Rong, The influence of the activation function in a convolution neural network model of facial expression recognition, Applied Sciences 10(5) (2020) 1897.

[217] M.A. Mercioni, S. Holban, The most used activation functions: Classic versus current, 2020 International Conference on Development and Application Systems (DAS), IEEE (2020) pp. 141-145.

[218] M.M. Adnan, M.S.M. Rahim, A.R. Khan, T. Saba, S.M. Fati, S.A. Bahaj, An improved automatic image annotation approach using convolutional neural network-slantlet transform, IEEE Access 10 (2022) 7520-7532.

[219] D. Scherer, A. Müller, S. Behnke, Evaluation of pooling operations in convolutional architectures for object recognition, International conference on artificial neural networks, Springer (2010) pp. 92-101.

[220] S.S. Basha, S.R. Dubey, V. Pulabaigari, S. Mukherjee, Impact of fully connected layers on performance of convolutional neural networks for image classification, Neurocomputing 378 (2020) 112-119.

[221] N.M. Nawi, R.S. Ransing, M.N.M. Salleh, R. Ghazali, N.A. Hamid, An improved back propagation neural network algorithm on classification problems, International Conferences, DTA and BSBT 2010, Proceedings, Springer (2010) pp. 177-188

[222] M.R. Rezaei-Dastjerdehei, A. Mijani, E. Fatemizadeh, Addressing imbalance in multi-label classification using weighted cross entropy loss function, 2020 27th national and 5th international iranian conference on biomedical engineering (ICBME), IEEE (2020) pp. 333-338.

[223] M. Kohler, S. Langer, Statistical theory for image classification using deep convolutional neural networks with cross-entropy loss, arXiv preprint arXiv, (2020) 2011.13602.

[224] E.M. Dogo, O. Afolabi, N. Nwulu, B. Twala, C. Aigbavboa, A comparative analysis of gradient descent-based optimization algorithms on convolutional neural networks, 2018 international conference on computational techniques, electronics and mechanical systems (CTEMS), IEEE (2018) pp. 92-99.

[225] D. Ueda, A. Yamamoto, T. Takashima, N. Onoda, S. Noda, S. Kashiwagi, T. Morisaki, T. Honjo, A. Shimazaki, Y. Miki, Training, validation, and test of deep learning models for classification of receptor expressions in breast cancers from mammograms, JCO Precision Oncology 5 (2021) 543-551.

[226] T. Eelbode, P. Sinonquel, F. Maes, R. Bisschops, Pitfalls in training and validation of deep learning systems, Best Practice & Research Clinical Gastroenterology 52 (2021) 101712.

[227] A. Vabalas, E. Gowen, E. Poliakoff, A.J. Casson, Machine learning algorithm validation with a limited sample size, PloS one 14(11) (2019) e0224365.

[228] E. Hoffer, I. Hubara, D. Soudry, Train longer, generalize better: closing the generalization gap in large batch training of neural networks, Advances in neural information processing systems, (2017) 30.

[229] Z.S. Aaraji, H.H. Abbas, Automatic Classification of Alzheimer's disease using brain MRI data and deep Convolutional Neural Networks, arXiv preprint arXiv, (2022) 2204.00068.

[230] Y. Guo, Deep learning for visual understanding, Doctoral dissertation (2017).

[231] M.D. Zeiler, R. Fergus, Visualizing and understanding convolutional networks, Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part I 13, Springer (2014) pp. 818-833.

[232] R.R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, D. Batra, Grad-cam: Visual explanations from deep networks via gradient-based localization, Proceedings of the IEEE international conference on computer vision (2017) pp. 618-626.

[233] G.E. Hinton, S. Roweis, Stochastic neighbor embedding, Advances in neural information processing systems (2002) 15.

[234] L. Van der Maaten, G. Hinton, Visualizing data using t-SNE, Journal of machine learning research 9(11) (2008) 2579-605.

[235] C. Xiao, S. Hong, W. Huang, Optimizing graph layout by t-SNE perplexity estimation, International Journal of Data Science and Analytics 15(2) (2023) 159-171.

[236] H. Sahoo, Optical spectroscopic and microscopic techniques, Springer (2022).

[237] C. Dysli, S. Wolf, M.Y. Berezin, L. Sauer, M. Hammer, M.S. Zinkernagel, Fluorescence lifetime imaging ophthalmoscopy, Progress in retinal and eye research 60 (2017) 120-143.

[238] L.A. Bagatolli, Fluorescence spectroscopy: basic foundations and methods, Analytical techniques in the pharmaceutical sciences (2016) 29-59.

[239] A.H. Alami, S. Alasad, H. Aljaghoub, M. Ayoub, A. Alashkar, A. Mdallal, R. Hasan, Characterization Techniques for Photovoltaics Manufacturing, PV Technology and Manufacturing, Springer (2023) pp. 139-153.

[240] J.R. Lakowicz, J.R. Lakowicz, Instrumentation for fluorescence spectroscopy, Principles of fluorescence spectroscopy (1999) 25-61.

[241] B. Valeur, M.N. Berberan-Santos, Molecular fluorescence: principles and applications, John Wiley & Sons (2013).

[242] H. Bisswanger, Enzyme kinetics: principles and methods, John Wiley & Sons (2017).

[243] J. Chan, S.C. Dodani, C.J. Chang, Reaction-based small-molecule fluorescent probes for chemoselective bioimaging, Nature chemistry 4(12) (2012) 973-984.

[244] J.C. Vickerman, D. Briggs, ToF-SIMS_an overview, surface analysis by mass spectrometry, (2001) 1-40.

[245] I.S. Gilmore, SIMS of organics-Advances in 2D and 3D imaging and future outlook, Journal of Vacuum Science & Technology A 31(5) (2013) 050819.

[246] N. Tsuyama, H. Mizuno, T. Masujima, Mass spectrometry for cellular and tissue analyses in a very small region, Analytical Sciences 27(2) (2011) 163-163.

[247] R.G. Messerschmidt, M.A. Harthcock, Infrared microspectroscopy. Theory and applications, (1988).

[248] B.C. Smith, Fundamentals of Fourier transform infrared spectroscopy, CRC press (2011).

[249] W. Kemp, Organic spectroscopy, Bloomsbury Publishing (2017).

[250] T.M. Hafshejani, Atomic-Scale Investigation of Strain Effect on Surface Properties of Silicon and Mineral Materials, Doctoral dissertation (2021).

[251] B.E. Hayden, Reflection absorption infrared spectroscopy, Vibrational spectroscopy of molecules on surfaces, Springer (1987) pp. 267-344.

[252] A. Bradshaw, E. Schweizer, Infrared reflection-absorption spectroscopy of adsorbed molecules, Advances in spectroscopy (1986) 16 (1988) 413-483.

[253] M.D. Porter, T.B. Bright, D.L. Allara, T. Kuwana, Quantitative aspects of infrared external reflection spectroscopy: polymer/glassy carbon interface, Analytical Chemistry 58(12) (1986) 2461-2465.

[254] J.S. Wong, Y.-S. Yen, Intriguing absorption band behavior of IR reflectance spectra of silicon dioxide on silicon, Applied spectroscopy 42(4) (1988) 598-604.

[255] H. Samuel, E. Etim, U. Nweke-Maraizu, B. Bako, J. Shinggu, Advances in Experimental Techniques for Corrosion Inhibition Studies: Insights and Applications, Journal of Applied Sciences and Environmental Management 27(12) (2023) 2957-2966.

[256] https://facultyweb.kennesaw.edu.

[257] A.J. Miles, B.A. Wallace, Circular dichroism spectroscopy of membrane proteins, Chemical society reviews 45(18) (2016) 4859-4872.

[258] http://www.physicsbootcamp.org/section-polarization-of-light.html.

[259] S. Subadini, P.R. Hota, D.P. Behera, H. Sahoo, Circular Dichroism Spectroscopy: Principle and Application, Optical Spectroscopic and Microscopic Techniques: Analysis of Biological Molecules, Springer (2022) pp. 19-33.

[260] L. Errede, M. Szwarc, Chemistry of p-xylylene, its analogues, and polymers, Quarterly Reviews, Chemical Society 12(4) (1958) 301-320.

[261] S. Vaez, B. Dadfar, M. Koenig, M. Franzreb, J. Lahann, Deep Learning-Based Classification of Histone–DNA Interactions Using Drying Droplet Patterns, Small Science (2024) 2400252.

[262] A.J. Geall, I.S. Blagbrough, Rapid and sensitive ethidium bromide fluorescence quenching assay of polyamine conjugate–DNA interactions for the analysis of lipoplex formation in gene therapy, Journal of pharmaceutical and biomedical analysis 22(5) (2000) 849-859.

[263] D.C. Montgomery, G.C. Runger, Applied statistics and probability for engineers, John wiley & sons (2010).

[264] F.-Y. Chou, T.C. Ramli, C.-Y. Lee, S.-M. Hu, J. Christy, H.-Y. Chen, Vapor-Deposited Polymer Films and Structure: Methods and Applications, Organic Materials 5(02) (2023) 118-138.

[265] R. Kothinti, N.M. Tabatabai, D.H. Petering, Electrophoretic mobility shift assay of zinc finger proteins: competition for Zn2+ bound to Sp1 in protocols including EDTA, Journal of inorganic biochemistry 105(4) (2011) 569-576.

[266] Y. Zhang, P.S. Cremer, Interactions between macromolecules and ions: the Hofmeister series, Current opinion in chemical biology 10(6) (2006) 658-663.

[267] J. Zhang, Protein-protein interactions in salt solutions, Protein-protein interactions-computational and experimental tools 6 (2012) 359-376.

[268] A. Pal, A. Gope, A. Sengupta, Drying of bio-colloidal sessile droplets: Advances, applications, and perspectives, Advances in Colloid and Interface Science 314 (2023) 102870.

[269] L. Hamadeh, S. Imran, M. Bencsik, G.R. Sharpe, M.A. Johnson, D.J. Fairhurst, Machine learning analysis for quantitative discrimination of dried blood droplets, Scientific reports 10(1) (2020) 3313.

[270] R.R. Selvaraju, A. Das, R. Vedantam, M. Cogswell, D. Parikh, D. Batra, Grad-CAM: Why did you say that?, arXiv preprint arXiv:1611.07450 (2016).

[271] K. Faust, Q. Xie, D. Han, K. Goyle, Z. Volynskaya, U. Djuric, P. Diamandis, Visualizing histopathologic deep learning classification and anomaly detection using nonlinear feature space dimensionality reduction, BMC bioinformatics 19 (2018) 1-15.

[272] R. Rohs, S.M. West, A. Sosinsky, P. Liu, R.S. Mann, B. Honig, The role of DNA shape in protein–DNA recognition, Nature 461(7268) (2009) 1248-1253.

[273] R. Rohs, X. Jin, S.M. West, R. Joshi, B. Honig, R.S. Mann, Origins of specificity in protein-DNA recognition, Annual review of biochemistry 79 (2010) 233-269.

[274] F. Aviles, T. Diez-Caballero, J. Palau, A. Albert, On the interaction of histone H1 and H1 peptides with DNA: Sedimentation, thermal denaturation and solubility studies, Biochimie 60(5) (1978) 445-451.

[275] Y. Ünal, Ş. Öztürk, M.N. Dudak, M. Ekici, Comparison of current convolutional neural network architectures for classification of damaged and undamaged cars, Advances in Deep Learning, Artificial Intelligence and Robotics: Proceedings of the 2nd International Conference on Deep Learning, Artificial Intelligence and Robotics,(ICDLAIR) 2020, Springer (2022) pp. 141-149.

[276] S. Sarkar, S. Kundu, Protein (BSA) adsorption on hydrophilic and hydrophobic surfaces, Materials Today: Proceedings (2023) 1.

[277] M.K. Braun, A. Sauter, O. Matsarskaia, M. Wolf, F. Roosen-Runge, M. Sztucki, R. Roth, F. Zhang, F. Schreiber, Reentrant phase behavior in protein solutions induced by multivalent salts: strong effect of anions Cl–versus NO3–, The Journal of Physical Chemistry B 122(50) (2018) 11978-11985.

[278] S.E. Whang, Y. Roh, H. Song, J.-G. Lee, Data collection and quality challenges in deep learning: A data-centric ai perspective, The VLDB Journal 32(4) (2023) 791-813.