steel research international

steel research international

RESEARCH ARTICLE OPEN ACCESS

Decreasing the Environmental Impact of the Electric Steelmaking Route Through Advanced Modelling Techniques

Valentina Colla[1] | Antonella Zaccara[2] | Stefano Dettori[1] | Laura Laid[1] | Ismael Matino[1] | Silvia Cateni[1] | Teresa Annunziata Branca[1] | Lorenzo Vannini[1]

[1]TeCIP Istitute, Scuola Superiore Sant'Anna, Pisa, Italy | [2]Dipartimento di Ingegneria Industriale, Università di Padova, Padua, Italy

Correspondence: Lorenzo Vannini (lorenzo.vannini@santannapisa.it)

Received: 1 October 2025 | **Revised:** 19 February 2026 | **Accepted:** 25 February 2026

Keywords: electric arc furnace | Ladle furnace | machine learning | modelling | scrap | sterile | sustainability

ABSTRACT

Although electric arc furnace (EAF)-based steelworks produce steel from recycled ferrous scrap and inherently implement the concept of circularity, they are challenged to reduce their overall environmental impact, reduce CO_2 emissions, and maximize energy and resource efficiency. The paper exemplary shows how advanced digital technologies, including artificial intelligence-based techniques, can support decarbonization and sustainability improvement of electric steelmaking. The paper presents computationally efficient machine learning models estimating sterile content in different types of scrap reaching the scrap yard as well as steel chemical composition and temperature at the exit of the Ladle furnace. The models are designed to be included in an innovative software platform based on federated learning (FL) helping industrial staff in decision-making by estimating energy consumption and other parameters affecting the environmental impact according to the material mix fed to the EAF. The paper describes the rationale behind models' design, the approach for selecting their hyperparameters, and the results achieved on data gathered from two different steelworks, the first one exploited as reference for models' first setup, the second one used to assess models' usability in the considered FL context. The performances are satisfactory in both cases, and key issues for implementation and further improvement are discussed.

1 | Introduction

The European steel industry is facing the challenges posed by the European Green Deal, which aims at cutting CO_2 emissions by 55% by 2030 and achieving a complete reduction in net emissions of green house gases (GHGs) by 2050. The overall goal is to enhance energy and resource efficiency, while ensuring an inclusive and sustainable growth and development of the European society [1]. In addition, the more recent Green Deal industrial plan targets an acceleration of the industrial transition especially in terms of climate neutrality [2, 3]. To this aim, among other actions, the European Union aims at increasing the circularity of industrial processes and decreasing waste generation, by doubling the use of recycled materials by 2030 to reduce primary raw materials consumptions and related environmental impacts [4, 5]. Electric steelworks implement the concept of circular economy since decades, far before this concept became so popular and widely used. In effects, they recover steel from end-of-life goods and demolition wastes by melting scrap in the electric arc furnace (EAF) by saving approximately 72% of energy compared to steel production from primary raw materials [6]. However, a lot of progress can still be made, e.g., improving use of secondary raw materials, characterization, pretreatment, and sorting of scrap [7, 8] and used mix of primary and secondary raw materials. Furthermore, the entire production process, considering the entire value chain, can be optimized. Digital technologies, including artificial intelligence (AI), can provide a substantial

contribution in this sense [9–11]. For instance, advanced modelling techniques and optimization strategies can be combined to support decision-making for exploitation of energy and material streams in the industrial context [12], with consistent benefits in terms of energy and material efficiency, such as discussed in the exemplary works of Larsson and Dahl [13] and Fernández Martínez et al. [14].

A comprehensive review of physics- and AI-based numerical models and related to the EAF process is provided in [15], while an interesting numerical simulation model of scrap melting in the EAF is discussed in [16]. A further example of efficient model for the optimization of bulk density of shredded and hammer milled scrap balancing productivity and metallurgical loss is provided in [17]. In particular, machine learning (ML) and deep learning (DL)-based modelling approaches are adopted in the context of optimization problems to lower the computing effort compared to complex physics-based models and traditional derivative-based optimization algorithms. A representative example in this sense is provided in [18], where DL and reinforcement learning are combined to optimize maintenance in scrap-based steelmaking.

A potential barrier to the wide adoption of AI-based models is the lack of transparency and interpretability, which can be however mitigated by suitable solutions, such as, for instance, the Shapley additive explanations adopted in [19] for a model estimating EAF energy consumption. Nonetheless, in the context of ML-based process models that are suitable to the integration in through-process optimization frameworks, solutions specifically designed for the steel sector and conceived to be inherently transferable are still missing. Moreover, transferability of AI-based models, even if they have already been validated in another company, is often hampered by the reluctance of companies in sharing sensitive production data with technology providers and/or by the need to implement costly and time-consuming ad hoc data collections.

An exemplar solution to overcome this barrier is elaborated within the European project entitled "*Data and decentralized Artificial intelligence for a competitive and green European metallurgy industry*" (ALCHIMIA). This project aims at optimizing the EAF charge mix and the entire EAF-based steel production process by providing a platform based on federated learning (FL) [20] and continual learning (CL) [21]. This system supports the industrial staff in decision-making procedures by estimating energy consumption and other parameters affecting the environmental impact of the process according to the input material mix.

FL allows a server to train ML models across multiple decentralized clients that privately store their own training data, by saving computational resources while avoiding the clients to outsource their private data to the server [22]. So doing, many clients can train the model in a collaborative way, underneath the central-server orchestration while data are stored in a decentralized way. Therefore, FL mitigates some of the major privacy issues and reduces costs compared to conventional centralized ML approaches [23]. FL trains models that are customized to the demands of each steelworks in a decentralized way, contributing to the creation of an optimized global model without creating a centralized dataset, such as schematically depicted in Figure 1. On the other hand, CL enables continuous improvement of ML model accuracy, ensuring long-term resilience and continuous upgrading of the solution.

The models at the core of the ALCHIMIA platform cover all the different stages of the production process, from the purchasing of input materials to the exit of secondary metallurgy processes. The present work focuses on some of these models, i.e., ensemble models estimating the sterile content in scrap, and feed-forward neural networks (FFNN)-based models estimating steel chemical composition and temperature at the exit of the Ladle furnace (LF). The capability to estimate with sufficient accuracy these variables is fundamental for a complex through-process optimization that, for a given target product, identifies the scrap mix minimizing the environmental impact of the production cycle while preserving steel quality. In particular,

- Soil contamination in scrap is a major source of efficiency loss in the EAF, as it strongly affects the metal yield of the heat. The possibility to exploit also low quality and very contaminated scrap is of utmost importance considering the foreseen increasing trend of scrap demand linked to the decarbonization of

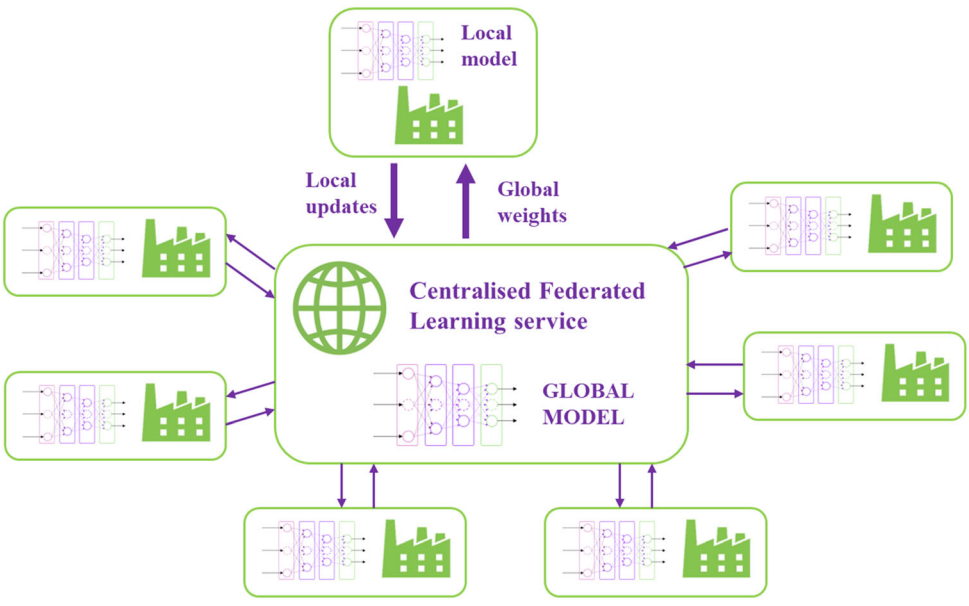

FIGURE 1 | Schematic representation of the federated learning approach.

steel production processes [24], such as attested by the numerous recent research activities on novel solutions also based on DL-based image processing for automatic scrap sorting, classification, rating and purification. [8, 25–29] Therefore, an estimate of soil contaminations in scrap can strongly support scrap yard operators in improving the planning of scrap processing and cleaning operations. This information is also fundamental for EAF simulation and optimization. As soil contamination depends on various factors that are difficult to model in an analytical form, ML represents a viable solution to get a first rough estimate of the level of sterile in a scrap lot based on historical data.

- Steel refining in the LF is a crucial phase of steelmaking, as it optimizes steel chemical composition and strongly affects the physical–mechanical properties of the product. Numerous mathematical models are available to simulate the LF process [30], including AI-based models often focusing on specific aspects, such as prediction of steel temperature, production optimization, slag formation, desulphurization, and alloying [31]. However, the ALCHIMIA platform demands for a holistic and computationally efficient approach to estimate the main aspects of quality and energy-related final outcomes of the LF process (i.e., the steel composition and final steel temperature) make ML the most viable solution for this task.

One of the main features of the developed models is that their input variables are selected among data that are already available as a standard in all electric steelworks, thus they do not need ad hoc measuring campaigns to be tuned. In this work, the models are trained and optimized by using industrial data from two different steelworks, by showing good performances that fit the purpose of the considered application and by demonstrating their full transferability, which is empowered by their design for inclusion in an FL platform.

2 | Materials and Methods

2.1 | Industrial Production Process and Data

The steel production process considered in the ALCHIMIA project is a standard electric steelmaking route as schematically depicted in Figure 2.

For the purpose of the present work, data were provided by two steelworks (in the following named “Steelworks A” and “Steelworks B” for confidentiality reasons) located in two European countries but belonging to the same multinational steelmaking group. They produce carbon steels for a large variety of applications ranging from billets, structural and merchant bars to reinforcing steels, wire rods, electro-melted meses, and lattice grids.

No grade-based selection was made on the data to be used for model design: the data cover several months of production for both steelworks.

In particular, the range of steel grades produced by Steelworks A is wider compared to Steelworks B, which implies a wide range of “recipes” and operating practices for steelmaking. For this reasons, Steelworks A has been taken as a reference in the present work, based on the consideration that having a more continuous distribution of the data across the possible variability range of the potential input variables of the model can be beneficial for the generalization capability of the model, also in the perspective of its transferability to different steelworks. Therefore, models design and optimization, including selection of input parameters, was carried out using only the data provided by Steelworks A, while the data of Steelworks B were used only for retraining and customization of the previously optimized models to assess their actual transferability in the context of an FL approach.

For the sterile contents estimate, the industrial database coming from Steelworks A contain data related to 14 538 scrap truckloads including 55 features, which refer, e.g., to storage location, arrival date, weight (entry, exit and net weight), quality, supplier, transport and other information. On the other hand, the database of Steeworks B refers to 16 369 truckloads for which 16 features are recorded, related to material quality, storage location, arrival date and timing weight as well as information on the supplier.

As far as the LF process is concerned, the available industrial databases provide for each heat a set of measurements and global features concerning timestamps and process duration, chemical analysis of different samples of each heat, materials additions in LF (in terms of ferroalloys, deoxidizers and slag formers), temperature measurement of different samples of each heat, argon volume consumption, electricity consumption for heating purposes, total scrap weight loaded in the EAF, number of personnel

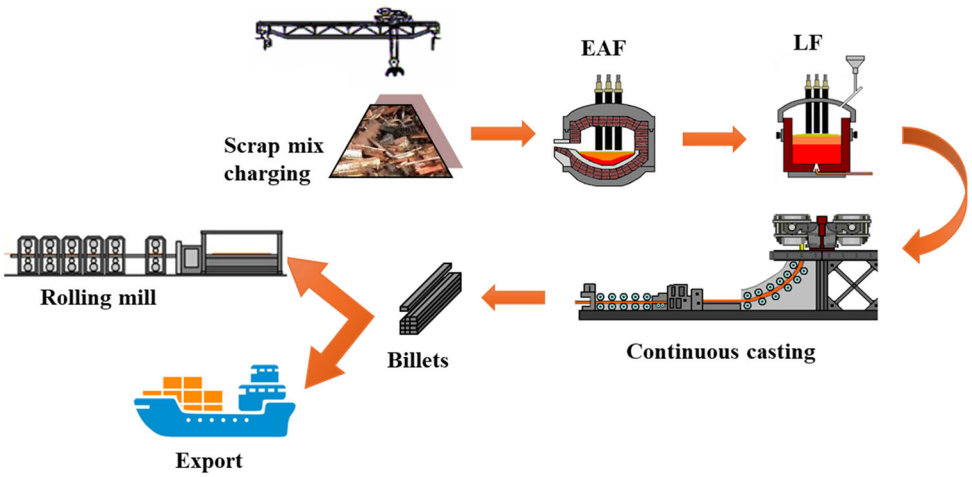

FIGURE 2 | Electric steel production chain considered in the project.

shifts and steel family and quality. In particular, the database provided by Steelworks A refers to more than 3000 heats, while the database provided by Steelworks B refers to about 2700 heats.

2.2 | Ensemble Models

An ensemble method is a technique that combines the predictions from multiple ML algorithms to provide more accurate prediction compared to any individual model, such as schematically depicted in Figure 3a, by de facto implementing a complex model comprised of many (usually simpler), models named *ensemble model*. In general, ensemble models are more robust and accurate than each of the single models, at the price of a higher computational costs, as they imply training of multiple models and management of their interactions.

In ML there are two main types of ensemble learning:

- *Bootstrap aggregation*, or *bagging*, an ensemble meta-learning technique that trains many classifiers on different partitions of the training data and uses a combination of the outputs of the classifiers to form the final output for a given input vector.
- *Boosting*, an approach that weights both steps of training and prediction. During the learning phase, the boosting procedure trains a new model several times, by adjusting the parameters of the new model based on the errors of the so far existing boosted model. During the prediction phase, it provides a prediction based on a weighted combination of the outputs provided by the models.

To estimate the sterile content in scrap, both strategies were tested via the following algorithms:

- *Random forest* (RF) regression, i.e., a supervised learning algorithm belonging to the bagging category using ensemble learning for regression: RF trees are executed in parallel, and the result is aggregated into a global one [32].
- *Gradient-boosting decision trees* implemented via XGBoost (i.e., extreme gradient boosting library), a combination of several simple regression trees, executed sequentially in a more robust model, so that each tree is trained on the error residuals of the previous sequence of trees [33, 34].

Tree-based modelling methodologies are characterized by high explainability and interpretability of the results, as they provide a direct measure of the input feature importance and because of their structure and the explicit feature usage in decision logic (a tree is equivalent to a set of IF-THEN rules).

In the present work, XGBoost and RF-based models have been developed in a Python environment using the Sklearn library. A third version of the RF-based model has also been developed by tuning the RF parameters via the grid-search optimization technique.

2.3 | Feed-Forward Neural Networks

FFNNs are a class of artificial neural networks that are widely adopted in regression and classification tasks due to their ability to model complex, nonlinear numerical relationships by mapping a dataset of inputs-target pairs through a composition of nonlinear surfaces [35]. FFNN are usually structured according to three types of layers connected in series, as shown in Figure 3b.

- the *input layer* receives numerical data streams and forwards it to the following layer after a set of transformation (e.g., normalization).
- A set of *hidden layers*, each consisting of a set of neurons, often identical in each layer, which communicate through a computational graph.
- A *readout layer* that synthetizes the output information by calculating a set of outputs.

In an FFNN, information flows in one direction, from the input to the readout. Each neuron performs a non-linear weighting operation on the inputs through a so-called activation function. The learning process is based on supervised training paradigm, which is performed by optimization algorithm that adjusts the trainable weights of the network to minimize an objective function. Such a learning procedure involves three phases: a *forward propagation* where the network is simulated with a set of input samples, the calculation of a function quantifying the difference between network outputs and target values (also named loss), and the *backpropagation of errors*, in which the neurons weights are updated in the opposite direction of the gradients of the loss function with respect to each weight. FFNN are very versatile but need careful tuning of the hyper-parameters and large datasets to reach good performance and are quite prone to overfitting.

With respect to other ML approaches, FFNN may perform with higher accuracy, with low computational burden if regression task is relatively simple. However, high accuracy is achieved at the cost of a low level of model explainability, as NN results are typically less interpretable.

In the present work, FFNN-based models are developed in the Python environment through the Keras library, with Adam

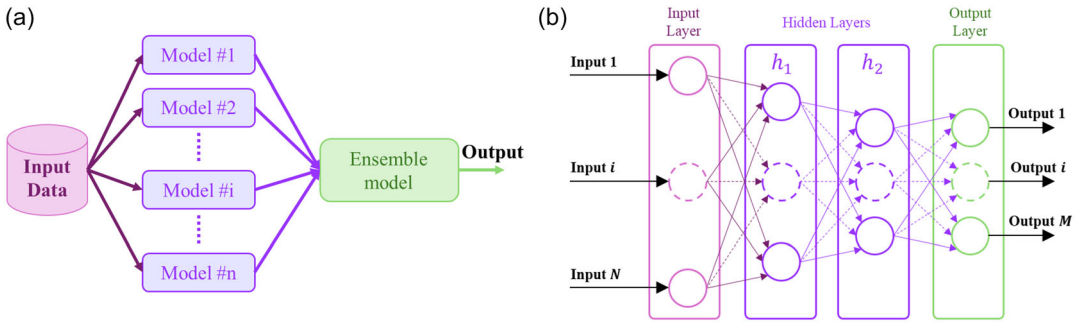

FIGURE 3 | Schematic representation of (a) an ensemble model; (b) a feed-forward neural network.

[36] optimizer implementing the learning process. The FFNNs are implemented with ReLU activation function in the hidden layers and a linear readout.

2.4 | Model Selection, Data Preprocessing and Hyper-Parameter Optimization

The first stage of the work was devoted to selecting the modelling architecture, aiming at achieving high accuracy and low computational burden. The balance between these two often counteracting performance indicators highly depends on the purpose of the model in the final software implementation. Typically, very accurate models may request higher computational burden if the models must describe a wide set of operating conditions (i.e., large data distributions). Additionally, the maximum computational burden of the model must also be constrained depending on the frequency of its inference. Within the ALCHIMIA platform, scrap models are used for monitoring purposes, while LF models are exploited within a model-based optimization system whose purpose is optimizing the LF operations in terms of energy and argon consumption, and added materials. The former requires high accuracy with relaxed constraints on computational burden, due to the very low frequency of inference of the model (i.e., one simulation for each scrap arrival). The latter requires high accuracy and constrained real-time requirements, with the scope of inferring a very efficient model with high frequency. Additionally, also requirements on model explainability drove the modelling selection in function of the modelling objective. In this context, for the scrap models, tree-based approaches are adopted, as RF and XGBoost show a good trade-off between explainability and accuracy, at the cost of higher computational complexity. Conversely, for modelling LF operations, FFNN is selected to achieve high predictive accuracy together with computational efficiency, which is required for real-time implementation. Additionally, linear models trained through ridge regression are used as a valid baseline to compare and assess FFNN results. As FFNNs, linear models can be easily federated and show significantly low computational burden for their inference, as required for real-time optimization approaches.

A mandatory preliminary step in the design of any ML-based model is a careful pre-processing of the dataset that is used to design and train the model itself. Unreliable or anomalous, also named *outliers*, must be removed. Moreover, to suitably set the models' dimension and limit the risk of overfitting, the dimension of the input vector must be kept limited via knowledge-drive or automatic identification of the variables mostly affecting the target. This stage also helps gaining a deeper understanding of the system/phenomenon to reproduce and favours model explainability. Finally, the hyperparameters of the ML-based models need to be optimized.

The datasets provided by both steelworks underwent the cleansing stage to eliminate outliers, while model structural optimization was carried out only using data from Steelworks A, as already mentioned in Section 2.1.

For the estimate of the sterile content in scrap (expressed in ton), based on the knowledge of process operators many redundant variables were eliminated, and the following 5 variables were selected as input:

- *origin*, namely the territorial reference of scrap origin;
- *supplier*, i.e., the provider of the scrap;
- *weight*, i.e., net weight of scrap, which is connected to the weight of soil contaminant;
- *month*, i.e., the cardinal month of arrival of the scrap to the scrapyard, which is considered as indicator of seasonal information;
- *scrap type*, i.e. the category of scrap.

In the raw datasets, *origin* and *suppliers* are not numerical variables, thus they were translated into numbers through a hot encoding methodology. Outliers' detection was carried out by using the

TABLE 1 | Input and target variables of the FFNN-based models of related to the LF process.

	Metallurgical model		Thermal model	
	Variable	**Unit**	**Variable**	**Unit**
Inputs	Steel temperature at the beginning of the LF process	°C	Steel temperature at the beginning of the LF process	°C
	Electric energy consumption at LF	MWh	Electric energy consumption at LF	MWh
	Argon consumption	Nm^3	Argon consumption	Nm^3
	Total scrap weight loaded in the EAF	ton	Total scrap weight loaded in the EAF	Ton
	Steel chemical composition at the beginning of the LF process: [C], [S], [Mn], [Si], [P], [Ca], [Al]. [Cu], [V], [Pb], [B], [Nb], [Sn], [Ni], [Cr], [Mo], [Ti], [Fe] [N]	wt% ppm	Steel chemical composition at the beginning of the LF process in terms of [Ca], [V], [B], [Nb], [Sn], [Ni], [Cr], [Fe]	wt%
	Materials additions in the LF (all)	kg	Materials additions in the LF (all)	Kg
Target	Final steel chemical composition: [C], [S], [Mn], [Si], [P], [Ca], [Al]. [Cu], [V], [Pb], [B], [Nb], [Sn], [Ni], [Cr], [Mo], [Ti], [Fe] [N]	wt% ppm	Steel temperature at the end of the LF process	°C

boxplot methodology [37]. Finally, the cleansed datasets were divided into two subsets: the training dataset comprising 70% of the available data, while the remaining data were used for models testing and assessment.

For the LF process models, a first variables selection stage was carried out based on the knowledge of the process operators to remove useless, redundant, and noisy measurements, resulting in 54 significative variables. As both the model estimating the final steel chemistry and the one estimating the final steel temperature learn mass and energy balance of the process, the preliminary selected inputs variables depict mass and energy flows characterizing the process. Afterwards, further outliers were identified based on quartiles selection, by using box plot methodology [37]. This led to a cleansed database of about 2900 heats, which can be considered a medium-sized dataset in the context of industrial and data-driven applications. To ensure a reliable and unbiased evaluation, the dataset was divided following a two-stage splitting strategy. First, the original dataset was randomly separated into 80% training-validation data and 20% test data. Subsequently, the training-validation subset was further split into 80% training data and 20% validation data. This resulted in an overall distribution of 64% training, 16% validation, and 20% test samples with respect to the original dataset. Such data partition scheme is commonly adopted in ML applications involving medium-sized datasets to maintain a statistically meaningful test set and preserve adequate data availability for training as well as similar data distribution across the partitions. The training dataset is used in the learning stage. The validation dataset is used to implement the so-called "early stopping" that helps to partially overcome overfitting issues, by enhancing the model generalization capability. In this work, the early stopping algorithm monitors the trend of the *mean square error* (*MSE*) evaluated on the validation dataset with a patience of 10 epochs before stopping the training session. The test dataset is used to evaluate the results of the variable selection (if applied), the optimization of hyperparameters and the final optimal model training.

As final steel chemical composition depends on a numerous chemical reaction, the model forecasting steel chemical composition exploits all the 54 input variables. On the other hand, variable selection was carried out to reduce the dimension of the input vector of the model forecasting the final steel temperature. Such selection removed several components of the starting chemical analysis, leading to a final number of input variables equal to 42.

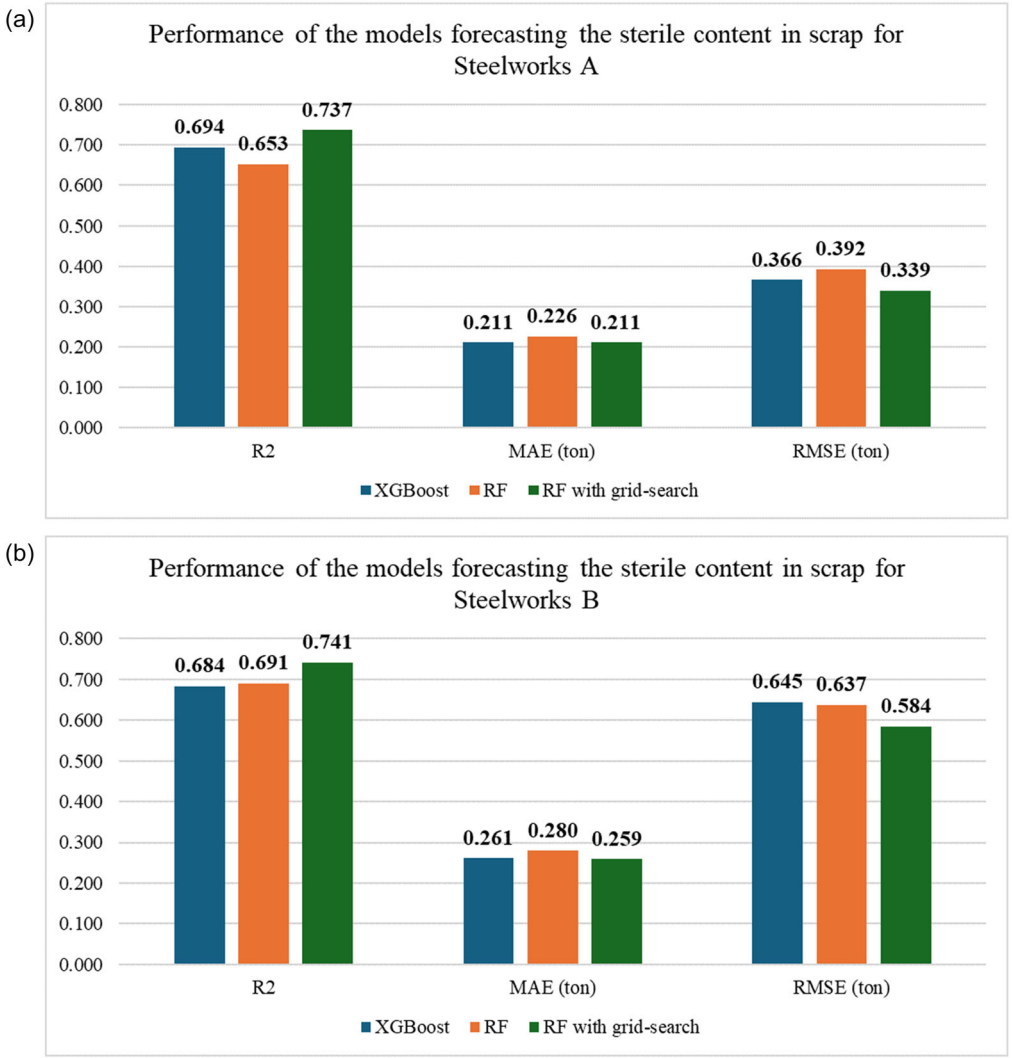

FIGURE 4 | Comparison of the performance of the different ensemble models to estimate the sterile contents in steel scrap assessed on the data provided by: (a) Steelworks A, (b) Steelworks B.

Table 1 provides the list of input and target variables of the two LF models: these variables refer to the single heat.

The material additions differ with respect to the plant, according to the target steel qualities produced. An exhaustive list of ferro-alloys, deoxidizers and slag formers cannot be reported for confidentiality reasons. Common added materials used in the plants are anthracite, SiMn, FeSi, FeV, CaSi, and CaO.

Variable selection and optimization of the FFNN hyperparameters (i.e., number of hidden layers N_{HL}, number of neurons in the hidden layers N_n, and learning rate l_r) of each LF-related model have been carried out through Bayesian algorithms [38] implemented in the Optuna library. To define the search range for the optimal hyper-parameters, a manual trial-and-error methodology has been followed to explore the performance of the networks in function of their value. This results in the following

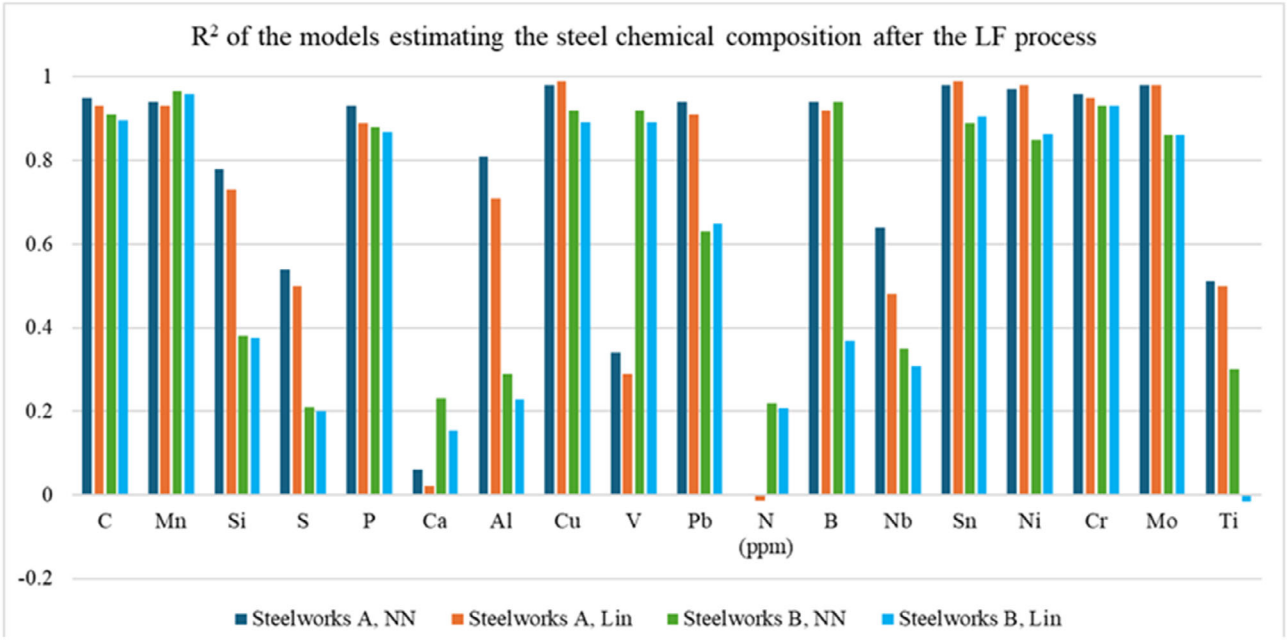

FIGURE 5 | R^2 values achieved by the FFNN-based model estimating the steel chemical composition after LF on the test datasets provided by the two considered steelworks.

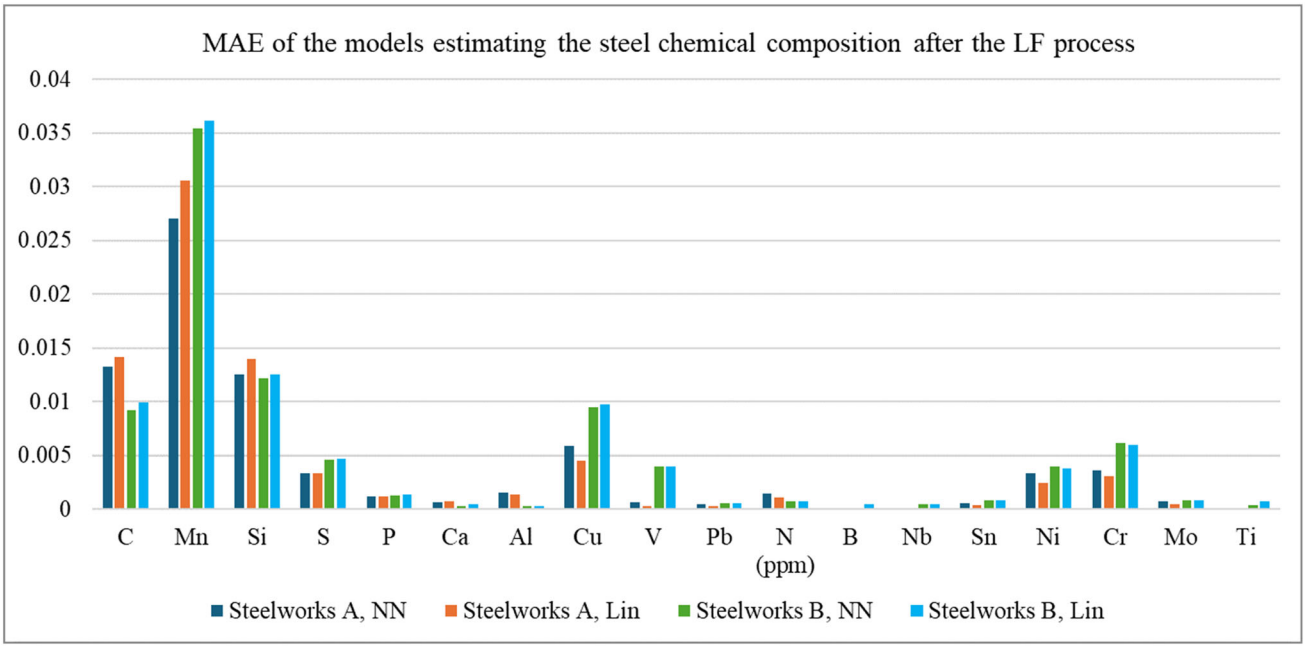

FIGURE 6 | *MAE* values achieved by the FFNN-based model estimating the steel chemical composition after LF on the test datasets provided by the two considered steelworks.

plausible optimal ranges: $N_{HL} \in [1,4]$, $N_n \in [10,100]$, $lr \in [10^{-6}, 10^{-3}]$ have been considered for the optimization.

3 | Results and Discussion

The models were trained and tested using two industrial datasets coming from two EAF-based steelworks as described in Section 2.1 to evaluate the accuracy of the provided estimates.

To assess and compare the performance of the different models, the following 3 indicators were adopted:

- *coefficient of determination* (R^2), which captures the proportion of the variance in a target variable y that is explained by the model. If N is the number of data in the test dataset, μ is the average value of the target, and $\hat{y}$ is the estimate of the target provided by the model, R^2 is computed as:

$$R^2 = \frac{\sum_{i=1}^{N} (\hat{y}_i - \mu)^2}{\sum_{i=1}^{N} (y_i - \mu)^2} \quad (1)$$

High values of R^2 imply a good predictive ability of the model.

- *Mean absolute error* (*MAE*), an easy interpretable index that measures the average error in predictions without considering their direction and is computed as

$$MAE = \frac{1}{N} \sum_{i=1}^{N} |y_i - \hat{y}_i| \quad (2)$$

- *Root mean square error* (*RMSE*), which is computed as

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2} \quad (3)$$

This index penalizes larger errors more than the MAE and gives an indication of the overall variance of the forecast. Low RMSE values indicate that the models provide stable and consistent forecasts.

3.1 | Estimate of the Sterile Content in Scrap

Figure 4 compares the performances of the 3 models predicting the sterile contents in scrap, i.e., XGBoost, RF, and RF optimized via grid-search assessed on the data provided by Steeworks A and Steelworks B.

The RF-based model optimized via grid-search shows the highest R^2 value in both cases ($R^2 = 0.737$ for Steelworks A and $R^2 = 0.741$ for Steelworks B), XGBoost slightly outperforms the baseline RF model for Steelworks A (XGBoost leads to $R^2 = 0.694$; baseline RF provides $R^2 = 0.691$), while the opposite situation arises for steelworks B (XGBoost leads to $R^2 = 0.684$; baseline RF provides $R^2 = 0.691$), but the difference between the two methods is minimal.

In terms of *MAE*, for both steelworks XGBoost and RF-based model optimized via grid-search show equivalent results (XGBoost leads to $MAE = 0.211$ *ton* for Steelworks A and $MAE = 0.262$ *ton* for Steelworks B; RF optimized via grid-search provides $MAE = 0.211$ *ton* for Steelworks A and $MAE = 0.259$ *ton* for Steelworks B), while the basic RF model provides slightly worst performance ($MAE = 0.226$ *ton* for Steelworks A and $MAE = 0.280$ *ton* for Steelworks B), but again the difference between these two methods is very limited.

In terms of *RMSE*, the optimized RF-based model provides again the best results for both steelworks ($RMSE = 0.339$ *ton* for Steelworks A and $RMSE = 0.584$ *ton* for Steelworks B), XGBoost slightly outperforms the baseline RF model for Steelworks A (XGBoost leads to $RMSE = 0.366$ *ton*; baseline

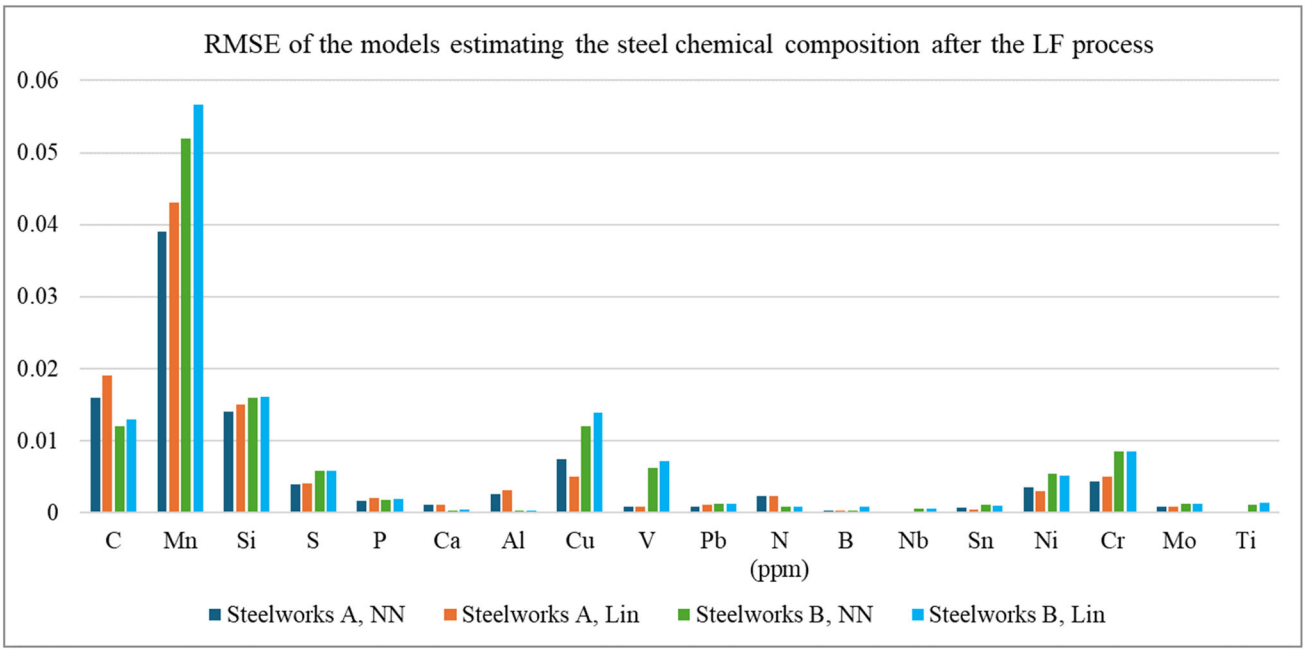

FIGURE 7 | *RMSE* values achieved by the FFNN-based model estimating the steel chemical composition after LF on the test datasets provided by the two considered steelworks.

RF provides *RMSE* = *0.392 ton*), while the opposite situation arises for Steelworks B (XGBoost leads to *RMSE* = *0.645 ton*; baseline RF provides *RMSE* = *0.637 ton*), although the difference is still limited.

To sum up, since contaminant presence in the scraps is a stochastic multifactorial phenomenon, which effect can be partially forecasted, all the models show encouraging results. All models exhibit high R^2 combined with low RMSE and MAE. Considering the complexity of the industrial sector, these modelling results are effective for supporting the decision-making of scrapyard operators on the scheduling of scrap processing. The comparison of modelling approaches shown the RF-based model optimized via grid-search outperforms the other ones on both datasets. Noticeably, while the observed R^2 and *MAE* values are comparable for both steelworks, the *RMSE* values for Steelworks B are sensibly higher than the ones observed for Steelworks A. This can be explained by the different scrap classification method used in Steelworks B, where a higher number of scrap classes (41) are registered compared to Steelworks A (11). Although the numbers of data provided by the two Steelworks are similar, the higher variability of the inputs negatively affects the performance.

3.2 | Models Related to the LF Treatment

The optimized FFNN-based model forecasting the contents of the different chemical components in steel at the end of the LF holds 2 hidden layers with 50 and 40 neurons, respectively.

Figures 5, 6 and 7 show models performance evaluated on the test datasets of the two steelworks in terms of R^2, *MAE* and *RMSE*, evaluated for the FFNN and linear model. In particular, the FFNN model show $R^2 > 0.9$ for C, Mn, P, Cu, Pb, B, Sn, Ni, Cr, Mo for Steelworks A while for Steelworks B $R^2 > 0.9$ for C, Mn, Cu, B, Cr, Mo and $0.85 \leq R^2 \leq 0.89$ for P, Sn, Ni, Mo. Overall, the linear model performs slightly worse than the FFNN, with a notable R^2 degradation for Si, Ca, Al, and Nb, for the steelworks A, and with a notable R^2 degradation for Ca, Al, B, Nb, and Ti, for the steelworks B. Moreover, linear model results for Sn and Ni are slightly superior to FFNN.

For FFNNs, the highest *MAE* values are observed for Mn (*RMSE* = *0.027 wt%* for Steelworks A and *MAE* = *0.035 wt%* for Steelworks B), C (*MAE* = *0.013 wt%* for Steelworks A and *MAE* = *0.009 wt%* for Steelworks B) and Si (*MAE* = *0.0125 wt%* for Steelworks A and *MAE* = *0.012 wt%* for Steelworks B). These are also the elements whose contents are highest in the

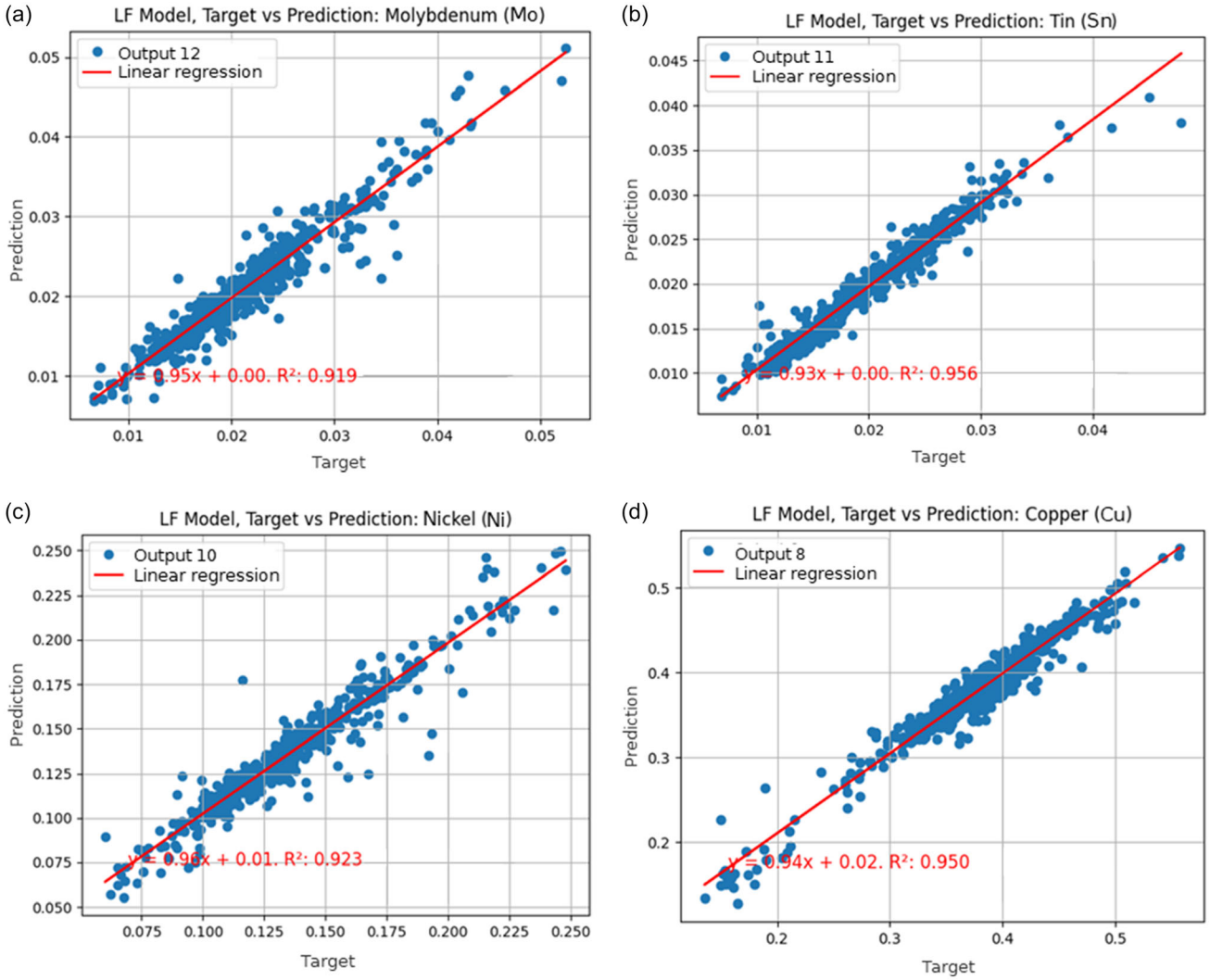

FIGURE 8 | Comparison of real and estimated content (expressed in wt%) of (a) molybdenum, (b) tin, (c) nickel, (d) copper in steel produced by Steelworks A at the end of the LF process.

produced steels. Also, the *RMSE* shows its highest values for Mn (*RMSE* = *0.039 wt%* for Steelworks A and *RMSE* = *0.052 wt%* for Steelworks B), C (*RMSE* = *0.016 wt%* for Steelworks A and *RMSE* = *0.012 wt%* for Steelworks B) and Si (*RMSE* = *0.014 wt%* for Steelworks A and *RMSE* = *0.016 wt%* for Steelworks B).

Although linear models exhibit trends similar to those observed for FFNNs, their overall performance is generally lower, as reflected by higher *RMSE* and *MAE* metrics for most of the elements. For Sn, Ni, and Mo, *RMSE* and *MAE* metrics are slightly better in the case of the linear model.

Figures 8 and 9 exemplary compare the real and estimated values of some chemical components related to Steelworks A and B in the case of the best model based on FFNNs. The regression straight line is also reported. The diagrams show the high generalization capability of FFNN models and the precision of the forecasting. A clear linear trend is observed in the target vs. prediction scatter plots, where data points exhibit minimal dispersion and remain close to the identity line, confirming the reliability of the model and its low error magnitude. This visual evidence is further supported by quantitative metrics, with low *RMSE* and *MAE* values and consistently high values of R^2, confirming the strong predictive accuracy of the model.

To sum up, the FFNN-based model shows good approximation and generalization capability. The obtained *RMSE* and *MAE* values are low for all the variables. R^2 values are high for the less reacting components (e.g., Cu, Cr, Mo, Sn, Ni), while they are lower when the model must reproduce complex chemical or pick-up phenomena determining, for instance, the calcium and nitrogen contents in steel. The estimation of N content can be affected also by its lower content compared to the other considered elements. Moreover, sensors sensitivity and related data reliability also impacts on the regression task of the model for some elements such as titanium and niobium. Additionally, since steel qualities with higher Ti or Nb contents are less frequently produced, the results on these elements are also affected by the poorer data distribution. The same reason is valid for the V content for the steelworks A. sulphur and aluminium contents are well estimated for Steelwork A, while for Steelworks B the accuracy decreases. This behaviour was expected, as the model structure was optimized using only the data coming from Steelworks A. Also, the additions that are made by Steelworks B are more limited compared to Steelworks A, which implies that a different structure of the model could better fit this specific dataset. However, such a structure would have a lower generalization capability and, thus, is not advisable in the context of the ALCHIMIA

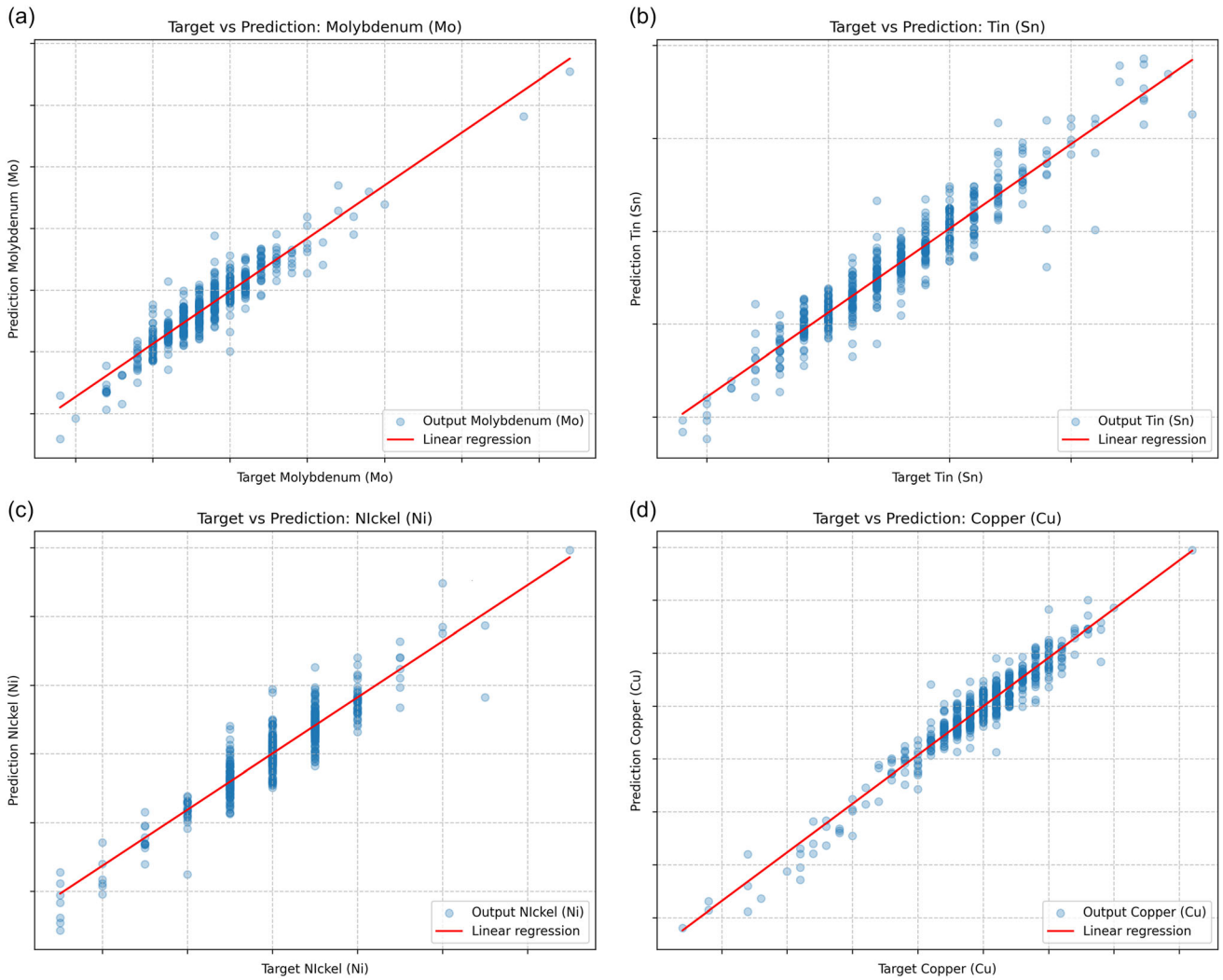

FIGURE 9 | Comparison of real and estimated content (expressed in wt%) of (a) molybdenum, (b) tin, (c) nickel, (d) copper in steel produced by Steelworks B at the end of the LF process (labels are not shown on the axes for industrial confidentiality constraints).

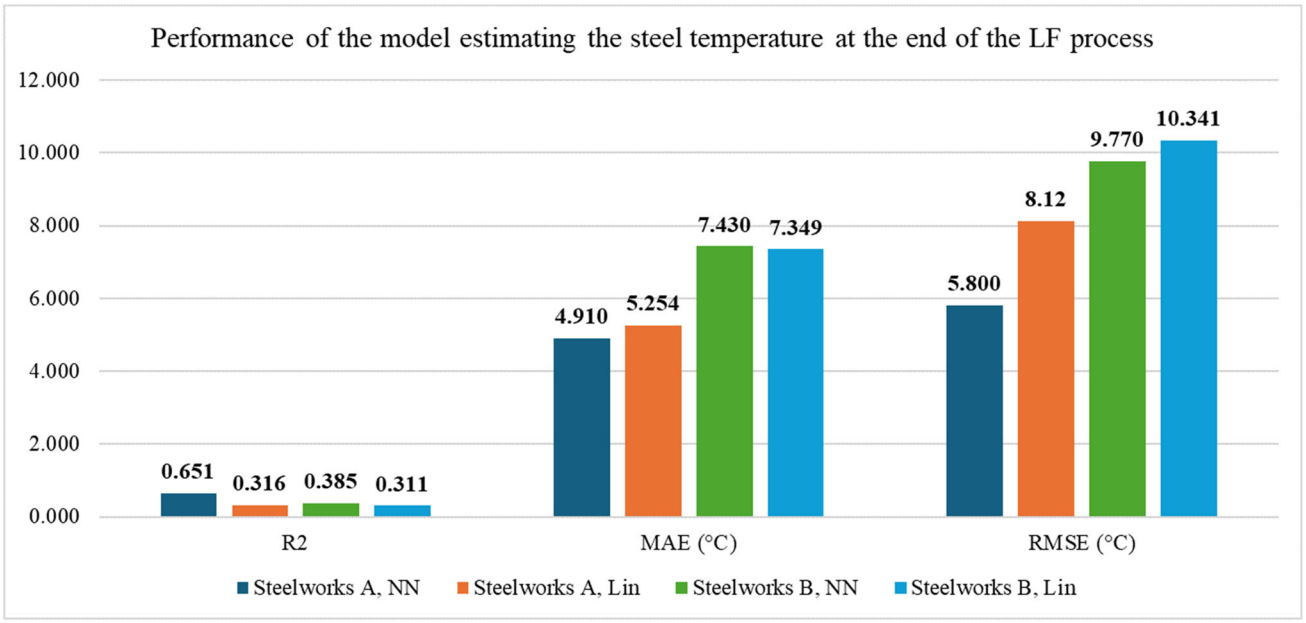

FIGURE 10 | Performance of the FFNN-based model estimating the steel temperature at the end of the LF treatment on the test datasets provided by the two considered steelworks.

solution, which is conceived to gradually and continuously adapt to the evolution of production.

Overall, linear models provide predictions comparable to FFNNs, but with evident performance gaps. The decreased performance observed on some of the target variables on the test database provided by Steelworks B does not represent a major obstacle to using the model. The reason is the mechanism of CL that is implemented within the more complex optimization platform in which the model is included, that enables continuous refining of the model performance. Therefore, as an initial step, the model performance can be considered satisfactory and suitable for monitoring and optimization purposes in the specific context for which the model has been developed.

On the other hand, the optimized FFNN-based model estimating the final steel temperature holds 2 hidden layers with 20 and 5 neurons, respectively. Figure 10 shows its performance assessed on the test datasets of both steelworks.

Also in this case, the FFNN results are good, although the performance is worse on the test datasets coming from Steelworks B ($R^2 = 0.385$, $MAE = 7.430°C$, and $RMSE = 9.770°C$) compared to Steelworks A ($R^2 = 0.651$, $MAE = 4.910°C$, $RMSE = 5.800°C$). In this case, however, the discrepancy in the results is mostly due to the difference in the cleanliness and quality of the measurements of the target variable, which is lower for Steelworks B compared to Steelworks A. However, the resulting accuracy is still sufficient to use this model for both monitoring and control purposes. Linear models show a significant decrease in performance for all three metrics, highlighting the effect of nonlinear behaviours on the thermal dynamics that are instead numerically well represented by FFNNs.

4 | Conclusion

Holistic ML-based models are proposed to estimate the sterile content in the scrap reaching the scrapyard as well as the steel chemical composition and temperature at the end of the LF process. These models are part of a digital twin of the EAF-based steelmaking route, which is used to support process operators in minimizing environmental and economic impacts of the EAF-based steelmaking route, while ensuring the quality of the steel feeding the continuous casting machine.

The models estimating the sterile contents in scrap based on scrap origin, supplier, weight, and month of arrival in the scrap yard show encouraging results. In particular, among the 3 tested models, the one based on RF optimized via grid-search provides the best results ($R^2 = 0.737$, $MAE = 0.211$ *ton* and $RMSE = 0.339$ *ton* for Steelworks A and $R^2 = 0.741$, $MAE = 0.259$ *ton* and $RMSE = 0.584$ *ton* for Steelworks B). However, the results also show that the scrap classification used in Steelwork A results in significant variability in the output when applied to Steelwork B, which in turn could affect the reliability and transferability of the models. Indeed, these models can be improved in the future considering the stochastic nature of the presence of these contaminants and the low number of input variables. Computer vision-based methodologies for scrap inspection can help improve the information conveyed by the data with a positive impact on the model performance.

The models related to the LF process are implemented via FFNNs whose hyperparameters are optimized through Bayesian algorithms and estimate of the final chemical components of the molten steel and its temperature based on initial steel temperature and composition, electric energy and argon consumption and total scrap weight loaded in the EAF. In particular, the model related to the steel composition provides acceptable performances for most components, showing $R^2 > 0.9$ for C, Mn, P, Cu, Pb, B, Sn, Ni, Cr, and Mo for Steelworks A (whose data were used for model structure optimization), while for Steelworks B $R^2 > 0.9$ for C, Mn, Cu, B, Cr, Mo and $0.85 \leq R^2 \leq 0.89$ for P, Sn, Ni, and Mo. The model related to the final steel temperature provides good results for Steelworks A ($R^2 = 0.651$, $MAE = 4.910°C$, $RMSE = 5.800°C$), while the performance is worse for Steelworks B ($R^2 = 0.385$, $MAE = 7.430°C$,

RMSE = *9.770°C*) due to the far lower cleanliness and quality of the temperature measurements in this steelworks. This confirms that ML-based models, such as any data-driven model, is sensitive to the quality of the data that are used for its training. However, the accuracy of these models is still sufficient to the purpose of the present work, as they can provide useful insights for decreasing the use of energy and added materials.

Ongoing and future work concern improvement of models' accuracy by enhancing the data collection phase and development of other models to estimate further target variables concerning by-products (e.g., slag) that can be included in the decision support system to provide more comprehensive feedback and a deeper insight on process quality, productivity and environmental impact. The LF model will also be improved by exploiting new data coming from the two considered steelworks and other ones. This improvement is possible thanks, respectively, to the CL mechanisms, which enable continuous adaptation and fine tuning of the models, and to the FL platform, which preserves data privacy and security while enhancing user's experience. One further topic to explore in future work consists in embedding in the available CL platform the capability to self-adapt the overall structure of the model, i.e., its hyperparameters.

Author Contributions

Valentina Colla: formal analysis (equal), funding acquisition (lead), investigation (equal), methodology: (equal), project administration (lead), resources (lead), supervision (lead), validation (supporting), visualization (supporting); writing – original draft (equal). **Antonella Zaccara**: conceptualization (equal), data curation (equal), formal analysis (equal), methodology (equal), Visualization (equal), writing – review & editing (equal). **Stefano Dettori**: conceptualization (equal), data curation (equal), formal analysis (equal), investigation (equal), methodology (equal), supervision (equal), validation (equal), visualization (equal), writing – review & editing (equal). **Laura Laid**: data curation (equal), investigation (equal), software (equal), validation (equal). **Ismael Matino**: conceptualization (equal), formal analysis (supporting), methodology (equal), validation (supporting), writing – review & editing: (equal). **Silvia Cateni**: conceptualization (equal), data curation (lead), investigation (equal), methodology (equal), validation (lead), writing – review & editing (equal). **Teresa Annunziata Branca**: conceptualization (equal), formal analysis (equal), investigation (equal), supervision (equal), validation: (equal), visualization (supporting), writing – review & editing (equal). **Lorenzo Vannini**: conceptualization (equal), data curation: (supporting), investigation (equal), software (lead), validation (equal), visualization (equal), writing – original draft: (lead).

Acknowledgments

The work described in the present paper has been developed within the project entitled "*Data and decentralized Artificial intelligence for a competitive and green European metallurgy industry*" (Ref. ALCHIMIA, Grant Agreement No. 101070046) that has received funding from the Horizon Europe research and innovation program of the European Union, which is gratefully acknowledged. The sole responsibility for the issues treated in the present paper lies with the authors; the Commission is not responsible for any use that may be made of the information contained therein.

Open access publishing facilitated by Scuola Superiore Sant'Anna, as part of the Wiley - CRUI-CARE agreement.

Funding

This work was supported by the Horizon Europe framework programme of the European Union, Grant Agreement No. 101070046.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

Data associated with this article cannot be disclosed due to confidentiality constraints of the involved steelmaking company.

References

1. *European Commission*, "The European Green Deal, last access September 26," (2025), https://ec.europa.eu/info/strategy/priorities-2019-2024/european-green-deal_en.

2. European Commission, "Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions. A Green Deal Industrial Plan for the Net-Zero Age," *COM2023*, (final, 2023),62.

3. International Energy Agency - IEA, "Achieving Net Zero Heavy Industry Sectors in G7 Members, Last Access September 26," (2025), https://www.iea.org/reports/achieving-net-zero-heavy-industry-sectors-in-g7-members.

4. European Commission, "Energy Intensive Process Industries, Last Access September 26," (2025), https://research-and-innovation.ec.europa.eu/research-area/industrial-research-and-innovation/energy-intensive-process-industries_en.

5. European Environment Agency, "Circular Material use Rate in Europe, Last Access September 26," (2025), https://www.eea.europa.eu/en/analysis/indicators/circular-material-use-rate-in-europe.

6. AISBL. E.U.R.I.C., "Recycling: Bridging Circular Economy & Climate Policy, Last Access September 26," (2025), https://circulareconomy.europa.eu/platform/sites/default/files/euric_metal_recycling_factsheet.pdf.

7. G. F. Porzio, V. Colla, B. Fornai, M. Vannucci, M. Larsson, and Håkan Stripple, "Process integration analysis and some economic-environmental implications for an innovative environmentally friendly recovery and pre-treatment of steel scrap," *Applied Energy* 161 (2016): 656, https://doi.org/10.1016/j.apenergy.2015.08.086.

8. J. Rieger, M. Leitner, V. Colla, A. Petrucciani, L. Sandberg, and J. Petersson, "Improved Analysis of Post-Consumer Scrap," *Steel Times International* 32 (2024): 35.

9. V. Colla, C. Pietrosanti, E. Malfa, and K. Peters, "Environment 4.0: How Digitalization and Machine Learning Can Improve the Environmental Footprint of the Steel Production Processes," *Matériaux & Techniques* 108 (2020): 5–6, https://doi.org/10.1051/mattech/2021007.

10. S. Dettori, I. Matino, V. Colla, and R. Speets, "A Deep Learning-Based Approach for Forecasting Off-Gas Production and Consumption in the Blast Furnace," *Neural Computing & Applications* 34, no. 2 (2022): 911, https://doi.org/10.1007/s00521-021-05984-x.

11. P. Gailhofer, A. Herold, J. P. Schemmel, et al., *The Role of Artificial Intelligence in the European* (European Parliament, 2021).

12. T. A. Branca, B. Fornai, V. Colla, et al., "Schröder: Industrial Symbiosis and Energy Efficiency in European Process Industries: A Review," *Sustainability* 13 (2021): 9159, https://doi.org/10.3390/su13169159.

13. M. Larsson and J. Dahl, "Reduction of the Specific Energy Use in an Integrated Steel Plant—the Effect of an Optimisation Model," *ISIJ International* 43, no. 10 (2003): 1664, https://doi.org/10.2355/isijinternational.43.1664.

14. A. Fernández Martínez, S. Muiños-Landín, A. Gordini, et al., "A Machine Learning Framework for Improving Resources, Process, and Energy Efficiency Towards a Sustainable Steel Industry,"*Lecture Notes in Networks and Systems* 1028 (2024): 3.https://doi.org/10.1007/978-3-031-61905-2_1.

15. M. M. Abadi, H. Tang, and M. M. Rashidi, "A review of simulation and numerical modeling of electric arc furnace (EAF) and its processes," *Heliyon* 10, no. 11 (2024): e32157, https://doi.org/10.1016/j.heliyon.2024.e32157.

16. M. Liu, G. Ma, X. Zhang, and D. Zheng, "Numerical Simulation on the Melting Kinetics of Steel Scrap in Iron-Carbon Bath," *Case Studies in Thermal Engineering* 34 (2022): 101995, https://doi.org/10.1016/j.csite.2022.101995.

17. C. Mapelli, G. Dall'Osto, S. Scolari, S. Bazri, V. Zanaglio, and D. Mombelli, "Modelling the Influence of Scrap Size on Charge-to-Melt Time and Metallic Loss in EAFs," *Ironmaking Steelmaking* 52 (2025): 1–8, https://doi.org/10.1177/03019233251387764.

18. W. Alves Ferreira Neto, C. A. Virgínio Cavalcante, and P. Do, "Deep Reinforcement Learning for Maintenance Optimization of a Scrap-Based Steel Production Line," *Reliability Engineering & System Safety* 249 (2024): 110199, https://doi.org/10.1016/j.ress.2024.110199.

19. L. S. Carlsson, P. B. Samuelsson, and P. G. Jönsson, "Interpretable Machine Learning—Tools to Interpret the Predictions of a Machine Learning Model Predicting the Electrical Energy Consumption of an Electric Arc Furnace," *Steel Research International* 91 (2020): 2000053, https://doi.org/10.1002/srin.202000053.

20. L. Li, Y. Fan, M. Tse, and K. Y. Lin, "A Review of Applications in Federated Learning," *Computers & Industrial Engineering* 149 (2020): 106854, https://doi.org/10.1016/j.cie.2020.106854.

21. R. Hadsell, D. Rao, A. A. Rusu, and R. Pascanu, "Embracing Change: Continual Learning in Deep Neural Networks," *Trends in Cognitive Sciences* 24 (2020): 12, https://doi.org/10.1016/j.tics.2020.09.004 1028.

22. A. Blanco-Justicia, J. Domingo-Ferrer, S. Martinéz, D. Sánchez, A. Flanagan, and K. E. Tan, "Achieving Security and Privacy in Federated Learning Systems: Survey, Research Challenges and Future Directions," *Engineering Applications of Artificial Intelligence* 106 (2021): 104468, https://doi.org/10.1016/j.engappai.2021.104468.

23. R. Gupta and T. Alam, "Survey on Federated-Learning Approaches in Distributed Environment," *Wireless Personal Communications* 125, no. 2 (2022): 1631, https://doi.org/10.1007/s11277-022-09624-y.

24. T. Gyllenram, N. Arzpeyma, W. Wei, and P. G. Jönsson, "Driving Investments in ore Beneficiation and Scrap Upgrading to Meet an Increased Demand from the Direct Reduction-EAF Route," *Mineral Economics* 35 (2022): 203, https://doi.org/10.1007/s13563-021-00267-2.

25. B. de la Peña, A. Iriondo, A. Galletebeitia, et al., "Toward the Decarbonization of the Steel Sector: Development of an Artificial Intelligence Model Based on Hyperspectral Imaging at Fully Automated Scrap Characterization for Material Upgrading Operations," *Steel Research International* 94, no. 11 (2023): 2200943, https://doi.org/10.1002/srin.202200943.

26. W. Xu, P. Xiao, L. Zhu, et al., "Classification and Rating of Steel Scrap Using Deep Learning," *Engineering Applications of Artificial Intelligence* 123 (2023): 106241, https://doi.org/10.1016/j.engappai.2023.106241.

27. P. Henrique dos Santos and V. de Carvalho Santos, "E.J. da Silva Luz: Towards Robust Ferrous Scrap Material Classification with Deep Learning and Conformal Prediction," *Engineering Applications of Artificial Intelligence* 140 (2025): 109724, https://doi.org/10.1016/j.engappai.2024.109724.

28. M. Schäfer, U. Faltings, and B. Glaser, "CLRiuS: Contrastive Learning for Intrinsically Unordered Steel Scrap," *Machine Learning with Applications* 17 (2024): 100573, https://doi.org/10.1016/j.mlwa.2024.100573.

29. G. Koinig, N. Kuhn, T. Fink, et al., "Deep Learning Approaches for Classification of Copper-Containing Metal Scrap in Recycling Processes," *Waste Manage* 190 (2024): 520, https://doi.org/10.1016/j.wasman.2024.10.022.

30. D. You, S. K. Michelic, and C. Bernhard, "Modeling of Ladle Refining Process Considering Mixing and Chemical Reaction," *Steel Research International* 91, no. 11 (2020): 2000045, https://doi.org/10.1002/srin.202000045.

31. Z. Wang, Y. Xie, L. Wang, et al., "Applications and Progress of Machine Learning Techniques in the Ladle Furnace Refining Process: A Review," *Steel Research International* 96, no. 7 (2025): 2400551, https://doi.org/10.1002/srin.202400551.

32. E. K. Sahin, "Assessing the Predictive Capability of Ensemble Tree Methods for Landslide Susceptibility Mapping Using XGBoost, Gradient Boosting Machine, and Random Forest," *SN Applied Sciences* 2 (2020): 1308, https://doi.org/10.1007/s42452-020-3060-1.

33. A. B. K. Didavi, R. G. Agbokpanzo, and M. Agbomahena, "Comparative Study of Decision Tree, Random Forest and XGBoost Performance in Forecasting the Power Output of a Photovoltaic System," (4th Int. Conf. Bioengineering for Smart Technologies (BioSMART), 2021): 1, https://doi.org/10.1109/BioSMART54244.2021.9677566.

34. T. Chen and C. Guestrin, XGBoost: A Scalable Tree Boosting System, (Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16), 2016): 785, https://doi.org/10.1145/2939672.2939785.

35. S. De Backer, A. Naud, and P. Scheunder, "Non-Linear Dimensionality Reduction Techniques for Unsupervised Feature Extraction," *Pattern Recognition Letters* 19, no. 8 (1998): 711, https://doi.org/10.1016/S0167-8655(98)00049-X.

36. Z. Zhang, "Improved Adam Optimizer for Deep Neural Networks," (26th IEEE/ACM Int. Symp. Quality of Service–IWQoS, 2018): 1.

37. A. Mazarei, R. Sousa, J. Mendes-Moreira, S. Molchanov, and H. M. Ferreira, "Online Boxplot Derived Outlier Detection," *International Journal of Data Science and Analytics* 19 (2025): 83, https://doi.org/10.1007/s41060-024-00559-0.

38. S. Watanabe, "Tree-Structured Parzen Estimator: Understanding Its Algorithm Components and Their Roles for Better Empirical Performance," (2023), https://doi.org/10.48550/arXiv.2304.11127.