sustainability

Article

Towards Cost-Optimal Zero-Defect Manufacturing in Injection Molding: An Explainable and Transferable Machine Learning Framework

Lucas Greif *, Jonas Ortner , Peer Kummert, Andreas Kimmig, Simon Kreuzwieser, Jakob Bönsch and Jivka Ovtcharova

Institute for Information Management in Engineering, Karlsruhe Institute of Technology, 76133 Karlsruhe, Germany; jonas.ortner1@gmail.com (J.O.); jakob.boensch@kit.edu (J.B.)
* Correspondence: lucas.greif@kit.edu

Abstract

In the era of Industry 4.0, Zero-Defect Manufacturing is critical for injection molding but faces three major hurdles: severe class imbalance, the "black-box" nature of AI models, and the lack of scalability across machines. This study presents a comprehensive framework addressing these challenges. Using industrial datasets, we evaluated state-of-the-art supervised algorithms. Results show that CatBoost outperforms other architectures. Crucially, we demonstrate that maximizing accuracy is insufficient; instead, we introduce a cost-sensitive threshold optimization that minimizes economic risk, identifying an optimal classification threshold significantly lower than the standard. To enhance trust, SHAP analysis reveals that motor power and specific nozzle temperatures are the primary defect drivers. Finally, we validate a transfer learning approach using LightGBM, proving that models can be adapted to new datasets with minimal retraining. The implementation of cost-sensitive thresholding reduces total failure costs by over 75% compared to standard classification, while the transfer learning approach cuts the data requirements for new machine adaptation by more than half, providing a high-impact, scalable solution for sustainable smart manufacturing.

Keywords: Zero-Defect Manufacturing; injection molding; explainable artificial intelligence; transfer learning; cost-sensitive learning; imbalanced data; smart manufacturing

1. Introduction

The global competitive landscape requires the manufacturing industry to seamlessly integrate design and production processes to enhance both product quality and sustainability [1–3]. To address these demands, Zero-Defect Manufacturing (ZDM) has emerged as a key strategy within the Industry 4.0 and the emerging Industry 5.0 paradigms. ZDM synthesizes traditional quality management with advanced digital technologies to create comprehensive "digital manufacturing" ecosystems that aim to minimize waste and operational expenses [4–6]. Although the theoretical benefits of such systems are well-documented, their practical implementation faces significant hurdles, particularly regarding the need for human-centric interaction and interpretability in automated decision making [7].

In this context, the injection molding sector serves as a compelling case study. As a crucial process for the mass production of precise, complex geometries [8,9], it is inherently

Check for updates

Academic Editor: Manuela Ingaldi

Received: 12 January 2026
Revised: 8 February 2026
Accepted: 12 February 2026
Published: 15 February 2026

https://doi.org/10.3390/su18042001

susceptible to resource-intensive scrap generation when the process parameters deviate. Consequently, the industry requires robust, smart quality control systems to ensure consistency [10,11]. However, data-driven defect detection in this domain is complicated by three persistent challenges: First, the data is characterized by a strong class imbalance, as defective parts occur much less frequently than non-defective ones. Second, standard algorithms often optimize for accuracy rather than economic impact; yet, in mass production, a false negative (delivering a defect) causes significantly higher asymmetric costs than a false positive. Third, high-performing "black-box" models often lack the transparency required for operator trust.

Building on these challenges, this paper proposes a unified, cost-sensitive, and explainable framework for defect detection. We make the following key contributions:

- We provide a reproducible workflow, ranging from sensor data fusion and correlation-based feature selection to hyperparameter optimization.
- We present a comparative analysis of state-of-the-art supervised methods to detect defects, explicitly addressing class imbalance through a cost-sensitive optimization approach that minimizes the economic risk of false negatives.
- We incorporate explainable artificial intelligence (XAI) techniques, specifically SHAP analysis, to illuminate how critical parameters drive defect outcomes. This improves transparency and fosters trust in data-driven decisions.
- We demonstrate the scalability of the proposed framework by showing how knowledge gained from one dataset can be successfully transferred to a second dataset with fewer parameters via Transfer Learning.

The remainder of this paper is organized as follows. Section 2 outlines the theoretical foundations of ZDM and the role of XAI. Section 3 details the data set and the proposed methodological framework. Section 4 presents the results of the performance evaluation and cost optimization. Section 5 discusses the industrial implications of the findings, and Section 6 concludes the study.

2. Related Work and Theoretical Foundation

2.1. Zero-Defect Manufacturing (ZDM)

ZDM has evolved from a niche concept to a cornerstone strategy for enhancing product and process quality in Industry 4.0. Unlike traditional quality control, which often focuses on post-production screening, ZDM advocates for a holistic framework comprising four paired strategies: *detection* and *repair*, as well as *prediction* and *prevention* [12,13].

Extensive reviews of the field reveal a clear trajectory but also significant unevenness in implementation. Hundreds of articles were analyzed in recent studies, concluding that *detection* remains the dominant strategy, accounting for the vast majority of implementations [12,14], while industries such as semiconductors have reached high maturity in automated inspection, other sectors often lack integrated multi-strategy approaches. Consequently, the transition from purely reactive detection to proactive prevention remains a primary challenge [15].

A critical gap identified across recent literature is the lack of economic integration, while ZDM is theoretically linked to sustainability and waste reduction, it is argued that technical solutions are frequently decoupled from rigorous business model innovations [13]. Most studies optimize for technical metrics (e.g., accuracy or recall) but neglect cost-benefit analyses that quantify the financial trade-offs of implementation. This is particularly relevant in heavy industries; for instance, the efficacy of Gradient-Boosting Machines was demonstrated in the steel sector, achieving high predictive power ($R^2 > 0.97$) [16].

However, even such advanced data-driven approaches often fail to explicitly model the asymmetric costs of false positives versus false negatives, a key factor in economic viability.

Furthermore, as manufacturing moves towards Industry 5.0, the focus is expanding beyond technical robustness to include human-centric factors and broader sustainability goals. It is highlighted that future ZDM frameworks must integrate auxiliary processes, such as root cause analysis and operator training, to ensure that systems are not only accurate but also interpretable and manageable by humans [15].

In summary, while the technological foundations of ZDM (specifically detection algorithms) are well-established, there is a distinct need for research that: (1) moves beyond isolated detection to holistic closed-loop systems; (2) explicitly integrates economic cost models into technical optimization; (3) enhances transparency to support human decision making in complex manufacturing environments.

2.2. AI-Driven Quality Assurance and the Necessity of Explainability

In the specific context of injection molding, the application of AI has transitioned from simple statistical monitoring to complex, self-optimizing systems. A primary stream of research focuses on adaptive process control to mitigate variations. For instance, a back-propagation neural network (BPNN) was developed that adjusts switchover positions and injection speeds in real-time, reducing part weight variation significantly [17]. Similarly, a predictive control algorithm using CNNs and regression models was implemented [18]. Quality metrics were successfully stabilized within two cycles via an OPC UA communication platform, although surface quality prediction remained challenging. Beyond control strategies, detecting deviations is critical. It was demonstrated that ensemble methods, specifically Random Forests, outperform standalone classifiers in real-time fault detection for high-precision parts [19]. Extending this to signal processing, Wavelet transforms and principal component analysis (PCA) combined with ANNs were utilized to diagnose faults from cavity pressure profiles with high accuracy [20]. More recently, the ability to detect subtle trends that traditional control charts miss was shown through hybrid approaches combining statistical process control (SPC) with Deep Learning [21]. A recurring debate in this domain concerns the trade-off between model complexity and data availability. The superiority of Deep Learning (Autoencoders) over tree-based models like LightGBM for detecting subtle anomalies was argued [22]. Conversely, it was found that simpler linear models, when enriched with domain knowledge, can rival complex Deep Learning architectures [23]. Crucially, it was noted that expensive cavity sensors offer only marginal gains over standard machine data, supporting the use of cost-effective, machine-internal sensor data for scalable solutions, while predictive accuracy has improved, the black-box nature of these models remains a hurdle. This was addressed by applying SHAP analysis to ANNs, identifying that initial mold temperature and packing pressure are dominant quality drivers [24]. As AI systems are increasingly integrated into high-stakes decision making, the opacity of complex models undermines human–AI collaboration. Black-box predictions can lead to either overreliance or skepticism by operators, ultimately degrading decision quality [25]. XAI bridges this gap by rendering decisions transparent and interpretable. Evidence from diverse high-stakes fields validates this utility. In healthcare, it was shown that explaining AI rationale significantly improves physician trust and accuracy [26]. In finance, it was demonstrated that SHAP values not only satisfy regulatory transparency but also enhance risk assessment accuracy [27]. Similarly, in cybersecurity and education, visual explanations allow experts to discern when to trust the AI, significantly improving joint task performance [28,29].

2.3. Research Gap and Contribution

Despite notable advances in data-driven defect detection [17,23,24], existing solutions face three critical limitations that hinder industrial scalability.

First, the vast majority of studies focus on maximizing technical metrics such as accuracy or F1-score, disregarding the economic asymmetry of errors. In mass production, a false negative (delivering a defect) incurs significantly higher costs than a false positive (re-inspecting a good part). Current algorithms rarely incorporate this cost-sensitivity, leading to models that are technically accurate but economically suboptimal. Second, the reliance on "black-box" models (e.g., Deep Neural Networks) erodes operator trust, while XAI was introduced in prior work, widespread adoption is lacking, and models are often deployed without providing actionable root-cause insights for process engineers [24]. Third, scalability remains underexamined. Most approaches are trained and tested on isolated datasets [23]. The potential of *Transfer Learning* to adapt models to new machines or products with minimal retraining is largely unexplored in the injection molding domain.

To address these gaps, this study proposes a holistic, cost-optimal, and transparent framework. Our specific contributions are:

- We conduct a rigorous comparison of modern algorithms, including gradient boosting variants (CatBoost, LightGBM, XGBoost) and tabular transformers (SAINT), identifying CatBoost as the superior performer for this domain.
- Unlike previous works, we shift the optimization objective from pure accuracy to economic risk minimization. By implementing a cost-sensitive threshold calibration combined with SMOTE, we demonstrate how to minimize the financial impact of defective parts.
- We integrate SHAP to decode model decisions, revealing that specific physical parameters (e.g., nozzle temperature, motor power) are the primary drivers of defects, thus enabling targeted process adjustments.
- We provide empirical evidence that the developed models can be successfully transferred to secondary datasets, proving their adaptability to changing production environments without extensive data collection.

3. Methodology

3.1. Sample, Sources, and Data

The data for this study were collected from a high-volume industrial production line under real-world manufacturing conditions. The molded component is a high-precision spherical shell used in automotive assemblies, requiring exceptional dimensional stability with tolerances in the micrometer range. To achieve this level of precision, a state-of-the-art fully electric injection molding machine was utilized. The material processed is a specialized high-performance composite polymer reinforced with carbon fibers. The inclusion of carbon fibers significantly increases process complexity due to their influence on flow behavior, thermal conductivity, and fiber orientation, which are critical factors in the formation of quality defects.

This study utilizes two industrial datasets of injection molding cycles, each enriched with sensor information and binary quality labels (acceptable vs. defective). We divided the two datasets into one main dataset and one transfer dataset to transfer the models and findings from the main dataset. The main dataset contains a broader array of parameters (e.g., temperature, energy), while the transfer dataset focuses on core process control and counting parameters.

Table 1 provides a comparison of the main and transfer datasets in terms of feature categories.

Table 1. Comparison of feature categories in both datasets.

Category	M	T	Examples
Time Series Attributes	1	1	*M:* Cycle Start *T:* Cycle Start
Quality Attributes (Label)	1	1	*M:* Binary classification (0/1) *T:* Binary classification (0/1)
Temperature Parameters	5	0	*M:* Mold Temperature (1,3), Nozzle Temperature (1,2), Reference Temperature *T:* –
Energy Parameters	6	0	*M:* Energy Consumption Total, Energy Total, Heating Energy, Heating Power, Motor Power, Power Total *T:* –
Process Control Parameters	5	5	*M:* Injection Time, Injection Speed (1,3), Switching Time, Dosing Time *T:* Injection Time, Switching Pressure, Dosing Time, Mass Pad, Flow Rating
Material Properties	2	0	*M:* Shear Rate, Viscosity *T:* –
Process Monitoring	1	0	*M:* Switchover Monitoring *T:* –

While Table 1 delineates the qualitative differences in the available feature sets, the quantitative composition of the datasets is equally critical. For the main dataset, 6122 injection molding cycles were originally recorded from industrial machines at 20 s intervals. After discarding the first 210 timestamps due to misaligned time indices between the two datasets, 5912 time stamps remain for analysis. These samples span 23 parameters, including temperature, energy consumption, and process control features, along with a binary quality label. The transfer dataset consists of 7290 cycles and focuses primarily on 9 process control parameters, omitting temperature and energy measurements. Despite the reduced feature set, quality labels and time information remain consistent. In both cases, the datasets exhibit a significant class imbalance, which is a common challenge in zero-defect manufacturing. The resulting sample distributions and label ratios for both datasets are summarized in Table 2.

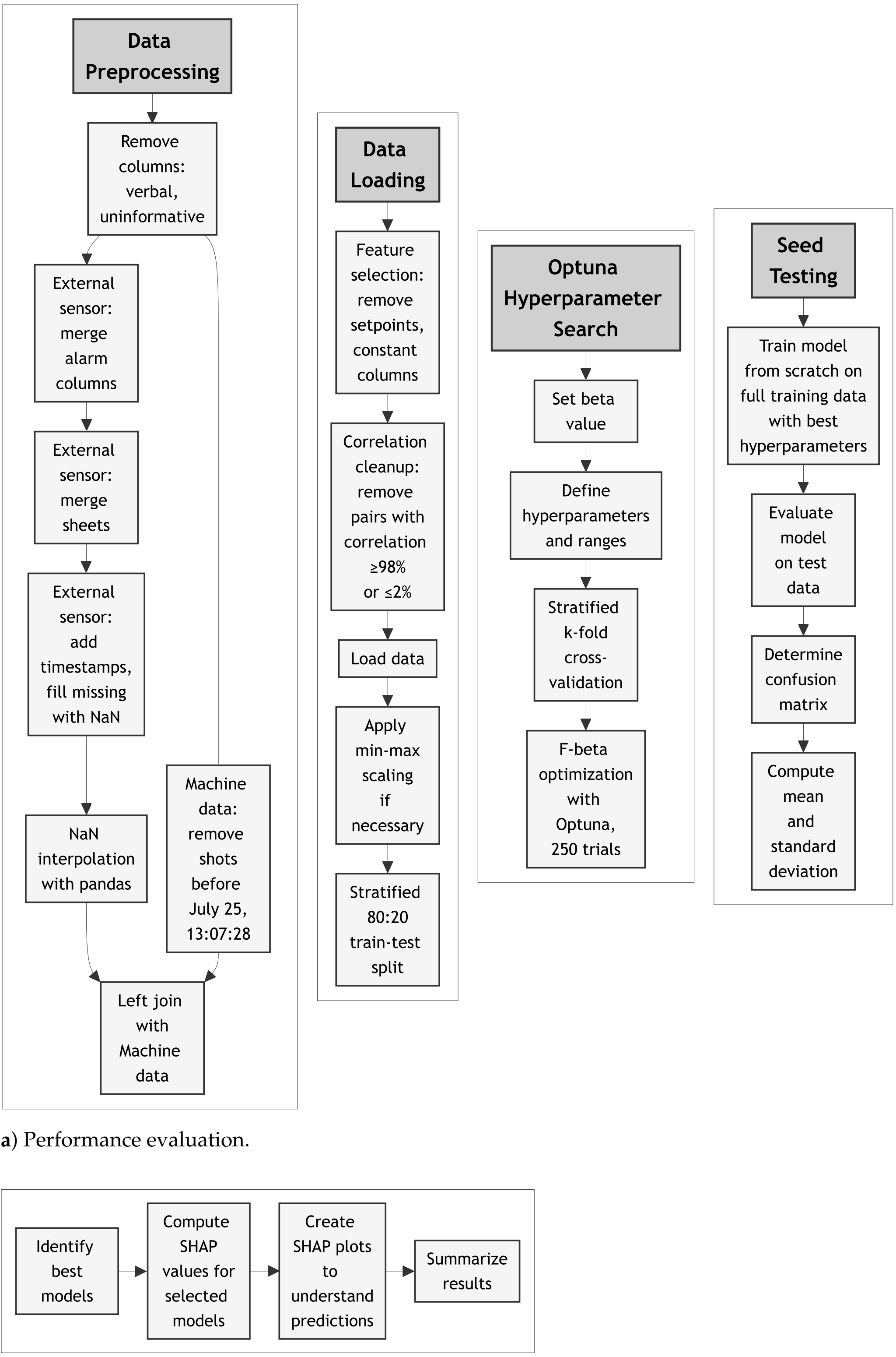
Table 2. Summary of good vs. defective parts of the two datasets.

Dataset	Total Samples	Good Parts	Defective Parts	Dimensionality
Main	5912	5715 (96.67%)	197 (3.33%)	22
Transfer	7290	6959 (95.46%)	331 (4.54%)	7

The limited proportion of defective parts in both datasets (3.33% and 4.54%, respectively) risks biasing standard classification models toward the majority class. Consequently, the subsequent modeling phases incorporate cost-sensitive calibration and resampling techniques to ensure that defective parts are accurately identified despite their rarity.

3.2. Data Analysis Procedure

The methodology, depicted in Figure 1 of the study, encompasses a structured sequential approach that spans five main stages: data preprocessing and merging, data loading, hyperparameter optimization, seed testing and XAI analyses.


Data Preprocessing
Remove columns: verbal, uninformative
External sensor: merge alarm columns
External sensor: merge sheets
External sensor: add timestamps, fill missing with NaN
NaN interpolation with pandas
Machine data: remove shots before July 25, 13:07:28
Left join with Machine data
Data Loading
Feature selection: remove setpoints, constant columns
Correlation cleanup: remove pairs with correlation ≥98% or ≤2%
Load data
Apply min-max scaling if necessary
Stratified 80:20 train-test split
Optuna Hyperparameter Search
Set beta value
Define hyperparameters and ranges
Stratified k-fold cross-validation
F-beta optimization with Optuna, 250 trials
Seed Testing
Train model from scratch on full training data with best hyperparameters
Evaluate model on test data
Determine confusion matrix
Compute mean and standard deviation
Identify best models
Compute SHAP values for selected models
Create SHAP plots to understand predictions
Summarize results

(**a**) Performance evaluation.

(**b**) Explainable AI.

Figure 1. Framework of the study.

In the first stage, data preprocessing and merging, initial cleansing involves the removal of irrelevant features, specifically text-based columns, and uninformative variables such as counters or set points. Subsequently, the external sensor data undergoes a preprocessing routine, which begins by merging multiple alarm columns into a unified feature. Afterward, sheets containing external sensor data are merged, time-stamps of machine-generated data are integrated, and missing values are explicitly filled with NaNs. To address data gaps, NaN values were interpolated using pandas-based time interpolation. Crucially, this interpolation was applied strictly to continuous sensor features (e.g., temperatures) and never to the quality labels, thereby preventing look-ahead bias or data leakage. In injection molding, process parameters are characterized by high thermal and mechanical inertia. Therefore, values from cycles $t-1$ and $t+1$ provide a physically bounded and accurate estimate of the state at time t without revealing future defect outcomes. Concurrently, the machine data itself undergoes preprocessing by discarding records before a specified cutoff date (25 July 13:07:28) due to irregularities observed in the early timestamps. The resulting cleaned sensor data are finally joined with the machine data, which is designated as the ground truth.

In the second stage, data loading, further refinement of features is conducted through correlation-based feature selection. Specifically, any pair of columns that exhibit an extremely high correlation (98%) or a negligible correlation (2%) with the target variable are removed, along with columns containing constant values. These specific cutoffs were implemented as a heuristic to eliminate extreme multicollinearity—which can destabilize feature importance estimates—and to filter out stochastic noise that lacks any statistical relationship with the defect. Sensitivity analysis conducted during the pilot phase showed that the framework is robust to minor shifts in these boundaries. Following this step, the processed dataset is loaded into the modeling pipeline. Min–max scaling is optionally applied to ensure that features share comparable value ranges. The data set is then partitioned into training and testing subsets using a stratified approach (80% training, 20% testing) to maintain representative distributions between subsets.

The third stage, hyperparameter optimization, employs Optuna, an automated hyperparameter-tuning framework. Hyperparameters and their permissible ranges are defined, after which a stratified k-fold cross-validation strategy is adopted to ensure robust evaluation across subsets of data. An extensive optimization process of 500 trials is carried out to determine the hyperparameters that yield optimal model performance according to the predefined F-1 metric.

The fourth stage, seed testing, assesses the model’s generalizability and stability. A total of 500 random seeds are selected from the range 0–10,000, and multiple training runs are executed independently (each time starting from scratch) on the complete training dataset, leveraging the optimal hyperparameters identified previously. Each trained model is evaluated against the validation set, and performance metrics, primarily confusion matrices, are calculated for each run. Ultimately, the evaluation concludes with the calculation of the mean and standard deviation of the performance metrics, ensuring a comprehensive assessment of the consistency and reliability of the model.

In the final and fifth stage, XAI methods are employed to undertake an attempt to evaluate the reasoning processes underlying the model predictions. Consequently, this allows for the extraction of valuable insights that can be utilized by the operators.

3.2.1. Supervised Learning

To differentiate between good and defective parts in injection molding processes based on machine and sensor data, a set of powerful and diverse methods was employed. In particular, the following were evaluated:

- **Gradient-boosting methods**: XGBoost [30], LightGBM [31], and CatBoost [32];
- **Random Forest** [33];
- **Explainable Boosting Machine** (EBM) [34];
- **Automated machine learning** frameworks: AutoGluon [35] and AutoSKLearn [36];
- **SAINT** (Self-Attention-based network for tabular data) [37];

- **CatBoost Time Series** variant (incorporating temporal dependencies).

Below, a concise explanation of each approach is provided, highlighting why these models are well-suited to the injection molding data used in this study.

Gradient boosting refers to an ensemble paradigm where *weak learners* (decision trees) are iteratively added to minimize a given loss function. The general form of a gradient-boosted objective is defined as:

$$\mathrm{Obj}(\Theta) \;=\; \sum_{i=1}^{n} \ell\big(y_i, \hat{y}_i\big) \;+\; \sum_{k=1}^{K} \Omega\big(f_k\big), \tag{1}$$

where $\ell(\cdot)$ is a differentiable loss (e.g., logistic for classification), and $\Omega(f_k)$ is a regularization term for each tree f_k. At each boosting iteration, the next tree is fitted to the *pseudo-residuals* derived from the current model's predictions.

- **XGBoost**: An efficient tree learning algorithm is introduced with explicit handling of sparse features, approximate split finding, and cache-aware block structures for speed. It often excels on tabular datasets and is characterized by robust feature-engineering capabilities.
- **LightGBM**: A histogram-based learning approach is used, reducing the computational overhead by bucketing continuous features into discrete bins. This allows for faster training and lower memory usage while maintaining competitive performance on large datasets.
- **CatBoost**: Specialized handling of categorical features via *ordered target encoding* is provided, alleviating overfitting issues typically associated with naive label-encoding. CatBoost is also robust to hyperparameter tuning when dealing with mixed data.

Random Forest is an ensemble of decision trees trained on bootstrap samples of the data, with a random subset of features considered at each split. This bagging-based method reduces variance while maintaining bias similar to a single decision tree. For classification tasks, each tree produces a class prediction, and the overall prediction is typically decided by a majority vote:

$$\hat{y}_i \;=\; \arg\max_{c \in \mathcal{C}} \sum_{t=1}^{T} \mathbf{1}\big(\hat{y}_i^{(t)} = c\big), \tag{2}$$

where $\hat{y}_i^{(t)}$ is the prediction of the t-th tree, and $\mathbf{1}(\cdot)$ is an indicator function. In many industrial applications, Random Forest serves as a strong baseline due to its stability and relative ease of use.

While tree-based models can be opaque, the **Explainable Boosting Machine** provides a more transparent alternative by learning a *generalized additive model*:

$$g(\mathbb{E}[Y]) \;=\; \beta_0 \;+\; \sum_{j=1}^{m} h_j\big(x_j\big), \tag{3}$$

where each $h_j(\cdot)$ is a shape function for feature x_j. This additive form allows domain experts to perceive how each individual sensor or machine parameter influences the probability of producing a good part. By examining these shape functions, inputs most affecting product quality can be identified, offering greater interpretability compared to standard "black-box" ensembles.

A key requirement in this study was the reduction in manual effort during model tuning. Therefore, two *AutoML* frameworks were employed:

- **AutoGluon**: An end-to-end toolkit is utilized that iteratively trains and refines multiple models, automatically managing hyperparameters based on validation performance;

https://doi.org/10.3390/su18042001

- **AutoSKLearn**: An AutoML solution built on scikit-learn is applied, using Bayesian optimization to explore model configurations.

SAINT (*Self-Attention and Intersample Transformers*) is a neural network architecture specifically tailored for tabular data. Unlike traditional Transformers, SAINT applies self-attention in two dimensions: among *features* (columns) and among *samples* (rows). The self-attention mechanism for a single head can be written as:

$$\text{Attention}(\mathbf{Q},\mathbf{K},\mathbf{V}) \;=\; \text{softmax}\!\left(\frac{\mathbf{Q}\mathbf{K}^{\top}}{\sqrt{d_k}}\right)\mathbf{V}, \tag{4}$$

where $\mathbf{Q}$, $\mathbf{K}$, and $\mathbf{V}$ are the query, key, and value matrices, respectively, and d_k is the dimension of the key vectors.

This approach is particularly beneficial when dealing with correlated variables in manufacturing scenarios (e.g., injection molding), where multiple sensor readings must be analyzed together to predict part quality accurately. The self-attention weights, as determined by the softmax operation in Equation (4), also provide insight into which features or samples contribute most to each prediction, potentially aiding in interpretability and fault diagnosis.

3.2.2. Hyperparameter Optimization

To optimize the performance of each machine learning algorithm used in this study, we used the Optuna framework [38] for hyperparameter tuning. Optuna is a Bayesian Optimization framework that organizes its search process in a structure called a study, where each distinct hyperparameter configuration is treated as a trial. By automating the exploration of the hyperparameter space, it significantly reduces the time and computational resources typically required for model optimization. In practice, an objective function is defined that integrates the hyperparameter suggestions generated by the Tree-structured Parzen Estimator, which is Optuna's default sampler, and updates its probabilistic model based on the performance of previously completed Trials. This Bayesian approach directs the search toward more promising regions of the hyperparameter space, enabling faster convergence than classical search methods such as Grid Search or Random Search. Moreover, Optuna's Pruner mechanism monitors intermediate results and halts unpromising trials early, thus conserving computational resources. The framework also supports parallel and distributed optimization, integrates seamlessly with common machine learning libraries, and provides visualization tools that track optimization history, reveal hyperparameter importance, and illustrate parameter interactions. The specific ranges for the hyperparameters tuned for each algorithm are listed in the Appendix A (Table A1).

3.2.3. Performance Evaluation

To evaluate the performance of our machine learning models, we applied stratified k-fold cross-validation [39] during the hyperparameter optimization phase with Optuna.This method splits the dataset into k equally sized folds, ensuring that each fold serves as a test set exactly once while the remaining $k-1$ folds are used for training. The resulting performance metrics were averaged across folds to estimate generalization performance for each hyperparameter configuration. In particular, during this phase, each configuration was evaluated using a single random seed and no seed variation was applied within the Optuna trials.

After identifying the best-performing hyperparameter configurations, we performed a separate seed testing procedure to assess model robustness. In this step, the final models were retrained from scratch without cross-validation, using the full training set, and

https://doi.org/10.3390/su18042001

evaluated on a fixed validation set of 20%. This process was repeated in 500 different random seeds to account for the variance introduced by initialization and data shuffle.

We report the mean and standard deviation of key performance metrics—including accuracy, precision, recall, and F1-score—across these 500 seed runs. This provides a reliable and comprehensive view of model stability and generalization under realistic production conditions.

3.2.4. Dealing with the Class Imbalance

The serious imbalance in the dataset might pose a problem to the classifiers. To account for this, we have employed oversampling techniques and threshold modification to further investigate the effect of class imbalance. The idea of oversampling methods is to increase the size of the minority class or shrink the majority class to achieve a balanced data set. SMOTE is a technique introduced by [40] to increase the sample size of the minority class. Instead of resampling the same instances multiple times, new synthetic samples are created. The new samples are created on the lines that connect the other samples in the feature space. By utilizing k-nearest neighbors for interpolation, SMOTE ensures that synthetic defects are generated within the multi-dimensional clusters of real-world failures, thereby preserving the physical correlations between temperature, energy, and control parameters. Another approach that addresses the issues of class imbalance is cost-sensitive learning, which puts weights on the four outcomes of the predictions: true positives, false negatives, false positives, and true negatives. However, it can be shown that cost-sensitive learning is equivalent to modifying the prediction threshold [41]. If the output neuron yields a value larger than the threshold, the part is classified as defective, and vice versa. Therefore, as a second approach to address the challenges of the imbalanced dataset, we introduce the threshold as an additional hyperparameter. The threshold was optimized simultaneously with the other hyperparameters.

In industrial quality control, not all misclassifications have the same practical impact. Although false positives result in unnecessary rejections, false negatives imply defective parts passing undetected, which is often associated with significantly higher costs. Therefore, instead of relying solely on accuracy or F1-score, we approximate a cost function that incorporates the asymmetric damage of false positives and false negatives. The specific cost parameters ($C_{FN} = 100$, $C_{FP} = 0.5$) were selected to reflect a typical industrial scenario in high-volume injection molding, while a FP merely results in minor costs for manual re-inspection, a FN represents a critical failure. If a defective part is integrated into a higher-level assembly before being detected, the resulting costs for disassembly, sorting, and warranty claims far outweigh the individual part's value. To identify the optimal operating point, we evaluate the expected cost across a range of classification thresholds using cross-validation. Plotting the total cost as a function of the threshold, along with standard metrics such as accuracy, F1-score, recall, and precision, allows us to visualize the trade-offs and determine the threshold that minimizes the overall economic impact. Using multiple random seeds and repeated cross-validation ensures that this analysis captures variability due to both model initialization and the stochastic nature of SMOTE resampling, providing a robust and reliable estimate of the cost-optimal threshold. This approach ensures that the classifier is not only optimized for statistical performance but also aligned with the real-world objectives of the injection molding process.

3.2.5. SHAP Analysis

To interpret the predictions of our machine learning models and understand the influence of each input feature, we used SHAP analysis [42]. SHAP is a unified approach based on cooperative game theory that assigns each feature an importance value for a particular prediction.

The Shapley value for a feature i is defined as:

$$\phi_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!\,(|N| - |S| - 1)!}{|N|!} [v(S \cup \{i\}) - v(S)],$$

where

- N is the set of all features;
- S is a subset of features not containing feature i;
- $v(S)$ is the value function representing the model prediction using the feature subset S.

The term $\frac{|S|!\,(|N|-|S|-1)!}{|N|!}$ is a weighting factor that accounts for all possible permutations of features, ensuring a fair contribution from each feature regardless of the order in which features are added.

By calculating Shapley values, SHAP quantifies the contribution of each feature to the model's output, allowing for a comprehensive understanding of feature impact. This analysis enabled us to identify the sensor parameters that most significantly affect the forecasting and classification results, providing valuable insights into the underlying processes.

3.2.6. Transfer Learning

Transfer learning is a machine learning technique that leverages knowledge gained from a source domain to improve learning in a target domain [43]. By reusing and adapting models or representations originally trained on related tasks, transfer learning helps address challenges such as limited labeled data and domain shifts, ultimately enhancing performance in new scenarios.

Transfer learning frameworks can be defined in terms of their respective domains and tasks. Let:

- $\mathcal{D}_S = \{\mathcal{X}_S, P(X_S)\}$ be the source domain with input feature space $\mathcal{X}_S$ and distribution $P(X_S)$;
- $\mathcal{T}_S = \{\mathcal{Y}_S, f_S(\cdot)\}$ be the source task with label space $\mathcal{Y}_S$ and predictive function $f_S(\cdot)$;
- $\mathcal{D}_T = \{\mathcal{X}_T, P(X_T)\}$ be the target domain with input feature space $\mathcal{X}_T$ and distribution $P(X_T)$;
- $\mathcal{T}_T = \{\mathcal{Y}_T, f_T(\cdot)\}$ be the target task with label space $\mathcal{Y}_T$ and predictive function $f_T(\cdot)$.

The goal is to improve the learning of $f_T(\cdot)$ in the target domain $\mathcal{D}_T$ by exploiting the knowledge of $\{\mathcal{D}_S, \mathcal{T}_S\}$. Therefore, transfer learning can be expressed as:

$$f_T(\cdot) \leftarrow \arg\min_{f_T} \mathcal{L}\big(\mathcal{D}_T, \mathcal{T}_T, f_T(\cdot)\big) \quad \text{subject to knowledge from} \quad \{\mathcal{D}_S, \mathcal{T}_S\},$$

where $\mathcal{L}$ is a loss function reflecting the performance on the target domain-task pair $\{\mathcal{D}_T, \mathcal{T}_T\}$. By transferring shared patterns and representations, transfer learning mitigates the need for large labeled datasets in the target domain and facilitates effective model adaptation to new conditions. In our study, we applied transfer learning to reuse the representational power of pretrained models, thereby improving robustness and reducing development time when dealing with differing sensor data distributions or new task requirements.

To quantify the effectiveness of transfer learning relative to the training of new models from scratch, we conducted a systematic sample size analysis outlined in Algorithm 1.

https://doi.org/10.3390/su18042001

Algorithm 1 Transfer Learning Sample Size Analysis

```
Preprocess source $(X_s, y_s)$ and transfer $(X_t, y_t)$ datasets
Split transfer data 80/20 into $(X_{t\_train}, y_{t\_train})$, $(X_{t\_test}, y_{t\_test})$
Align features between domains
Initialize sample sizes $P \leftarrow \{50, 55, 60, \dots\}$

Train base model $M_{base}$ on $(X_s, y_s)$
for each $p \in P$ do
    Sample subset $D_p \leftarrow p$ examples from $(X_{t\_train}, y_{t\_train})$
    Transfer Learning:
    Initialize $M_{transfer}$ from $M_{base}$
    Fine-tune on $D_p$ (50 rounds)
    $acc_{transfer} \leftarrow$ test on $(X_{t\_test}, y_{t\_test})$

    New Model:
    Train new model $M_{new}$ from scratch on $D_p$
    $acc_{new} \leftarrow$ test on $(X_{t\_test}, y_{t\_test})$

    Store $(p, acc_{transfer}, acc_{new})$
end for

return Sample sizes and corresponding accuracies
```

4. Results

4.1. Hyperparameter Optimization

The impact of hyperparameter tuning on model performance is summarized in Table 3, illustrating the relative improvements achieved across various models.

Table 3. Relative Improvement due to Hyperparameter-Tuning.

Model	Relative Improvement
CatBoost	16.40%
CatBoostTimeSeries	0.92%
EBM	13.38%
LightGBM	0.95%
Random Forest	10.62%
SAINT	61.07%
XGBoost	2.84%

Particularly noteworthy are the substantial performance gains observed for SAINT (61.07%) and CatBoost (16.40%), indicating the importance of hyperparameter optimization to improve predictive capabilities. Across various models, the most crucial hyperparameter to fine-tune is the learning rate.

The final hyperparameter of the different models can be found in the Appendix A (Table A1).

4.2. Performance Evaluation

Table 4 presents the confusion matrices for the models evaluated, the row labeled *Ground Truth* indicating the actual class distribution in the dataset (i.e., total positives and negatives).

Among the approaches tested, CatBoost exhibits the highest accuracy. The next four best-performing classifiers are AutoGluon, LightGBM, Random Forest and XGBoost. Due to their strong predictive performance and robustness, these five classifiers were selected for further analyses, specifically XAI investigations and transfer learning experiments.

Table 4. Confusion Matrix for the different Machine Learning Methods.

Model	TN	FP	FN	TP	F1
XGBoost	1141.92 ± 0.28	2.08 ± 0.28	6.80 ± 0.40	32.20 ± 0.40	0.8788 ± 0.0060
LightGBM	1142.05 ± 0.75	1.95 ± 0.75	6.25 ± 0.68	32.75 ± 0.68	0.8887 ± 0.0124
EBM	1139.74 ± 0.52	4.26 ± 0.52	6.04 ± 0.34	32.96 ± 0.34	0.8649 ± 0.0072
SAINT	1141.87 ± 4.62	2.13 ± 4.62	16.78 ± 3.53	22.22 ± 3.53	0.7015 ± 0.0725
AutoGluon	1143.00 ± 0.00	0.00 ± 0.00	7.00 ± 0.00	32.00 ± 0.00	0.9014 ± 0.0000
AutoSKLearn	1143.00 ± 0.00	1.00 ± 0.00	9.00 ± 0.00	30.00 ± 0.00	0.8571 ± 0.0000
CatBoost	1143.14 ± 0.48	0.86 ± 0.48	**5.52 ± 0.69**	**33.48 ± 0.69**	**0.9130 ± 0.0106**
CatBoostTS	1140.00 ± 0.00	3.00 ± 0.00	9.00 ± 0.00	30.00 ± 0.00	0.8333 ± 0.0000
Random Forest	**1143.86 ± 0.34**	**0.14 ± 0.34**	7.00 ± 0.04	32.00 ± 0.04	0.8996 ± 0.0043
Ground Truth	1144.00	0.00	0.00	39.00	1.0000

4.3. Dealing with Class Imbalance

Table 5 shows results for the balancing approaches. Additional approaches, that were sorted out due do their lack of performance, can be found in Figure A1 as well as according runtimes. The Model with a fine-tuned threshold clearly surpasses the basic CatBoostModel.

Table 5. Results of the Balancing.

Model	Mean F1-Score
Basic	0.9130 ± 0.011
Threshold tuning	0.9091 ± 0.019
SMOTE	0.9027 ± 0.005
SMOTE & Threshold tuning	**0.9263 ± 0.005**

The results were obtained by training the CatBoost model with fixed hyperparameters, which had been optimized in advance using 500 Optuna trials and named approaches. For each balancing configuration, the finalized model was retrained 500 times with varying random seeds. In the case of SMOTE, the seed that controls the generation of synthetic samples was also varied. This setup ensures that variability due to initialization and data resampling is averaged out, so that differences in performance can be attributed solely to the balancing strategy itself. As shown in Table 5 and Figure 2, the baseline model already achieves a mean F1-score of 0.9130 ± 0.011. Threshold tuning alone reaches a level comparable to the higehr variance and a score 0.9091 ± 0.019, while SMOTE in isolation performs slightly worse at 0.9027 ± 0.005. The best performance is achieved when SMOTE is combined with threshold tuning, resulting in a mean F1-score of **0.9263 ± 0.005**. The corresponding box plots in Figure 2 illustrate the distribution of F1 in box plots, providing a robust overview of the variability introduced by both model initialization and the SMOTE resampling process.

To ensure the robustness of the selected approach, several additional baselines—including class weighting, focal loss, and balanced ensembles—were evaluated, while these methods demonstrated competitive accuracy, they were found to be suboptimal in terms of the target F1-score compared to the proposed hybrid SMOTE and threshold tuning approach. A comprehensive performance comparison of these additional strategies is provided in Figure A1 in the Appendix A.

4.4. Cost Optimization

To approximate the cost-optimal classification threshold, we performed a 10-fold cross-validation on the training data, repeated across five different random seeds to account for variability in data partitioning and SMOTE resampling. For each fold, SMOTE was applied exclusively to the training subset to prevent data leakage, after which a CatBoost model with hyperparameters obtained from 500 Optuna trials was trained on the balanced training data and evaluated on the original, unbalanced validation split. The evaluation was carried out on 50 thresholds in the range $[0, 1]$. At each threshold, the cost was calculated based on a weighted scheme, assigning a high penalty to false negatives and a small penalty to false positives. The resulting costs from all folds and seeds were aggregated to obtain mean and percentile estimates. As shown in Figure 3, the analysis yields an optimal threshold at 0.02 that minimizes the expected cost, significantly improving over the default value of 0.5. Since the SMOTE & threshold tuning approach had shown the best overall performance, it was selected as the basis for this cost optimization. The shaded area in the plot reflects the 25–75 percentile interval.

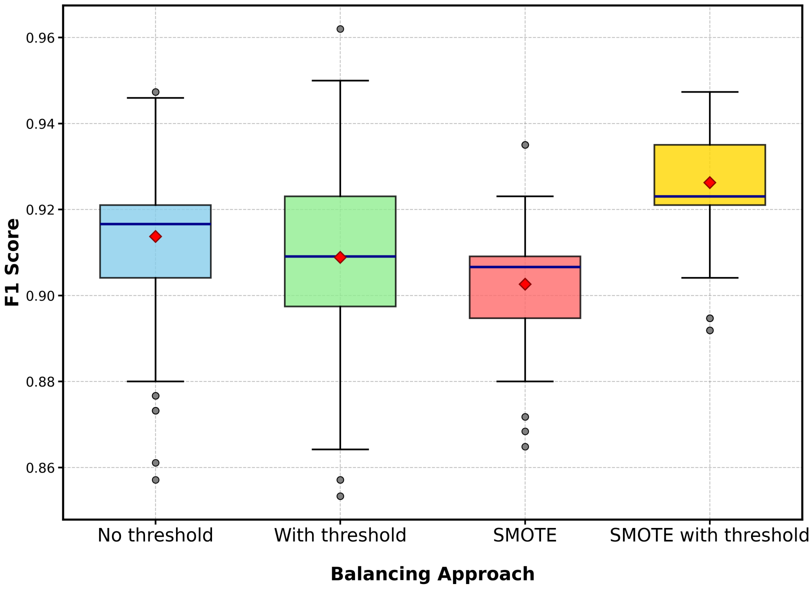


Figure 2. Balancing Approaches compared by F1.

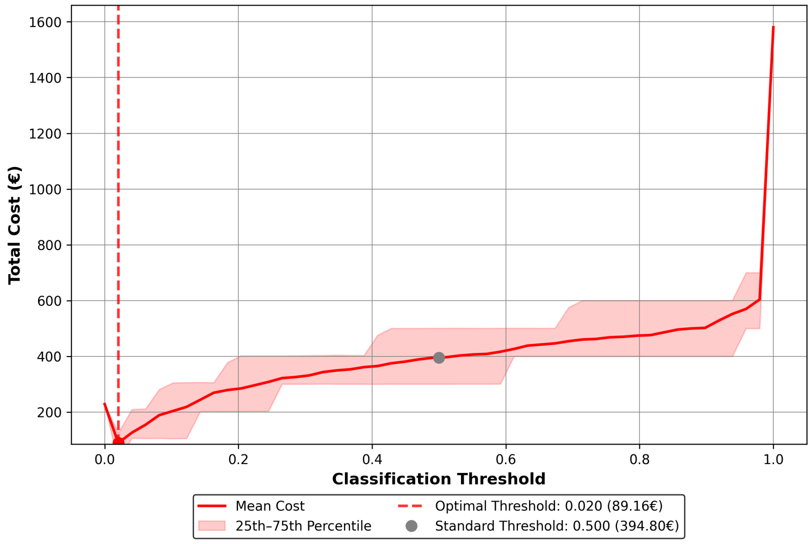


Figure 3. Finding optimal threshold.

Table 6 reports the average entries in the confusion matrix and the corresponding costs for the different balance strategies. The plain CatBoost baseline shows a high number of true positives, but its relatively large variation in false negatives leads to a higher expected cost. Threshold tuning alone further reduces false positives, but at the expense of more false negatives, resulting in the highest overall cost. SMOTE in isolation achieves a stable reduction in false negatives but introduces a higher number of false positives. Nevertheless, despite the worse F1-Score before it is a more cost-effective approach than just using the baseline. The combined SMOTE & threshold tuning approach provides the most favorable trade-off: it keeps false negatives at a very low level (1.96 ± 0.40) while maintaining acceptable false positives, resulting in the lowest average cost of EUR 226.76. It is worth noting that approaches with low false negatives tend to have a high variation in the number of their false positives.

Table 6. Performance Metrics of Different Approaches.

Approach	TP	FP	FN	TN	Cost
SMOTE Threshold	1058.48 ± 151.57	85.52 ± 151.57	1.96 ± 0.40	37.04 ± 0.40	226.76 ± 23.56
CatBoost	1120.42 ± 6.25	23.58 ± 6.25	3.22 ± 1.76	35.78 ± 1.76	333.79 ± 175.06
Threshold	1133.12 ± 3.71	10.88 ± 3.71	4.58 ± 2.12	34.42 ± 2.12	463.44 ± 211.50
SMOTE	1018.78 ± 145.60	125.22 ± 145.60	2.04 ± 0.20	36.96 ± 0.20	256.26 ± 19.80

Figure 4 illustrates the cost-sensitive evaluation of the best-performing approach, which is SMOTE & threshold tuning. The model was trained with hyperparameters obtained from 500 Optuna trials and subsequently evaluated on the test set across 50 different thresholds over 50 random seeds varying model and SMOTE seed. In addition to the total cost, standard performance metrics such as accuracy, F1-score, recall, and precision are reported in the plot. The shaded regions capture the 25–75 percentile interval, providing a measure of variability between seeds. The red dashed line marks the cost-optimal threshold at 0.02, corresponding to the minimum mean cost of EUR 226.76, confirming that substantial cost reductions can be achieved compared to the conventional threshold of 0.5.

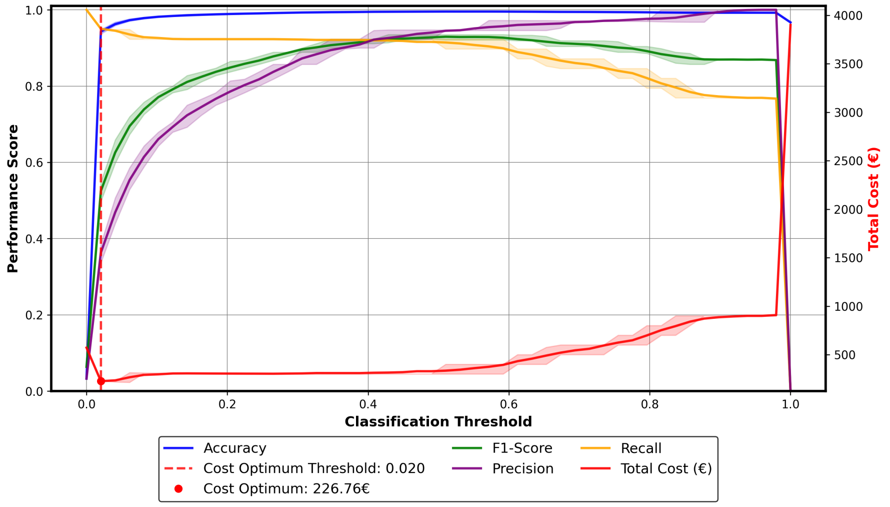


Figure 4. Threshold to cost and accuracy plot of our best approach.

To ensure the robustness of our conclusions beyond this specific case, Figure 5 provides a sensitivity analysis on a range of cost ratios. It illustrates how the optimal threshold adapts to varying economic priorities.

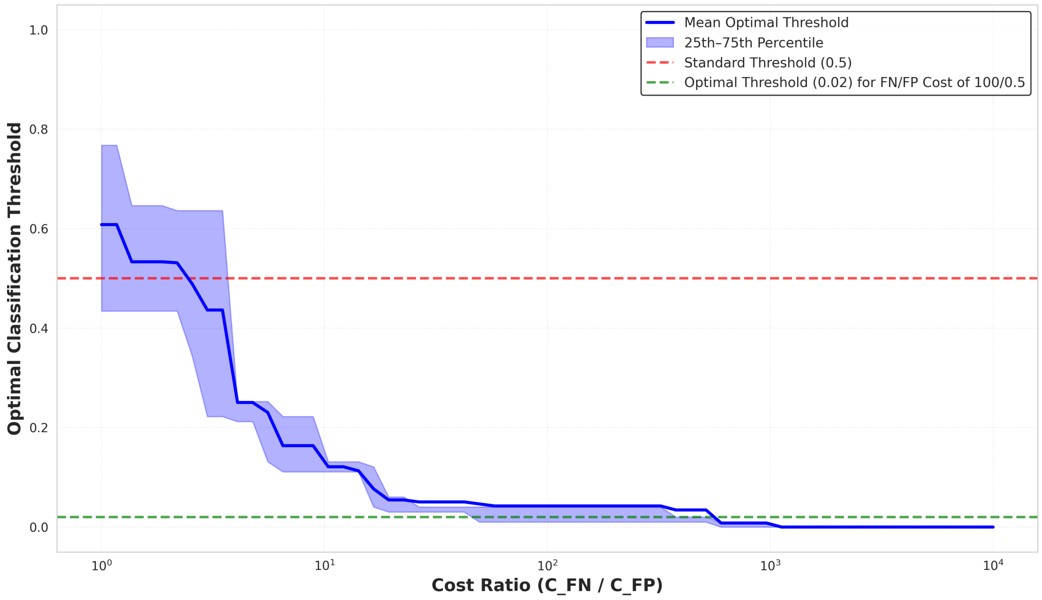


Figure 5. Optimal classification threshold as a function of the cost ratio between false negatives and false positives. The green dashed line shows the analytically derived optimal threshold of 0.02 for a cost ratio of 100/0.5. The threshold decreases monotonically as the relative cost of false negatives increases, converging to values near zero for high cost ratios, demonstrating the necessity of threshold adjustment in imbalanced classification scenarios.

4.5. Model Explanation

Figure 6 illustrates the 10 highest normalized features importance of the five best performing machine learning methods.

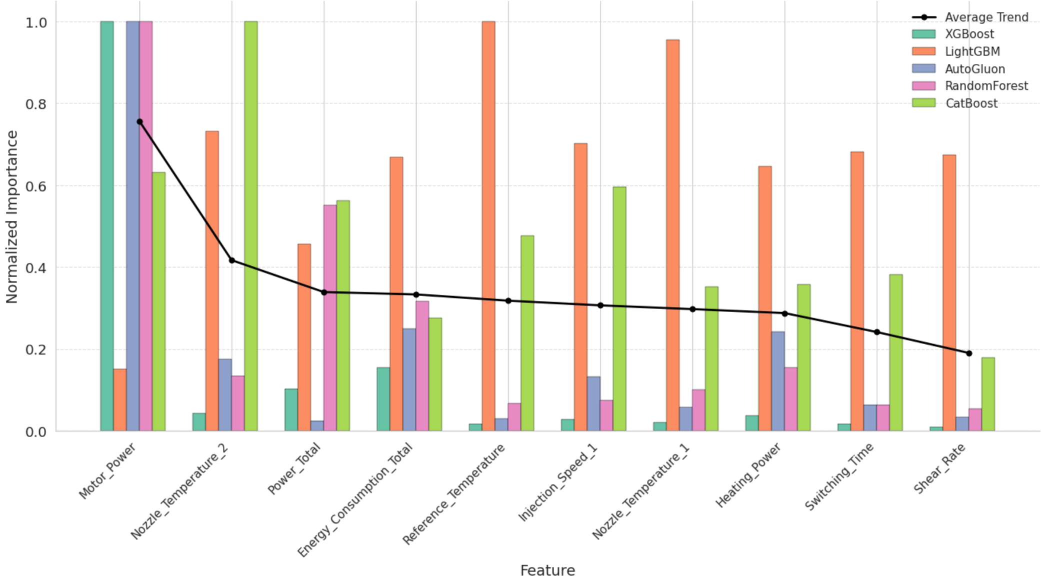


Figure 6. Feature Importance of the best–performing models.

The features, sorted by importance from left to right as motor power, nozzle temperature 2, power total, energy consumption total, reference temperature, injection speed 1, nozzle temperature 1, switching time, injection time, and heating power, were normalized to allow a direct comparison across models. The black overlaid trend line represents the average importance in all models.

The analysis reveals that Motor Power exhibits the highest normalized importance, particularly in models such as XGBoost, AutoGluon, and RandomForest, indicating its pivotal role in predicting outcomes. Nozzle temperature 2 follows closely, which underscores the significance of thermal conditions at the nozzle. Energy Consumption Total and Power Total, ranked third and fourth, respectively, show moderately high importance, suggesting that energy-related metrics are also critical for model performance. Mid-range features such as Reference Temperature and Nozzle Temperature 1 demonstrate moderate contributions, reflecting a secondary, yet still relevant, influence on the predictive accuracy of the models. In contrast, features located on the right side of the spectrum, switching time, injection speed, injection time, and especially heating power, display progressively lower normalized importance. The declining trend of the black average line from left to right corroborates this hierarchy, indicating a diminishing impact on model performance as one moves from power-related and nozzle temperature metrics to process dynamics and heating parameters.

These results highlight the predominance of power and thermal parameters in the injection molding process, emphasizing their central role in fault detection and overall predictive capability.

Figure 7 illustrates the visualization of SHAP bee swarm plots, detailing the directional influence on the predicted probability that parts are defective. It should be emphasized that the features are organized based on an alphabetical arrangement rather than being prioritized according to their significance. Another important aspect is that the displayed features differ across models, as each graph individually represents the 10 most significant features specific to that particular model. This approach is implemented to assist in the comparison between various models.

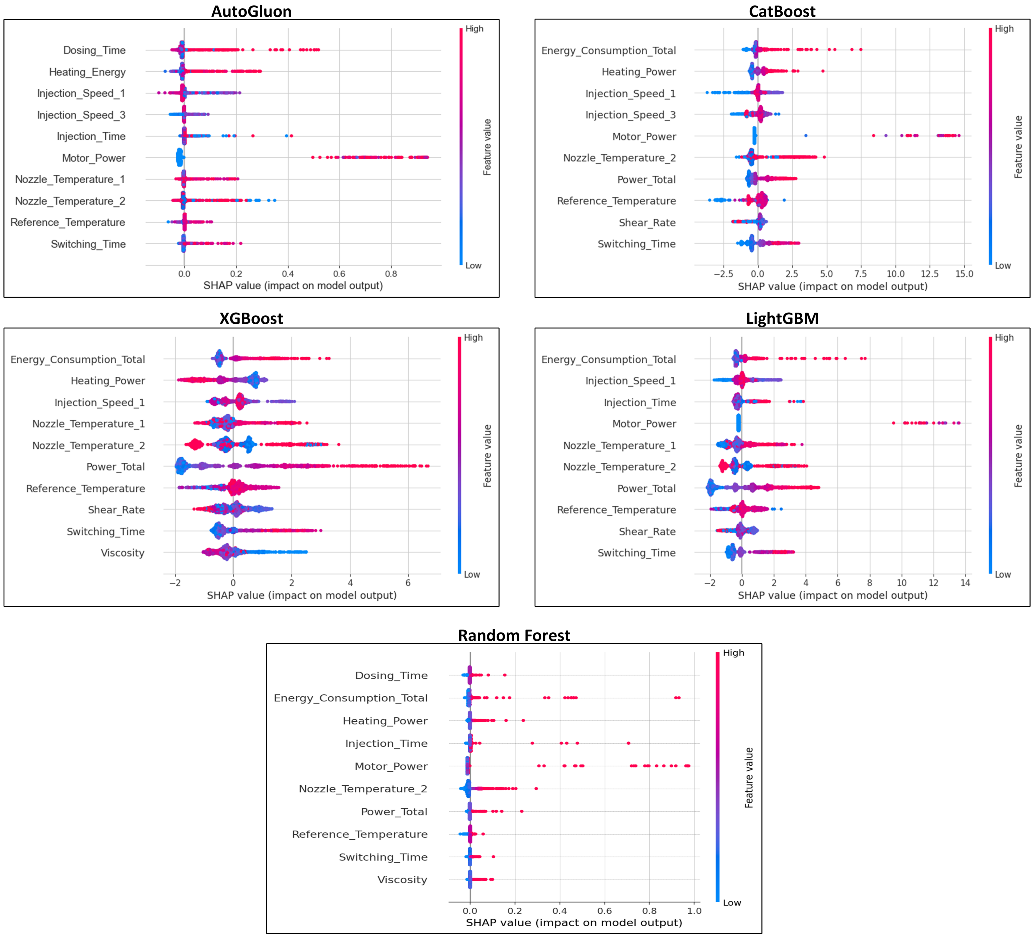


Figure 7. SHAP bee swarm plots visualizing the impact of top 10 features considering Feature Importance on each model.

Across all models evaluated, Motor Power emerges as a key determinant: a higher motor power (red points) consistently shift SHAP values into the positive range, indicating an elevated risk of defects, while lower power levels (blue points) generally align with negative SHAP values and therefore a prediction of ′good′. Similarly, Heating Energy, Power Total and Energy Consumption Total exhibit a similar pattern. All algorithms place higher values of the features on the positive SHAP side, suggesting a propensity toward defects. With respect to Heating Power, CatBoost and XGBoost exhibit divergent behaviors: XGBoost attributes higher values to good predictions, while CatBoost assigns higher values to defective predictions.

Injection Speed 1 generally shows that higher speeds promote negative SHAP values, that is, a tendency toward ′good′ predictions. In contrast, Injection Speed 3 shows that lower speeds promote negative SHAP values.

Dosing time, while only ranking among the top 10 key features of AutoGluon, indicates that longer dosing intervals (red) often correspond to positive SHAP values (favoring "defective").

Among temperature-related features (e.g., Nozzle Temperature, Reference Temperature), extreme high or low values often correlate with an increased risk of defects, evident in positive SHAP shifts. In contrast, precisely-optimized mid-range temperatures often result in negative SHAP values, marking those conditions as potentially favorable for quality outcomes. Certain models exhibit a bimodal SHAP distribution for temperature, indicating that the effects may interact combinatorially with other parameters.

The influence of shear rate and viscosity is likewise evident across the algorithms. At higher shear rates or viscosities, SHAP values shift to the negative domain, suggesting a higher likelihood that parts are satisfactory. In contrast, lower feature values are associated with higher probabilities of the parts being defective.

Lastly, lower switching time values increase the likelihood that the part is satisfactory, whereas higher values enhance the probability of the part being defective.

The details of both XAI analyses are summarized in Table 7.

Table 7. Aggregated insights of XAI analyses.

Feature	Importance	Insight
Motor Power	1	Higher motor power leads to more defective parts.
Nozzle Temperature 2	2	Extreme nozzle temperatures increase defect risk.
Energy Consumption Total	3	High energy consumption results in more defects.
Power Total	4	Higher power totals are linked to defective outcomes.
Reference Temperature	5	Extreme temperatures raise defect risk; optimal mid-range values are preferable.
Nozzle Temperature 1	6	Extreme values increase defect likelihood; mid-range conditions are more favorable.
Switching Time	7	Longer switching times are associated with defects.
Injection Speed	8	The effect is nuanced—optimal speeds minimize defects.
Injection Time	9	Has a comparatively lower impact on defects.
Heating Power	10	Impact varies by model, showing inconsistent trends.

4.6. Transfer Learning and Scalability Analysis

To evaluate the scalability of the proposed framework, we tested various transfer learning strategies, including feature-based transfer and model blending. Fine-tuning—adapting a pretrained model by updating weights on the target dataset—emerged as the most effective approach. In LightGBM, this was achieved via continued boosting by extending a pretrained ensemble of 100 trees with an additional 50 trees while keeping the original model weights fixed. This approach not only reduces training time but also mitigates the risk of overfitting compared to training a complete 100-tree model from scratch.

Figure 8 illustrates the learning curves for CatBoost and LightGBM. A striking disparity is evident: CatBoost exhibits no significant gain from transfer learning, performing identically to a model trained from scratch. In contrast, LightGBM demonstrates strong domain adaptation.

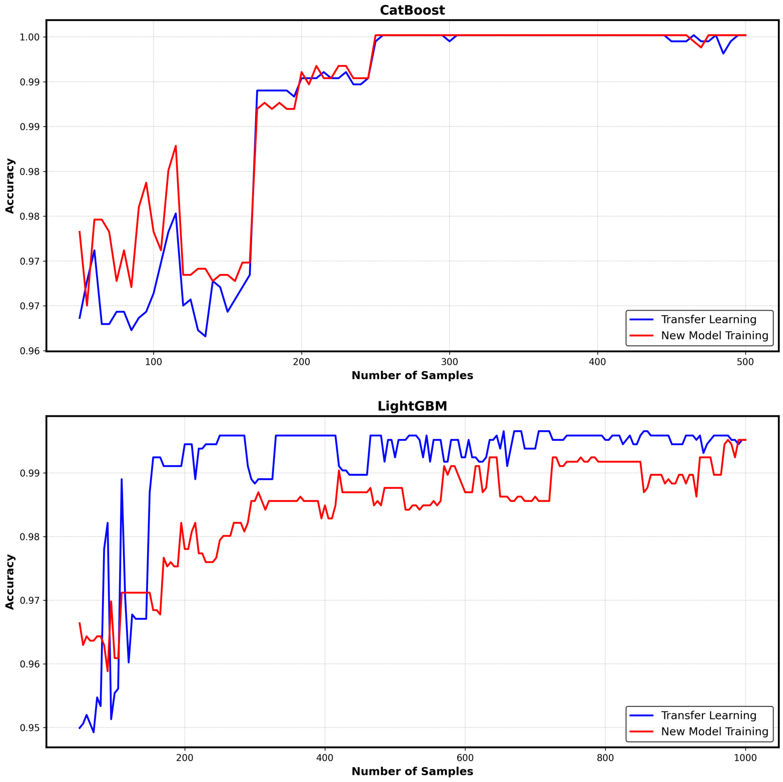


Figure 8. Learning curves comparing Transfer Learning (TL) vs. New Model (NM). LightGBM shows a significant "lift" in the early phase (<200 samples).

As quantified in Table 8, the transfer-learning (TL) variant of LightGBM reaches stability with approximately 200 samples, whereas the new model (NM) requires over 450 samples to achieve comparable performance. This indicates that transfer learning can reduce the "cold-start" data requirements by more than 50%.

Table 8. Efficiency Comparison: Transfer Learning (TL) vs. New Model (NM).

Model	Phase	TL (Samples)	NM (Samples)	Key Observation
CatBoost	Stability Plateau	180–260 >260	190–260 >260	Minimal impact; TL offers no data efficiency advantage. Convergence is identical.
LightGBM	Stability Plateau	**170–200** >200	185–450 >450	**TL stabilizes significantly earlier.** NM requires ≈ 2.5× more data to converge.

The divergent behaviors of the models can be explained by their reliance on shared features. The source and target datasets only share two common parameters: *Injection Time* and *Dosing Time*. For transfer learning to succeed, the model must rely primarily on these persistent features. Table 9 reveals that LightGBM assigns a cumulative importance of **87.98%** to these shared features. Consequently, the knowledge learned from the source domain is directly applicable to the target. Conversely, CatBoost (28.95%) and XGBoost

(6.89%) distribute importance to source-specific features that do not exist in the target dataset, rendering the pretrained knowledge largely irrelevant.

Table 9. Normalized Feature Importance of Shared Features (Injection & Dosing Time).

Feature	AutoGluon	CatBoost	LightGBM	Random Forest	XGBoost
Injection Time	0.3511	0.2234	0.3529	0.1291	0.0607
Dosing Time	0.1278	0.0661	**0.5269**	0.0298	0.0082
Total Sum	0.4789	0.2895	**0.8798**	0.1589	0.0689

The technical efficiency of LightGBM directly translates into economic benefits for industrial deployment. By reducing the data requirements for new machines, manufacturers can significantly reduce setup costs and time-to-value (see Table 10).

Table 10. Estimated Resource Savings via Transfer Learning (LightGBM).

Metric	Data Savings	Time Reduction	Energy Impact
Improvement	≈55%	≈55% faster	Up to 50% reduction

In conclusion, while CatBoost performed best on the single static dataset LightGBM is the superior choice for scalable, multi-machine environments due to its ability to leverage limited shared features for effective transfer learning.

5. Discussion

This study aimed to bridge the gap between theoretical ZDM concepts and their industrial application in injection molding. By benchmarking state-of-the-art algorithms and introducing a cost-sensitive, explainable framework, we address the critical challenges of scalability, trust, and economic viability.

5.1. Algorithmic Superiority: GBDTs vs. Deep Learning

Our results corroborate the "No Free Lunch" theorem but clearly favor Gradient-Boosted Decision Trees (GBDTs) over Deep Learning architectures for this domain. CatBoost emerged as the superior model for the primary dataset, effectively handling categorical process parameters without extensive preprocessing. Interestingly, while Transformer-based architectures like SAINT show promise in tabular domains, they failed to outperform CatBoost in our experiments. This aligns with recent findings in the broader ML literature suggesting that for structured, tabular industrial data with limited sample sizes (compared to image/text data), well-tuned GBDTs remain the gold standard [23]. Conversely, simple baseline models like Random Forests, while easy to deploy, lacked the precision required for ZDM standards. This suggests that the complexity of modern injection molding processes necessitates the advanced regularization and boosting mechanisms found in CatBoost and LightGBM.

5.2. From Accuracy to Profitability: The Cost-Sensitive Paradigm

A critical contribution of this work is the shift from technical accuracy to economic risk minimization. Standard classification metrics often mask the financial impact of False Negatives (delivering defective parts). Our analysis of balancing strategies reveals a clear trade-off:

- Threshold Tuning alone reduces risk but introduces high variance, making the system unstable across production runs.

https://doi.org/10.3390/su18042001

- SMOTE stabilizes the learning process by constructing a robust decision boundary around the minority class.
- The Hybrid Approach (SMOTE + Threshold Moving) proved optimal. It creates a synergy where SMOTE provides the necessary support vectors for the minority class, while threshold calibration fine-tunes the sensitivity to align with the specific cost matrix of the manufacturer.

This finding implies that industrial ZDM systems must be "economically calibrated." A model with 99% accuracy is worthless if the 1% error consists exclusively of expensive False Negatives.

Regarding industrial production viability, a runtime comparison of the different balancing strategies was conducted (see Table A2). Although the hybrid SMOTE and threshold tuning approach increases the training overhead compared to the native model, the total training time of approximately 1727 s on standard hardware remains well within industrial requirements. More importantly, the inference latency remains in the millisecond range, which is negligible compared to the 20-s machine cycle time, confirming the framework's suitability for real-time deployment.

5.3. Enablers for Industry 4.0: Explainability and Transferability

For AI to be adopted on the shop floor, it must be trustworthy and scalable.

Our SHAP analysis confirms that the models are not learning spurious correlations but are driven by physically relevant parameters—specifically motor power and nozzle temperature. This consistency with domain knowledge (e.g., the influence of melt temperature on viscosity [17]) validates the model's decision-making process. Unlike "black-box" approaches criticized in early ZDM literature, our framework empowers operators to understand *why* a defect is predicted, facilitating targeted root-cause analysis rather than blind trust.

Finally, our Transfer Learning experiments highlight a crucial nuance: While CatBoost was superior in static learning, LightGBM excelled in transferability. Because LightGBM heavily prioritized features shared between datasets (Injection and Dosing Time), it allowed for a data efficiency gain of over 50% when adapting to a new machine. This addresses the "Cold Start" problem identified by [16]. For the industry, this means that a generalized "Global Model" (LightGBM) can be deployed to new lines to gather initial data, which can later be replaced by a specialized "Local Model" (CatBoost) once sufficient data is available.

While promising, our study relies on offline datasets. Future research should focus on the deployment of these models in a streaming edge-computing environment to validate real-time inference latencies. Additionally, integrating the cost-sensitive feedback loop directly into the machine controller (predictive control) represents the next logical step towards fully autonomous ZDM.

5.4. Limitations and Future Work

Despite the robustness of the proposed framework, it is important to acknowledge the limitations regarding the sample size of defective parts. With 197 defective cycles in the main dataset and 331 in the transfer dataset, the study operates in a small-data regime. Although our extensive seed testing and stratification provide high confidence in the stability of the results, larger industrial datasets would be advantageous to further reduce the variance of the performance estimates. Future research should aim to aggregate data across multiple production years or diverse polymer types to assess the long-term drift of the cost-optimal thresholds and the potential for even more complex, data-hungry deep learning architectures. A second limitation concerns the transfer learning analysis, which relied on a narrow overlap of only two shared features: injection and dosing time.

We interpret this specific setup as a minimalist proof-of-concept. Since these two parameters represent the fundamental "physical DNA" of the specific injection molding cycle, achieving a stability gain of over 50% with such a sparse feature set demonstrates the high sensitivity of the methodology. It is expected that the efficiency gains of transfer learning will scale proportionally with the degree of feature overlap. Consequently, our results represent a lower-bound scenario. In larger industrial environments with standardized sensor sets and broader feature alignment, the reduction in cold-start data requirements is likely to be even more substantial. Future work should therefore explore the framework's scalability in multi-machine environments with varying degrees of sensor standardization. In future research, the current sensor-based framework could be significantly extended through the integration of inline camera vision systems to detect surface defects on a per-cycle basis. By fusing high-resolution visual data with machine sensor information, a more robust and cost-optimized classification would be achieved. The potential of such multi-modal integration is demonstrated by robust vision-based models in other high-precision domains. For instance, pixel-level semantic segmentation for histopathological diagnosis was achieved using a deep convolutional neural network based on DeepLab v3, which allowed for the precise detection of cancerous regions with high sensitivity [44]. Similarly, labor-intensive monitoring was successfully replaced by an efficient computer vision framework, where a custom Mask R-CNN with Feature Pyramid Networks (FPN) was leveraged to provide real-time, high-confidence estimations of material properties [45]. The adaptation of these advanced vision architectures to the injection molding environment is viewed as the next logical step toward a holistic, AI-supported ZDM ecosystem.

6. Conclusions

This study presents a comprehensive, cost-optimal framework for Zero-Defect Manufacturing (ZDM) in high-precision injection molding. By evaluating various state-of-the-art machine learning architectures, it was determined that Gradient-Boosted Decision Trees, specifically CatBoost [32], deliver superior predictive performance for industrial sensor data, effectively handling the high-dimensional and imbalanced nature of the process [23]. A central finding of this research is that maximizing technical accuracy is insufficient for industrial viability [13]. Instead, the implementation of a cost-sensitive threshold calibration led to a reduction in total failure costs by over 75% compared to standard classification approaches. This demonstrates the necessity of aligning model objectives with the asymmetric economic risks of manufacturing [41]. Furthermore, the integration of SHAP analysis [42] successfully decoded the "black-box" predictions, identifying motor power and nozzle temperatures as the primary drivers of quality defects. Finally, the transfer learning experiments showed that models could be adapted to new machines with significantly reduced data requirements, representing a data efficiency gain of over 50% compared to training from scratch [43].

Practical Implications and Sustainability Impact

The proposed framework offers significant sustainability benefits by directly contributing to resource optimization and waste reduction [2,3]. By accurately predicting defects before they reach the consumer, material waste is minimized [12], and the energy-intensive processing of defective parts is avoided. From a practical perspective, the high data efficiency of the transfer learning approach enables small and medium-sized enterprises (SMEs) to adopt AI-driven quality assurance without the need for extensive, costly data collection [23]. Ultimately, this transparency-driven and economically optimized approach facilitates a more sustainable, resilient production environment in the context of Industry 5.0 [7,15].

https://doi.org/10.3390/su18042001

Author Contributions: Conceptualization, L.G.; Methodology, L.G., J.O. (Jonas Ortner), P.K., A.K., S.K. and J.B.; Software, L.G., J.O. (Jonas Ortner), P.K., A.K., S.K. and J.B.; Validation, L.G. and J.O. (Jonas Ortner); Formal Analysis, L.G.; Investigation, L.G.; Resources, J.O. (Jivka Ovtcharova); Data Curation, L.G., P.K. and A.K.; Writing—Original Draft Preparation, L.G.; Writing—Review and Editing, L.G., J.O. (Jonas Ortner), P.K., A.K., S.K., J.B. and J.O. (Jivka Ovtcharova); Visualization, L.G.; Supervision, J.O. (Jivka Ovtcharova); Project Administration, J.O. (Jivka Ovtcharova). All authors have read and agreed to the published version of the manuscript.

Funding: The research was partially funded by the German Federal Ministry for Economic Affairs and Climate Action (BMWK) through a research project (No.03LB3058B).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data will be available upon request.

Conflicts of Interest: The authors declare no competing or financial interests.

Appendix A

Table A1. Hyperparameter search ranges and optimal values identified via Optuna.

Model	Hyperparameter	Range	Best Value
CatBoost	iterations	[100, 1000]	970
	learning_rate	[1×10^{-3}, 0.1]	0.0753
	depth	[4, 10]	4
	l2_leaf_reg	[1×10^{-8}, 10]	5.7635
	border_count	[32, 255]	110
	min_data_in_leaf	[1, 50]	39
CatBoostTimeSeries	sequence_length	[1, 100]	19
	iterations	[100, 1000]	867
	learning_rate	[1×10^{-3}, 0.1]	0.0309
	depth	[4, 10]	4
	l2_leaf_reg	[1×10^{-8}, 10.0]	3.7892×10^{-7}
	border_count	[32, 255]	32
	min_data_in_leaf	[1, 50]	4
EBM	max_bins	[32, 128]	114
	max_interaction_bins	[16, 64]	19
	learning_rate	[0.01, 0.1]	0.017
	min_samples_leaf	[10, 50]	42
	max_leaves	[3, 30]	3
LightGBM	max_depth	[−1, 15]	7
	learning_rate	[1×10^{-3}, 0.4]	0.3924
	n_estimators	[50, 500]	97
	subsample	[0.4, 1.0]	0.4526
	colsample_bytree	[0.5, 1.0]	0.5552
	num_leaves	[20, 150]	48
	min_child_samples	[1, 75]	59
	lambda_l1	[0, 10]	0.0285
	lambda_l2	[0, 10]	8.8419
RandomForest	n_estimators	[50, 1000]	130
	max_depth	[3, 30]	18
	min_samples_split	[2, 20]	3
	min_samples_leaf	[1, 20]	1
	max_features	[sqrt, log2, None]	sqrt

Table A1. *Cont.*

Model	Hyperparameter	Range	Best Value
SAINT	mlp_ratio	[2.0, 6.0]	2.989
	threshold	[0.01, 0.5]	0.477
	embedding_dim	[96, 160]	96
	num_heads	[2, 8]	4
	num_layers	[0, 1]	1
XGBoost	max_depth	[3, 10]	7
	learning_rate	[1×10^{-3}, 0.3]	0.1724
	n_estimators	[50, 500]	476
	subsample	[0.4, 1.0]	0.8339
	colsample_bytree	[0.5, 1.0]	0.8311

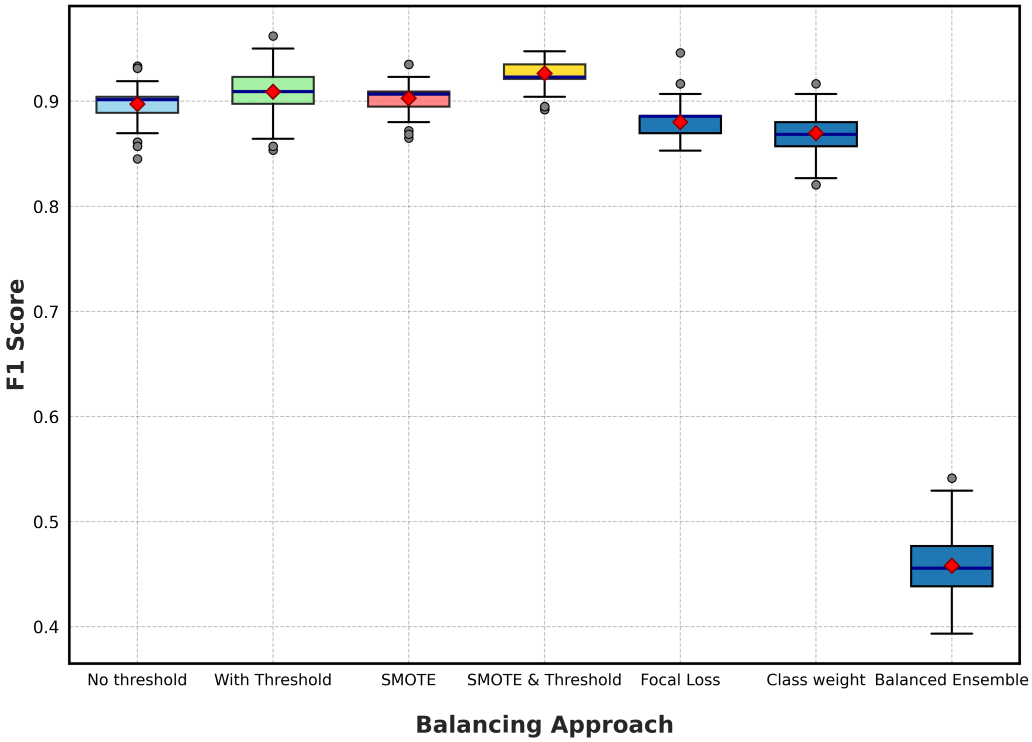


Figure A1. Additional balancing approaches that were sorted out due to insufficient performance. It is worth noting that the accuracy of Balanced Ensembles matched the other approaches, although all training was focused on the F1-score.

Table A2. Runtime Comparison of Different Approaches.

Approach	Runtime [s]
Native	308
ClassWeights	403
Threshold	328
SMOTE	1441
SMOTEwithThreshold	1727
BalancedEnsemble	1852
FocalLoss	1869

References

1. Kim, D.W.; Yoon, S. Special issue on smart automation and manufacturing. *Int. J. Comput. Integr. Manuf.* **2018**, *31*, 675–676. [CrossRef]

2. Gajšek, B.; Stradovnik, S.; Hace, A. Sustainable Move towards Flexible, Robotic, Human-Involving Workplace. *Sustainability* **2020**, *12*, 6590. [CrossRef]
3. Abubakr, M.; Abbas, A.T.; Tomaz, Í.; Soliman, M.; Luqman, M.; Hegab, H. Sustainable and Smart Manufacturing: An Integrated Approach. *Sustainability* **2020**, *12*, 2280. [CrossRef]
4. Sousa, J.; Nazarenko, A.; Grunewald, C.; Psarommatis, F.; Fraile, F.; Meyer, O.; Sarraipa, J. Zero-defect manufacturing terminology standardization: Definition, improvement, and harmonization. *Front. Manuf. Technol.* **2022**, *2*, 947474. [CrossRef]
5. Psarommatis, F. A generic methodology and a digital twin for zero defect manufacturing (ZDM) performance mapping towards design for ZDM. *J. Manuf. Syst.* **2021**, *59*, 507–521. [CrossRef]
6. Li, Y. Digital transformation and pathways for promoting global value position: An empirical study in Chinese manufacturing industries. *Int. J.-Low-Carbon Technol.* **2025**, *20*, 119–128. [CrossRef]
7. Yang, J.; Liu, Y.; Morgan, P.L. Human–machine interaction towards Industry 5.0: Human-centric smart manufacturing. *Digit. Eng.* **2024**, *2*, 100013. [CrossRef]
8. Gao, H.; Zhang, Y.; Zhou, X.; Li, D. Intelligent methods for the process parameter determination of plastic injection molding. *Front. Mech. Eng.* **2018**, *13*, 85–95. [CrossRef]
9. Lai, H. Study on Improving the Life of Stereolithography Injection Mold. *Adv. Mater. Res.* **2012**, *468–471*, 1013–1016. [CrossRef]
10. Kumar, S.; Park, H.S.; Lee, C. Data-driven smart control of injection molding process. *Cirp J. Manuf. Sci. Technol.* **2020**, *31*, 439–449. [CrossRef]
11. Kurasov, D. Injection Molding Technology. *Mater. Res. Proc.* **2022**, *21*, 251–254. [CrossRef]
12. Psarommatis, F.; May, G.; Dreyfus, P.A.; Kiritsis, D. Zero defect manufacturing: State-of-the-art review, shortcomings and future directions in research. *Int. J. Prod. Res.* **2020**, *58*, 1–17. [CrossRef]
13. Psarommatis, F.; Sousa, J.; Mendonça, J.P.; Kiritsis, D. Zero-defect manufacturing the approach for higher manufacturing sustainability in the era of industry 4.0: A position paper. *Int. J. Prod. Res.* **2022**, *60*, 73–91. [CrossRef]
14. Azamfirei, V.; Psarommatis, F.; Lagrosen, Y. Application of automation for in-line quality inspection, a zero-defect manufacturing approach. *J. Manuf. Syst.* **2023**, *67*, 1–22. [CrossRef]
15. Psarommatis, F.; May, G.; Azamfirei, V. Zero defect manufacturing in 2024: A holistic literature review for bridging the gaps and forward outlook. *Int. J. Prod. Res.* **2024**, 1–37. [CrossRef]
16. Getachew, M.; Beshah, B.; Mulugeta, A.; Kitaw, D. Application of artificial intelligence to enhance manufacturing quality and zero-defect using CRISP-DM framework. *Int. J. Prod. Res.* **2024**, 1–25. [CrossRef]
17. Tsai, M.H.; Fan-Jiang, J.C.; Liou, G.Y.; Cheng, F.J.; Hwang, S.J.; Peng, H.S.; Chu, H.Y. Development of an online quality control system for injection molding process. *Polymers* **2022**, *14*, 1607. [CrossRef]
18. Aminabadi, S.S.; Tabatabai, P.; Steiner, A.; Gruber, D.P.; Friesenbichler, W.; Habersohn, C.; Berger-Weber, G. Industry 4.0 in-line AI quality control of plastic injection molded parts. *Polymers* **2022**, *14*, 3551. [CrossRef]
19. Rousopoulou, V.; Nizamis, A.; Vafeiadis, T.; Ioannidis, D.; Tzovaras, D. Predictive maintenance for injection molding machines enabled by cognitive analytics for industry 4.0. *Front. Artif. Intell.* **2020**, *3*, 578152. [CrossRef]
20. Zhang, J.; Alexander, S. Fault diagnosis in injection moulding via cavity pressure signals. *Int. J. Prod. Res.* **2008**, *46*, 6499–6512. [CrossRef]
21. Tayalati, F.; Boukrouh, I.; Azmani, A.; Azmani, M. Implementation of Digital Twin and Deep Learning for Process Monitoring: Case Study in Injection Molding Manufacturing. In *Proceedings of the 10th World Congress on Electrical Engineering and Computer Systems and Sciences (EECSS'24)*; International Aset Inc.: Ottawa, ON, Canada, 2024.
22. Jung, H.; Jeon, J.; Choi, D.; Park, J.Y. Application of machine learning techniques in injection molding quality prediction: Implications on sustainable manufacturing industry. *Sustainability* **2021**, *13*, 4120. [CrossRef]
23. Rønsch, G.Ø.; Kulahci, M.; Dybdahl, M. An investigation of the utilisation of different data sources in manufacturing with application in injection moulding. *Int. J. Prod. Res.* **2021**, *59*, 4851–4868. [CrossRef]
24. Gim, J.; Turng, L.S. Interpretation of the effect of transient process data on part quality of injection molding based on explainable artificial intelligence. *Int. J. Prod. Res.* **2023**, *61*, 8192–8212. [CrossRef]
25. Senoner, J.; Schallmoser, S.; Kratzwald, B.; Feuerriegel, S.; Netland, T. Explainable AI improves task performance in human–AI collaboration. *Sci. Rep.* **2024**, *14*, 31150. [CrossRef]
26. Tonekaboni, S.; Joshi, S.; McCradden, M.D.; Goldenberg, A. What clinicians want: Contextualizing explainable machine learning for clinical end use. In Proceedings of the Machine Learning for Healthcare Conference, Ann Arbor, MI, USA, 8–10 August 2019; pp. 359–380.
27. Bussmann, N.; Giudici, P.; Marinelli, D.; Papenbrock, J. Explainable AI in credit risk management. *Comput. Econ.* **2021**, *57*, 203–216. [CrossRef]
28. Bodria, F.; Barbiero, A.; Giudici, P. Benchmarking explainable artificial intelligence methods for intrusion detection systems. *Expert Syst. Appl.* **2021**, *168*, 114241. [CrossRef]

29. Holstein, K.; McLaren, B.M.; Aleven, V. Co-designing AI for teacher assistance: A case study with student learning data. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*; ACM: New York, NY, USA, 2019; pp. 1–15. [CrossRef]
30. Chen, T.; He, T.; Benesty, M.; Khotilovich, V.; Tang, Y.; Cho, H.; Chen, K.; Mitchell, R.; Cano, I.; Zhou, T. *Xgboost: Extreme Gradient Boosting*, R Package Version 0.4-2. 2015. Available online: https://CRAN.R-project.org/package=xgboost (accessed on 10 December 2025).
31. Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; Liu, T.Y. Lightgbm: A highly efficient gradient boosting decision tree. *Adv. Neural Inf. Process. Syst.* **2017**, *30, 3146–3154.*.
32. Dorogush, A.V.; Ershov, V.; Gulin, A. CatBoost: Gradient boosting with categorical features support. *arXiv* **2018**, arXiv:1810.11363. [CrossRef]
33. Rigatti, S.J. Random forest. *J. Insur. Med.* **2017**, *47*, 31–39. [CrossRef]
34. Nori, H.; Jenkins, S.; Koch, P.; Caruana, R. Interpretml: A unified framework for machine learning interpretability. *arXiv* **2019**, arXiv:1909.09223. [CrossRef]
35. Erickson, N.; Mueller, J.; Shirkov, A.; Zhang, H.; Larroy, P.; Li, M.; Smola, A. Autogluon-tabular: Robust and accurate automl for structured data. *arXiv* **2020**, arXiv:2003.06505.
36. Feurer, M.; Eggensperger, K.; Falkner, S.; Lindauer, M.; Hutter, F. Auto-sklearn 2.0: The next generation. *arXiv* **2020**, arXiv:2007.04074.
37. Somepalli, G.; Goldblum, M.; Schwarzschild, A.; Bruss, C.B.; Goldstein, T. Saint: Improved neural networks for tabular data via row attention and contrastive pre-training. *arXiv* **2021**, arXiv:2106.01342. [CrossRef]
38. Akiba, T.; Sano, S.; Yanase, T.; Ohta, T.; Koyama, M. Optuna: A next-generation hyperparameter optimization framework. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*; Association for Computing Machinery: New York, NY, USA, 2019; pp. 2623–2631.
39. Zeng, X.; Martinez, T.R. Distribution-balanced stratified cross-validation for accuracy estimation. *J. Exp. Theor. Artif. Intell.* **2000**, *12*, 1–12. [CrossRef]
40. Chawla, N.V.; Bowyer, K.W.; Hall, L.O.; Kegelmeyer, W.P. SMOTE: Synthetic Minority Over-sampling Technique. *J. Artif. Intell. Res.* **2002**, *16*, 321–357. [CrossRef]
41. Ling, C.; Sheng, V. Cost-Sensitive Learning and the Class Imbalance Problem. In *Encyclopedia of Machine Learning*; Springer: Berlin/Heidelberg, Germany, 2010.
42. Lundberg, S.M.; Lee, S.I. A unified approach to interpreting model predictions. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 4765–4774.
43. Pan, S.J.; Yang, Q. A Survey on Transfer Learning. *IEEE Trans. Knowl. Data Eng.* **2010**, *22*, 1345–1359. [CrossRef]
44. Song, Z.; Zou, S.; Zhou, W.; Huang, Y.; Shao, L.; Yuan, J.; Gou, X.; Jin, W.; Wang, Z.; Chen, X.; et al. Clinically applicable histopathological diagnosis system for gastric cancer detection using deep learning. *Nat. Commun.* **2020**, *11*, 4294. [CrossRef]
45. Kabir, H.; Wu, J.; Dahal, S.; Joo, T.; Garg, N. Automated estimation of cementitious sorptivity via computer vision. *Nat. Commun.* **2024**, *15*, 9935. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.