Using Machine Learning for Robust Target Prediction in a Basic Oxygen Furnace System

JUHEE BAE, YURONG LI, NICLAS STÅHL, GUNNAR MATHIASON, and NIKLAS KOJOLA

The steel-making process in a Basic Oxygen Furnace (BOF) must meet a combination of target values such as the final melt temperature and upper limits of the carbon and phosphorus content of the final melt with minimum material loss. An optimal blow end time (cut-off point), where these targets are met, often relies on the experience and skill of the operators who control the process, using both collected sensor readings and an implicit understanding of how the process develops. If the precision of hitting the optimal cut-off point can be improved, this immediately increases productivity as well as material and energy efficiency, thus decreasing environmental impact and cost. We examine the usage of standard machine learning models to predict the end-point targets using a full production dataset. Various causes of prediction uncertainty are explored and isolated using a combination of raw data and engineered features. In this study, we reach robust temperature, carbon, and phosphorus prediction hit rates of 88, 92, and 89 pct, respectively, using a large production dataset.

I. INTRODUCTION

THE Basic Oxygen Furnace (BOF) process for the decarburization of hot metal in primary steel making is very complex from a process control perspective since it takes place at very high temperatures and includes turbulent multi-phase mass flow and chemical reactions. The fundamental process is that solid steel scrap and hot metal are charged into a BOF converter vessel, forming a production batch (heat). To reduce the carbon content through oxidization, oxygen is blown onto the heat by supersonic injection through a vertical lance. Fluxes such as lime and calcined dolomite are added to form a slag and remove impurities. During the blow, an operator fine-tunes the chemical reactions and the energy and mass balances by additions such as ore, coke or ferrosilicon, and by adjusting the position of the oxygen lance relative to a predefined movement pattern. During heat design and during the oxygen blow period, operators are usually assisted by some computer-based process guidance system. Such systems typically propose process parameters and operator actions for every heat execution, and are based on physical principles such as calculations of mass and energy balances and thermodynamic calculations.

A complex interaction between various influential factors determines the final outcome of the BOF process. The outcome is measured in terms of a number of target values such as the temperature, carbon content, and phosphorus content of the final melt and metallic losses to the slag. The process (the blow) takes about 20 minutes, and it is desirable to predict the optimal duration of individual blows by predicting the cut-off point. The evolution of critical parameters during the process is shown schematically in Figure 1. If the process is ended too early, targets are not reached and the process must be re-started, causing a decrease in productivity and high cost. If the process is ended too late, the temperature becomes too high and the metallic losses to the slag become large, which significantly increases the environmental impact and cost. Since the BOF process is the world's most commonly used process for ore-based steel production, tools to improve cut-off prediction could significantly contribute to a more sustainable and efficient global steel production. BOF cut-off prediction has resulted in many studies attempting to model the BOF process, by using advanced simulations (*e.g.*, mass and energy flows in the vessel) and by approaches based on modeling the process from process data. The literature contains several machine learning (ML)-based approaches to predict target temperature (T), carbon content (pct C), and phosphorus content (pct P) with 80 to 100 pct prediction accuracy.

JUHEE BAE, YURONG LI, NICLAS STÅHL, and GUNNAR MATHIASON are with the School of Informatics, University of Skövde, 54134 Skövde, Sweden. Contact e-mail: juhee.bae@his.se NIKLAS KOJOLA is with the Group function R&I, SSAB AB, 11164 Stockholm, Sweden.
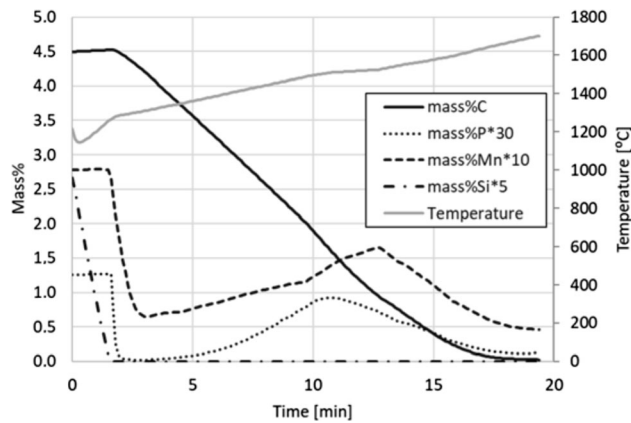
Fig. 1—Element composition during the BOF converter process (SSAB).

However, many of these studies make use of limited datasets or a limited number of ML approaches so the accuracy and robustness of these can, therefore, be questioned (refer to Table IV). To the best of our knowledge, there is no investigation published that compares a large set of commonly used ML algorithms with a full-size production dataset.

It is difficult to monitor and collect data in the vessel during a blow due to the high temperatures, aggressive chemical environment, and the violent motion of the mass, which quickly consumes most sensory equipment. In many cases, the only well-defined measurement available consists of the final temperature and an analysis of a melt sample for the final element composition of the heat. These measurements are collected at the end of the process. For some BOF converters, an additional measurement of the same parameters may be taken in the form of pre-final melt samples using a sub-lance at an estimated point of approximately 90 pct of the blow time. The element composition of the melt sample is determined by a lab analysis which usually takes several minutes. Consequently, acting on the final measurements can only result in re-blows if targets are not met, while the time to act based on the 90 pct measurement is very limited.

To investigate and expand on existing published work, we evaluate target predictéhms with fullé our work includes key engineered features based on the collected data that use well-known thermodynamic principles, chosen in collaboration with domain experts. Standard ML models are able to capture the complex behavior of the BOF process with high prediction accuracy using a real, full-size production dataset with a large set of features and high variability in the data. For further prediction accuracy improvements, it is likely that additional data sources and even more expressive ML models must be used, such as level 1 time-series process data and ML algorithms for sequential models, and we intend to explore them in subsequent work.

Moreover, in this paper, we raise the question as to whether the actual BOF process, i.e., the events taking place inside the vessel, can be modeled using data collected about the entire production unit including the BOF converter. The BOF process is partly automated, partly manually controlled, and typically the production results are within specification. Problems during the blow are often, but not always, manually mitigated during the process, e.g., resulting in a strong correlation between features that lack a causal relation. Therefore, we claim that the trained prediction model captures the behavior of the observable data about the production unit, and that a trained model based on such data does not model the actual BOF process itself, but rather models the entire production unit's behavior. The observed data are a consequence of a mix of the BOF process itself and current process control. Thus, the trained model includes the dynamics of the BOF converter, the dynamics of the current control and prediction systems in place, and also the dynamics of the actions of the operators acting on their estimations of the current process state. This has a consequence when evaluating an analysis of which features influence the predictions of the trained model, since these features are actually influential predictors of the production unit's behavior rather than the behavior of the actual BOF.

II. RELATED WORK IN MACHINE-LEARNING-BASED PREDICTION MODELS

A sizable body of work has been produced in recent years where authors use different machine learning approaches to capture the relation between process parameters (features in data) and their targets, such as a prediction of the BOF target for a more accurate process cut-off time point. Some ML approaches are more frequently used than others and several publications claim to have achieved very high prediction accuracy. Most existing work predicts only one target—usually the end temperature. There is a preference for interpretable approaches with the aim being to explain the process, which is why several methods are based on first-principles[1] and simulations.[2]

The existing related studies differ from each other in several aspects: the size and fidelity of the dataset, the robustness and expressiveness of the chosen ML algorithm, the variability and size of the data used, the target error range, the number of direct and calculated features used, and finally how the result is validated. These are factors that affect the findings of the studies. The complex ML algorithms, for example, overfit easily small datasets. It is also easier to achieve high prediction accuracy with a simple ML algorithm when using a small dataset that has little variability. Furthermore, a small dataset may not represent the actual data distribution of a long production period very well, making limited approaches unsuitable for actual production models. In this paper, we compare and contrast a number of comparable related studies that make use of a large dataset or use many features.

There are three main types of ML approaches used in related studies: Case-based reasoning (CBR), Support Vector Machines (SVM), and Artificial Neural Networks (ANN). In CBR, instances are grouped by their behavior as expressed by the data. The assumption is that instances that have similar recorded features behave similarly, and are therefore put in the same behavioral group. They are expected to behave the same way during the process, resulting in a similar outcome. The goal of using CBR is to get a comprehensibly small set of behavioral groups representing different types of heats. These groups can then act as blueprints for how to interpret and control future heats.[3–7] The drawback of a CBR-based approach is that the discovered groups only correspond to a rough estimate of the behavior of such types of heats. It is therefore impossible to study any granular behavior differences and, thus, impossible to gain any deep understanding of the process. Still, under this limitation, CBR has been proven to be a useful method when it comes to predicting the outcome of a BOF or other similar processes. Wang et al.,[3] for example, used four different CBR methods and three other ML methods with a dataset of 460 heats from an Argon Oxygen Decarburization (AOD) converter. With one of these CBR approaches and 12 chosen features, Wang et al.[3] managed to predict the outcome temperature within ± 15 °C from the measured temperature. The AOD process is somewhat similar to BOF, but it is not the same since the oxygen is injected into the liquid steel below the bath surface level. This is why the prediction accuracy between the two processes cannot be directly compared. Another use of CBR is presented by Feng et al.[4] that combined CBR with ANN to model the Ruhrstahl Heraeus (RH) vacuum degasser using 2500 training heats with nine features. Feng et al.[4] reported that 95 pct of all predictions are less than ± 10 °C from the measured value and that it can be beneficial for the accuracy to combine several methods such as CBR and ANNs. The RH process is different from BOF in many aspects. Their ML methodologies may be compared, but the targets are different from the RH process.

SVM, and more specifically support vector machines for regression (SVR), can effectively capture nonlinear relationships in data, and SVMs are therefore used for nonlinear classification and regression. The effectiveness of an SVM typically depends on the kernel function used, where a common choice is the Radial Basis Function (RBF) kernel. SVMs are sometimes used in combination with other techniques to improve prediction accuracy further. Related studies that use models based on SVMs have good prediction results, but it is difficult to generalize the findings from most of the studies due to the small size of datasets used.[8–11] An exception is found in Schlüter et al.[12] who used an SVM approach with a large dataset of 1400 heats with 50 to 60 features predicting four targets: temperature (T), carbon (pct C) and phosphorus (pct P) content, and the iron content of the slag. The SVM model is claimed to outperform traditional metallurgical models. However, the study lacks detail so their work cannot be replicated. Gao et al.[13] used an improved twin support vector regression approach for T and pct C prediction, using 300 selected samples. They achieved hit rates of 96 and 94 pct within the error bound ± 15 °C and ± 0.005 pct for temperature and carbon content, respectively. The original data contain 2000 heats, so most heats were disqualified and removed. This pruning of the data may bias the final result and hence the high predictive power could be subject to some doubt.

There are many variants of using ANNs for the modeling of the BOF process.[4,8,9,14–25] For all of these, only between 6 and 18 features are used, and the number of used samples varies from 17 to 2500 with a majority at the lower end. The prediction accuracy typically reaches around 90 pct, but the data used are typically selectively chosen, and it is therefore likely that the result is biased in a positive direction. The target temperature range varies between ± 10 °C, but a prediction is typically denoted as correct if it is within ± 15 °C of the measured target. A majority of the papers do not report the ANN configurations used in the reported experiments, such as the number of layers and neurons, which makes replication of the experiments difficult. Some authors also pre-process the data in addition to filtering. He et al.,[25] for example, applied principal component analysis (PCA) to the initial data to reduce the number of features and to get uncorrelated features. By reducing the original dataset of 18 interrelated features to only seven components and then applying an ANN, He et al.[25] achieved a phosphorus (pct P) predictive accuracy of 93.33 pct, with an error tolerance of ± 0.005 pct.

There are several combination methods, besides the ones mentioned above, which have been used for predictions of the BOF process. In He et al.,[25] multiple linear regression was used. There are also some studies that consider combinations or modifications of the techniques described above.[9,17,22,23] In Wang et al.,[18] the authors applied weighted K-means clustering and a group method of data handling (GMDH), combined with a separate ANN for each cluster. Each cluster is hence treated separately. Using this method, Wang et al.[18] reported a pct P prediction accuracy of 96.40 pct, with an error tolerance of ± 0.006 pct. Yuan et al.[26] combined principal components regression (PCR) with multiple support vector machine (MSVM) and least square support vector machine (LS-SVM) for Electric Arc Furnace (EAF) temperature, carbon, and phosphorus targets. With 82 heats and 10 features, this approach achieves 93 pct accuracy, but this study used only simulated data. Han et al.[27] used a radial basis function neural network (RBF-NN) model, combined with a particle swarm optimization algorithm and independent component analysis, to predict BOF end-point temperature and carbon content based on only 60 observations. The authors claim the hit rate to be 100 pct when the range of the temperature error is ± 15 °C and 92 pct when the range of carbon content error is ± 0.05 pct. In Laha et al.,[21] random forest (RF), ANN, dynamic evolving neuro-fuzzy inference system (DENFIS), and support vector regression (SVR) were compared for BOF end-point prediction based on 54 heat samples and 10

features. SVR showed the best performance, with a mean value for R^2 at 0.821138 over 50 trials, and used less CPU time than ANN and DENFIS.

In summary, we found a few comparable studies that claim to predict the targets well for a BOF process. In Table IV, we summarize a comparison of their approaches and configurations as a prequel to a comparison with our experiments and results.

III. PROBLEM

Due to the harsh conditions in the vessel, the complex BOF process cannot be easily measured by sensors during the process execution. In particular, the high temperature is challenging since most sensors are destroyed by the heat. In some earlier studies, indirect sensor readings are used to mitigate this. With the lack of direct time-series-based data, it is advantageous to simulate the process, or to let ML algorithms predict the targets based on the data that are actually available. In fact, the aim of this article is to reduce the target prediction uncertainty compared with the current process control. In addition, augmenting the available data with data generated or calculated by experts allows the ML model prediction to make use of expert knowledge. This paper aims to compare ML-based prediction models that have been previously shown to be effective, but now using a complete production dataset, without any reduction in variance originating from less well-controlled process variables. We examine the expression capability of such standard ML algorithms for these data and investigate the information content for combinations of partitions of the dataset. The aim is to see how standard ML algorithms can make use of such rich information content.

Most of the existing work for predicting temperature, carbon, and phosphorus contents claims high prediction accuracy. However, the models used are often based and evaluated on data which are collected during a short time frame, so there is a low variance among the heats (data samples). Such approaches cannot feasibly be generalized and applied in a production environment since samples are much more diverse in a full production setting. Therefore, we use a large dataset comprising three years of production data. Moreover, since most of the existing studies only cover a few algorithms and typically use a limited dataset with less variation (refer to Table IV), it is difficult to compare existing studies to find usable real-world approaches. Therefore, it is necessary to compare a large set of commonly used ML approaches using full data, and then conclude how a valid and useful production approach should be designed. This will also establish a baseline for our future work where we aim to apply a time-series-based approach.

Furthermore, we examine the addition of expert-based engineered features based on these collected data with standard ML algorithms. One promise of deep neural networks is that there is no need for feature engineering by experts and that algorithms can discern useful features automatically from raw data.[28] In contrast, there are also previous BOF target prediction approaches that rely on engineered features to improve prediction accuracy, or use methods that choose a smaller subset of features.[3,21,22,25] Therefore, we also evaluate the effect of using expert-based features for each of the ML approaches we have chosen to compare.

In the next sections, we describe the dataset and ML algorithms we used and then the two experiments. Our results show how different ML algorithms work on a real-world steel manufacturing dataset using significantly more features and heats than existing studies. Moreover, we verify whether engineered features, that is, the usage of a simplified thermodynamic approach, would achieve an improvement compared to the straightforward use of the collected data.

IV. METHOD

A. Data Overview

The dataset consists of production data from an industrial BOF converter for steel production operated by the SSAB group in Sweden. The data were collected between 2014 and 2017 and contains data for around 20 to 30 heats of continuous production per day, resulting in a dataset of 17,000 heats with 33 parameters describing each heat, with one value stored per parameter and per heat. These 33 parameters include 20 parameters stored before the oxygen blow and 13 parameters stored during the oxygen blow. In addition, five parameters are collected as time-series throughout the oxygen blow. For each heat, the resulting target outcomes are stored as one single value per heat and target for temperature, carbon, and phosphorus (T, pct C, pct P).

From the original 17,000 heats collected, a number of heats were removed that suffered from corrupt or missing values which were determined by the process experts, resulting in a final heat dataset of 9708 heats. The heats were removed due to corrupt or missing values in pre-sample temperature, last tap temperature, pre-sample carbon, and hot metal carbon. Also, the first four heats after a furnace relining were removed since they can be regarded as tuning heats for the furnace, and are therefore outliers for describing continuous production. In addition, heats with exceptional values were removed, such as those with a too short or too long blow time (thresholds set for blow time less than 12 minutes. or longer than 30 minutes) and the apparent sensor breakdowns.

Furthermore, the five time-series for each heat were aggregated into 64 statistical descriptors (Section IV–A–2) to represent the time-series using one value per heat. An additional set of 17 'expert' or 'engineered' parameters (Section IV–A–3) were calculated from the originally collected parameters using thermodynamic principles to capture expected dependencies between parameters.

We normalize the dataset by calculating the mean and standard deviation and for each observed value of a feature, subtract the mean and divide by the standard deviation. The data are split for 10-fold cross-validation into a training set of 80 pct of the 9708 heats, a 10 pct validation set for model selection, and a 10 pct hold-out test set for model evaluation.

Figure 2 shows a view of the complexity of this large and high-dimensional real-world dataset, as visualized by a t-SNE scatter plot (t-distributed Stochastic Neighbor Embedding[29]) that embeds a high-dimensional problem space in a low-dimensional similarity representation. The distance of clusters in this plot indicates high-dimensional separability. The interleaving of red and blue data points in Figure 2, respectively, representing the classes 'above' and 'below' heat outcomes, clearly shows that simple dimensionality reducers cannot easily separate classes directly from these data. This motivates the use of algorithms that can find the more complex nonlinear dependencies. To separate the effect of using knowledge of different parameter types, such as parameters known before and during the oxygen blow and the effect of expected dependencies between original parameters, we partition the data for our experiments into feature groups I, II, and III (Table I).

1. Feature group I: features collected before the process

Out of the 33 single-value parameters (features) collected for each heat, 20 features describe the heat before the blow is started, while the remaining features are collected during the blow of the heat. These pre-blow features include data such as the element composition of the hot metal (with essential elements such as carbon, silicon, and manganese). Other significant pre-blow features are the amount of hot metal and scrap used for the heat; the temperature of the hot metal; the waiting time between subsequent heats (where both the material and the equipment cool off); the final heat tap temperature of the previous heat; furnace ID and lance ID; the lifetime in number of heats of the currently used oxygen lance; the number of heats processed so far in the current furnace body since last body relining; and the duration between the sample time of the hot metal's temperature and the start time of the blow. We evaluate the prediction accuracy of these pre-blow features in isolation since some information that determines the results is already available before the blow.

2. Feature group II: features collected during the process

Out of the 33 single-value features collected for each heat, 13 features are collected during the blow. These features include the amount of nitrogen and argon used; the total amount of oxygen blown; the accumulated amount of material additions to the heat during the blow (such as dolomite, ferrosilicon, and lime); duration of heat and blow periods; the time point of the final measurements of steel temperature; and analysis of the hot metal sample.

In this feature group, we also include aggregations of time-series that describe the actual blow execution. Five time-series were collected during the blow including outdoor temperature; carbon monoxide (CO) and carbon dioxide (CO_2) levels of the blow exhaust gas; lance cooling water temperature; and lance movements. The five time-series are aggregated into 64 statistical descriptors. More specifically, for outdoor and cooling water temperatures, we calculate the mean, standard deviation, and the maximum and minimum values, resulting in 8 features. For the CO and CO_2 time-series of the exhaust gas composition, we calculate the mean, standard deviation, and the integral, resulting in six features. The lance movement is treated in more detail. Depending on the process phase, the distance between the oxygen lance and the liquid metal surface varies following a preset movement profile through the

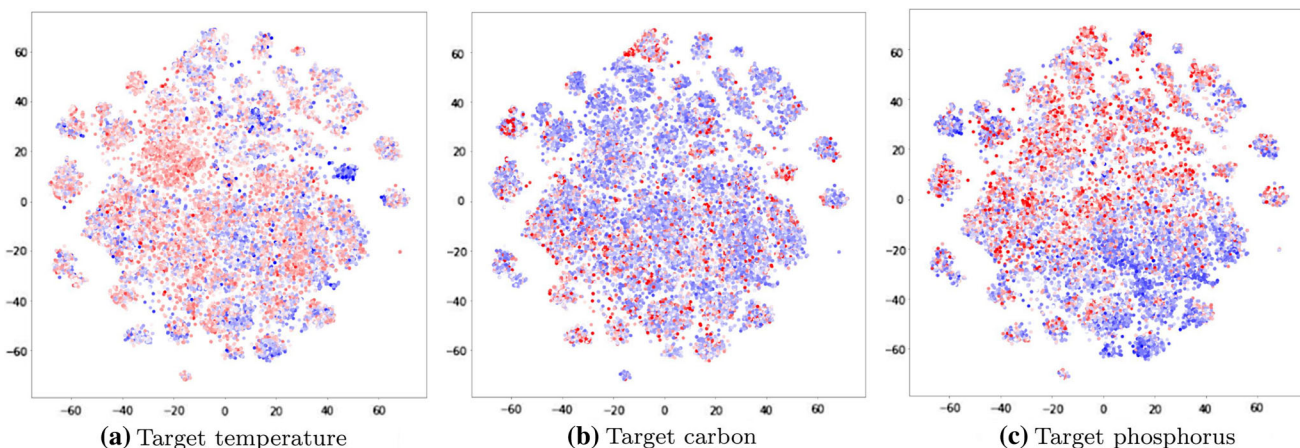**(a)** Target temperature   **(b)** Target carbon   **(c)** Target phosphorus

Fig. 2—Our multidimensional dataset, reduced into a two-dimensional space by the t-SNE algorithm. Higher than average values are represented in red (brighter), while lower than average values are represented in blue (darker). There are some clusters of samples with similar values, but the overall picture is that there is no consistency in how the samples are distributed. Hence, predicting the outcome of a heat targeting (a) temperature, (b) carbon, and (c) phosphorus would require methods that can capture complex dependencies in the feature space (Color figure online).

Table I. Feature Groups (A Total of 114 Features)		
Feature Group	Description	Number of Features
I	Features collected before BOF process	20
II	Features collected during BOF process + aggregated time-series features	77 (13 + 64)
III	Thermodynamic calculation features	17

process. The operator may, for optimization purposes, adjust this movement slightly, deviating from the preset profile. In an early stage of the process, the lance position is relatively high over the steel bath. In the middle stage, the lance position is lowered to accelerate the decarburization, while at an end stage, the position is raised again. Since the lance height affects the decarburization rate in the heat over time, we assume that the lance height influences the heat composition in a way that affects the process targets, and that the influence varies during the process. Therefore, we create features based on evenly spaced time slots (bins) of the lance movement program. Each such bin is described by features aggregating the mean, standard deviation, maximum, minimum, and integral. Each lance sequence is divided into 10 times 144-second bins, with the first bin starting from the blow start time. In total, this lance program aggregation results in 50 features. The reason for elaborating on the lance program this way is that lance movements directly influence the oxidation during the blow, and that seemingly insignificant operator lance adjustments may matter much for the final result. This feature group II is evaluated in combination with feature group I (*i.e.*, feature group combination I + II, refer to Section V–B), and the significant difference in prediction accuracy between them is expected to reveal the effect of during-process features.

3. *Feature group III: thermodynamic calculation data*

We assume that a simplified thermodynamic and kinetic calculation of the thermal balance of the system would capture properties of the data that could give a significant prediction improvement compared with only using the collected features. Excluding the possible influence of mechanical stimulus to the heat and assuming constant pressure, the thermal energy balance of a process can be approximated by Eq. [1].

$$H_{out} = H_{in} - H_{loss} \qquad [1]$$

H_{out} is the final enthalpy of the system, H_{in} is the enthalpy of the in-going raw materials, and H_{loss} is the enthalpy losses from mass transfer, radiation, and conduction. If the reference level of enthalpy is defined as the pure elements at room temperature, the terms of this thermal balance equation can be approximated by combining some of the logged features. The H_{out} can be described as the sum of the thermal energy of the total outgoing mass and reaction product enthalpies. The former is taken as proportional to the product of the total mass and the final temperature, and the latter is taken as proportional to the respective masses of

reacting elements. H_{in} can be described as the thermal energy contained in the hot metal, here taken as proportional to the product of hot metal temperature and hot metal mass. H_{loss} is the term that is mostly difficult to simplify, as the energy flux is described by a differential equation containing nonlinear, time-dependent temperature terms. In this case, the time steps between the tapping of the last heat and blow start, between hot metal temperature reading and blow start, and between blow start and final temperature reading have been chosen to describe the time-dependent losses. Now, leaving the term containing the final temperature aside and dividing both sides of the equation by the total mass produces the 17 new features that describe the process. This feature group is evaluated both in isolation and in combination with the feature group combination I + II (*i.e.*, feature group combination I + II + III, refer to Section V–B), and the significant difference in prediction accuracy is expected to reveal the information content in these informed expert-designed features based on current domain knowledge.

B. *Hyperparameter Tuning*

We perform a grid search using a range of hyperparameters for the chosen ML methods using 10-fold cross-validation. Table V in Appendix A lists the parameters used for combinations in the grid search and also marks the parameters that achieved the best result. The chosen methods are ANN, SVR, XGBoost, random forest (RF), k-nearest neighbors (KNN), and decision tree (DT). With linear regression, we use the default settings provided by scikit-learn. For SVR, we use a radial basis function kernel. The scoring metric used is 'mean_squared_error' for the grid search with SVR, XGB, random forest, KNN, decision tree, as well as the loss function for ANN. The mean-squared error regression loss (MSE) calculates the average of squared differences between the predicted and actual values and is often used in a regression problem.

With ANN, we evaluate different combinations of neurons in the hidden layer with input and output layers to find the best-fit combination. We use LeakyReLU (Rectified Linear Units) since their performance is better than other activation functions in the hidden layer. For example, we find the optimal hyperparameters for target temperature to be as follows. The ANN is configured with three hidden layers with 128, 512, 64 units, respectively, set up with the Python Keras package.[30] To avoid ANN overfitting, we add a dropout layer with rate 0.5 after the output of each hidden layer. In

addition, it is trained with weight decay using the most common type of regularization L2 with 0.001, which is a typically cited value. The model output is given by a linear combination of the output signals of the neurons in the last hidden layer. In order to train the model to the given data, we minimize the mean square error between the output and the observation, using Adam optimization with a learning rate of 0.0005. The model is trained for 250 epochs using a batch size of 512.

C. Target Evaluation

The end-point steel temperature (T), steel carbon content (pct C), and steel phosphorus content (pct P) are the three target variables to predict.

In order to evaluate the performance of using these data with commonly used ML methods, we apply seven different machine learning algorithms to predict the hit rate (HR) of T, pct C, and pct P. We define the hit rate (HR) as

$$HR = \frac{N(|\text{Prediction} - \text{RealValue}| < \varepsilon)}{N(\text{Heats})} \quad [2]$$

where $\varepsilon = 15\,°C$ in temperature (T) prediction, $\varepsilon = 0.02$ pct in carbon (pct C) prediction, and $\varepsilon = 0.003$ pct in phosphorus (pct P) prediction.

D. ML Algorithms Evaluation

The seven ML algorithms used are artificial neural networks (ANN),[31] support vector regression (SVR),[32] extreme gradient boosting (XGB),[33] linear regression (LR),[34] random forest (RF),[35] k-nearest neighbors (KNN),[36] and decision tree (DT).[37] For a baseline comparison, we use a mean-guess, which is calculated by

$$\frac{N|\text{mean}(\text{RealValue\_Train}) - \text{RealValue\_Test}| < \varepsilon}{N(\text{Heats}(\text{Test}))} \quad [3]$$
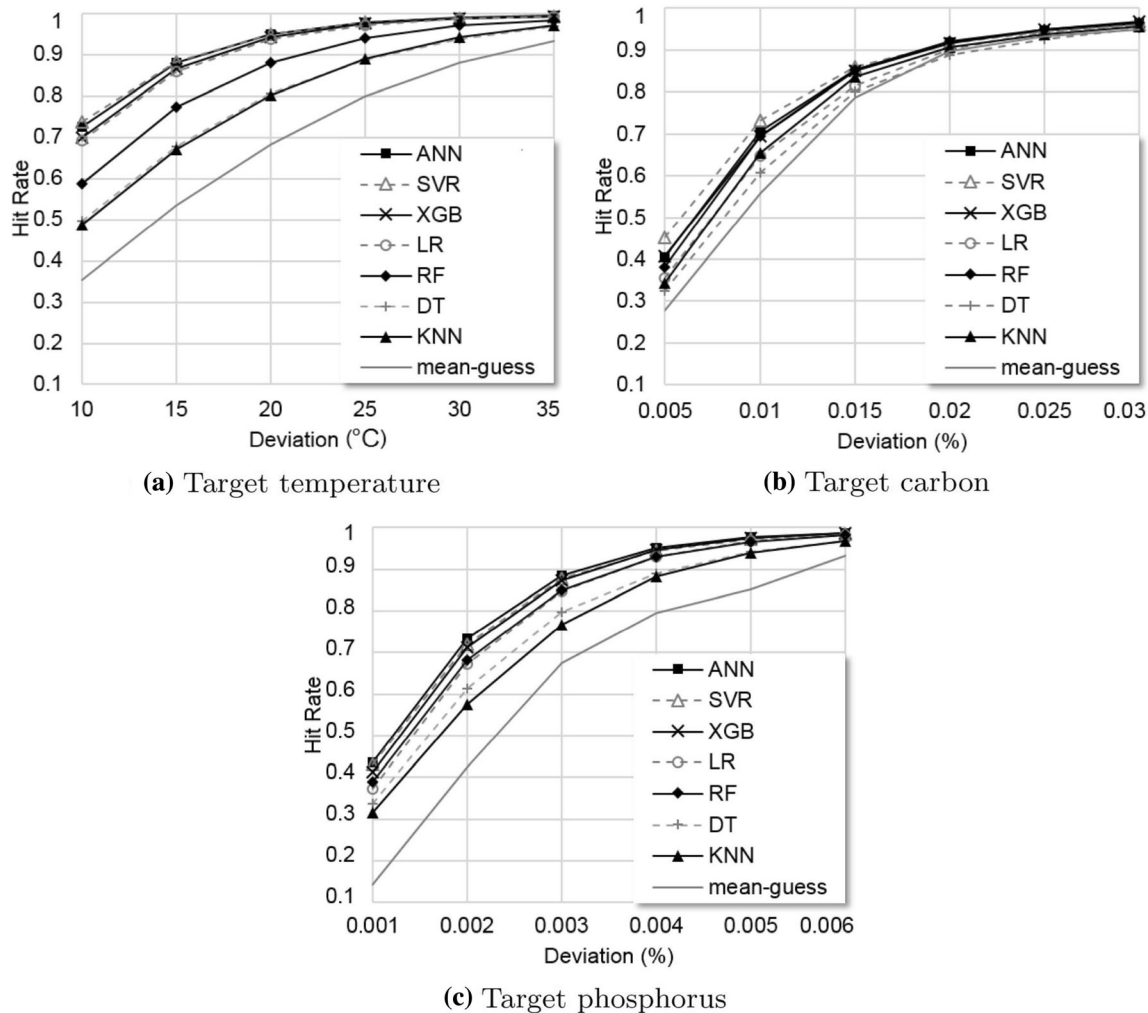

(a) Target temperature

(b) Target carbon

(c) Target phosphorus

Fig. 3—Hit rates of different ML algorithms targeting (*a*) temperature, (*b*) carbon, and (*c*) phosphorus.

Table II. Hit Rates for Whole Dataset for Target Temperature (T, ± 15 °C), Carbon (C, ± 0.02 pct), and Phosphorus (P, ± 0.003 pct)

Target	ANN	SVR	XGB	LR	RF	KNN	DT	mean-guess
T	0.88	0.88	0.87	0.86	0.77	0.67	0.68	0.53
C	0.92	0.92	0.92	0.9	0.92	0.91	0.89	0.9
P	0.89	0.88	0.87	0.85	0.85	0.77	0.8	0.68

Table III. Hit Rates of Different Combinations of Feature Groups: I + II + III (114 Features), I + II (97 Features), III (17 Features), and I (20 Features) for the Targets Temperature (T, ± 15 °C), Carbon (C, ± 0.02 pct), and Phosphorus (P, ± 0.003 pct)

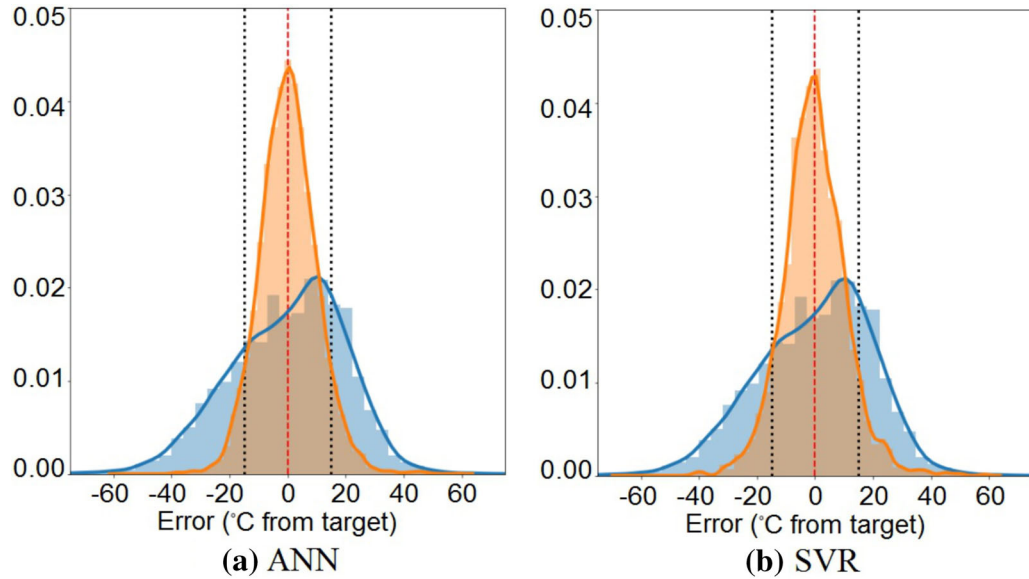| Feature Group Combination	Target	ANN	SVR	LR	Mean-Guess
Pre-process + During-process +	T	**0.88**	0.88	0.86	0.53
Thermodynamic features	C	0.92	0.92	0.9	0.9
(Feature Groups I + II + III)	P	**0.89**	0.88	0.85	0.68
Without Thermodynamic-	T	**0.86**	0.86	0.84	0.53
Calculation Features	C	0.92	0.91	0.9	0.9
(Feature Groups I + II)	P	**0.85**	0.84	0.8	0.68
Only Thermodynamic-	T	0.81	0.8	0.79	0.53
Calculation Features	C	0.9	0.89	0.9	0.9
(Feature Group III)	P	0.84	0.81	0.79	0.68
Only Pre-process	T	0.68	0.67	0.66	0.53
Features	C	0.9	0.89	0.89	0.9
(Feature Group I)	P	0.73	0.73	0.71	0.68

Fig. 4—The distribution of the predictive error for the two best performing methods, (*a*) ANN and (*b*) SVR, for target temperature. The error distribution from a mean-guess approach is added in the background for comparison (Color figure online).

V. RESULTS

For evaluation, we perform two main experiments. The first experiment compares the hit rate of alternative ML algorithms and a mean-guess baseline (Figure 3, Table II). For a subset of algorithms, we further evaluate the hit rates of the combinations of different feature groups (Table 3). Moreover, the prediction error distribution of the two algorithms, ANN and SVR, are illustrated in Figure 4. The second experiment further examines the use of informed features based on thermodynamic calculations (Figure 5, Table III).
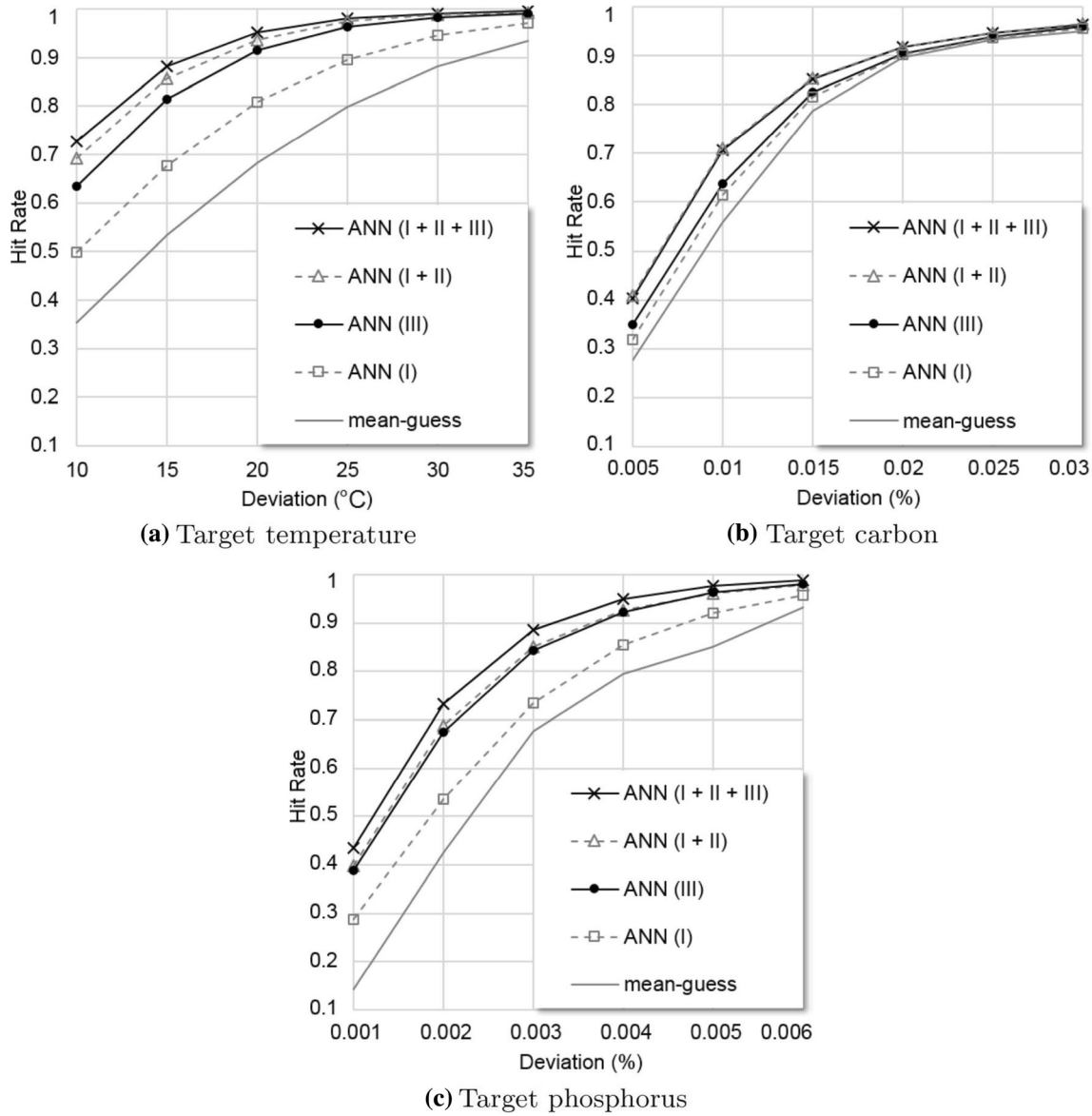
(a) Target temperature

(b) Target carbon

(c) Target phosphorus

Fig. 5—Hit rates of ANN prediction for target (*a*) temperature, (*b*) carbon, and (*c*) phosphorus. Refer to Sections IV–A–1, IV–A–2, and IV–A–3 for more details of the different feature groups (I, II, III).

A. *Experiment 1: Prediction Performance Comparison*

In order to compare seven chosen ML algorithms' performance on a large dataset we have, we train all the algorithms with 10-fold cross-validation based on the 80/10/10 data splits with 114 features and 9708 heats. For each target, a separate model is trained for each of the algorithms. Table II lists the hit rates for each method and target combination, as well as the mean-guess baseline. We conclude that it is possible to achieve high hit rates when using standard ML algorithms, also with a full-sized and rich dataset that has a mixture of heats with a wide range of process uncertainty captured by the data. This shows that most of these algorithms are capable of capturing the dependencies, even for more complex data with the variance in a full

production dataset. ANN and SVR have higher hit rates than other algorithms and Figure 3 plots the hit rates for each algorithm over different target deviation ranges.

In addition, we compare the distribution of the target error for the top-two algorithms (Figure 4) to understand more about the prediction uncertainty. In this figure, the predictive errors of the top-two algorithms are shown as the difference between the predicted and actual temperatures. The distribution of the target error for these methods is shown in a brighter distribution plot (in orange). The distribution of the differences between the target temperature and the measured outcome during the production is shown as a darker distribution plot (in blue). The dashed line (in red) shows the temperature target center, and two dotted

lines (in black) mark the temperature target range. The hit rate values of Table II are the integral of the area between these two dotted lines.

B. *Experiment 2: The Effect of Using Thermodynamic Features and Features During the process*

To understand how informed features such as features based on thermodynamic calculations influence the hit rate, we evaluate combinations of the feature groups. The feature group combinations I; III; I + II; and I + II + III (refer to Sections IV−A−1, IV−A−2, and IV−A−3) compare hit rates when using or not using the thermodynamic calculations and features during the process. The hit rates of the two commonly used and best performing algorithms, ANN and SVR, are further described in Table III. The result of a linear regression model, which is much simpler and easier to interpret, as well as the selected baseline of mean-guessing are also reported in this table to contrast the results of the two best models.

Figure 5 plots the hit rates of target temperature, carbon, and phosphorus. In general, ANN and SVR perform better than other algorithms. Using all features to predict end-point temperature, the hit rates of ANN and SVR are 88 pct, XGB is 87 pct, and linear regression is 86 pct.

When excluding the thermodynamic-based features (compare I + II + III and I + II), the hit rate is lower for temperature (88 *vs*. 86 pct) and phosphorus (89 *vs*. 85 pct). However, using only the thermodynamic-based features, we still achieve a high hit rate around 81 pctpct for temperature while the baseline is 53 pct. Thus, these 17 features capture important information for the temperature and phosphorus predictions. Furthermore, the thermodynamic calculation seems to simplify the prediction problem towards a more linear dependency between data and prediction, since all three algorithms including linear regression (LR) have a hit rate around 80 pct when using only 17 thermodynamic features of target temperature. The differences in absolute hit rate values seem small, so we test for significance.

To investigate the significance of these hit rates, we perform an analysis of variance (ANOVA) on ANN with 10-run hit rates for the targets we used. There are statistically significant differences between different feature group combinations for temperature as determined by repeated measures ANOVA ($F(3,27) = 1230.88$, $p < 0.00001$). A Tukey *post hoc* test[38] reveals that the hit rate is statistically significantly higher at 'all features (I + II + III)' or 'with thermodynamic data' (0.882 ± 0.0085) compared with 'without thermodynamic data (I + II)' (0.857 ± 0.0076), 'only thermodynamic data (III)' (0.815 ± 0.015), and 'only pre-process data (I)' (0.677 ± 0.0136). The condition 'without thermodynamic data (I + II)' (or 'pre- and during-process data (I + II)') is significantly higher compared with the 'only pre-process data (I).' In general, four feature group combinations were found to be significantly different from each other for target temperature in Figure 5(a).

Moreover, there are statistically significant differences between different feature group combinations for phosphorus ($\pm$ 0.003 pct) as determined by a repeated measures ANOVA ($F(3,27) = 818.37$, $p < .00001$). The condition 'all features (I + II + III)' (0.886 ± 0.0133) is significantly higher than 'without thermodynamic data (I + II)' (0.851 ± 0.0146), 'only thermodynamic data (III)' (0.843 ± 0.0133), and 'only pre-process data (I)' (0.734 ± 0.0121). The condition 'without thermodynamic data (I + II)' (or 'pre- and during-process data (I + II)') is significantly higher compared with the 'only pre-process data (I).' However, 'without thermodynamic data (I + II)' and 'only thermodynamic data (III)' are not significantly different ($p = 0.494$) for phosphorus (refer to Figure 5(c)).

Carbon ($\pm$ 0.02 pct) shows a different result compared with temperature and phosphorus based on repeated measures ANOVA ($F(3,27) = 17.97$, $p < .00001$) as in Figure 5(b). A Tukey *post hoc* test shows that the conditions 'all features (I + II + III)' and 'without thermodynamic data (I + II)' are significantly higher compared with 'only thermodynamic data (III)' and 'only pre-process data (I).' The conditions 'all features (I + II + III)' (0.917 ± 0.0082) and 'without thermodynamic data (I + II)' (0.917 ± 0.0081) are not significantly different ($p = .999$) from each other. In addition, 'only thermodynamic data (III)' (0.904 ± 0.0111) and 'only pre-process data (I)' (0.902 ± 0.0109) are not significantly different ($p = .96$) from each other. However, the condition 'without thermodynamic data (I + II)' (or 'pre- and during-process data (I + II)') is significantly higher compared with the 'only pre-process data (I).'

VI. DISCUSSION

We find from the results of Experiment 1 (Section V–A) that certain ML algorithms perform better than others. The results of all three targets (end-point temperature, end-point content of carbon, and phosphorus) are shown in Tables II and III. Specifically, we perform 10 independent runs to examine whether the hit rates from various algorithms are significantly different from each other with a Tukey *post hoc* test.[38] The results show that the prediction hit rates of ANN and SVR are not significantly different from each other and both are significantly better than other machine learning techniques for all three targets.

In Experiment 2 (Section V–B), with end-point temperature and phosphorus, we find that thermodynamic data (III) improve the prediction performance compared with only running with pre- and during-process features (I + II) (compare I + II *vs*. I + II + III). For all targets, we find that during-process features (II) improve the prediction performance compared with only using the pre-process features (I) (compare I *vs*. I + II). Although we find carbon not showing significant difference between 'all features (I + II + III)' and 'without thermodynamic data (I + II),' this is theoretically explainable, since the 17 calculated features primarily describe the thermal balance. That is,

phosphorus and temperature have strong relationships with the 17 thermodynamic features, while carbon, as it continuously leaves the system, does not. In the case of carbon (Figure 3(b)), mean-guessing showed even higher hit rates than decision tree (DT) for deviations more than 0.02 pct. In order to fully understand the results that we achieved, we plan to work on explaining the differences among the target predictions.

Overall, there is a significant difference between the best and lowest performing ML algorithms and, thus, it is beneficial to evaluate many algorithms when conducting a similar study as presented. The decision tree (DT) algorithm was found to be the worst performing algorithm for carbon, KNN for phosphorus, and both decision tree (DT) and KNN for temperature (not significantly different) in this set of algorithms. This came as a surprise to us since the decision tree algorithm is often recommended to be used in industrial settings, since it is easy to use and understand and is also claimed to perform tolerably well.[39,40] The explanatory properties of decision trees are useful in an applied setting, so again, it is needed to explain classifications and predictions, and we aim to pursue this in subsequent work.

The performance of the ANN and SVR methods are compared to how well the process is controlled today in the actual production unit and how much the measured value deviates from the desired target, which is shown in Figure 4. The error of the predictions is both smaller and less volatile compared to the currently observed differences. This would suggest that the current process could be improved by incorporating the presented ML methods into the process as a decision support system for the operators.

One common claim about deep learning is that there is no need for feature extraction since this would be conducted automatically by the algorithm. However, in this paper we show that in our particular case, this assumption does not hold. Instead, Table III (bold text) shows that the evaluated algorithms perform better at target temperature and phosphorus with a few features added that are engineered by experts. Thus, in this particular case, it is more beneficial to add features that are engineered by experts than to simply use the gathered data. Hence, in this particular case which considers a very dynamic and complex system, it is important to emphasize that feature engineering can be used to achieve a significant improvement over only using the raw data.

A comparison with related studies (Section II) shows that most related studies use less data or address different types of problems. Comparable studies of BOF cut-off prediction are listed, with their parameters, in Table IV. The basic problem is the same, but there are still differences in algorithms and methods, target error range, number of heats and features, final production, procedure, and also the aim of the process. In particular, many studies use a limited number of features (or do not specify), likely due to limitations on what data the authors were able to collect or considered prominent. Using a production-size dataset, we find our results to be competitive and useful, considering the number of heats, features, and the combination of targets used. Our study differs in that it considers all types of heats and incorporates the prediction of all three significant targets used in actual production.

Our initial attempt to identify a ranked list of influential factors out of the 114 features has resulted in a list that is not immediately obvious. It was expected that top factors influencing the prediction hit rate would correspond to the factors that operators expect to influence the BOF process. We conjecture that a ranked list of factors corresponds instead to what influences the learned model based on available data, rather than a model that captures the factors that influence the BOF outcome. Our current explanation is that the ML model is based on the observable data, rather than the direct BOF process data. The collected data represent the overall behavior of the entire production unit, so the ML model captures just the overall behavior. The production unit that is actually measured includes the existing temperature prediction support system, which the operators use for process treatment during the process and for the operators themselves. Thus, the current production unit with all its constituent parts is capable of mitigating a deviating process, in many instances successfully bringing it back to a good process state. The result ends up being highly dependent on the operators' level of experience, in terms of long-term training in tacit skills and in transferable process knowledge.

Table IV. Other Prominent Methods, Configuration, Deviations, and Hit Rates (Pct) for Temperature (T), Carbon (C), and Phosphorus (P) on BOF

Ref.	T(°C)	C (Pct)	P(Pct)	Method	Heat#	Feature#	Misc.
Ours	±15	±0.02	±0.003	ANN	9708	114	
	88	92	89				
Gao et al.[41]	±15	±0.005	—	SVR	300	?	
	96	94	—				
Han et al.[27]	±15	±0.05	—	RBF-NN	60	?	
	100	92	—				
He et al.[25]	—	—	±0.005	ANN	1978	7	used pca
	—	—	93.3				
Wang et al.[18]	—	—	±0.006	ANN	2000	?	used k-means
	—	—	96.4				

Our conclusion is that we cannot model the BOF process directly given the data collected, but that we can model the overall production system. It seems to be common that data-based process models capture an "outer system" only. Another example of this is an industrial ammonia scrubber. There is limited knowledge about the reacting components, and the system is known to be nonlinear and time-dependent. This is fundamentally based on chemical equilibrium, reaction kinetics, and thermodynamics.[42] Thus, reacting components are unknown in the actual process. From what we find from steel-making processes, the collected data mainly reflect what has occurred as a result of the optimizing operations by the operators during the process (*e.g.*, amount of initial scrap, updating the lance position). The relationship of input features and the output of the process is not only complicated by heat and mass transfers and chemical reactions, but also by the noise of the dataset. Moreover, there is a limitation on capturing solely the behavior of the internal reaction simply from the collected dataset, which necessitates separating the internal reactions from external phenomenon as seen in the collected data.

Given that our process model captures an "outer system," process experts are able to explain our ranked list of influential features with rather high accuracy. Assuming that these features are influential on the prediction of the overall system rather than the BOF process alone, these features would represent those that are not yet fully controlled by the current control system. If so, the feature list would be an actionable list of features to control to improve the current process control. Such knowledge would be of great use, and thus, we plan to further investigate the features influencing this process in order to understand those that have larger effects on the outcome of the model's predictions.

We recognize some limitations of our study. Most obviously, the data are collected from only one BOF furnace. We have collected data from two more BOF furnaces, but have not yet evaluated that data in the same way, so we do not yet know how generalizable our results are. In addition, our data cleaning in collaboration with steel experts may still have introduced some bias, even though we intended to use a dataset with all types of heat executions. Moreover, for the aggregated time-series, various other aggregation methods exist which may better capture features relevant to BOF target prediction. In fact, it may be that binning might not be a suitable approach for capturing relevant process properties. There are multiple approaches to BOF cut-off prediction which we plan to use in future work, in particular, the use of multivariate time-series models.

From an industrial implementation point of view, there is an immediate value and potential in improved optimization possibilities from detecting previously unknown relationships between features. Also, within a relatively short time frame, a machine learning-based prediction model with high accuracy could be installed as a real-time application in parallel to existing process systems used by the operators, continuously predicting the process end-points based on the log of actual events and the planned events during the heat. Such a tool would be especially valuable for operators with limited experience and when the process is running under unstable conditions.

VII. SUGGESTIONS FOR FUTURE WORK

The potential energy, material, and environmental gains from accurate cut-off predictions are clear. We find that there had been various attempts to improve performance in the past, but they seem limited due to a lack of details and transparency and were limited by the quantities of data and features. It seems that authors have determined the influencing factors based on a weight method[18] and *a priori* reasoning[21] and then applied this to neural networks or other selected methods. Moreover, there is no standard for the error ranges of targets, probably due to different steel production procedures for the different cases reported. The variation can be seen in Table IV. In this paper, we selected the deviation for phosphorus of 0.003 pct and get an 89 pct hit rate with ANN. However, if we selected a deviation for phosphorus of 0.005 pct, we would get a 98 pct hit rate with ANN which is much higher than what others have reported. Thus, we suggest a general and standardized protocol of configuration and reporting the results within the BOF processing domain in different steel production procedures. In addition, it would motivate more researchers if there were public benchmarking datasets for steel manufacturing that everyone can use.

Domain experts and researchers are especially interested in discovering the influencing factors and understanding the industrial process as a whole. As our next step, we plan to use SHAP[43] and other statistical analyses with this real-world dataset to find the influencing factors and enhance the interpretation of BOF cut-off prediction models.

Yet, to our knowledge, there is no dynamic sequential model that can predict the targets at any given time during the process, for instance using deep learning-based time-series modeling. Besides, it is even more challenging to determine the influencing factors and automate the parameters in real time. It is also necessary to generalize the model and resolve the uncertainty issues that are caused by sources such as measuring devices, modeling error, human error, and inner *vs.* outer systems. Ultimately, there is a need for a closed loop system that optimizes itself based on a time-series.

VIII. CONCLUSION

In this paper, we examine the prediction accuracy of a number of standard machine learning algorithms using a large production dataset with the aim of predicting the BOF cut-off point. We achieve high and robust accuracy for temperature, carbon, and phosphorus prediction while using significantly more and diverse production data compared to previous studies. We also engineer

features based on thermodynamic approaches in collaboration with domain experts. Furthermore, we address some limitations of the previous research, such as the dataset size and data selection, and evaluate all of the most commonly used machine learning approaches. We find that algorithms based on neural networks and support vector machines perform better than other standard machine learning techniques, similar to what has been described in existing research. Moreover, we show that it is possible to have the same performance even with a full and more complex production dataset, in spite of the variance in data caused by realistic process uncertainty. The BOF process targets can still be predicted with high accuracy.

The engineered features enable further improvements in prediction accuracy. For this particular case, we find that it is beneficial to use a number of well-crafted, domain-specific, and informative features. This results in more process knowledge for standard ML algorithms to learn from, compared to merely using all available data to train standard ML algorithms. Although it is challenging to estimate the balance of mass and energy of a BOF heat using sensor readings and implicit variables, our results support the importance of feature engineering and the combination of engineered features with the raw data collected.

ACKNOWLEDGMENTS

OPEN ACCESS

APPENDIX A: HYPERPARAMETERS USED IN GRID SEARCH

See Table V.

Table V. Hyperparameters and the Best Parameters for the Targets, Temperature (T), Carbon (C), and Phosphorus (P)

Method	Parameters	T (Best)	C (Best)	P (Best)	Tested Parameters
ANN	learning_rate	0.0005	0.0005	0.001	0.001, 0.0005, 0.0001
	layer_sizes	[128, 512, 64]	[64, 128, 16]	[128, 512, 64]	[512, 1024, 128],[128, 512, 64],[64, 128, 16]
	drop_out	0.5	0.5	0.5	0, 0.2, 0.5
	batch_sizes	512	256	512	128, 256, 512
SVR	epsilon	0.05	0.1	0.1	0.001, 0.005, 0.01, 0.05, 0.1
	C	10	50	10	1, 10, 50, 100
	gamma	0.001	0.001	0.001	0.0001, 0.001, 0.01, 0.1
XGBoost	n_estimators	700	300	500	100, 300, 500, 700
	learning_rate	0.1	0.1	0.1	0.01, 0.5, 0.1
Random Forest (RF)	n_estimators	500	500	500	100, 300, 500
	min_samples_leaf	1	5	1	1, 2, 5
	min_samples_split	6	6	6	4, 6, 8
	max_depth	20	20	30	10, 20, 30
K-Nearest Neighbor (KNN)	n_neighbors	25	50	25	3, 5, 10, 25, 50
Decision Tree (DT)	min_samples_leaf	15	20	15	1, 5, 10, 15, 20
	min_samples_split	200	200	200	2, 10, 50, 100, 200
	max_depth	20	5	20	5, 10, 20, 30

REFERENCES

1. B.M. Dyrdal: Predictive Monitoring of Basic Oxygen Steel Refining. Master's thesis, Norwegian University of Science and Technology, 2018.
2. L. Cao, Y. Wang, Q. Liu, and X. Feng: *ISIJ Int.*, 2018, vol. 58 (4), pp. 573–84.
3. H.. Wang, A. Xu, L. Ai, N. Tian, and X. Du: *ISIJ Int.*, 2012, vol. 52 (1), pp. 80–86.
4. K. Feng, H..B. Wang, A..J. Xu, and D.F. He: *Int. J. Miner. Metall. Mater.*, 2013, vol. 20 (12), pp. 1148–54.
5. K. Feng, D. He, A.. Xu, and H. Wang: *Steel Res. Int.*, 2016, vol. 87 (1), pp. 79–86.
6. X. Wang, J. Xing, J. Dong, and Z. Wang: in *2017 36th Chinese Control Conference (CCC)*, 2017, pp. 9708–13.
7. Y. Cheng, J. Xing, J. Dong, Z. Wang, and X. Wang: in *2018 Eighth International Conference on Information Science and Technology (ICIST)*, 2018, pp. 472–77.
8. X. Wang, M. Han, and J. Wang: *Eng. Appl. Artif. Intell.*, 2010, vol. 23 (6), pp. 1012–18.
9. Z. Yu-xian, L. Tong, and W. Jian-hui: in *2013 25th Chinese Control and Decision Conference (CCDC)*, 2013, pp. 1900–903.
10. S. Klanke, M. Löpke, N. Uebber, H.J. Odenthal, J. Van Poucke, and A. Van Yperen-De Deyne: *Association for Iron & Steel Technology 2017 Proceedings*, 2017, vol. 2, pp. 1307–13.
11. C. Liu, L. Tang, J. Liu, and Z. Tang: *IEEE Trans. Autom. Sci. Eng.*, 2019, vol. 16 (3), pp. 1097–1109.
12. J. Schlüter, H.J. Odenthal, N. Uebber, H. Blom, and K. Morik: in *Association for Iron & Steel Technology Conference, Pittsburgh, USA*, 2013, vol. 6.
13. C. Gao, M. Shen, X. Liu, L. Wang, and M. Chen: *Trans. Indian Inst. Met.*, 2019, vol. 72 (1), pp. 257–70.
14. I.J. Cox, R.W. Lewis, R.S. Ransing, H. Laszczewski, and G. Berni: *J. Mater. Process. Technol.*, 2002, vol. 120 (1), pp. 310–15.
15. N. Rajesh, M.R. Khare, and S.K. Pabi: *Mater. Res.*, 2010, vol. 13, pp. 15–19.
16. A. Das, J. Maiti, and R.N. Banerjee: *Exp. Syst. Appl.*, 2010, vol. 37 (2), pp. 1075–85.
17. C. Bing-yao, Z. Hui, and Y. You-jun: *2011 International Conference on Mechatronic Science, Electric Engineering and Computer (MEC)*, 2011, pp. 2278–81.
18. H.B. Wang, A.J. Xu, L.X. Ai, and N.Y. Tian: *J. Iron Steel Res. Int.*, 2016, vol. 19 (1), pp. 11–16.
19. L. Qu, X. Zhang, and Y. Qu: *2012 24th Chinese Control and Decision Conference (CCDC)*, IEEE, Taiyuan, China, 2012, pp. 4110–13.
20. F. He, D.F. He, A.J. Xu, H.B. Wang, and N.Y. Tian: *J. Iron Steel Res. Int.*, 2014, vol. 21 (2), pp. 181–90.
21. D. Laha, Y. Ren, and P.N. Suganthan: *Exp. Syst. Appl.*, 2015, vol. 42 (10), pp. 4687–96.
22. M. Han and L. Jiang: *2010 International Conference on Intelligent Control and Information Processing*, 2010, pp. 388–93.
23. W. Li, X. Wang, and H. Wang: *Chem. Eng. Trans.*, 2016, vol. 51, pp. 475–80.
24. M.A. Viana Junior, C.A. Silva, and I.A. Silva: *REM Int. Eng. J.*, 2018, vol. 71 (4), pp. 587–92.
25. F. He and L. Zhang: *J. Process Control*, 2018, vol. 66, pp. 51–58.
26. P. Yuan, Z.Z. Mao, and F.l. Wang: *J. Iron Steel Res. Int.*, 2007, 14(2), 20–29.
27. M. Han and C. Liu: *Appl. Soft Comput.*, 2014, vol. 19, pp. 430–37.
28. Y. LeCun, Y. Bengio, and G. Hinton: *Nature*, 2015, vol. 521 (7553), pp. 436–44.
29. L. van der Maaten: *J. Mach. Learn. Res.*, 2014, vol. 15, pp. 3221–45.
30. F. Chollet. Keras. https://keras.io, 2015.
31. W.S. McCulloch and W. Pitts: *Bull. Math. Biophys.*, 1943, vol. 5, pp. 115–33.
32. H. Drucker, C.J.C. Burges, L. Kaufman, A. Smola, and V. Vapnik: *Proceedings of the 9th International Conference on Neural Information Processing Systems*, MIT Press, 1996, pp. 155–61.
33. T. Chen and C. Guestrin: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ACM, 2016, pp. 785–794.
34. F. Galton: *J. R. Anthropol.*, 1886, vol. 15, pp. 246–63.
35. L. Breiman: *Mach. Learn.*, 2001, vol. 45 (1), pp. 5–32.
36. N.S. Altman: *Am. Stat.*, 1992, vol. 46 (3), pp. 175–85.
37. J.R. Quinlan: *Mach. Learn.*, 1986, vol. 1, pp. 81–106.
38. J.W. Tukey: *Exploratory Data Analysis*, Addison-Wesley, Boston, 1977.
39. L. Rokach and O. Maimon: *Data Mining With Decision Trees: Theory and Applications*, 2nd ed., World Scientific Publishing Co.Inc, Hackensack, 2014.
40. H.S. Kang and C.H. Jun: *ISIJ Int.*, 2019, vol. 59 (9), pp. 1582–90.
41. C. Gao, M. Shen, X. Liu, L. Wang, and M. Chen: *Trans. Indian Inst.* Met. 2018.
42. J.P. Ylén: Measuring, modelling and controlling the pH value and the dynamic chemical state. Ph.D. thesis, Helsinki Univ. of Tech., Control Engineering Lab, 2001.
43. S.M. Lundberg and S.I. Lee: *Proceedings of the 31st International Conference on Neural Information Processing Systems*, Curran Associates Inc., 2017, pp. 4768–77.