applied sciences

MDPI

Article

Predictive Modelling of Nitrogen Content in Molten Metal During BOF Steelmaking Processes via Python-Based Machine Learning: A Benchmarking of Statistical Techniques

Jaroslav Demeter *, Branislav Buľko and Martina Hrubovčáková

Faculty of Materials, Metallurgy and Recycling, Institute of Metallurgical Technologies and Digital Transformation, Technical University of Košice, Letná 1/9, 042 00 Košice, Slovakia; branislav.bulko@tuke.sk (B.B.); martina.hrubovcakova@tuke.sk (M.H.)

* Correspondence: jaroslav.demeter@tuke.sk; Tel.: +421-55-602-2407

Abstract

This study benchmarks eight Python-based machine learning models for predicting nitrogen content across four sequential stages of BOF steelmaking. A dataset of 291 metallic samples from 76 heats was employed, covering pig iron desulfurization (PHASE #1), crude steel before BOF tapping (PHASE #2), and secondary metallurgy start (PHASE #3) and completion (PHASE #4). Linear regression, polynomial regression, ridge regression, decision tree, random forest, feedforward neural networks (FNNs), Gaussian Process Regression (GPR), and Support Vector Regression (SVR) were implemented in Python 3 with Z-score normalization and an 80/20 train–test split, and evaluated via MAE, MSE, MAPE, and R^2. Ridge regression achieved the highest accuracy in PHASE #1 (84.59%) and PHASE #4 (84.04%); FNNs excelled in PHASE #2 (78.27%) with consistent cross-phase performance; linear regression was optimal for PHASE #3 (79.06%). The advanced kernel-based methods demonstrated competitive performance, with GPR achieving 84.73% in PHASE #1 and SVR attaining 77.10% in PHASE #3 and 83.40% in PHASE #4, confirming their suitability for limited industrial datasets with a nonlinear structure. A hybrid strategy remains recommended: ridge regression for PHASES #1 and #4, FNNs for PHASES #2 and #4, and linear regression for PHASE #3, with SVR as a robust alternative in phases with moderate nonlinearity.

Keywords: machine learning; nitrogen content prediction; BOF steelmaking; modelling; process optimization; digitalization

Academic Editors: Janis Arents, Vytautas Bucinskas, Andrius Dzedzickis and Feiyang Zhao

Received: 26 February 2026
Revised: 9 April 2026
Accepted: 10 April 2026
Published: 12 April 2026

1. Introduction

Controlling the nitrogen content during steelmaking is a critical challenge for modern metallurgical operations, as nitrogen significantly affects the mechanical properties and overall quality of steel products. Precise prediction and management of nitrogen content throughout the basic oxygen furnace (BOF) steelmaking process is crucial for meeting stringent international quality standards and optimizing production efficiency [1]. Despite advances in traditional mechanistic models based on thermodynamic equilibrium calculations, conventional approaches often fail to capture the complex, multivariable interactions inherent in industrial steelmaking operations [2,3]. These operations involve multiple simultaneous reactions, variable raw material compositions and rapidly evolving process conditions, creating a highly nonlinear system [4,5]. In recent years, machine learning methodologies have emerged as transformative technologies, capable of identifying hidden

patterns in historical process data and enabling more accurate real-time predictions than traditional empirical or purely thermodynamic models [6].

It has been demonstrated that elevated nitrogen concentrations adversely affect the deep drawability and age resistance of steel. They also diminish the extent of recrystallisation and compromise critical mechanical properties, including formability and tensile strength [7–9]. Furthermore, the presence of excessive nitrogen in conventional steel grades has a detrimental effect on weldability [10,11], particularly when nitrogen levels exceed 0.4 wt.%. This negatively affects the core loss value of electrical steel [12]. Inadequate nitrogen control increases the tendency for cold cracking [13], particularly in high-strength steels, as interstitial nitrogen atoms generate stress concentrations that cause cracks to form and spread [13]. With increased nitrogen absorption, plastic deformability decreases, causing premature failure under load—a critical issue in structural applications where formability is essential for safety [14].

Steel produced via the basic oxygen furnace (BOF) process typically exhibits nitrogen concentrations ranging from 20 to 60 ppm. Nitrogen can be found in steel either in its elemental form or as a constituent of chemical compounds, such as nitrides [15]. The nitrogen content of steel is influenced by a variety of factors whose intricate interactions during the steelmaking process determine the final nitrogen concentration. Since the solubility of gaseous species in molten metals is typically minimal, these systems are usually treated as infinitely dilute solutions in thermodynamic analyses [16]. Equation (1) describes the dissolution mechanism of nitrogen in molten metal, while Equation (2) represents the corresponding Gibbs free energy change for this process [17–19]. The equilibrium constant for the reaction described in Equation (1) is expressed in the form given by Equation (3) [18–20]. A comprehensive explanation of all the nomenclature used in this manuscript can be found in the "Abbreviations" section at the end of the article.

$$^{1}/_{2}\ N_{2(g)} = [N] \tag{1}$$

$$\Delta G^{\circ} = -3590 - 23.89\ T\ [J{\cdot}mol^{-1}] \tag{2}$$

$$K_N = \frac{a_N}{(p_{N_2})^{\frac{1}{2}}} = \frac{f_N[\%N]}{\sqrt{p_{N_2}}} \tag{3}$$

Since nitrogen constitutes a dilute solution in molten metal, the activity coefficient f_N can be assumed to equal unity [21]. Since (1) represents an endothermic process, the solubility of nitrogen in the molten state increases with temperature. The solubility of nitrogen in liquid metal depends strongly on gas pressure, temperature and the metal's chemical composition, all of which vary considerably under heat. The relationship between the concentration of dissolved elemental nitrogen and gas pressure at a constant temperature is described by Sievert's law (4) [20,22,23].

$$[\%N] = K_N{\cdot}(p_{N_2})^{\frac{1}{2}} \tag{4}$$

Strict adherence to Sievert's law indicates that the gas exists in elemental form within the liquid metal. In practice, however, gas solubility often exhibits a more complex dependence on pressure, suggesting that gases may exist in non-atomic forms. Furthermore, Sievert's law only applies to lower gas pressures above the molten metal surface [24]. The temperature dependence of the concentration of dissolved elemental gas at constant pressure is given by Equation (5) [25]. The amount of dissolved nitrogen in molten metal is directly related to the equilibrium constant K_N (3). According to van 't Hoff's isobaric equation, this equilibrium constant varies with temperature, as expressed in Equation (6). A substantial increase in nitrogen solubility in iron occurs at 907 °C. Subsequently, the

solubility of nitrogen in γ-Fe decreases with increasing temperature due to the formation of nitrides as shown in Equations (7) and (8) [26].

$$[\%N] = C \cdot e^{\frac{-\Delta H}{2kT}} \quad (5)$$

$$\frac{dlnK_N}{dT} = \frac{\Delta H}{RT^2} \quad (6)$$

$$2\ \mathrm{N} + 8\ \mathrm{Fe} \leftrightarrow 2\ \mathrm{Fe_4N} \quad (7)$$

$$\mathrm{N} + 2\ \mathrm{Fe} \leftrightarrow \mathrm{Fe_2N} \quad (8)$$

High nitrogen levels can cause gas bubbles and porosity in steel because nitrogen has low solubility in molten steel, and this solubility drops sharply during solidification. When the nitrogen content exceeds its solubility limit, nitrogen bubbles form inside the steel [27]. These bubbles form in accordance with Sieverts' law, which describes the relationship between gas concentration in a material and the applied pressure [28].

Nitrogen in molten steel originates from multiple sources [29]: atmospheric air entrainment during the high-velocity oxygen lance blowing process, nitrogen-containing raw materials including ores and scrap, and intentional addition in certain high-strength steels where nitrogen enhances hardness [30]. The nitrogen content in steel is governed by complex thermodynamic and kinetic relationships involving temperature, pressure, chemical composition, and interfacial phenomena at gas-metal boundaries [31]. During the BOF process, nitrogen can be both absorbed and removed depending on process conditions, making its control particularly challenging [32].

While excess nitrogen is harmful, controlled additions benefit several steel grades. In austenitic stainless steels, nitrogen increases strength without sacrificing ductility. In duplex and super-duplex steels, it stabilizes and strengthens austenite, boosting chloride corrosion resistance. Certain high-temperature steels also use nitrogen for solid-solution strengthening and better creep resistance [33].

In practical industrial settings, the ideal relationships listed above are often violated due to the presence of surface-active elements, complex slag-metal equilibria and interfacial phenomena which cannot be easily characterized mathematically. Consequently, although purely mechanistic approaches are theoretically sound, they frequently fail to predict nitrogen evolution with sufficient accuracy for real-time process control without substantial empirical calibration.

This limitation has prompted the development and use of data-driven machine learning models that bypass the need for an explicit mechanistic formulation by learning the underlying relationships directly from industrial data [34]. Traditional approaches to nitrogen prediction have relied predominantly on linear regression models, which, although providing good interpretability and computational efficiency, often fail to capture the nonlinear relationships between process variables and nitrogen content [1]. Recent advances have shown that ensemble methods, neural networks and hybrid approaches that combine mechanistic knowledge with machine learning can substantially improve predictive accuracy. Some studies have reported accuracies ranging from 77% to 95% across different production stages [5,35,36].

This study addresses a research gap by developing and comparing predictive models based on various statistical methods. The study focuses on predicting the nitrogen content of pig iron after desulfurization, of crude steel before tapping from a basic oxygen furnace (BOF) and of steel at the beginning and end of secondary metallurgy. The primary objective was to determine the most effective model for a given dataset in terms of accuracy, generalization, and computational efficiency. To achieve this goal, several regression techniques were implemented, including linear regression, polynomial regression, ridge regression,

decision tree regressor, random forest regressor and neural networks. Models were trained and evaluated under standardized conditions to ensure a fair comparison.

As all data comes from U. S. Steel Košice, s.r.o. (Košice, Slovak Republic), the models reflect its specific processes, materials and operating conditions. Once adapted through knowledge transfer, these models can support process control by enabling proactive parameter optimization and reducing laboratory delays. This helps to prevent excessive nitrogen content in the final steel.

2. Materials and Methods

The metal production process employed in this investigation consisted of the following sequential phases. First, pig iron from the blast furnace was pretreated by desulfurization using a vertical refractory lance containing a mixture of metallic Mg and CaO (PHASE #1). This mixture was then injected into the pig iron using nitrogen as the carrier gas. The desulphurized pig iron was subsequently charged into a top-blown basic oxygen furnace (BOF) vessel with a maximum capacity of 170 tonnes that had already been pre-charged with selected steel scrap (PHASE #2). A water-cooled oxygen lance was then inserted into the vessel and high-purity oxygen was blown through it at supersonic velocity for approximately 17 min. This process oxidized impurities such as silicon, carbon, manganese and phosphorus. If the required specifications were not met by the chemical composition or temperature, oxygen reblow was performed. The crude steel was subsequently tapped into a ladle containing coke at the bottom to promote mixing through CO_2 bubble generation (PHASE #3). The addition of aluminum to the crude steel stream during tapping for deoxidation was followed by ferroalloys. Throughout secondary metallurgy, fine ferroalloys were added to complete the steel composition. Initially, the melt was stirred in the ladle using argon, and then gently bubbled with argon blown through a porous plug at the bottom of the ladle (PHASE #4). The treated steel was then prepared for continuous casting. None of the monitored heats underwent treatment in an RH vacuum degasser. As shown in Figure 1, the production process via the BOF steelmaking route is outlined schematically, along with the sequence of the individual phases.

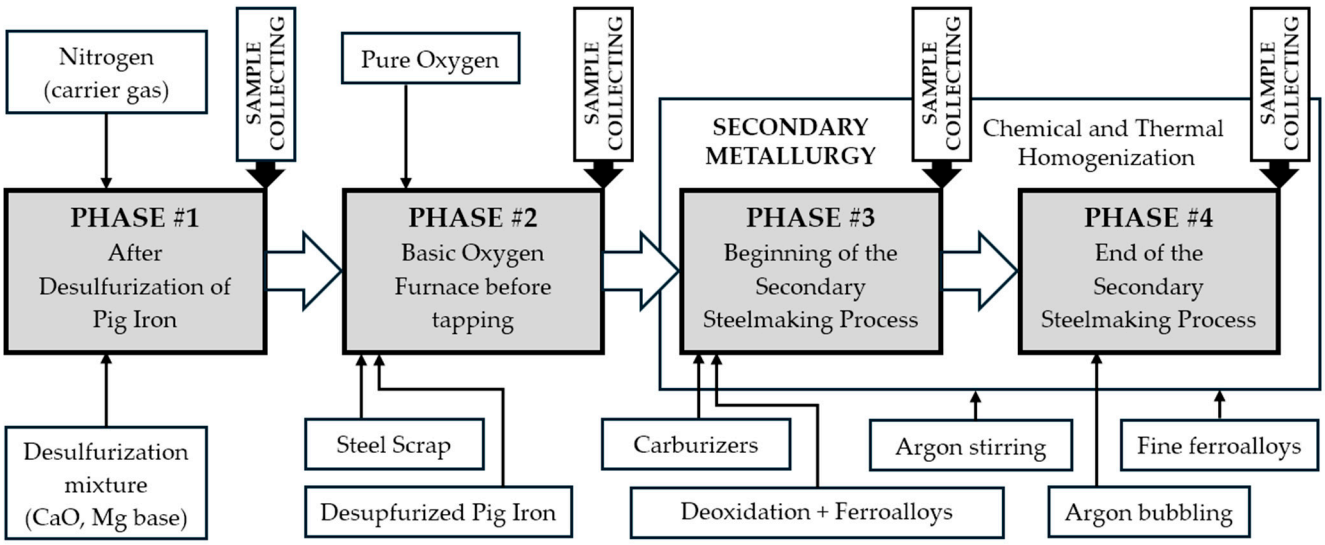

Figure 1. Schematic representation of BOF steel production and processing, showing the sequence of individual additions and operations within the monitored stages. The moment of sample collection for chemical analysis is indicated.

From 17 to 22 May 2025, a systematic sampling campaign was conducted, during which 291 metallic samples from 76 individual heats were collected and analyzed for

nitrogen content. The sampling methodology was designed to cover all four production stages within each heat, enabling continuous monitoring of nitrogen concentration variations throughout the steelmaking process. In particular, 76 samples were taken from desulphurised pig iron in the ladle, 68 from crude steel before tapping from the BOF, 75 from molten steel at the initiation of secondary metallurgy and 72 at its conclusion. The deviation from the target of 76 samples per production phase was due to sampling failures, whereby certain specimens could not be evaluated reliably for nitrogen content. The research used industrial data collected at four phases of the steelmaking process, each of which is a critical control point for managing nitrogen levels. The dataset covered a wide set of process parameters, chemical compositions and operational conditions across the entire production workflow. This approach enabled the investigation of nitrogen evolution across sequential phases of the production process within identical heats, providing insights into the dynamics of nitrogen absorption and removal during steel production (Figure 1).

The primary objective was to create regression models to predict the nitrogen content of metal in the individual phases of BOF-based steel production (PHASE #1−#4). The predictive ability of the models was then compared and evaluated in terms of accuracy, generalization and computational efficiency. The following methods were implemented: linear regression, polynomial regression, ridge regression, decision tree, random forest and neural networks. Data was preprocessed and to ensure a fair comparison across all models for each process, every model was trained and evaluated under standardized conditions as part of the machine learning process.

2.1. Analyzed Materials

The dataset comprised samples from two steel grades. The first was a structural steel containing more than 0.80% manganese with a specified minimum aluminum content (Product #1). The second was a deep-drawing, aluminum-killed steel (Product #2). Both grades conformed to the chemical composition requirements summarized in Table 1.

Table 1. Chemical composition specifications for the analyzed steel grades included in the study.

Chemical Composition	Grade of Steel	
	Product #1	**Product #2**
C (%)	0.07–0.21	0.02–0.1
Mn (%)	0.8–1.6	0.1–0.55
Si (%)	0.03–0.6	max. 0.08
Al (%)	min. 0.02	0.02–0.07
P (%)	max. 0.025	0.01–0.07
S (%)	max. 0.020	max. 0.020
Nb (%)	-	0.004–0.0075

Nitrogen concentrations in pig iron and steel were measured at the Quantometric Laboratory (Labortest, s.r.o., Košice, Slovak Republic) using an ELTRA ON 900 combustion analyzer operating with thermal conductivity detection, in accordance with ASTM E1019. The instrument provides a measurement ranging from 0.0001 to 0.03% of nitrogen, with an accuracy of ±0.1 ppm or ±1% [37]. Routine manufacturer calibration, annual servicing, hourly checks with a standard sample and duplicate (primary and control) analyses for each specimen were included in the quality assurance procedures [38]. Daily measurements of certified reference materials were performed to update calibration factors and correct drift. With appropriate maintenance and drift adjustment, ELTRA analyzers generally achieve a standard deviation of 1–3% and inter-day precision of ±0.1–2 ppm at low nitrogen levels,

and a relative standard deviation of less than 1.5% at higher concentrations, which meets ASTM E1019-18 requirements [39].

2.2. Analyzed Parameters

The desulfurization of pig iron (PHASE #1) involved treating the iron to reduce its sulfur content before steelmaking. The dataset included 15 parameters covering the chemical composition (C, Mn, Si and P), the levels of sulfur before and after treatment, temperature readings, mass-balance data and process-timing parameters. The target variable was the nitrogen percentage measured after desulfurization.

The BOF phase (PHASE #2) represents the main steelmaking operation, whereby molten pig iron is converted into crude steel through the introduction of oxygen. The dataset contained 32 input variables, including steel and slag chemical composition, time-related parameters, temperature and oxygen-activity measurements, mass-balance data for raw materials and fluxes, and other operational factors. The target variable was the nitrogen percentage in the crude steel prior to BOF tapping.

The third phase (PHASE #3) marked the start of secondary steelmaking in the ladle, emphasizing initial composition adjustment and temperature control. The dataset contained 12 input variables, including the composition of the steel after tapping, the initial temperature of the secondary steelmaking process, the duration of the tapping process, mass-related parameters and factors related to the geometry of the tapping. The nitrogen percentage in the metal at the start of secondary steelmaking was the target variable.

The last PHASE #4 covered the conclusion of secondary steelmaking, incorporating the final stages of alloying, deoxidation and argon stirring. This dataset contained 35 input variables, including deoxidation additives (in varying forms of aluminum), the full chemical composition, process duration parameters, argon stirring conditions, mass balance data and ferroalloy addition practices. The target variable was the final nitrogen percentage in the steel following secondary metallurgy.

2.3. Applied Software

The nitrogen level was determined using heat and sample identification numbers, which were then matched with the relevant database entries. These synchronized datasets provided information on the relevant chemical composition, temperature, weight and other process-specific parameters associated with each production stage.

This consolidated dataset was initially compiled using Microsoft Excel 365 (version 2601, build 16.0.19628.20166) together with the Lumivero XLSTAT 2019 statistical add-in (version 2019.2.2; Lumivero Inc., Denver, CO, USA). The data analysis workflow was implemented in Python 3 (version 3.13) utilizing the Jupyter Notebook (version 7.5.1) environment. Data manipulation and exploratory analysis were conducted using pandas (version 2.3.2) for tabular data operations and NumPy (version 2.4.0) for numerical computations. Visualization was achieved through matplotlib.pyplot (version 3.10.8) for basic plotting and seaborn for advanced statistical graphics, including heatmaps and distribution plots. Data preprocessing employed scikit-learn's (version 1.8) StandardScaler for Z-score normalization of both input features and target variables. Processed datasets and fitted scalers were serialized using Python's pickle module for reproducibility.

The modelling phase utilized multiple machine learning frameworks. Conventional regression models (linear, polynomial, and Ridge regression with L2 regularization) were implemented via scikit-learn's linear_model module (version 7.0). Tree-based ensemble methods, including Random Forest (100 estimators) and Decision Tree regressors, were deployed from scikit-learn's tree (version 2.2.1) and ensemble modules (version 0.1.3).

Decision tree visualization was attempted using the graphviz library (version 0.21), though execution was hindered by system path configuration issues.

Deep learning models were constructed using PyTorch (version 2.10.0), implementing fully connected feedforward neural networks with sequential architecture. The network comprised three hidden layers (256, 128, and 32 neurons) with ReLU activation functions and dropout regularization (p = 0.3). Training optimization employed the AdamW algorithm with adaptive learning rate scheduling (ReduceLROnPlateau) and early stopping mechanisms. Model states persisted using PyTorch's native state dictionary format.

Model evaluation consistently employed scikit-learn's metrics module, calculating mean squared error (MSE), mean absolute error (MAE), root mean squared error (RMSE), coefficient of determination (R^2), mean absolute percentage error (MAPE) and accuracy (100-MAPE) across training and testing partitions. All experimental results, including performance metrics and trained models, were systematically archived in CSV and binary formats for subsequent analysis.

Two additional advanced regression models were incorporated to address the limitations of simpler parametric methods under constrained sample sizes. Gaussian Process Regression (GPR) was implemented via scikit-learn's GaussianProcessRegressor class (version 1.8) using a composite kernel composed of a Matérn covariance function (ν = 2.5) combined with a WhiteKernel for noise estimation. Kernel hyperparameters were optimized by maximizing the log marginal likelihood using the L-BFGS-B algorithm with ten random restarts (n_restarts_optimizer = 10) to mitigate local optima. Target normalization was enabled (normalize_y = True) to improve numerical conditioning. Support Vector Regression (SVR) was implemented via scikit-learn's SVR class with a radial basis function (RBF) kernel. Hyperparameter optimization was conducted using five-fold cross-validated grid search (GridSearchCV, cv = 5) over the regularization parameter $C \in \{0.1, 1, 10, 100\}$, the epsilon tube width $\varepsilon \in \{0.0001, 0.001, 0.01, 0.1\}$ and the kernel coefficient $\gamma \in$ {'scale', 'auto', 0.1, 1.0}. Both models were trained and evaluated under identical standardized conditions applied to all other models, ensuring methodological consistency.

2.4. Exploratory Analysis

An Exploratory Data Analysis (EDA) was conducted to gain a thorough understanding of the characteristics of the dataset to identify the underlying relationships between the process parameters and to detect any potential anomalies or patterns within the steelmaking process data. This preliminary investigation was crucial for making informed decisions about which features to select and how to develop models, as it revealed data distribution properties, variable interdependencies and potential quality issues that could affect predictive performance.

The dataset comprised four distinct subsets corresponding to sequential stages of the steelmaking process: Desulphurization of pig iron (PHASE #1; 15 parameters), BOF steel production (PHASE #2; 32 parameters), Secondary metallurgy beginning (PHASE #3; 12 parameters), and Secondary metallurgy end (PHASE #4; 35 parameters). Each subset contained 77 observations representing individual production heats.

A statistical characterization was performed using descriptive metrics, including minimum, maximum, arithmetic mean and standard deviation, for all numerical variables. These summary statistics were systematically computed and exported to structured documentation for comprehensive review. Distribution analysis used histogram visualization for each parameter to enable assessment of normality, skewness and outliers. Selected histograms for PHASE #1–#4 are shown in Figure 2a–d.

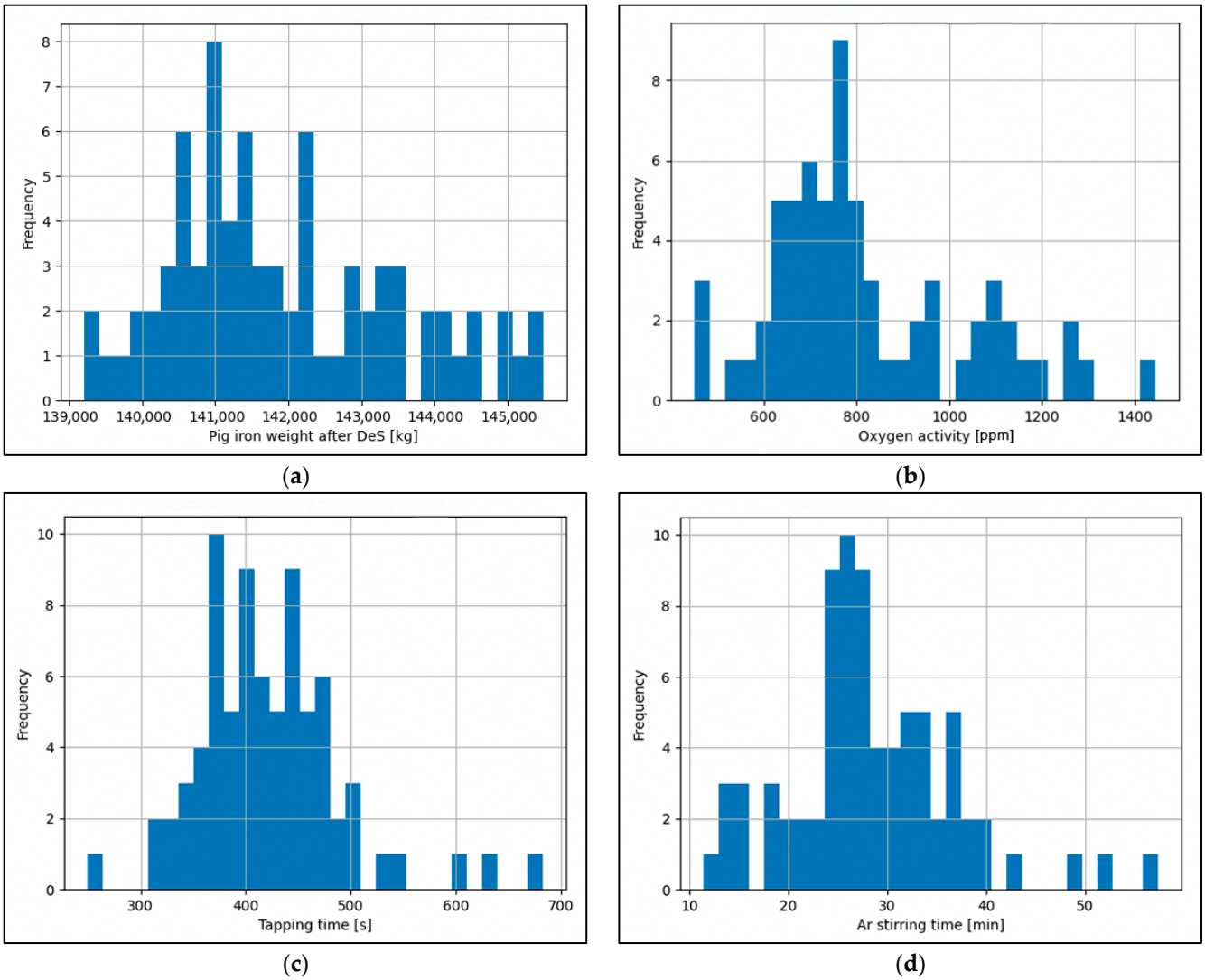

Figure 2. Selected histograms for given phases of steelmaking process: (**a**) PHASE #1: Distribution of pig iron weight after its desulfurization; (**b**) PHASE #2: Distribution of oxygen activity in crude steel; (**c**) PHASE #3: Distribution of tapping time from BOF; (**d**) PHASE #4: Distribution of argon stirring time during secondary steelmaking.

A correlation analysis was performed to quantify the linear relationships between the process variables and the target nitrogen content. The correlation matrices were visualized as annotated heatmaps using a diverging color scheme to facilitate the identification of multicollinearity and relevant predictive features. Rendering these matrices at high resolution provided visual evidence of parameter interactions across production stages. Correlation heatmaps generated for PHASE #1−#4 are depicted in Figure 3a–d.

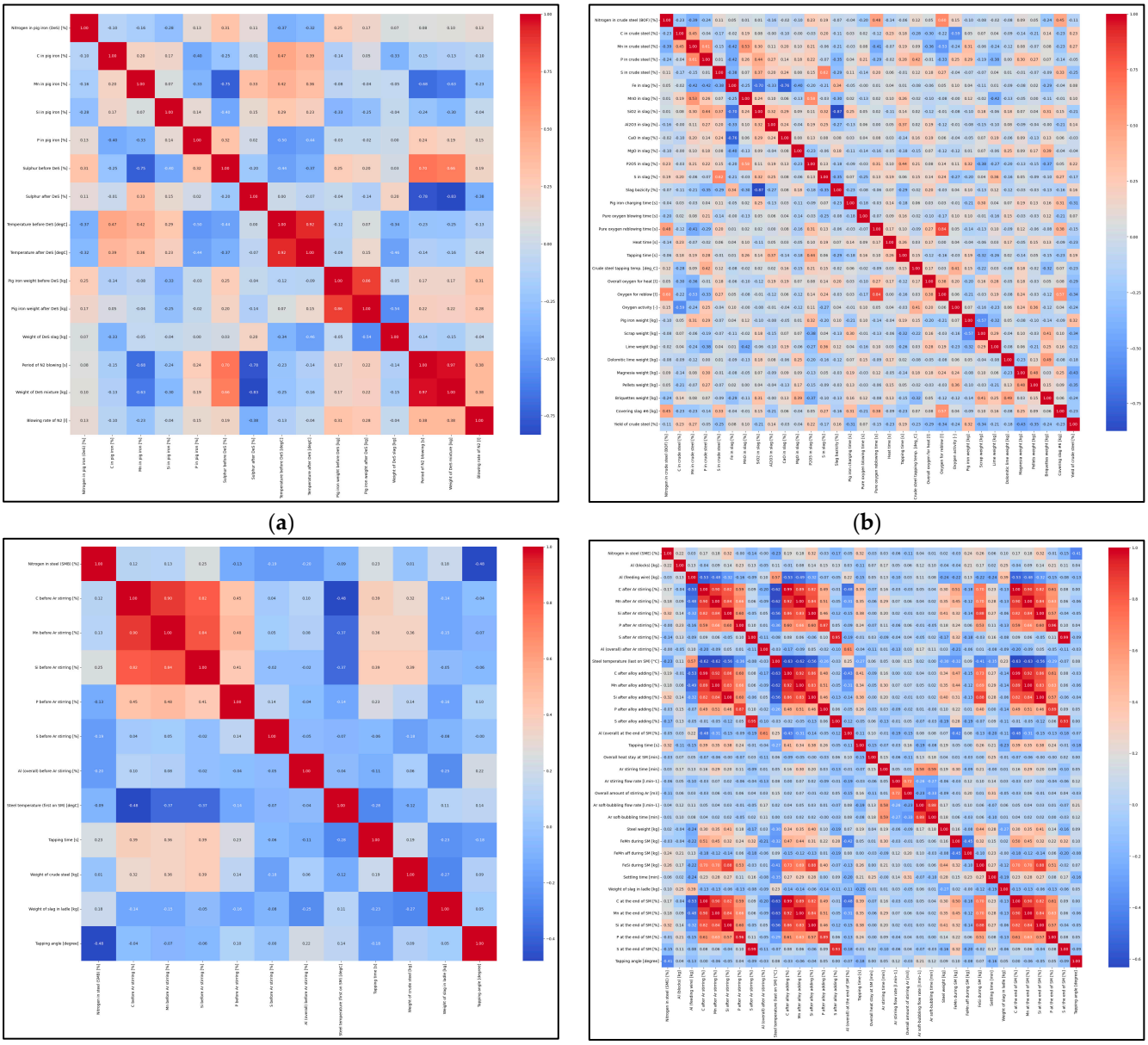

Figure 3. Correlation heatmaps for particular phases of production process: (**a**) PHASE #1: Correlation matrix of 15 parameters after desulfurization of pig iron; (**b**) PHASE #2: Correlation matrix of 32 parameters prior to tapping of crude steel from BOF; (**c**) PHASE #3: Correlation matrix of 13 parameters at the beginning of secondary steelmaking; (**d**) PHASE #4: Correlation matrix of 35 parameters at the end of secondary steelmaking.

2.5. Data Preprocessing

Following exploratory analysis, data preprocessing prepared the datasets for modelling through standardization and systematic partitioning. Z-score normalization was applied independently to input features and target variables using scikit-learn's StandardScaler, transforming all parameters to zero mean and unit variance. This standardization ensures numerical stability during model training and enables fair comparison of feature importance across parameters with disparate scales.

Critically, separate scaler objects were fitted and preserved for both predictor and response variables, enabling inverse transformation of model predictions to original measurement units for practical interpretation. The fitted scalers and normalized datasets were serialized using Python's pickle protocol, ensuring reproducibility and facilitating

deployment in production environments. Data partitioning employed a simple random 80:20 train–test split with a fixed random seed (random_state = 42) applied independently to each production phase. This approach ensures reproducibility and enables fair cross-model comparison within each phase, as all models are trained and evaluated on identical data partitions.

2.6. Prediction Methods

Eight distinct modelling approaches were systematically evaluated for nitrogen content prediction across four steelmaking process stages, progressing from classical parametric methods to advanced non-parametric techniques and deep learning architectures. Particular methods can be divided into Classical Regression Methods (represented by linear regression, polynomial regression and ridge regression), Tree-based Ensemble Methods (represented by decision tree regressor and random forest regressor) and Deep Learning Architecture (represented by neural network).

2.6.1. Linear Regression

This method served as the baseline model, establishing a linear relationship between process parameters and nitrogen content through ordinary least squares (OLS) estimation. The model assumes the functional form (9) [40]. The "Abbreviations" section of the article explains nomenclature.

$$f(x) = \beta_0 + \beta_1 \cdot x_1 + \cdots + \beta_n \cdot x_n \quad (9)$$

The strengths of linear regression include high interpretability, low computational complexity and its suitability as a reference model. However, it cannot directly capture nonlinearities and complex interactions, and it may be ineffective in highly nonlinear processes [41].

2.6.2. Polynomial Regression

Method extends the linear framework by introducing higher-order polynomial terms and interaction effects amongst predictors. The transformation generates an augmented feature space comprising quadratic terms and pairwise interactions, expressed mathematically as (10) [42,43].

$$f(x) = \beta_0 + \sum_{i=1}^{n} \beta_i x_i + \sum_{i=1}^{n} \beta_{ii} x^2 + \sum_{1 \le i < j \le n} x_i x_j \quad (10)$$

Whilst maintaining linearity in parameters, this approach captures curvature and interactive effects characteristic of complex industrial processes. The polynomial degree constitutes the primary hyperparameter, with second-degree transformations offering a pragmatic balance between model complexity and interpretability. However, feature space expansion introduces elevated risks of overfitting, particularly in high-dimensional contexts with limited observations, necessitating careful regularization strategies [44,45].

2.6.3. Ridge Regression

This method addresses multicollinearity and model instability through L2 regularization, incorporating a penalty term proportional to the squared magnitude of coefficients. The sum of squares of errors with the added term $\alpha \cdot \Sigma \beta_i^2$ is minimized. This penalty shrinks coefficient estimates towards zero, stabilizing parameter estimation in the presence of correlated predictors whilst retaining all features in the model [46]. The hyperparameter α (in this study $\alpha = 100$) requires tuning: higher values increase regularization intensity, approaching coefficient nullification, whilst lower values approximate standard OLS estimation. Ridge regression proves particularly valuable for metallurgical datasets characterized by inherently correlated process parameters [47].

2.6.4. Decision Tree Regressor

It employs recursive binary partitioning of the feature space, constructing a hierarchical structure through successive conditional splits. Each internal node represents a decision rule of the form $x_i \leq t$, selected to maximise variance reduction in descendant nodes, typically measured by mean squared error minimization [48]. Terminal leaves contain constant predicted values, commonly the arithmetic means of training observations within that partition. This non-parametric approach inherently captures non-linear relationships and higher-order interactions without requiring explicit feature transformation or distributional assumptions [49]. The method proves intuitive for domain experts, as decision paths constitute interpretable logical rules. However, unconstrained trees exhibit high variance and propensity for overfitting, necessitating regularization through depth limitation, minimum samples per split, or pruning strategies [50].

2.6.5. Random Forest Regressor

Method constitutes an ensemble method combining multiple decision trees through bootstrap aggregating (bagging). Each constituent tree trains on a bootstrap sample—a random subset drawn with replacement from the training data—whilst additionally employing feature randomization at each split point, considering only a random subset of predictors [51]. Dual randomization mechanism substantially reduces model variance compared to single trees, enhancing generalization performance whilst preserving non-linearity and interaction modelling capabilities [52]. The method proves robust to noise and outliers, requires minimal hyperparameter tuning, and provides implicit feature importance rankings through mean decrease in impurity metrics. Primary hyperparameters include the number of estimators, maximum tree depth, feature subset size, and minimum leaf sample requirements [53]. In this study, *n-estimator* = 100 trees; *random_state* = 42; *max_depth* = None; *min_samples_split* = 2; *min_samples_leaf* = 1; *max_features* = 1.0 (all features considered at each split in scikit-learn 1.8); *bootstrap* = True.

2.6.6. Feedforward Neural Networks (FNNs)

The FNN method is the simplest and most widely used neural architecture. FNNs consist of an input layer, one or more hidden layers, and an output layer, with information flowing strictly forward through weighted connections and nonlinear activations [54]. A feedforward neural network contains no closed loops in its topology. Its input nodes have no incoming arcs, and its output nodes have no outgoing arcs [55]. The architecture consists of fully connected (dense) layers arranged sequentially, each implementing an affine transformation followed by a non-linear activation function. In the present study, the network follows the architecture input_dim $\rightarrow$ 256 $\rightarrow$ 128 $\rightarrow$ 32 $\rightarrow$ 1, where input_dim corresponds to the number of process parameters specific to each production stage. The number of training parameters was as follows: PHASE #1: 40,353; PHASE #2: 44,897; PHASE #3: 39,585; PHASE #4: 45,665.

Rectified linear unit (ReLU) activation functions $(ReLU, \sigma(x) = max(0, x))$ are used between hidden layers to improve computational efficiency and alleviate vanishing-gradient issues. Regularization is applied via dropout (p = 0.3) after the first two hidden layers, randomly deactivating neurons to reduce feature co-adaptation. The output layer contains a single neuron without an activation function, producing a continuous regression output. Training uses mini-batch stochastic gradient descent via PyTorch's DataLoader (batch size = 4), balancing gradient stability with the small dataset size. Parameter updates employ AdamW (lr = 0.001, weight_decay = 10^{-4}), which applies decoupled weight decay for more consistent regularization. The model is optimized using the L1 loss (MAE), selected for its robustness to outliers typical of industrial process measurements.

Adaptive learning-rate scheduling is handled by ReduceLROnPlateau, which lowers the rate by a factor of 0.1 after 10 epochs without validation-loss improvement. This supports coarse updates early in training and finer adjustments near convergence. Early stopping (patience = 20) further stabilizes training by halting optimization when validation performance no longer improves and restoring the best model state. Although the maximum training length was 300 epochs, early stopping typically ended training well before this limit. Together, adaptive scheduling and early stopping help prevent overfitting while ensuring reliable convergence, which is crucial for small metallurgical datasets.

2.6.7. Gaussian Process Regression

Gaussian Process Regression (GPR) is a non-parametric Bayesian method that defines a prior distribution directly over functions, yielding a closed-form posterior with a predictive mean μ_* (11) and variance σ_*^2 (12) [56]:

$$\mu_* = k_* \left(K + \sigma_n^2 I \right)^{-1} y \tag{11}$$

$$\sigma_*^2 = k(x_*, x_*) - k_* \left(K + \sigma_n^2 I \right)^{-1} k_* \tag{12}$$

A composite Matérn ($\nu = 2.5$) + WhiteKernel was employed, with hyperparameters optimised by maximising the log-marginal likelihood ($n_{\text{restarts}} = 10$). GPR was implemented via *sklearn.gaussian_process. GaussianProcessRegressor* (scikit-learn 1.8). Its principal advantage over all other methods in this study is the simultaneous provision of a calibrated predictive uncertainty estimate σ_* alongside each point prediction, whilst Bayesian regularisation of model complexity makes it particularly well-suited to limited-data regression [56].

2.6.8. Support Vector Regression

Support Vector Regression (SVR) seeks a function $f(\mathbf{x})$ that deviates from observed targets by at most ε, whilst minimising model complexity through the following primal optimisation (13) [57]:

$$min_{w,b,\xi,\xi*} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^{n} (\xi_i + \xi_i^*) \tag{13}$$

The dual formulation admits kernel substitution; in the present study, the Radial Basis Function (RBF) kernel $k(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma \parallel \mathbf{x}_i - \mathbf{x}_j \parallel^2)$ was employed [58], which is well-suited to the nonlinear thermodynamic interactions governing nitrogen solubility in molten steel. The three hyperparameters C, ε, and γ were jointly optimised via exhaustive grid search with 5-fold cross-validation ($C \in \{0.1, 1, 10, 100\}$; $\varepsilon \in \{0.0001, 0.001, 0.01, 0.1\}$; $\gamma \in \{\text{scale}, \text{auto}, 0.1, 1\}$). SVR was implemented using *sklearn.svm.SVR* (scikit-learn 1.8) under the same standardized training and evaluation conditions as all other methods (Section 2.5).

3. Results

The comparative evaluation of eight regression models—Linear Regression, Polynomial Regression, Ridge Regression, Decision Tree Regressor, Random Forest Regressor, Feedforward Neural Networks (FNNs), Gaussian Process Regression (GPR) and Support Vector Regression (SVR) —was conducted across four critical phases of the BOF steelmaking production route, designated as PHASE #1 (after desulfurization of pig iron), PHASE #2 (crude steel prior to tapping from the BOF), PHASE #3 (steel at the beginning of secondary metallurgy), and PHASE #4 (steel at the end of secondary metallurgy). All models were trained on Z-score normalized industrial data with an 80/20 train–test split and a fixed random seed to ensure reproducibility. The single train–test split employed in

this study ensures reproducibility and fair cross-model comparison. Model performance on the held-out test set was evaluated using four metrics: MAE, MSE, R^2, MAPE, and prediction accuracy (100% − MAPE), which provides an intuitive percentage-based comparison across models and phases. Tables 2–9 present both training and test performance metrics for all model-phase combinations, enabling direct assessment of overfitting magnitude. Figures 4–9 employ parity plots (predicted vs. actual nitrogen content) with diagonal reference lines. Tables and figures are followed by an analysis of performance trends in relation to metallurgical process complexity and feature dimensionality. Two additional kernel-based methods—Gaussian Process Regression (GPR) with a Matérn ($\nu = 2.5$) covariance kernel and Support Vector Regression (SVR) with an RBF kernel and cross-validated hyperparameters—are incorporated into the same evaluation framework, extending the comparative analysis to eight models and providing a more rigorous assessment of nonlinear behavior and limited-data effects across PHASES #1–#4.

Table 2. Comparison of linear regression performance for phases in BOF production route.

Metrics	Dataset	Production Phases			
		PHASE #1	**PHASE #2**	**PHASE #3**	**PHASE #4**
MAE	Test	0.00066	0.00117	0.00076	0.00118
	Train	0.00052	0.00040	0.00056	0.00038
MSE	Test	7.03×10^{-7}	1.71×10^{-6}	1.27×10^{-6}	3.92×10^{-6}
	Train	4.12×10^{-7}	2.86×10^{-7}	6.86×10^{-7}	2.80×10^{-7}
RMSE (wt.%)	Test	0.00084	0.00131	0.00113	0.00198
	Train	0.00064	0.00053	0.00083	0.00053
MAPE (%)	Test	16.30289	62.70137	20.94005	30.60207
	Train	13.70989	18.72288	18.29465	12.13652
100 − MAPE (%)	Test	83.69711	37.29863	79.05995	69.39793
	Train	86.29011	81.27712	81.70535	87.86348
R^2	Test	0.69128	0.74286	0.62	0.87628
	Train	0.81930	0.81660	0.79430	0.92620

Table 3. Evaluation of polynomial regression accuracy for individual phases of the BOF steelmaking route.

Metrics	Dataset	Production Phases			
		PHASE #1	**PHASE #2**	**PHASE #3**	**PHASE #4**
MAE	Test	0.00123	0.00067	0.00498	0.00094
	Train	0.00000	0.00000	0.00000	0.00000
MSE	Test	2.99×10^{-6}	8.24×10^{-7}	4.29×10^{-5}	2.01×10^{-6}
	Train	5.25×10^{-36}	2.15×10^{-36}	2.95×10^{-36}	7.20×10^{-36}
RMSE (wt.%)	Test	0.00173	0.00091	0.00655	0.00142
	Train	2.29×10^{-18}	1.47×10^{-18}	5.43×10^{-18}	2.68×10^{-18}
MAPE (%)	Test	31.06468	36.90866	148.15750	23.33192
	Train	0.00000	0.00000	0.00000	0.00000
100 − MAPE (%)	Test	68.93532	63.09134	−48.15750	76.66808
	Train	100.00000	100.00000	100.00000	100.00000
R^2	Test	0.31279	0.47219	−11.87379	−0.46895
	Train	1.00000	1.00000	1.00000	1.00000

Table 4. Ridge regression accuracy is evaluated for the different phases of the BOF steelmaking route.

Metrics	Dataset	Production Phases			
		PHASE #1	**PHASE #2**	**PHASE #3**	**PHASE #4**
MAE	Test	0.00063	0.00066	0.00083	0.00066
	Train	0.00054	0.00048	0.00066	0.00060
MSE	Test	6.38×10^{-7}	6.33×10^{-7}	1.25×10^{-6}	1.09×10^{-6}
	Train	4.74×10^{-7}	5.51×10^{-7}	8.78×10^{-7}	7.39×10^{-7}
RMSE (wt.%)	Test	0.00080	0.00080	0.00112	0.00105
	Train	0.00069	0.00074	0.00094	0.00086
MAPE (%)	Test	15.40945	34.40710	25.22126	15.95519
	Train	14.22858	21.66274	21.91848	20.02565
100 − MAPE (%)	Test	84.59055	65.59290	74.77874	84.04481
	Train	85.77142	78.33726	78.08152	79.97435
R^2	Test	0.72000	0.59424	0.62644	0.71000
	Train	0.79170	0.64660	0.73700	0.80510

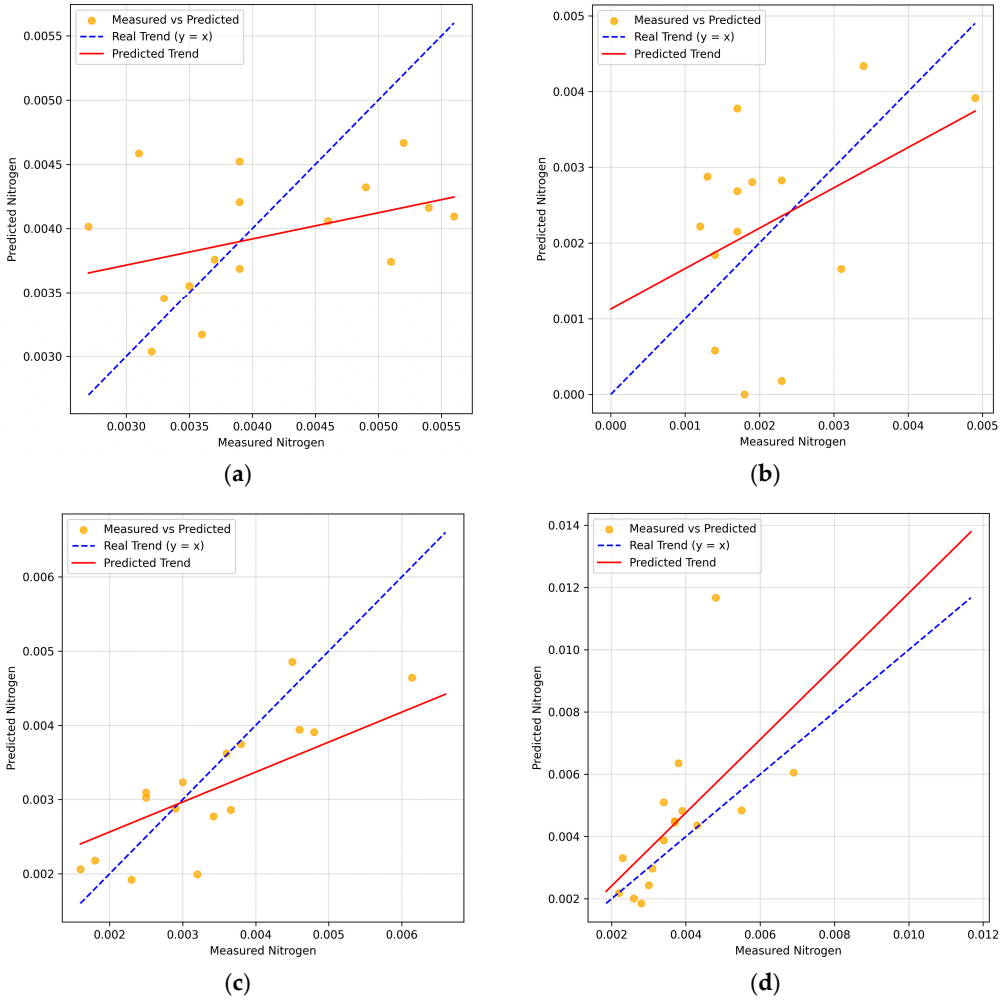

Figure 4. Parity plot of measured versus predicted nitrogen content in molten metal for the linear regression model (test dataset); the dashed diagonal line (y = x) represents perfect agreement between measured and predicted values: (**a**) PHASE #1: after desulfurization of the pig iron; (**b**) PHASE #2: crude steel in BOF prior to tapping; (**c**) PHASE #3: steel at the beginning of secondary steelmaking; (**d**) PHASE #4: steel at the end of secondary steelmaking.

Table 5. The predictive accuracy of the decision tree regressor model is assessed across the individual phases of the BOF steelmaking route.

Metrics	Dataset	Production Phases			
		PHASE #1	**PHASE #2**	**PHASE #3**	**PHASE #4**
MAE	Test	0.00063	0.00102	0.00096	0.00093
	Train	0.00000	0.00086	0.00096	0.00093
MSE	Test	6.30×10^{-7}	2.13×10^{-6}	1.62×10^{-6}	1.54×10^{-6}
	Train	0.00000	9.78×10^{-6}	1.62×10^{-6}	1.54×10^{-6}
RMSE (wt.%)	Test	0.00079	0.00146	0.00127	0.00124
	Train	0.00000	0.00000	0.00000	0.00000
MAPE (%)	Test	21.41843	62.43555	25.72360	23.24638
	Train	0.00000	21.41843	25.72360	23.24638
100 − MAPE (%)	Test	78.58157	37.56445	74.27640	76.75362
	Train	100.00000	78.58157	74.27640	76.75362
R^2	Test	0.72344	0.36726	0.51499	0.59401
	Train	1.00000	0.37320	0.51499	0.59401

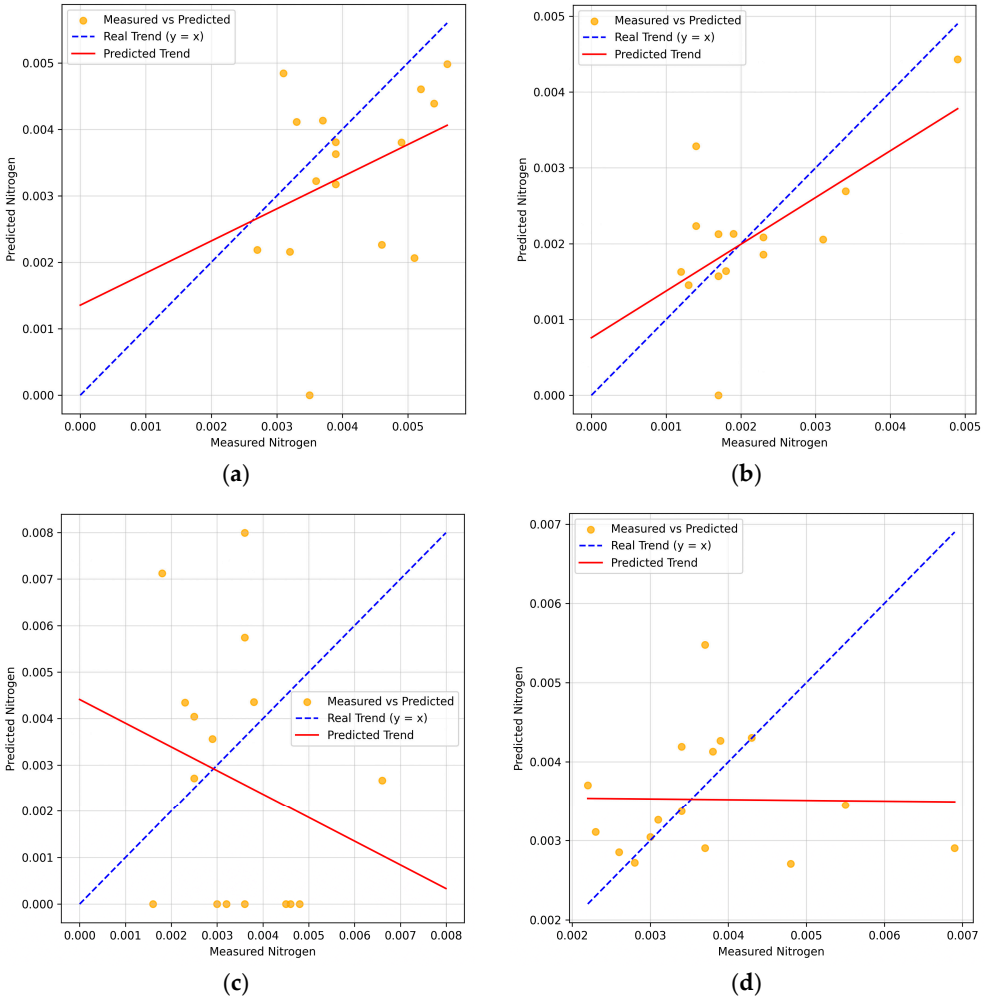

Figure 5. Measured versus predicted nitrogen content in molten metal for the polynomial regression model evaluated on the test dataset: (**a**) PHASE #1: after desulfurization of the pig iron; (**b**) PHASE #2: crude steel in BOF prior to tapping; (**c**) PHASE #3: steel at the beginning of secondary steelmaking; (**d**) PHASE #4: steel at the end of secondary steelmaking.

Table 6. The random forest regressor's predictive performance is assessed for each phase of the BOF steelmaking process.

Metrics	Dataset	Production Phases			
		PHASE #1	**PHASE #2**	**PHASE #3**	**PHASE #4**
MAE	Test	0.00070	0.00054	0.00082	0.00077
	Train	0.00023	0.00022	0.00027	0.00027
MSE	Test	7.31×10^{-7}	4.18×10^{-7}	1.27×10^{-6}	1.35×10^{-6}
	Train	8.90×10^{-8}	1.25×10^{-7}	1.28×10^{-7}	1.41×10^{-7}
RMSE (wt.%)	Test	0.00086	0.00065	0.00113	0.00116
	Train	0.00030	0.00035	0.00036	0.00038
MAPE (%)	Test	17.48432	29.14900	23.34319	18.69600
	Train	6.09263	9.99410	8.86757	9.21979
100 − MAPE (%)	Test	82.51568	70.85100	76.65681	81.30400
	Train	93.90737	90.00590	91.13243	90.78021
R^2	Test	0.67898	0.73224	0.61891	0.64393
	Train	0.96090	0.92000	0.96170	0.96290

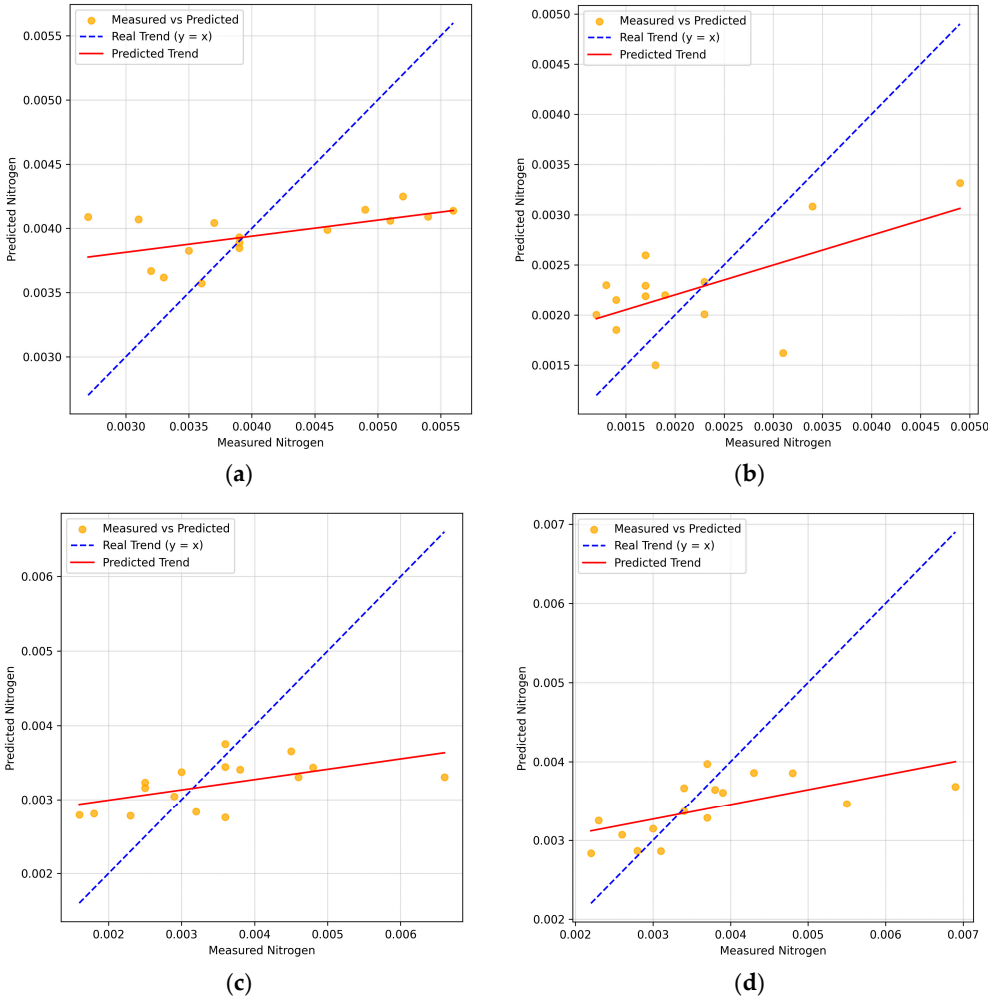

Figure 6. The predictive accuracy of nitrogen in metal is shown as a graph, using a model based on a test dataset of the ridge regression model: (**a**) PHASE #1: after desulfurization of the pig iron; (**b**) PHASE #2: crude steel in BOF prior to tapping; (**c**) PHASE #3: steel at the beginning of secondary steelmaking; (**d**) PHASE #4: steel at the end of secondary steelmaking.

Table 7. Evaluation of the feedforward neural network's predictive accuracy is conducted separately for all phases of the BOF production process.

Metrics	Dataset	Production Phases			
		PHASE #1	**PHASE #2**	**PHASE #3**	**PHASE #4**
MAE	Test	0.00071	0.00042	0.00077	0.00076
	Train	0.00030	0.00011	0.00026	0.00024
MSE	Test	7.74×10^{-7}	2.25×10^{-7}	8.83×10^{-7}	8.49×10^{-7}
	Train	1.80×10^{-7}	8.18×10^{-8}	1.53×10^{-7}	2.23×10^{-7}
RMSE (wt.%)	Test	0.00092	0.00050	0.00119	0.00075
	Train	0.00042	0.00029	0.00039	0.00047
MAPE (%)	Test	21.30315	20.22735	20.99344	19.65124
	Train	17.74781	14.19212	19.77758	17.67900
100 − MAPE (%)	Test	78.69685	79.77265	79.00656	80.34876
	Train	82.25219	85.80788	80.22242	82.32100
R^2	Test	0.63132	0.84000	0.57518	0.85074
	Train	0.70682	0.89284	0.67563	0.91352

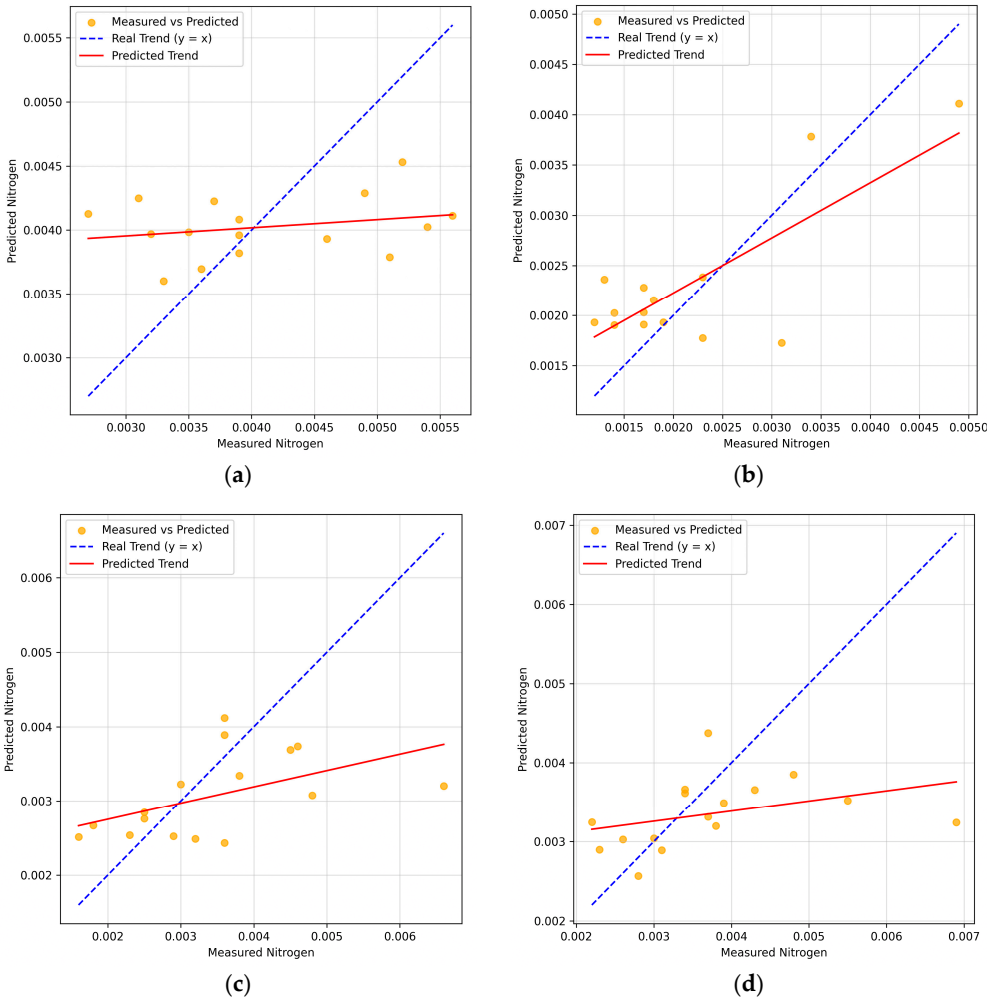

Figure 7. Parity plot of measured versus predicted nitrogen content in molten metal for the decision tree regressor model (test dataset); the dashed diagonal line (y = x) represents perfect agreement between measured and predicted values: (**a**) PHASE #1: after desulfurization of the pig iron; (**b**) PHASE #2: crude steel in BOF prior to tapping; (**c**) PHASE #3: steel at the beginning of secondary steelmaking; (**d**) PHASE #4: steel at the end of secondary steelmaking.

Table 8. Evaluation of GPR prediction accuracy for individual phases of the BOF steelmaking route (test set, 80/20 split, Matérn ν = 2.5 kernel, n_restarts_optimizer = 10).

Metrics	Dataset	Production Phases			
		PHASE #1	**PHASE #2**	**PHASE #3**	**PHASE #4**
MAE	Test	6.12×10^{-4}	8.21×10^{-4}	1.03×10^{-3}	1.06×10^{-3}
	Train	1.09×10^{-11}	9.09×10^{-12}	1.69×10^{-11}	1.38×10^{-11}
MSE	Test	6.28×10^{-7}	1.07×10^{-6}	1.61×10^{-6}	2.46×10^{-6}
	Train	2.04×10^{-22}	2.23×10^{-22}	8.58×10^{-22}	3.57×10^{-22}
RMSE (wt.%)	Test	7.93×10^{-4}	1.03×10^{-3}	1.27×10^{-3}	1.57×10^{-3}
	Train	4.52×10^{-11}	4.73×10^{-11}	9.27×10^{-11}	5.98×10^{-11}
MAPE (%)	Test	15.27	39.98	33.07	26.67
	Train	≈0.00	≈0.00	≈0.00	≈0.00
100 − MAPE (%)	Test	84.73	60.02	66.93	73.33
	Train	100.00	100.00	100.00	100.00
R^2	Test	0.186	−0.100	−0.063	−0.759
	Train	1.000	1.000	1.000	1.000

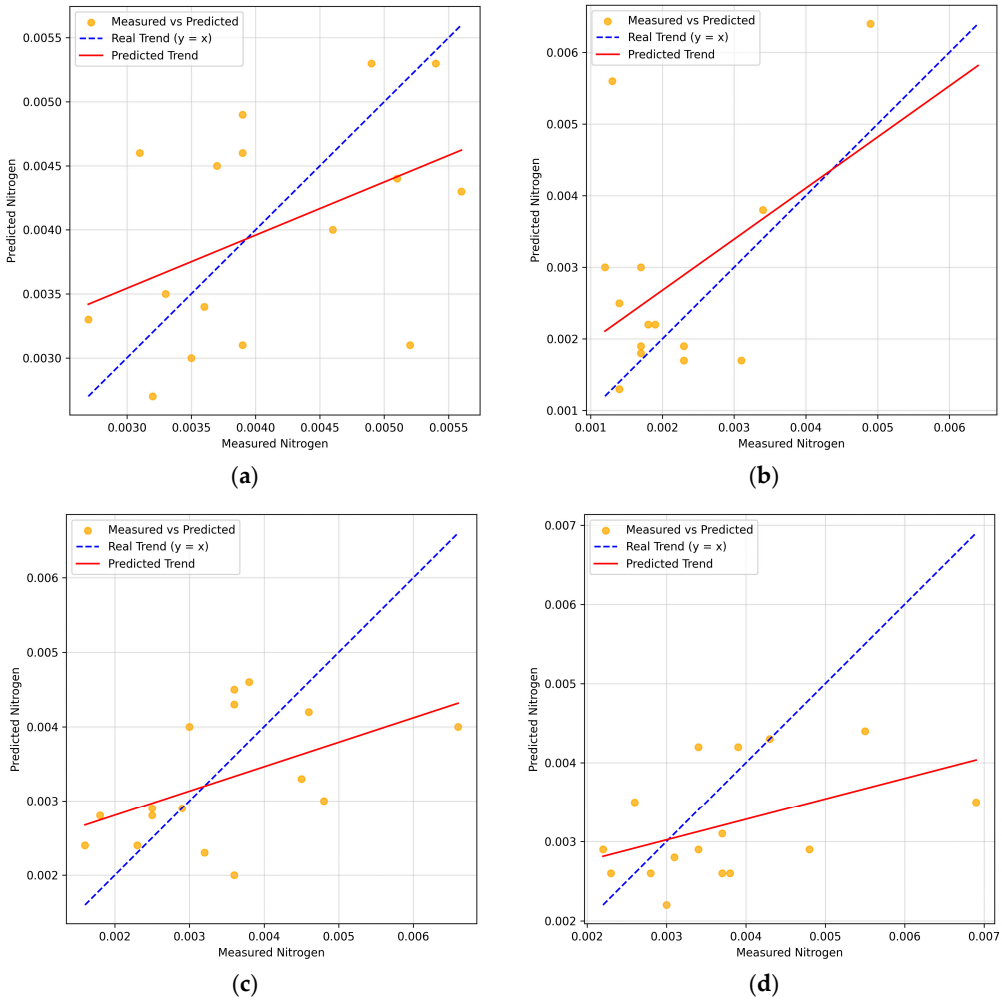

Figure 8. Measured versus predicted nitrogen content in molten metal for the random forest regressor model evaluated on the test dataset: (**a**) PHASE #1: after desulfurization of the pig iron; (**b**) PHASE #2: crude steel in BOF prior to tapping; (**c**) PHASE #3: steel at the beginning of secondary steelmaking; (**d**) PHASE #4: steel at the end of secondary steelmaking.

Table 9. Evaluation of SVR prediction accuracy for individual phases of the BOF steelmaking route (test set, 80/20 split, RBF kernel, GridSearchCV cv = 5).

Metrics	Dataset	Production Phases			
		PHASE #1	**PHASE #2**	**PHASE #3**	**PHASE #4**
MAE	Test	6.69×10^{-4}	6.29×10^{-4}	8.02×10^{-4}	6.81×10^{-4}
	Train	3.12×10^{-4}	2.04×10^{-4}	3.90×10^{-3}	2.83×10^{-3}
MSE	Test	6.28×10^{-7}	7.03×10^{-7}	1.27×10^{-6}	1.21×10^{-6}
	Train	2.71×10^{-7}	4.23×10^{-7}	5.15×10^{-7}	4.26×10^{-7}
RMSE (wt.%)	Test	7.93×10^{-4}	8.39×10^{-4}	1.13×10^{-3}	1.10×10^{-3}
	Train	5.21×10^{-4}	6.50×10^{-4}	7.18×10^{-4}	6.53×10^{-4}
MAPE (%)	Test	15.98	27.90	22.90	16.60
	Train	8.34	5.14	12.44	9.17
100 − MAPE (%)	Test	84.02	72.10	77.10	83.40
	Train	91.66	94.86	87.56	90.83
R^2	Test	0.186	0.277	0.157	0.136
	Train	0.518	0.548	0.551	0.604

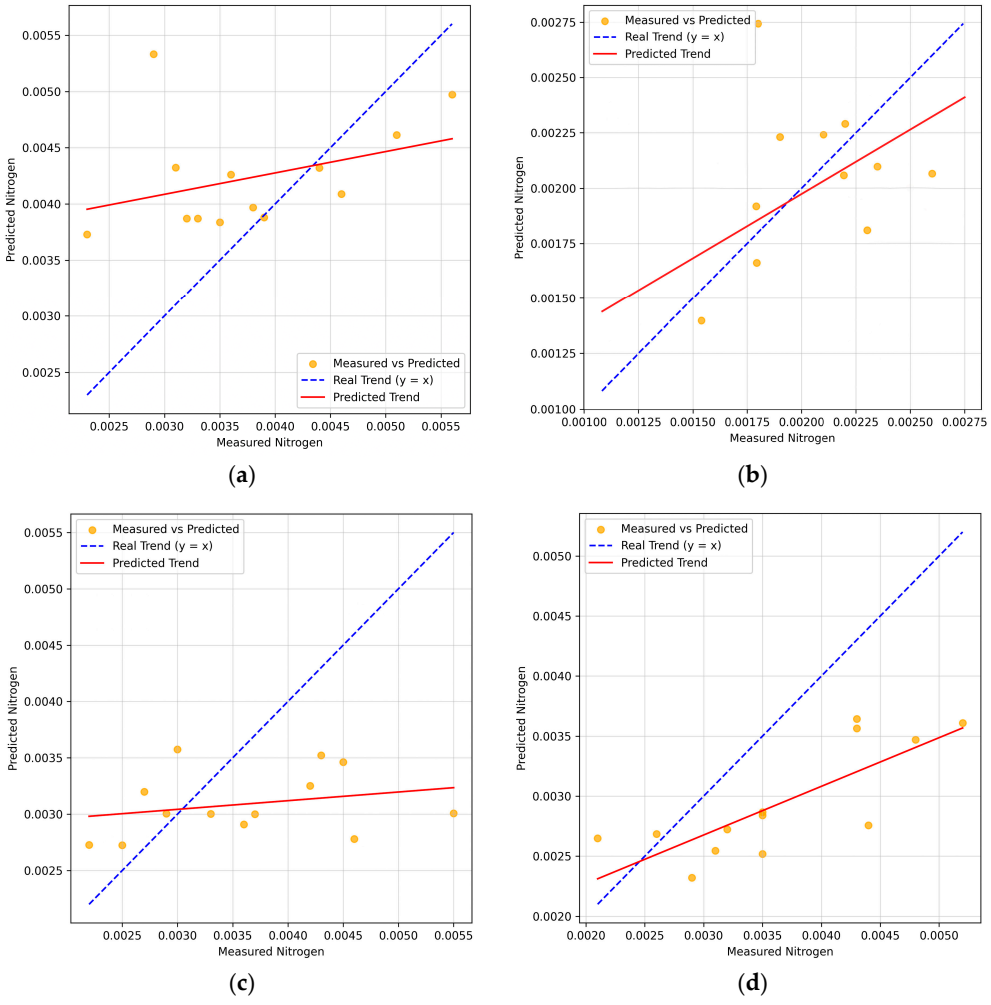

Figure 9. Parity plot of measured versus predicted nitrogen content in molten metal for the feed-forward neural network model (test dataset); the dashed diagonal line (y = x) represents perfect agreement between measured and predicted values: (**a**) PHASE #1: after desulfurization of the pig iron; (**b**) PHASE #2: crude steel in BOF prior to tapping; (**c**) PHASE #3: steel at the beginning of secondary steelmaking; (**d**) PHASE #4: steel at the end of secondary steelmaking.

3.1. Linear Regression Model

The results obtained for nitrogen prediction in molten metal for the PHASE #1−#4 of the BOF steelmaking production process using linear regression are shown in Table 2.

Figure 4a–d presents parity plots of measured versus predicted nitrogen content in molten metal for the linear regression model evaluated on the test dataset across each phase of the BOF steelmaking production process (PHASE #1–#4), illustrating the agreement between measured and predicted values.

3.2. Polynomial Regression Model

Table 3 presents the polynomial regression results for predicting nitrogen content in molten metal across PHASES #1–#4 of the BOF steelmaking process.

Figure 5a–d presents parity plots of measured versus predicted nitrogen content in molten metal for the polynomial regression model (test dataset) across each phase of the BOF steelmaking production process (PHASE #1–#4); the degree of scatter around the diagonal line (y = x) reflects the model's generalization capability on unseen data.

3.3. Ridge Regression Model

Table 4 presents the ridge regression results for predicting nitrogen content in molten metal across PHASES #1–#4 of the BOF steelmaking process.

Figure 6a–d shows the correspondence between measured and predicted nitrogen content in molten metal across each phase of the BOF steelmaking production process, as obtained by the ridge regression model evaluated on the test dataset.

3.4. Decision Tree Regressor Model

Table 5 presents the decision tree regressor results for predicting nitrogen content in molten metal across PHASES #1–#4 of the BOF steelmaking process.

Figure 7a–d depicts measured versus predicted nitrogen content in molten metal for the decision tree regressor model (test dataset) across each phase of the BOF steelmaking production process.

3.5. Random Forest Regressor Model

Table 6 summarizes the performance of the random forest regressor in predicting nitrogen content in molten metal across PHASES #1–#4 of the BOF steelmaking process.

Figure 8a–d shows the correspondence between measured and predicted nitrogen content in molten metal across each phase of the BOF steelmaking production process, as obtained by the random forest regressor model evaluated on the test dataset.

3.6. Feedforward Neural Networks (FNNs) Model

Table 7 presents the evaluation results of the neural network (FNN) model applied to nitrogen-content prediction in molten metal across the four BOF process phases.

Figure 9a–d depicts measured versus predicted nitrogen content in molten metal for the feedforward neural network model (test dataset) across each phase of the BOF steelmaking production process.

3.7. Gaussian Process Regression Model

Table 8 presents the evaluation metrics for the Gaussian Process Regression model on the held-out test set across all four steelmaking phases, and the corresponding parity plots are shown in Figure 10a–d.

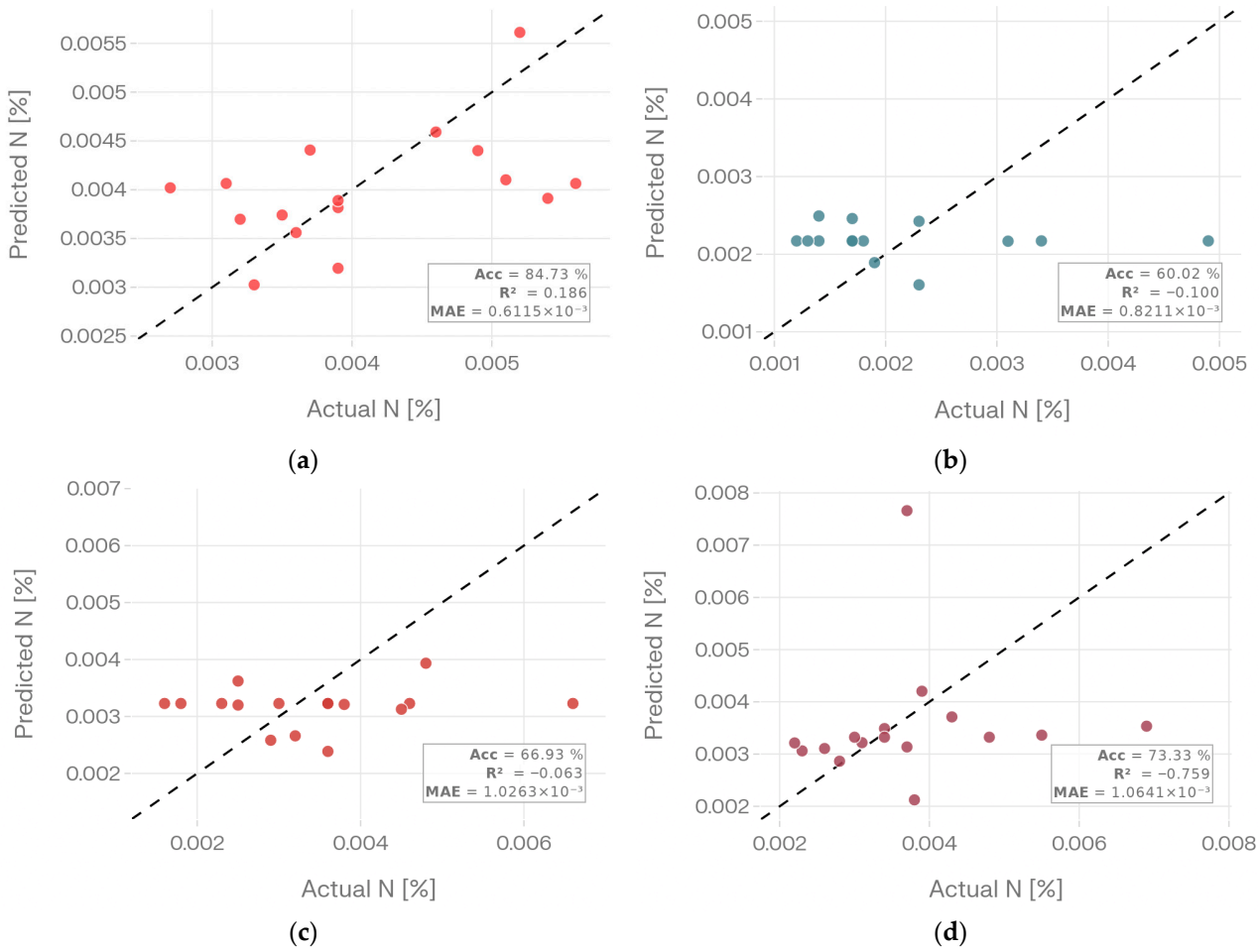

Figure 10. Predicted versus measured nitrogen content obtained using the Gaussian Process Regression model on the test dataset: (**a**) PHASE #1: after desulfurization of the pig iron; (**b**) PHASE #2: crude steel in BOF prior to tapping; (**c**) PHASE #3: steel at the beginning of secondary steelmaking; (**d**) PHASE #4: steel at the end of secondary steelmaking.

The GPR model exhibited characteristic Bayesian interpolation on the training subset, achieving near-perfect fits across all four phases ($R^2 = 1.000$, MAPE $\approx 0\%$). This behaviour is theoretically expected, as the Matérn GP posterior mean passes exactly through training observations in the absence of noise. On the held-out test data, performance varied markedly across phases. PHASE #1 returned the highest accuracy (84.73%, MAE = 6.12×10^{-4}%, $R^2 = 0.186$), which is the second-best result across all models in that phase and comparable to ridge regression (84.59%). This outcome is consistent with the lower feature-to-sample ratio of the desulfurization dataset (14 features, 77 heats), which allows the Matérn kernel length-scale parameters to be estimated more reliably. PHASE #2 and PHASE #3 yielded accuracies of 60.02% and 66.93% with negative R^2 values of -0.100 and -0.063, whilst PHASE #4 exhibited the weakest result (73.33%, $R^2 = -0.759$). The deterioration in PHASES #2 and #4 is mechanistically consistent with the high dimensionality of these datasets (31 and 34 features against approximately 55 and 64 training samples), which renders the multivariate Matérn kernel ill-conditioned and causes the model to interpolate training noise rather than underlying process structure. The parity plots in Figure 10 confirm that predictions in PHASE #1 cluster closely around the 1:1 line, whilst PHASES #2–#4 show systematic dispersion.

Figure 10a–d shows the comparison between the measured and predicted nitrogen content in molten metal for the Gaussian Process Regression model (test dataset) at each

stage of the BOF steelmaking production process. The dashed line represents the ideal 1:1 correspondence. Kernel: Matérn (ν = 2.5) + WhiteKernel; n_restarts_optimizer = 10; normalize_y = True.

3.8. Support Vector Regression Model

Table 9 presents the SVR model evaluation metrics for all four steelmaking phases. Parity plots for the test-set predictions are provided in Figure 11a–d.

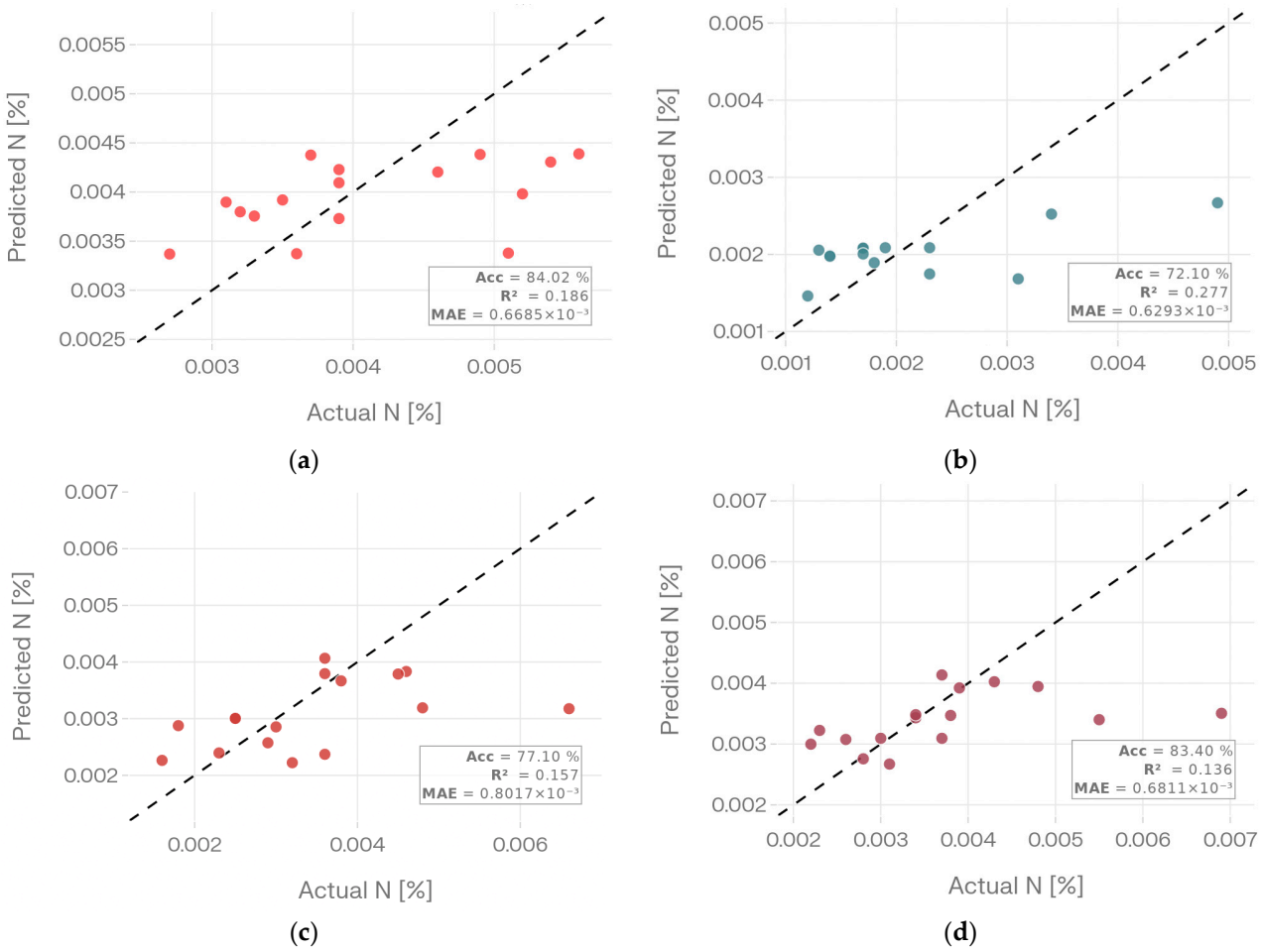

Figure 11. Predicted versus measured nitrogen content obtained using the Support Vector Regression model on the test dataset: (**a**) PHASE #1: after desulfurization of the pig iron; (**b**) PHASE #2: crude steel in BOF prior to tapping; (**c**) PHASE #3: steel at the beginning of secondary steelmaking; (**d**) PHASE #4: steel at the end of secondary steelmaking.

The SVR model with RBF kernel and cross-validated hyperparameters demonstrated a balanced bias–variance profile that contrasts markedly with the GPR behavior. On the training set, SVR achieved moderate R^2 values of 0.518–0.604 across phases, confirming that the margin-based regularization controlled by the optimal C = 1 and ε values deliberately avoids interpolating individual training observations, unlike GPR. On the test set, SVR produced positive R^2 values across all four phases (0.136–0.277), representing the most consistent variance-explaining performance of any model in the study and confirming that cross-validated hyperparameter selection effectively limited overfitting under constrained sample sizes. The highest test accuracy was achieved in PHASE #1 (84.02%, MAE = 6.69 × 10^{-4}%, R^2 = 0.186), closely comparable to GPR in that phase, confirming that PHASE #1 contains a learnable signal accessible to both kernel methods. In PHASE #2, SVR outperformed GPR substantially (72.10% vs. 60.02%, R^2 = 0.277 vs. −0.100), in-

dicating that the ε-insensitive loss and explicit regularisation through C better handle the high-dimensional, noisy BOF dataset. PHASE #3 yielded 77.10% (R^2 = 0.157) and PHASE #4 yielded 83.40% (R^2 = 0.136), with PHASE #4 representing the second-highest absolute accuracy across all SVR phases. The grid-search-identified optimal ε for PHASE #3 (0.1) was notably larger than for the remaining phases (0.0001), reflecting the greater target variability in the secondary metallurgy initiation phase. The parity plots in Figure 11 confirm closer alignment to the 1:1 line in PHASES #1 and #4 and somewhat greater scatter in PHASES #2 and #3. In Figure 11, the dashed line represents the ideal 1:1 correspondence. Kernel: RBF; hyperparameters selected by GridSearchCV (cv = 5); optimal parameters per phase are listed in Table 9.

4. Discussion

The study interprets the comparative results obtained from the eight regression models and the four production phases, considering the underlying metallurgical processes, the characteristics of the feature space, and the inherent constraints of the available industrial dataset. Rather than considering model performance in isolation, the analysis aims to clarify the physicochemical and methodological factors that determine the suitability of each modelling approach at a specific stage of the BOF steelmaking process.

4.1. Performance Metrics Interpretation of PHASE #1

The effectiveness of the individual statistical models used to predict the nitrogen content in molten pig iron after desulfurization (PHASE #1) is summarized in Table 10. This summary is based on the results of the individual statistical methods that were used to make these predictions.

Table 10. Evaluation of the effectiveness of individual statistical models for predicting the nitrogen content in molten pig iron after desulfurization (PHASE #1).

Model	MAE	MSE	RMSE (wt.%)	MAPE (%)	100 − MAPE (%)	R^2
Linear Regression	0.00066	7.033×10^{-7}	0.00084	16.30	83.70	0.6913
Polynomial Regression	0.00123	2.991×10^{-6}	0.00173	31.06	68.94	−0.3128
Ridge Regression	**0.00063**	$\mathbf{6.378 \times 10^{-7}}$	**0.00080**	**15.41**	**84.59**	**0.72**
Decision Tree	0.00086	9.781×10^{-7}	0.00079	21.42	78.58	0.7234
Random Forest	0.00070	7.313×10^{-7}	0.00086	17.48	82.52	0.6790
FNN	0.00071	7.737×10^{-7}	0.00092	21.30	78.70	0.6313
GPR	0.00061	6.28×10^{-7}	7.93×10^{-4}	15.27	84.73	0.186
SVR	0.00067	6.28×10^{-7}	7.93×10^{-4}	15.98	84.02	0.186

The best model for PHASE#1 is highlighted in bold.

Ridge regression achieved the highest accuracy of 84.59% in PHASE #1, closely followed by linear regression (83.70%) and Random Forest (82.52%). The strong performance of regularized and standard linear models indicates that the relationship between the 14 input features and nitrogen content in desulphurised pig iron is predominantly linear in nature. This finding aligns with the thermodynamic understanding of nitrogen dissolution during desulphurization, where the primary mechanisms—nitrogen carrier gas dissolution through the metal–gas interface, sulfur removal freeing active sites for nitrogen adsorption, and temperature-dependent solubility governed by Sievert's law—exhibit approximately linear dependencies within the observed operational ranges.

The superiority of ridge regression over standard linear regression (84.59% versus 83.70%) demonstrates the benefit of L2 regularization when working with a limited sample size of approximately 76 observations and 14 features. The regularization parameter $\alpha = 100$ effectively constrained coefficient magnitudes, preventing overfitting to noise in the training data while preserving the capacity to capture the dominant linear trends.

Polynomial regression performed notably worse (68.94%), despite achieving near-perfect training accuracy (≈100%). The second-degree polynomial transformation expanded the 14 original features into 119 polynomial terms (including interactions and squared terms), creating a severely underdetermined system given the small training set. This classic overfitting scenario—characterized by negligible training error but substantially elevated test error—confirms that the additional polynomial capacity captures noise rather than genuine nonlinear patterns at this production stage.

This interpretation is fully consistent with the RMSE values in Tables 4 and 8, where ridge regression yields the lowest test RMSE of approximately 0.00080 wt.% N, closely followed by linear regression and random forest, while polynomial regression exhibits the highest RMSE despite its almost perfect training fit. In addition, ridge regression achieves the highest prediction accuracy (100 − MAPE) of 84.59% in PHASE #1, indicating that most relative deviations between predicted and measured nitrogen contents remain well below 20%, which is substantially better than for the non-regularized linear and tree-based models. The combination of low RMSE and high accuracy is consistent with a high coefficient of determination R^2, confirming that ridge regression explains the dominant share of variance in the nitrogen content after desulfurization without incurring the overfitting behavior observed for the polynomial model.

The Decision Tree and FNN models yielded comparable accuracies of approximately 78.6–78.7%. For the Decision Tree, the absence of pruning likely resulted in overfitting to training data, while the FNN's moderate performance may reflect the difficulty of optimizing a 256-neuron architecture with only ~61 training samples.

The two kernel-based methods introduced in this revision produced competitive results for PHASE #1. GPR achieved an accuracy of 84.73% with MAE = 6.12×10^{-4}%, slightly surpassing ridge regression (84.59%) and ranking second overall in this phase. SVR attained 84.02%, yielding identical R^2 (0.186) to GPR. The convergence of both kernel methods to equivalent performance levels in PHASE #1 is attributable to the favorable feature-to-sample ratio (14 predictors, 77 heats), which enables reliable kernel estimation and limits the variance amplification typical of limited-data inference.

4.2. Performance Metrics Interpretation of PHASE #2

Table 11 summarizes the predictive performance of the statistical models applied to estimate nitrogen content in crude steel prior to tapping from BOF (PHASE #2). The summary consolidates the outcomes obtained from each individual statistical method employed in the prediction task.

PHASE #2 represents the most challenging prediction environment, as evidenced by the wide performance spread (37.30–79.77%) across models. The FNN achieved a decisively superior accuracy of 79.77% with the lowest MSE (2.250×10^{-7}) and MAE (0.00042) among all models. This result is attributable to the highly nonlinear and complex physicochemical interactions governing nitrogen behavior during BOF steelmaking.

Table 11. Evaluation of the performance of individual statistical modelling approaches used to predict nitrogen content in molten crude steel before tapping during BOF (PHASE #2).

Model	MAE	MSE	RMSE (wt.%)	MAPE (%)	100 − MAPE (%)	R^2
Linear Regression	0.00117	1.708×10^{-6}	0.00131	62.70	37.30	−0.0942
Polynomial Regression	0.00067	8.236×10^{-7}	0.00091	36.91	63.09	0.4722
Ridge Regression	0.00066	6.332×10^{-7}	0.00080	34.41	65.59	0.5942
Decision Tree	0.00102	2.134×10^{-6}	0.00146	62.44	37.56	−0.3673
Random Forest	0.00054	4.178×10^{-7}	0.00065	29.15	70.85	0.7322
FNN	**0.00042**	$\mathbf{2.250 \times 10^{-7}}$	**0.00050**	**20.23**	**79.77**	**0.84**
GPR	0.00082	1.07×10^{-6}	0.00103	39.98	60.02	−0.100
SVR	0.00063	7.03×10^{-7}	0.00084	27.90	72.10	0.277

The best model for PHASE#2 is highlighted in bold.

The BOF process involves simultaneous oxidation of carbon, manganese, silicon, and phosphorus, intense CO evolution creating turbulent metal–gas interfaces, slag–metal equilibria across 31 measured variables, and temperature-dependent reaction kinetics. The failure of linear regression (37.30%) and the Decision Tree (37.56%) demonstrate that these models cannot capture the intricate multi-variable interactions characteristic of this process stage. Specifically, linear regression assumes additive, linear feature contributions, which is fundamentally inadequate for modelling the complex interplay between oxygen blowing parameters, slag composition, and resulting nitrogen partition. The Decision Tree's equally poor performance, despite its capacity for nonlinear partitioning, likely stems from uncontrolled tree depth combined with only ~54 training samples across 31 features, producing extreme overfitting.

Ridge regression (65.59%) provided substantially better results than linear regression, as the L2 penalty with α = 100 effectively handled the high-dimensional feature space (31 features) by constraining unstable coefficient estimates that arise from collinearity among slag and steel composition variables. Nevertheless, the linear model family remains fundamentally limited by its inability to capture interaction effects and nonlinear kinetics.

The superiority of the FNN in PHASE #2 is further supported by the RMSE, which reaches only about 0.00050 wt.% N on the test set, compared with 0.00065 for the random forest and 0.00080 for ridge regression, whereas both linear regression and the decision tree display substantially higher RMSE values above 0.00130 wt.% N and 0.00146 wt.% N, respectively. In parallel, the FNN attains the highest prediction accuracy (100 − MAPE) of 79.77%, clearly outperforming all other models in this phase and indicating that most relative deviations remain within approximately 20% of the measured nitrogen content. Taken together with the low RMSE, this accuracy level is consistent with a high coefficient of determination R^2, implying that the neural network captures most of the variance in the BOF nitrogen data while avoiding the severe overfitting observed for the polynomial regression model. The FNN's architecture (256 → 128 → 32 → 1 with 30% dropout) proved particularly effective for this phase. The network's capacity to learn hierarchical nonlinear feature representations allowed it to capture complex interactions between oxygen activity, slag basicity, tapping temperature, and reblow parameters that govern nitrogen behavior during oxygen steelmaking. The use of dropout regularization and early stopping (triggered at epoch 21–38; well before the 300-epoch maximum) prevented overfitting despite the limited training set. The Random Forest (70.85%) also performed well, as its ensemble of 100 trees effectively captured nonlinear

decision boundaries through bootstrap aggregation, though it underperformed the FNN by approximately 9 percentage points.

In PHASE #2, the performance gap between the two kernel methods was pronounced. GPR attained only 60.02% accuracy with $R^2 = -0.100$, confirming that full Bayesian interpolation with a smooth Matérn kernel is ill-suited to the high-dimensional BOF dataset (31 features, 69 heats). In contrast, SVR with cross-validated regularization achieved 72.10% and $R^2 = 0.277$, representing the highest R^2 of any model in this phase and a practically meaningful improvement over ridge regression (65.59%). The superiority of SVR in this phase confirms that explicit margin-based complexity control via the ε-insensitive tube provides better generalization than kernel interpolation under an unfavorable feature-to-sample ratio.

4.3. Performance Metrics Interpretation of PHASE #3

Table 12 presents a summary of the predictive performance of the statistical models used to estimate nitrogen content in steel at the beginning of secondary steelmaking (PHASE #3). This summary integrates the results generated by each individual statistical method employed in the prediction analysis.

Table 12. Assessment of the performance of individual statistical modelling approaches used to predict nitrogen content in molten steel at the beginning of secondary steelmaking (PHASE #3).

Model	MAE	MSE	RMSE (wt.%)	MAPE (%)	100 − MAPE (%)	R^2
Linear Regression	**0.00076**	**1.268×10^{-6}**	**0.00113**	**20.94**	**79.06**	**0.62**
Polynomial Regression	0.00498	4.295×10^{-5}	0.00655	148.16	−48.16	−11.8738
Ridge Regression	0.00083	1.246×10^{-6}	0.00112	25.22	74.78	0.6264
Decision Tree	0.00096	1.618×10^{-6}	0.00127	25.72	74.28	0.5150
Random Forest	0.00082	1.271×10^{-6}	0.00113	23.34	76.66	0.6189
FNN	0.00077	8.830×10^{-7}	0.00119	20.99	79.01	0.5752
GPR	0.00103	1.61×10^{-6}	0.00127	33.07	66.93	−0.063
SVR	0.00080	1.27×10^{-6}	0.00113	22.90	77.10	0.157

The best model for PHASE#3 is highlighted in bold.

PHASE #3 exhibits a distinctive pattern: linear regression (79.06%) and FNN (79.01%) achieved virtually identical best performance, whilst polynomial regression significantly failed with a negative accuracy of −48.16%. The near-identical performance of the simplest and most complex models strongly suggests that the relationship between the 11 input features and nitrogen content at the beginning of secondary metallurgy is predominantly linear in character.

This linearity is physically plausible. At the beginning of the secondary steelmaking phase, the steel composition has been established during BOF processing, and the initial ladle conditions are relatively well-controlled. The compact feature set (11 variables including steel composition before Ar stirring, tapping time, crude steel weight, slag weight, and tapping angle) represents stable, well-defined process parameters without the complex reaction dynamics present during BOF processing.

The significant failure of polynomial regression (MAPE = 148.16%) constitutes the most extreme overfitting case in the entire study. The second-degree polynomial transformation expanded 11 features into 77 polynomial terms, yet the training set contained only approximately 60 samples. The notebook data confirm that polynomial regression achieved

essentially perfect training accuracy, indicating complete memorization of training noise. When applied to unseen test data, the highly unstable polynomial coefficients produced predictions that diverged dramatically from actual values, yielding predictions with errors exceeding 100%.

From the RMSE perspective, linear regression, ridge regression and random forest achieve very similar test errors around 0.00112–0.00113 wt.% N, whereas the polynomial model reaches by far the largest RMSE of approximately 0.00655 wt.% N, which quantitatively confirms its poor generalization performance in this phase. This pattern is mirrored in the accuracy (100 − MAPE), where linear regression and the FNN attain the highest values of 79.06% and 79.01%, respectively, while ridge regression and tree-based methods remain slightly lower and the polynomial model again performs worst with a negative accuracy of −48.16%. The combination of low RMSE and high accuracy for the linear model and the FNN is consistent with comparatively high coefficients of determination R^2, indicating that in PHASE #3, the underlying relationship between process variables and nitrogen content is predominantly linear and can be captured effectively without resorting to strongly nonlinear or heavily regularized models.

Notably, ridge regression (74.78%) underperformed standard linear regression (79.06%) at this stage. This counterintuitive result suggests that the strong regularization ($\alpha = 100$) excessively constrained the model coefficients for this particular phase, where the genuine linear relationships are sufficiently strong that regularization-induced bias outweighed the variance reduction benefit.

For PHASE #3, SVR yielded 77.10% accuracy (MAE = 8.02×10^{-4}%, $R^2 = 0.157$), closely approaching the best-performing model in this phase (linear regression, 79.06%). The cross-validation-selected $\varepsilon = 0.1$ for PHASE #3 was substantially larger than for the remaining phases, reflecting the wider nitrogen variance at the start of secondary metallurgy and suggesting that the model appropriately applied stronger insensitivity to the increased noise level. GPR underperformed significantly (66.93%, $R^2 = -0.063$), confirming that this low-dimensional but structurally variable phase does not provide sufficient support for reliable Matérn kernel estimation.

4.4. Performance Metrics Interpretation of PHASE #4

Table 13 provides a consolidated summary of the predictive performance of the statistical models used to estimate nitrogen content in steel at the conclusion of secondary steelmaking (PHASE #4). This summary integrates the results produced by each individual statistical method applied in the prediction analysis.

Ridge regression achieved the highest accuracy (84.04%) in PHASE #4, the most feature-rich stage with 34 input variables. This result highlights the critical importance of L2 regularization when the feature-to-sample ratio is unfavorable (34 features versus approximately 58 training samples). The strong regularization ($\alpha = 100$) effectively addressed the collinearity among chemical composition measurements taken at three distinct time points during secondary metallurgy (after Ar stirring, after alloy addition, and at SM completion), as well as the correlations between ferroalloy addition weights and resulting compositional changes.

The dramatic performance gap between linear regression (69.40%) and ridge regression (84.04%)—a difference of nearly 15 percentage points—provides compelling evidence for the presence of severe multicollinearity in the PHASE #4 feature space (Figure 3d). Without regularization, the OLS estimator produced highly unstable coefficients that amplified noise in the test predictions, whereas the ridge penalty stabilized the coefficient vector whilst retaining the linear model's interpretability.

Table 13. Analysis of the performance of individual statistical models employed to predict nitrogen levels in molten steel during the final stage of secondary steelmaking (PHASE #4).

Model	MAE	MSE	RMSE (wt.%)	MAPE (%)	100 − MAPE (%)	R^2
Linear Regression	0.00118	3.918×10^{-6}	0.00198	30.60	69.40	−0.0333
Polynomial Regression	0.00094	2.014×10^{-6}	0.00142	23.33	76.67	0.4689
Ridge Regression	**0.00066**	$\mathbf{1.100 \times 10^{-6}}$	**0.00105**	**15.96**	**84.04**	**0.71**
Decision Tree	0.00093	1.539×10^{-6}	0.00124	23.25	76.75	0.5940
Random Forest	0.00077	1.350×10^{-6}	0.00116	18.70	81.30	0.6439
FNN	0.00076	8.493×10^{-7}	0.00075	19.65	80.35	0.8507
GPR	0.00106	2.46×10^{-6}	0.00157	26.67	73.33	−0.759
SVR	0.00068	1.21×10^{-6}	0.00110	16.60	83.40	0.136

The best model for PHASE#4 is highlighted in bold.

The FNN achieved a strong 80.35% accuracy with the lowest MSE (8.493×10^{-7}) across all models for this phase. The neural network's capacity for automatic feature interaction learning proved valuable for capturing the complex relationships between argon stirring parameters, deoxidation chemistry, and nitrogen behavior during final secondary metallurgy. However, its accuracy fell below ridge regression by nearly 4%, suggesting that the dominant relationships at this stage are sufficiently linear to be captured by a well-regularized linear model, and the FNN's additional nonlinear capacity provided diminishing returns.

In PHASE #4, ridge regression again attains one of the lowest test RMSE values (approximately 0.00105 wt.% N), only slightly exceeded by the neural network (0.00075 wt.% N), while linear regression exhibits the largest RMSE of about 0.00198 wt.% N, highlighting the importance of regularization or non-linear modelling when dealing with a high-dimensional and strongly correlated feature space. This hierarchy is fully consistent with the accuracy (100 − MAPE), where ridge regression reaches the highest value of 84.04%, followed by the FNN with 80.35% and the random forest with 81.30%, whereas linear regression remains at a lower level of 69.40% despite its simplicity. The joint observation of low RMSE and high accuracy for ridge regression implies a comparatively high coefficient of determination R^2, confirming that this regularized linear model captures most of the variance in the final nitrogen content while avoiding the instability and loss of explanatory power that affect the unregularized linear model in this complex, multicollinear feature space.

Random Forest (81.30%) and polynomial regression (76.67%) occupied intermediate positions. Unlike PHASE #3, polynomial regression performed reasonably at this stage because the larger feature set (34 features) provided sufficient information to support some polynomial terms without significant overfitting, though the performance remained below regularized and ensemble methods.

In PHASE #4, SVR delivered an 83.40% accuracy (MAE = 6.81×10^{-4}%, R^2 = 0.136), the second-highest accuracy in that phase after ridge regression (84.04%), and outperformed FNNs (82.41%) despite using only 15 training features selected by the optimal SVR kernel. GPR degraded markedly in PHASE #4 (73.33%, $R^2 = -0.759$), further reinforcing the conclusion that the Bayesian interpolation mechanism becomes unreliable when the feature space is large (34 predictors) relative to the training set size (64 observations after the 80/20 split).

4.5. Cross-Phase Comparative Analysis

The cross-phase comparison confirms that nitrogen prediction performance is strongly phase-dependent, reflecting the interaction between process complexity, feature dimensionality, and the limited number of industrial observations available per stage. Across all four phases, ridge regression, random forest, and feedforward neural networks (FNNs) consistently provide high accuracies on the test set, with ridge regression reaching 84.59% and 84.04% in PHASE #1 and PHASE #4, respectively, and FNNs achieving 78.27% in PHASE #2 and 82.41% in PHASE #4. Linear regression remains surprisingly competitive in PHASE #3 (79.06%), indicating that the nitrogen evolution at the beginning of secondary metallurgy can be approximated adequately by a predominantly linear relationship under the present data regime. Polynomial regression, by contrast, exhibits pronounced overfitting behavior, with near-zero training errors but severely degraded generalization, especially in PHASE #3, where the accuracy drops to −48.16%.

The inclusion of GPR and SVR extends this benchmark to eight models and allows a more nuanced assessment of nonlinear behavior under limited data. GPR, implemented with a Matérn (ν = 2.5) covariance kernel and WhiteKernel noise component, achieved its strongest performance in PHASE #1 (84.73%, R^2 = 0.186), closely matching ridge regression and confirming that the desulfurization dataset contains a smooth nonlinear component that can be effectively captured by a Bayesian kernel method when the feature-to-sample ratio remains favorable (14 predictors, 77 heats). However, GPR performance deteriorated in PHASES #2–#4 (60.02–73.33%, R^2 between −0.100 and −0.759), where the number of predictors (31–34) approaches or exceeds the effective number of training samples. This pattern is consistent with the curse of dimensionality: the high-dimensional Matérn kernel becomes difficult to estimate reliably, leading to near-perfect interpolation of the training data but poor test-set generalization, as reflected by the negative R^2 values.

SVR with an RBF kernel and five-fold cross-validated hyperparameters (C, ε, γ) produced a distinctly different cross-phase profile. Its test-set accuracies reached 84.02% in PHASE #1, 72.10% in PHASE #2, 77.10% in PHASE #3, and 83.40% in PHASE #4, with positive R^2 values across all four phases (0.136–0.277). In PHASE #2, SVR delivered the highest R^2 (0.277) among all models while maintaining competitive accuracy, surpassing both ridge regression (65.59%) and random forest (70.85%) and only trailing FNNs in absolute accuracy. In PHASE #3, SVR nearly matched linear regression in accuracy (77.10% vs. 79.06%) whilst delivering a positive R^2 (0.157), which suggests that this phase exhibits a modest nonlinear component that benefits from kernel regularization but does not require the full flexibility of a neural network. In PHASE #4, SVR offered an accuracy of 83.40%, slightly below ridge regression but higher than random forest, again with a positive R^2 (0.136). These results highlight SVR as the most robust of the nonlinear methods when judged simultaneously by accuracy and variance explanation under constrained sample sizes.

The cross-phase evidence shows that no single model is universally optimal across all stages of the BOF route. For PHASE #1 and PHASE #4, ridge regression remains the primary recommendation, with GPR and SVR providing competitive kernel-based alternatives in PHASE #1 and PHASE #4, respectively. For PHASE #2, FNNs continue to deliver the highest accuracy, but SVR emerges as a particularly attractive choice when emphasis is placed on R^2 and model stability under high dimensionality. For PHASE #3, linear regression retains its leading position, while SVR offers a regularized nonlinear counterpart that improves variance explanation without compromising generalization. The comparative bar chart in Figure 12 summarizes the test-set accuracy of all eight models across the four phases, visually reinforcing that model selection should be

tailored to the phase-specific balance between process nonlinearity, feature dimensionality, and limited data.

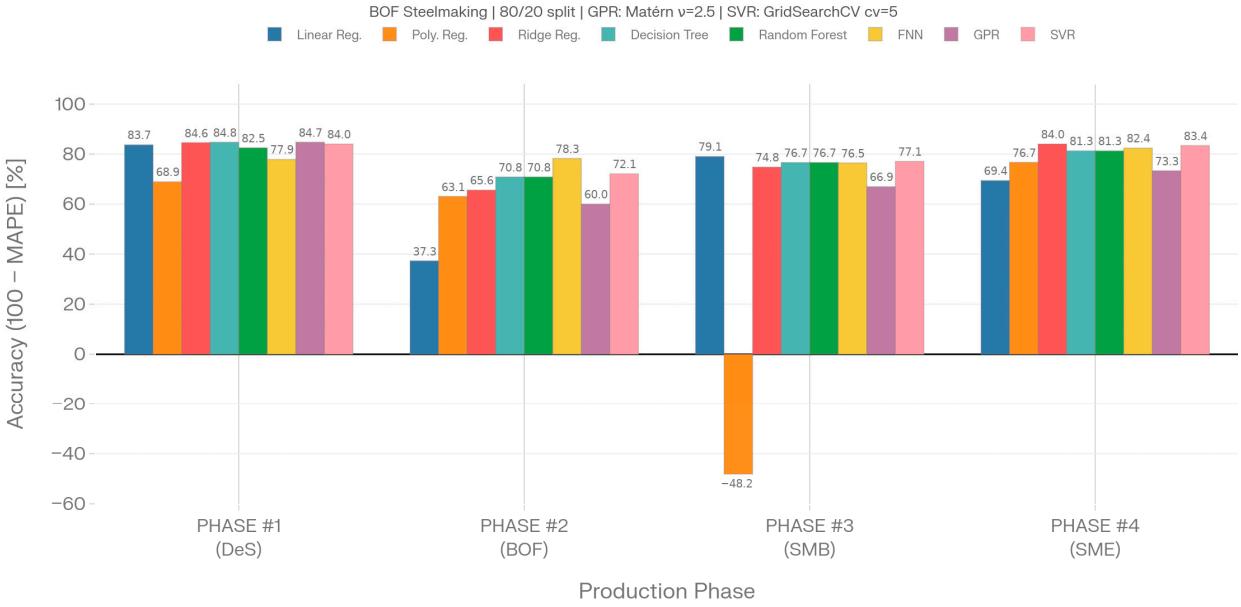

Figure 12. The summarization of prediction accuracy for all statistical models within PHASE #1–PHASE#4.

The prediction accuracy summarization for all statistical models across the analyzed phases is shown in Figure 12.

4.6. Interpretation of the Coefficient of Determination (R^2)

Unlike absolute error metrics (MAE, MSE) or relative error metrics (MAPE), R^2 offers a dimensionless measure of goodness of fit that facilitates direct comparison across models and production phases with different nitrogen concentration ranges.

Across the four production phases, R^2 values for the best-performing models ranged from 0.62 to 0.84 (Tables 10–13), indicating that the models capture a substantial proportion of nitrogen variability. The FNN achieved the highest R^2 of 0.84 in PHASE #2 and 0.85 in PHASE #4, confirming its ability to represent complex, highly nonlinear interactions in high-dimensional feature spaces associated with BOF steelmaking and secondary metallurgy. Ridge Regression showed strong and stable performance in PHASE #1 (R^2 = 0.72) and PHASE #4 (R^2 = 0.71), demonstrating its suitability for predominantly linear relationships under multicollinearity (Figure 3d). Linear Regression reached R^2 = 0.62 in PHASE 3, which, together with the FNN result (R^2 = 0.58), indicates that additional model complexity yields only marginal gains when nitrogen evolution is governed mainly by linear effects in a compact feature space. By contrast, Polynomial Regression displayed extreme overfitting, with highly negative R^2 values (−11.87 in PHASE #3), despite training fit, due to a dramatic expansion of the feature space relative to the limited sample size. Decision Tree and Random Forest models provided mixed results, with acceptable R^2 in some phases but clear signs of overfitting or limited generalization in others. Overall, the concordance between R^2 and error-based metrics (MAE, MSE, MAPE) supports a phase-dependent hybrid strategy: Ridge Regression in PHASES 1 and 4, FNN in PHASE #2, and Linear Regression or FNN in PHASE #3.

The extended benchmark including GPR and SVR further clarifies the limitations of R^2 under constrained industrial data. GPR with a Matérn (ν = 2.5) kernel attains R^2 = 1.000 on the training set for all phases but yields strongly negative R^2 values in PHASES #2–#4, with

the most pronounced case in PHASE #4 ($R^2 = -0.759$). Despite this, the corresponding MAE (1.06×10^{-3} wt.% N) and RMSE (1.57×10^{-3} wt.% N) remain of the same order as those of several other models and are comparable to the analytical resolution of the nitrogen measurement. This combination—near-perfect training fit, test errors in an acceptable absolute range, but heavily negative R^2—indicates that GPR in PHASE #4 is overfitting the limited and high-dimensional training data (34 predictors, ≈64 training samples) and is extremely sensitive to the particular composition of the small test partition (16 samples). In practical terms, the large negative R^2 signals that the GPR predictions fluctuate more around the test-set mean than the mean itself, even though the absolute deviations remain small; as a result, GPR cannot be recommended for PHASE #4, where ridge regression, FNNs, and SVR provide comparable or better MAE with non-negative or only mildly positive R^2.

By contrast, SVR with an RBF kernel and cross-validated hyperparameters achieves modest but consistently positive R^2 across all four phases (0.136–0.277), together with competitive MAE and MAPE values, thus offering a more stable balance between goodness-of-fit and generalization under the present data limitations. Overall, the concordance between R^2 and error-based metrics (MAE, MSE, RMSE, MAPE) for the more robust models supports a refined phase-dependent hybrid strategy: Ridge Regression in PHASES #1 and #4, FNN in PHASE #2, Linear Regression or SVR in PHASE #3, and SVR as a regularized nonlinear alternative wherever a positive and stable R^2 is required in addition to low absolute error.

4.7. Best Model Recommendations by Phase

Based on extensive previous analyses, the results and conclusions can be summarized in the form of recommendations for the individual phases of steel production (PHASE #1–PHASE #4). These recommendations are presented in Table 14.

Table 14. A summary of the recommendations for the most suitable statistical model for predicting nitrogen content in molten metal within specific phases of the steelmaking production process.

Phase	Best Model	R^2	100 − MAPE (%)	Reason
PHASE #1	Ridge Regression	0.72	84.59	Predominantly linear relationships: L2 regularization stabilizes 14-feature model
PHASE #2	FNN	0.84	79.77	Highly nonlinear 31-feature interactions require deep learning capacity
PHASE #3	Linear Regression	0.62	79.06	Compact 11-feature linear system; regularization unnecessary
PHASE #4	Ridge Regression	0.71	84.04	High dimensionality (34 features) with multicollinearity requires L2 penalty

4.8. Factors Influencing Model Performance

The relationship between feature count and optimal model type follows a clear pattern across the four phases. Low-dimensional phases (PHASE #1 with 14 features and PHASE #3 with 11 features) favor simpler linear models because the limited feature space is insufficient to support complex nonlinear models without overfitting. High-dimensional phases (PHASE #2 with 31 features and PHASE #4 with 34 features) benefit from regularization or ensemble methods that can manage collinearity and extract relevant signals from redundant features.

The fundamental limitation of all models is their small sample size (291 samples in total, distributed across four phases and yielding only 54–76 samples per phase after the test split). This constraint manifests in several ways: major overfitting in polynomial regression when feature space is expanded beyond the capacity of the training set, vulnerability to noise in the Decision Tree without pruning constraints, and reliance on aggressive regularization (dropout, early stopping and weight decay) in the FNN to prevent memorization.

The variation in model performance across phases reflects the fundamental differences in nitrogen thermodynamics and kinetics at each production stage. During desulphurization (PHASE #1), nitrogen dissolution follows well-characterized kinetic models governed by Sievert's law, with sulfur removal as the primary driver increasing nitrogen pickup through freed interfacial sites. BOF processing (PHASE #2) involves the simultaneous occurrence of carbon oxidation, intense CO evolution, slag–metal reactions and temperature fluctuations, creating a highly nonlinear system that requires complex modelling approaches. The secondary metallurgy stages (PHASE #3 and #4) represent progressively more controlled environments with established steel compositions, in which nitrogen behavior becomes more predictable and regularized linear models are favored.

5. Conclusions

A comparative evaluation of eight machine learning models across four steelmaking phases reveals that the optimal model selection is phase-dependent and is driven by the interplay between feature dimensionality, process complexity and sample size constraints. The single train–test split employed in this study ensures reproducibility and fair cross-model comparison. Ridge regression is the most effective technique overall, achieving the highest accuracy in two phases (PHASE #1: 84.59%; PHASE #4: 84.04%) and delivering competitive results elsewhere. The FNN provides the most consistent performance across phases and is the only model capable of adequately predicting nitrogen levels during the highly non-linear BOF stage (PHASE #2: 79.77%). To deploy these models in an industrial process control system, we recommend a hybrid strategy: use ridge regression for desulphurization and secondary metallurgy completion, an FNN for the BOF stage and either linear regression (PHASE #3: 79.06%) or an FNN for secondary metallurgy initiation. Polynomial regression without regularization should be avoided entirely for small industrial datasets. Future improvements should focus on expanding the dataset size, implementing k-fold cross-validation strategies with stratification by steel grade to provide robust generalizability assessment with confidence intervals. Future optimization should implement phase-specific hyperparameter tuning to establish performance upper bounds of each phase.

Author Contributions: Conceptualization, J.D.; methodology, J.D.; validation, J.D. and B.B.; formal analysis, J.D.; investigation, J.D.; resources, J.D. and B.B.; data curation, J.D. and B.B.; writing—original draft preparation, J.D.; writing—review and editing, B.B., and M.H.; visualization, J.D.; supervision, B.B. and M.H.; project administration, J.D., B.B. and M.H.; funding acquisition, J.D. All authors have read and agreed to the published version of the manuscript.

Funding: Funded by the EU NextGenerationEU through the Recovery and Resilience Plan for Slovakia under the project No. 09I03-03-V04-00047.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Restrictions apply to the availability of these data. The data were obtained from U. S. Steel Košice, s.r.o., Slovak Republic and are available from the authors with the permission of U. S. Steel Košice, s.r.o., Slovak Republic.

Acknowledgments: The authors sincerely acknowledge the anonymous reviewers for their insights and comments, which further improved the quality of the manuscript.

Conflicts of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as potential conflicts of interest.

Abbreviations

The following abbreviations are used in this article:

MAE	Mean Absolute Error
MSE	Mean Squared Error
MAPE	Mean Absolute Percentage Error
100 − MAPE	Accuracy
R^2	Coefficient of Determination
GPR	Gaussian Process Regression
SVR	Support Vector Regression
G°	Gibbs free energy [J·mol^{-1}]
K_N	Equilibrium constant of reaction [−]
a_N	Activity of elemental nitrogen dissolved in metal [−]
f_N	Activity coefficient of elemental nitrogen dissolved in metal [−]
p_{N_2}	Partial pressure of gas in molecular form in a gaseous atmosphere above molten metal [Pa]
[%N]	Equilibrium concentration of elemental nitrogen dissolved in metal [wt.%]
C	Constant that depends on the given gas $C_{N^{3-}} = 6.01 \cdot 10^{-13}$ [J·K^{-1}]
ΔH	Heat effect of dissolving 1 mole of gas [J·mol^{-1}]
k	Boltzmann constant k = $(1.380658 \pm 0.000012) \cdot 10^{-23}$ [J·K^{-1}]
T	Absolute temperature [K]
R	Universal (molar) gas constant; R = 8.314 [J·mol^{-1}·K^{-1}]
β_0	Constant/Intercept
β_1–β_n	Coefficient of variable
x_1–x_n	Value of an independent variable
n	Total number of input variables
β_i	Linear coefficient for the i-th variable
β_{ii}	Quadratic coefficient for the i-th variable
β_{ij}	Interaction coefficient for the pair of variables i and j
x_i	i-th input variable
x_i^2	Square of the i-th variable
$x_i x_j$	Product of two different variables
i, j	Indexes
α	Hyperparameter
t	Threshold value
σ	Activation function
x	Entry into the neuron
K	Kernel matrix of training observations
k_*	Covariance vector between the test point and the training set
σ_n^2	Noise variance
C	Trade-off between model flatness and tolerance
ξ_i, ξ_i^*	Slack variables

References

1. Demeter, J.; Buľko, B.; Demeter, P.; Hrubovčáková, M. Prediction Models for Nitrogen Content in Metal at Various Stages of the Basic Oxygen Furnace Steelmaking Process. *Appl. Sci.* **2025**, *15*, 9561. [CrossRef]
2. Patra, S.; Nayak, J.; Singhal, L.; Pal, S. Prediction of Nitrogen Content of Steel Melt during Stainless Steel Making Using AOD Converter. *Steel Res. Int.* **2017**, *88*, 1600271. [CrossRef]
3. Liu, C.; Tang, L.; Liu, J. A Stacked Autoencoder With Sparse Bayesian Regression for End-Point Prediction Problems in Steelmaking Process. *IEEE Trans. Autom. Sci. Eng.* **2020**, *17*, 550–561. [CrossRef]
4. Zhang, R.; Yang, J. State of the Art in Applications of Machine Learning in Steelmaking Process Modeling. *Int. J. Miner. Metall. Mater.* **2023**, *30*, 2055–2075. [CrossRef]
5. Nam, J.; Rout, B.; Chatterjee, S.; Van Ende, M.-A.; Jung, I.-H. Integrated Mathematical Model for Nitrogen Control in Oxygen Steelmaking Process. *Metall. Mater. Trans. B* **2025**, *56*, 2202–2223. [CrossRef]
6. Yoon, C.; Eom, C.; Jeon, Y.; Kim, K. Development of a Nitrogen Prediction Model for 320 Tonne Converter. In *12th International Conference of Molten Slags, Fluxes and Salts (MOLTEN 2024) Proceedings*; Springer: Berlin/Heidelberg, Germany, 2024. [CrossRef]
7. Zhao, F.; Liu, X.; Zhang, Z.; Xie, J. Effect of Nitrogen Content on the Mechanical Properties and Deformation Behaviors of Ferritic-Pearlitic Steels. *Mater. Sci. Eng. A* **2022**, *855*, 143918. [CrossRef]
8. Hänninen, H.; Romu, J.; Ilola, R.; Tervo, J.; Laitinen, A. Effects of Processing and Manufacturing of High Nitrogen-Containing Stainless Steels on Their Mechanical, Corrosion and Wear Properties. *J. Mater. Process. Technol.* **2001**, *117*, 424–430. [CrossRef]
9. Bazaleeva, K.O. Mechanisms of the Influence of Nitrogen on the Structure and Properties of Steels (A Review). *Met. Sci. Heat. Treat.* **2005**, *47*, 455–461. [CrossRef]
10. Liu, Z.; Fan, C.; Yang, C.; Ming, Z.; Lin, S.; Wang, L. Dissimilar Welding of High Nitrogen Stainless Steel and Low Alloy High Strength Steel under Different Shielding Gas Composition: Process, Microstructure and Mechanical Properties. *Def. Technol.* **2023**, *27*, 138–153. [CrossRef]
11. Woo, I.; Kikuchi, Y. Weldability of High Nitrogen Stainless Steel. *ISIJ Int.* **2002**, *42*, 1334–1343. [CrossRef]
12. Saxena, A.; Sengupta, A.; Chaudhuri, S.K. Effect of Absorbed Nitrogen on the Microstructure and Core Loss Property of Non-Oriented Electrical Steel. *ISIJ Int.* **2005**, *45*, 299–301. [CrossRef]
13. Duraipandi, R.; Nani Babu, M.; Moitra, A. Fatigue Crack Growth Behavior of Nitrogen-Alloyed Low-Carbon Austenitic Stainless Steel at Room Temperature. *JOM* **2023**, *75*, 478–487. [CrossRef]
14. Gu, J.; Li, J.; Chen, Y. Microstructure and Strengthening-Toughening Mechanism of Nitrogen-Alloyed 4Cr5Mo2V Hot-Working Die Steel. *Metals* **2017**, *7*, 310. [CrossRef]
15. Misra, S.; Fruehan, R.J. *Hydrogen and Nitrogen Control in Ladle and Casting Operations*; Carnegie Mellon University Pittsburgh: Pittsburgh, PA, USA, 2005; p. 62.
16. Turkdogan, E.T. *Fundamentals of Steelmaking*; Institute of Materials: London, UK, 2010; ISBN 978-1-907625-73-2.
17. Pitkälä, J.; Xia, J.; Jokilaakso, A. CFD Modeling of Nitrogen Dissolution into a Steel Bath During Gas Purging. In Proceedings of the Second International Conference on CFD in the Minerals and Process Industries, CSIRO, Melbourne, Australia, 6–8 December 1999; pp. 35–40.
18. Inomoto, T.; Kitamura, S.; Yano, M. Kinetic Study of the Nitrogen Removal Rate from Molten Steel (Normal Steel and 17 mass%Cr Steel) under CO Boiling or Argon Gas Injection. *ISIJ Int.* **2015**, *55*, 1822–1827. [CrossRef]
19. Slater, C.; Spooner, S.; Davis, C.; Sridhar, S. Observation of the Reversible Stabilisation of Liquid Phase Iron during Nitriding. *Mater. Lett.* **2016**, *173*, 98–101. [CrossRef]
20. Trotter, D.; Varcoe, D.; Reeves, R.; Hornby, S. Use of HBI and DRI for Nitrogen Control in Steel Products. *SEAISI Q.* **2002**, *31*, 39–50.
21. Seetharaman, S. *Fundamentals of Metallurgy*; CRC Press: Boca Raton, FL, USA, 2005; ISBN 978-1-85573-927-7.
22. Gupta, C.K. *Chemical Metallurgy: Principles and Practice*; John Wiley & Sons: Hoboken, NJ, USA, 2006; ISBN 978-3-527-60525-5.
23. Rosenqvist, T. *Principles of Extractive Metallurgy*; Tapir Academic Press: Trondheim, Norway, 2004; ISBN 978-82-519-1922-7.
24. Gavriljuk, V.G.; Berns, H. *High Nitrogen Steels: Structure, Properties, Manufacture, Applications*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 1999; ISBN 978-3-540-66411-6.
25. Derin, B.; Alan, E.; Suzuki, M.; Tanaka, T. Phosphate, Phosphide, Nitride and Carbide Capacity Predictions of Molten Melts by Using an Artificial Neural Network Approach. *ISIJ Int.* **2016**, *56*, 183–188. [CrossRef]
26. Liapina, T. Phase Transformations in Interstitial Fe-N Alloys. Ph.D. thesis, Max-Planck-Institut für Intelligente Systeme, Stuttgart, Germany, 2005.
27. Pitkälä, J.; Holappa, L.; Jokilaakso, A. A Study of the Effect of Alloying Elements and Temperature on Nitrogen Solubility in Industrial Stainless Steelmaking. *Metall. Mater. Trans. B* **2022**, *53*, 2364–2376. [CrossRef]
28. Zhang, Y.; Liu, J.; Yang, H.; Li, W.; Zhang, S.; Zhou, W. Kinetics of Nitrogen Absorption/Vacuum Denitrigenization and Precipitation Behavior of Nitrogen Bubbles in 42CrMoA Molten Steel. *Steel Res. Int.* **2025**, *96*, 2400071. [CrossRef]

29. Singh, M.K.; Keshari, K.K.; Prasad, A.; Das, A. Effect of Steel Making Parameters on Nitrogen Level in Steel. In *Processing and Characterization of Materials: Select Proceedings of CPCM 2020*; Pal, S., Roy, D., Sinha, S.K., Eds.; Springer: Singapore, 2021; pp. 63–73; ISBN 978-981-16-3937-1.
30. Yu, S.; Miettinen, J.; Louhenkilpi, S. Modeling Study of Nitrogen Removal from the Vacuum Tank Degasser. *Steel Res. Int.* **2014**, *85*, 1393–1402. [CrossRef]
31. Biswas, J.; Ghosh, S.; Ballal, N.B.; Basu, S. A Dynamic Mixed-Control Model for BOF Metal–Slag–Gas Reactions. *Metall. Mater. Trans. B* **2021**, *52*, 1309–1321. [CrossRef]
32. Li, H.; Xie, C.; Yu, F.; Zhang, J.; Zhang, Z. Study on the Technology of Controlling Nitrogen Content in Molten Steel. *J. Phys. Conf. Ser.* **2023**, *2539*, 012103. [CrossRef]
33. Klimova, M.; Shaysultanov, D.; Semenyuk, A.; Zherebtsov, S.; Salishchev, G.; Stepanov, N. Effect of Nitrogen on Mechanical Properties of CoCrFeMnNi High Entropy Alloy at Room and Cryogenic Temperatures. *J. Alloys Compd.* **2020**, *849*, 156633. [CrossRef]
34. Kuang, J.; Long, Z. Prediction Model for Corrosion Rate of Low-Alloy Steels under Atmospheric Conditions Using Machine Learning Algorithms. *Int. J. Miner. Metall. Mater.* **2024**, *31*, 337–350. [CrossRef]
35. Demeter, J.; Buľko, B.; Demeter, P.; Hrubovčáková, M.; Hubatka, S.; Fogaraš, L. Automated Machine Learning for Nitrogen Content Prediction in Steel Production: A Comprehensive Multi-Stage Process Analysis. *Appl. Sci.* **2025**, *16*, 441. [CrossRef]
36. Liu, S.; Li, Q.; Ye, Q.; Zhao, Z.; Dianyu, E.; Kuang, S. Enhanced Semi-Supervised Learning for Top Gas Flow State Classification to Optimize Emission and Production in Blast Ironmaking Furnaces. *Int. J. Miner. Metall. Mater.* **2026**, *33*, 204–216. [CrossRef]
37. ELTRA. GmbH Basic Application Information. Available online: https://www.eltra.com/files/446383/expert-guide-application-information.pdf (accessed on 22 January 2026).
38. ELTRA. GmbH Effective Quality Control of Steel and Iron Products with Combustion Analysis. Available online: https://www.eltra.com/files/14146/effective-quality-control-of-steel-and-iron-products-with-combustion-analysis.pdf (accessed on 22 January 2026).
39. *ASTM E1019-18*; Standard Test Methods for Determination of Carbon, Sulfur, Nitrogen, and Oxygen in Steel, Iron, Nickel, and Cobalt Alloys by Various Combustion and Inert Gas Fusion Techniques. ASTM: West Conshohocken, PA, USA. Available online: https://store.astm.org/e1019-18.html (accessed on 8 December 2025).
40. Osborne, J.W. Regression & Linear Modeling: Best Practices and Modern Methods. Available online: https://www.researchgate.net/publication/299508276_Regression_Linear_Modeling_Best_Practices_and_Modern_Methods (accessed on 14 July 2025).
41. Good, P. *Ordinary Least Squares*; Chapman and Hall/CRC: Boca Raton, FL, USA, 2012; pp. 141–162.
42. Maulud, D.; Abdulazeez, A.M. A Review on Linear Regression Comprehensive in Machine Learning. *J. Appl. Sci. Technol. Trends* **2020**, *1*, 140–147. [CrossRef]
43. Hassi, H.; El Mkhalet, M.; Lamdouar, N. Evaluating Regression Models for Project Forecasting: A Comparative Analysis of Linear and Polynomial Approaches with Monte Carlo Simulations. In Proceedings of the 2025 5th International Conference on Innovative Research in Applied Science, Engineering and Technology (IRASET), Fez, Morocco, 15–16 May 2025; pp. 1–8.
44. Ostertagová, E. Modelling Using Polynomial Regression. *Procedia Eng.* **2012**, *48*, 500–506. [CrossRef]
45. Bera, D.; Chatterjee, N.D.; Bera, S. Comparative Performance of Linear Regression, Polynomial Regression and Generalized Additive Model for Canopy Cover Estimation in the Dry Deciduous Forest of West Bengal. *Remote Sens. Appl. Soc. Environ.* **2021**, *22*, 100502. [CrossRef]
46. Dar, I.S.; Chand, S. Bootstrap-Quantile Ridge Estimator for Linear Regression with Applications. *PLoS ONE* **2024**, *19*, e0302221. [CrossRef] [PubMed]
47. Rokem, A.; Kay, K. Fractional Ridge Regression: A Fast, Interpretable Reparameterization of Ridge Regression. *Gigascience* **2020**, *9*, giaa133. [CrossRef] [PubMed]
48. Jena, M.; Dehuri, S. DecisionTree for Classification and Regression: A State-of-the Art Review. *Informatica* **2020**, *44*, 405–420. [CrossRef]
49. Dumitrescu, E.; Hué, S.; Hurlin, C.; Tokpavi, S. Machine Learning for Credit Scoring: Improving Logistic Regression with Non-Linear Decision-Tree Effects. *Eur. J. Oper. Res.* **2022**, *297*, 1178–1192. [CrossRef]
50. Mienye, I.D.; Jere, N. A Survey of Decision Trees: Concepts, Algorithms, and Applications. *IEEE Access* **2024**, *12*, 86716–86727. [CrossRef]
51. Biau, G.; Scornet, E. A Random Forest Guided Tour. *TEST* **2016**, *25*, 197–227. [CrossRef]
52. de Silva, J.L.S.; de Paula, M.V.; de Barros, J.S.G.; Barros, T.A.D.S. Anomaly Detection Workflow Using Random Forest Regressor in Large-Scale Photovoltaic Power Plants. *IEEE Access* **2025**, *13*, 54168–54176. [CrossRef]
53. Mrabet, Z.E.; Sugunaraj, N.; Ranganathan, P.; Abhyankar, S. Random Forest Regressor-Based Approach for Detecting Fault Location and Duration in Power Systems. *Sensors* **2022**, *22*, 458. [CrossRef]
54. Kriegeskorte, N.; Golan, T. Neural Network Models and Deep Learning. *Curr. Biol.* **2019**, *29*, R231–R236. [CrossRef]

55. Kumaraswamy, B. Neural Networks for Data Classification. In *Artificial Intelligence in Data Mining*; Academic Press: London, UK, 2021; pp. 109–131; ISBN 978-0-12-820601-0.
56. Rasmussen, C.E.; Williams, C.K.I. *Gaussian Processes for Machine Learning*; The MIT Press: Cambridge, MA, USA, 2005; ISBN 978-0-262-25683-4.
57. Smola, A.J.; Schölkopf, B. A Tutorial on Support Vector Regression. *Stat. Comput.* **2004**, *14*, 199–222. [CrossRef]
58. Support Vector Regression. In *Machine Learning*; Academic Press: Cambridge, MA, USA, 2020; pp. 123–140.