



Deep learning for robust forecasting of hot metal silicon content in a blast furnace

Cinzia Giannetti¹ · Eugenio Borghini¹ · Alex Carr² · James Raleigh² · Ben Rackham²

Received: 2 February 2024 / Accepted: 5 February 2024
© The Author(s) 2024

Abstract

The hot metal silicon content is a key indicator of the thermal state in the blast furnace and it needs to be kept within a pre-defined range in order to ensure efficient operations. Effective monitoring of silicon content is challenging due to the harsh environment in the furnace and irregularly sampled measurements. Data-driven approaches have been proposed in the literature to predict silicon content using process data and overcome the sparsity of silicon content measurements. However, these approaches rely on the selection of hand-crafted features and ad hoc interpolation methods to deal with irregular sampling of the process variables, adding complexity to model training and optimisation, and requiring significant effort when tuning the model over time to keep it to the required level of accuracy. This paper proposes an improved framework for the prediction of silicon content using a novel deep learning approach based on Phased LSTM. The model has been trained using 3 years of data and validated over a 1-year period using a robust walk-forward validation method, therefore providing confidence in the model performance over time. The Phased LSTM model outperforms competing approaches due to its in-built ability to learn from event-based sequences and scalability for real-world deployments. This is the first time that Phased LSTM has been applied to real-world datasets and results suggest that the ability to learn from event-based data can be beneficial for the process industry where event-driven signals from multiple sensors are common.

Keywords Deep learning · Steel making · Blast furnace · Phased LSTM · Irregular sampling

1 Introduction

The blast furnace is crucial to the global production of steel due to its ability to quickly and efficiently reduce large quantities of iron ore into molten pig iron, referred to as hot metal, which is subsequently processed into steel. The blast furnace is a large chemical reactor that consumes iron ore, carbon-based reductant (in the form of coal and coke), and fluxes. Carbon-based reductants are used to heat the furnace's burden material to the extreme temperatures required by the process and to reduce the oxides in the iron ore. The efficiency of a blast furnace is measured by the amount of reductant required to produce a tonne of hot metal. As a result

of economic and environmental considerations, modern iron production using blast furnaces requires them to operate very close to their limits in terms of efficiency. However, maintaining stable process operations is challenging due to the large number of variables that can affect the thermal level of the furnace, including quality of burden material at tuyere level, oxygen enrichment, blast volume, blast temperature. Additionally, the large scale and hostile nature of the furnace prevents direct measurement of its internal state and makes optimisation difficult.

In a blast furnace, the silicon concentration of the hot metal produced is used to monitor the process, as it is a key indicator of the process's thermal state [1]. An accurate prediction of silicon content trend in hot metal can improve operational optimisation and thermal control of the blast furnace iron making process by supporting early identification of abnormal conditions and subsequent corrective actions [2]. The prediction of silicon content has been widely studied in the literature. Blast furnace processes exhibit large amount of lag, strong coupling, and high nonlinearity, and therefore it is extremely challenging to develop computationally

Cinzia Giannetti and Eugenio Borghini contributed equally to this work.

The manuscript has not been published elsewhere and is not under consideration by other journals.

✉ Cinzia Giannetti
c.giannetti@swansea.ac.uk

Extended author information available on the last page of the article

efficient models [3]. For this reason, data-driven approaches have been proposed to overcome the limitation of on-line measuring technologies and the difficulty in modelling the behaviour of silicon content with physical models [2]. Data-driven models are directly derived from the data, without requiring extensive knowledge of the underlying process [4]. As detailed in Section 2, several data-driven methods have been proposed for the prediction of silicon content, namely those based on neural networks [5–8], support vector machines [4, 9], and ensemble trees [10]. However, when processing large datasets, these methods are inefficient to train and require fine-tuning of parameters. While the above-mentioned studies have successfully demonstrated the potential of data-driven approaches to overcome limitations of physical models in prediction of silicon content, their applicability for real-time prediction and control is limited as they have been validated on relatively small datasets over short time periods. To account for the high variability of blast furnace processes it is necessary to evaluate and compare methods across a much longer timescale using robust validation techniques for time series prediction such as walk-forward validation [11, 12]. Additionally, in current approaches a lot of time is spent in applying ad hoc interpolation methods and feature extraction, hence further hindering their deployment in real-world applications.

Deep Learning (DL) refers to techniques for learning high-level features from data in a hierarchical manner using stacked layer-wise architectures. DL has been used for modelling and prediction of manufacturing processes in several contexts, including anomaly detection [8], monitoring of machine health [13], fault diagnosis [14, 15], machine speed prediction [16, 17] and event detection [18]. DL can significantly reduce the complexity of the training process as it does not require extensive feature engineering since the features are directly learned as part of the training process. A recent study has investigated the application of LSTM (Long Short-Term Memory) to modelling and prediction of hot metal silicon content [19]. LSTM are a type of recurrent neural network that are commonly used for time sequential data. Although the LSTM model proposed in [19] is shown to achieve good performance, the model was trained and validated on simulated data. Furthermore, one of the major challenges in applying recurrent neural network to the prediction of silicon content in blast furnace processes is the fact that data are sampled at irregular intervals, hence requiring the use of interpolation methods as a pre-processing step. In a recent development, a novel method called Phased LSTM has been proposed to overcome this limitation. The Phased LSTM model can handle event-driven asynchronously sampled input data, eliminating the need to apply interpolation methods for irregularly sampled data before the training process [20].

Motivated by these recent developments, this paper proposes a novel approach, based on Phased LSTM, for prediction of hot metal silicon content which allows to handle the irregularity of both the input and response variables sampled at asynchronous intervals. The proposed novel architecture uses a Phased LSTM layer as a feature extraction step for exogenous variables, which are then merged with the auto-regressing variables (e.g. previous values of silicon content) and passed to a series of drop-out layers for the final classification. The performance of this approach is compared against two established DL methods such as Convolutional Neural Network, Long Short-Term Memory (CNN-LSTM), Auto Encoders (AE), and a baseline shallow Feed Forward Neural Network (FFNN) using a walk-forward validation approach. The models are trained and validated using real-world data of a blast furnace acquired over a 4-year period, hence significantly improving earlier works which were validated on relatively small datasets. The main contribution of the paper are summarised as follows:

- A novel method for prediction of HM silicon content is proposed based on Phased LSTM, which allows to handle the irregular sampling of silicon content and exogenous variables without the need of ad hoc interpolation methods;
- The method is the first example of application of Phased LSTM to modelling a real-world time series forecasting problem with asynchronous sampling for a complex manufacturing process;
- This is the first study that compares several DL models for forecasting of silicon content concentration in the HM using robust validation over a long time scale.

This paper is structured as follows. Section 2 presents a review of the existing approaches to silicon content forecasting in the literature. The technical description of deep learning layers used to build subsequent models is introduced in Section 3. Section 4 is devoted to the description of the models, data and the pre-processing steps, together with the precise problem formulation and the specification of the validation strategy. Section 5 presents and discusses the results while Section 6 closes the paper.

2 Literature review

The prediction of silicon content for the next casts in the blast furnace is a well-researched topic. In [5] a soft sensor approach to predict silicon content has been proposed based on Feed Forward Neural Network (FFNN). A major limitation of this work is the small data sample used for training and test. Also, the model assumed stable operation and no

changes in the charged material. Chen et al [6] used a combination of FFNN and genetic programming to develop a model to predict the silicon content for the next cast. While the results show that the method achieved good performance, the test dataset consisted of only 70 samples. The paper also lacks details on the methodology for variable selection to be further considered for deployment in industrial applications.

Recent studies have validated the models on larger datasets and focused on the challenges of data interpolation that arise from irregular sampling and large time lags in key parameters. In a typical furnace the effect of input parameters can take several hours to propagate. From a data modelling perspective this implies a careful selection of hand-crafted lagged features and use of interpolation methods to harmonise inconsistencies in data detection periods [2, 7, 8].

In the work of Cui et al. [21], the authors applied a FFNN trained via a genetic algorithm. The data were interpolated to 1 h frequency using cubic splines and consisted of 1500 samples, 1000 used for training and 500 for testing. For each of the selected control variables, a weighted average of the last 6 time steps was computed, where the weights were set to be the absolute values of the correlation with the target variable. The best predictions were obtained by correcting the errors of the original neural network via a separate network (similar to a two-step boosting approach). The authors of [22] proposed to combine a priori clustering on the data (both hard and fuzzy C-means) with support vector regression on each cluster as the predictive model. The data amounted to 400 training samples and 100 test samples, measured at 1 h intervals. Additionally, 4 time delays were considered for each variable. Similarly as in the previous work, the authors adopted a two-step approach [2]. First, clustering was performed using fuzzy C-means to apply afterwards an exogenous nonlinear autoregressive model (NARX) to forecast the silicon content on each cluster. A neural network was employed as the nonlinear component of their NARX. The data totalled 4550 training samples, 975 validation samples and 975 test samples. They were collected at 30 min intervals and the PCHIP algorithm was employed to fill the missing values. The method combining the clustering with the neural networks achieved a very high accuracy. It should be noted that the clusters were determined from the hot metal temperature and the silicon content (the target variable). The authors found 8 clusters under this methodology, which is conceptually interesting since it suggests the existence of several different *operating modes* of the blast furnace. However, in order to forecast the silicon content at the plant, it would still be necessary to establish under which of these operating modes the blast furnace is working. In a real-world situation the clustering step to determine the operating modes cannot be performed as it requires the temperature measures and

silicon content, which would be unknown at the time of prediction.

In [7] a novel method to predict silicon concentration at different future horizons (from 3 to 8 h) is proposed. The method uses the Maximal Overlap Discrete Wavelet Packet Transform (MODWPT) algorithm as a feature extraction step, followed by a Nonlinear Autoregressive (NAR) network. The method is validated on a period of 30 days. The advantage of this method compared to other proposed methods is that silicon content can be predicted on a longer time horizon, hence could be potentially more useful in determining preventive actions, considering that long latency of the process (typically 6–8 h). However, in order to assess its suitability for industrial applications, evaluation of this approach on a longer time scale would be required. This is due to the fact that over time there will be variations in silicon content caused by new operating conditions and changes in the quality of the charged material, leading to model degradation. Furthermore, the method requires extensive analysis of the dataset to determine the time lags of the input variables, hence posing an additional burden on the analyst when the model needs to be updated to take into account the dynamic nature of blast furnace processes.

This drawback can be overcome by using deep learning models based on LSTM networks. These methods embed the feature extraction step within the training process and are able to learn temporal features without the need to specify the time lag. The authors of [19] developed an LSTM model for predicting the silicon content of the next cast on simulated data. The simulation comprised 648 training samples and 72 testing samples and was generated by adding Gaussian white noise to mathematical equations formulated in [23, 24]. The model achieved high accuracy, however, it was only validated on simulated data.

The consensus that emerges from several sources is that the accurate prediction of the silicon content for future casts is a challenging task and there are still many limitations, hindering adoption of these models in industry. Although several methods have been proposed, they have been evaluated on relatively small datasets. Furthermore, several methods necessitate to apply extensive feature engineering and variable selection steps. Therefore, this paper addresses these limitations by exploring the application of deep learning to prediction of silicon content.

3 Deep learning layers

This paper uses different deep learning layers for the prediction of silicon content. The main deep learning layers used are described below, while the network architecture is described in Section 4.2.

3.1 CNN

Convolutional Neuron Networks (CNNs) are a kind of feed-forward neural networks that are generally employed in the field of image and video recognition [25]. Their architecture allows to train *kernels* which are simply small matrices that get multiplied and averaged with all the patches in the image or video of matching size. The idea behind this design is that it allows to capture local features of the images, such as corners and basic shapes. In the context of the present paper and more generally, the study of time series, they are used to discover patterns in the temporal dimension, for instance, local maxima, minima or abrupt changes.

3.2 Autoencoders (AE)

Autoencoders (AEs) are a type of feed-forward network specifically designed to learn a lower-dimensional representation of the input in an unsupervised way. They are composed by an *encoder*, which reduces the dimension of the input returning the *code* and finally a *decoder* that attempts to reconstruct the output from the compressed representation. In this paper, the considered autoencoders will be embedded in a larger neural network that has as a final goal the prediction of the silicon content. The rationale is that they act as a kind of regulariser while reducing the dimensionality of the input at the same time.

3.3 Long Short-Term Memory (LSTM) networks

LSTM models are recurrent neural networks that have been successfully applied to time series forecasting in recent years. The main feature of the LSTM architecture is a mechanism that can forget nonrelevant details during the learning process, while retaining information of long-term dependencies. Essentially, an LSTM cell keeps both a long-term and a short-term state that are updated every time new information is available. The memory cell of a standard LSTM contains three control gates: input, output and forget gates. These gates determine how much of the previous state is retained or forgotten.

3.4 Phased LSTM

The Phased LSTM model (introduced by Neil et al. [20]) is an extension of the LSTM model, consisting of the addition of a trainable time gate to the LSTM cell, which controls the time at which the cell's inner states are updated through a parametrised oscillation. Typically, the frequency is set in such a way that the updates of the memory cell occur very sparsely during the training cycle. The time gate allows to save information for a longer period to deal with event-based triggers. The proposed modification of the baseline LSTM

architecture is particularly useful when the considered time windows are long (in the order of hundreds or more), to deal with asynchronous sampling rates, and for event-based time series [20], making it a particularly promising to handle the asynchronous measurement of the blast furnace.

4 Problem formulation, training and validation strategy

The goal of this paper is the accurate prediction of the silicon content for the next cast using past and current information on the blast furnace status. In a real-world situation, this is desirable due to the delayed measures that are typically obtained using laboratory analysis. However, due to the sparsity and the varying sampling rates of the input variables, the formulation of the problem is not entirely straightforward. Furthermore, although the blast furnace is tapped on average every 2 h, the periods between taps are variable and hence the frequency at which the target variable is sampled is irregular. Given the volume of the data, forcing a regular sampling rate through interpolation, as in recent publication [2], would have entailed working with far more inferred values for the target variable than actual measurements. Although, from an industrial perspective, it would be desirable to be able to make predictions over a longer time horizon, due to the scale and challenges presented in the dataset, the scope of this work is limited to prediction of next cast.

4.1 Problem formulation

The forecasting of the silicon content was approached in an *event-based* fashion. The *events* in this case are the hot metal casts. Based on the common consensus that the blast furnace process is a nonlinear dynamical system with delayed effects, the silicon content y_t at time t is predicted as:

$$y_t = F((y_{s-1}, t_{s-1}), \dots, (y_{s-d_y}, t_{s-d_y}), (u_{t-1}, \dots, u_{t-d_u})) \quad (1)$$

where F is a nonlinear function to fit, t_j indicates the time at which the corresponding cast y_j was measured, u_i are the explaining variables, and d_y , d_u are the user-selected delays for the cast and the exogenous variables, accordingly. In the proposed work the selected explaining variables for the past 6 h are included, taken every 10 min. On the other hand, the previous three silicon content values together with their timestamps are recorded. Thus, each data point consists of a block of size 36×21 (36 time steps for 21 variables) and another one of size 3×2 (3 time steps for the silicon content together with the corresponding timestamp).

4.2 Deep learning models

In this paper three DL models are trained and compared against a baseline Feed Forward Neural Network (FFNN). The first two models use respectively Convolutional Neural Network Long Short-Term Memory (CNN-LSTM) and Convolution Auto Encoders (AE) layers. These models were chosen because they have shown good performance in the literature to solve time series forecasting problems across different domains. The third model uses a recently proposed variant of LSTM, called Phased LSTM, which can handle forecasting of event-based time series [20]. This is the first study where these models have been evaluated for the task of prediction of HM silicon concentration in blast furnace processes over a long time period.

Figure 1 shows the common schema for all the three models. The network include a feature extraction module for the exogenous multivariate time series part of the data points (that is, the block of measurements corresponding to the last 6 h for the explaining variables). Then the results are merged with the autoregressive part (that is, the previous values of the silicon content together with its timestamps) and topped with dense layers that output the final prediction. The main difference between them lies in the way they extract features. The process of feature extraction is detailed below for each model:

- **CNN-LSTM:** Two stacked 1-dimensional convolutional filters are applied, each followed by a max pooling layer that halves the dimensionality. The convolution is performed along the temporal dimension. The resulting filtered features are then passed through an LSTM layer.
- **AE:** A stacked convolutional autoencoder is applied to the exogenous variables, where the dimensionality reduc-

tion is achieved by max pooling layers. The resulting code is then merged with the autoregressive part of the data points. The network is trained to both reconstruct the input from the code and forecast the silicon content for the next cast.

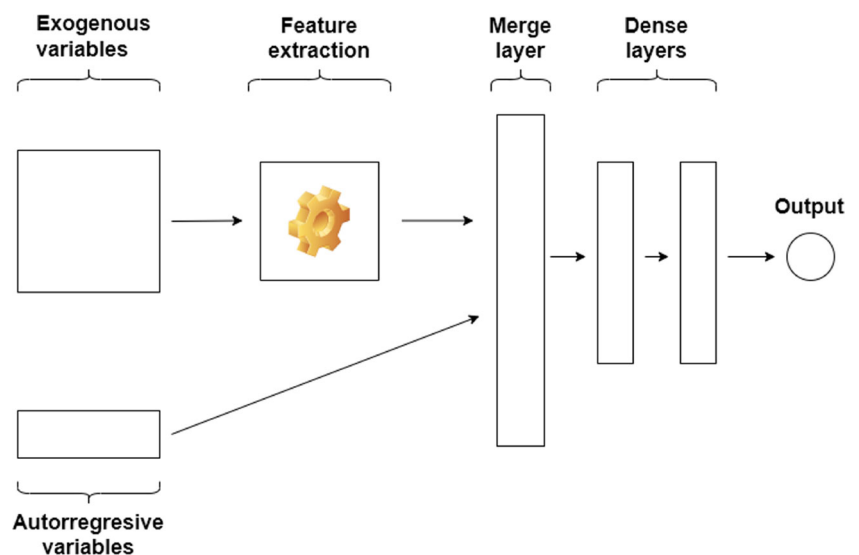
- **Phased LSTM:** A single Phased LSTM layer is employed to filter the exogenous variables, which requires additionally to pass the associated timestamps. The proportion of time in which updates are allowed is kept at the default value of 5%. As in the previously described models, the result is merged with the autoregressive variables and processed by dense layers to generate the final prediction.

Lastly, a simpler *baseline* model defined as a Feed-Forward Neural Network (FFNN) with 3 hidden layers is used to compare results and determine the value of some hyper-parameters.

4.3 Data description

The dataset was collected from a Blast Furnace at Tata Steel's Port Talbot Steelworks. In total, the dataset encompasses more than 4 years of operational data ranging from January 2017 to April 2021, amounts to 2,102,442 rows and 219 columns after a preliminary cleaning and contains in total 19,723 casts. The rows correspond to data sampled every minute from the blast furnace iron making process. Most of the variables reflect process parameters (such as volume, temperature, pressure) and features computed from these measurements. Thus, the dataset is appreciably large and rich compared to the ones that are usually considered in the literature.

Fig. 1 Common pattern of the proposed models



4.4 Data pre-processing and window size

The pre-processing of the dataset is conceptually divided in two steps. The first step deals with cleaning and filling missing values, while the second step is intertwined with the development of the data-driven models. In first place, the brief periods where the blast furnace was shut down were filtered out from the dataset. Regarding the missing values, they appear as a consequence of the data collection design, since most of the variables are recorded only when a change above a certain threshold with respect to the last registered measurement is observed. In some cases, the ratio of missing values was high (above 90%), so it was decided to apply forward fill for those columns; rolling averages were used instead to fill missing values for the remaining variables, in which the frequency of the blanks is lower.

It should be noted that due to both the data collection process and the filling of missing data, the differences between the explaining variables for two given samples are usually rather small. Two of the main challenges faced during the development of the models were the high dimensionality of the data and the irregular sampling of the target variable. With the guidance of domain experts, 21 variables out of the initial 219 were identified as the most influential in the silicon content. They include the following: the blast oxygen, the hot blast volume, pressure and temperature, the top gas temperature and volume, the η_{CO} and the total heat loss, among others. A further reduction in the volume of data was achieved by down-sampling the cleaned dataset to 10 min frequency from the original 1 min periods. This step not only decreased the computational burden required to train the models, but also led to less noisy predictions.

With respect to the irregular sampling issue for the target, it was decided to use only the casts values from the dataset, together with their timestamps and all the process variables during the previous 6 h. Due to the size of the data, this filtering still left a significant amount of data and avoided the need of interpolating the target, an approach that was taken in other work [2]. The optimal windows size of 6 h and sampling rate of 10 min was determined using a grid search approach as described in Section 4.5. Including information of the 6 h before the casting is also aligned with the process underlying physics as the blast furnace is known to react in a delayed way to variations in the control variables, with a cycle completing in between 6 and 8 h. It must be noted that due to the irregular sampling of the silicon content a consistent sliding window approach cannot be employed.

4.5 Hyper-parameter optimisation

The first hyper-parameters that needed to be set were the sampling rate and the amount of previous information to consider. To this end, the simple baseline model was trained in the grid

[10 min, 20 min, 30 min] \times [6h,7h,8h]. As a result, it was decided to reduce the sampling rate to 10 min and include the measurements for the past 6 h at each data point.

The number of neurons of the models was set to roughly match the problem's degrees of freedom, i.e. number of features \times number of samples. The learning rates and number of epochs for the CNN-LSTM and the AE were chosen simultaneously through grid search and inspection of the train-validation error curves. In the case of the Phased LSTM model, it was observed that the one-cycle policy developed by Smith [26] worked better than the usual training schedule; the initial learning rate was fixed through the learning rate finder described in the cited work. A summary of the hyper-parameters selected after optimisation is included in Table 1.

4.6 Validation strategy

The models were evaluated through the walk-forward validation technique. Approximately the last year of data (more precisely, the period comprised between May 2020 and April 2021, containing 4000 casts) was used to validate the models, the remaining part was used as the training set. The validation set (consisting of 4000 casts) was split in 40 consecutive test sets, each comprising 100 casts. Thus, the models first experience an initial training phase on the first 15,723 data points (approximately 3 years) with the hyper-parameters values determined as explained above. In the next step the model is trained on the second validation test and all the previous data is used for training. The process is repeated 40 times and during each one of them the models predict the silicon content for the forthcoming test set (see Fig. 2). After that, the model is fine-tuned in the complete dataset for a small number of epochs with a low learning rate (10^{-5}).

5 Results

In order to assess the models performance, the three following metrics are used: the coefficient of determination R^2 , the mean absolute error MAE and the root mean squared percentage error RMSPE. Table 2 displays the final results, where the errors refer to the predictions for the last 4000 casts obtained as explained in Section 4.6. To obtain more mean-

Table 1 Hyper-parameters of the DL models used following optimisation

Model	Epochs	Batch size	Learning rate
CNN-LSTM	30	64	$3e-4$
AE	24	32	$1e-5$
Phased LSTM	30	64	Obtained using [26]

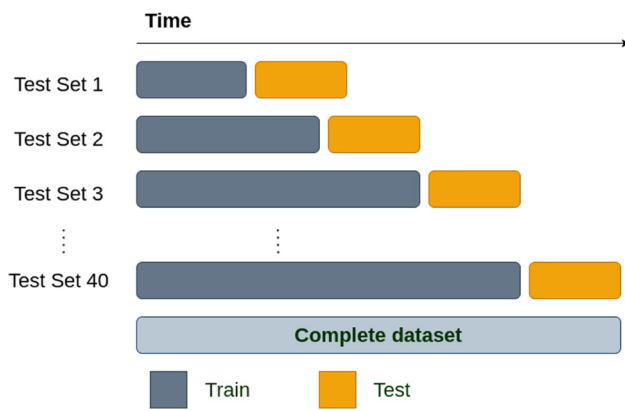


Fig. 2 Diagram illustrating the walk-forward validation strategy. There are 40 test sets in total, each containing 100 samples

ingful results, in the case of the RMSPE only the test cases with silicon content greater or equal than 0.2 were considered.

As shown by Table 2, the three proposed deep learning models outperform the baseline model, with the Phased LSTM model being the strongest one and obtaining an appreciable improvement over the simple FFNN. This is the first time such an exhaustive evaluation of DL models is conducted on blast furnace real data.

The convolutional filters in both the CNN-LSTM model and the AE model likely extract valuable information on the impact of changes in the explaining variables along the 6 h cycle, which allow them to outperform the baseline model. In the case of the AE model, the design of the architecture probably optimises the learned representation towards the prediction of the silicon content, which explains why it is slightly better than the CNN-LSTM model. The Phased LSTM model stands out as a promising technique to deal with event-based sequences. In this particular problem, it shows better results than more traditional models based on LSTM (even when paired with convolutional filters) and convolutional autoencoders.

Figure 3 shows the prediction for the last 150 casts. The model predictions are good overall; however, in some occurrences, the model seems to have some delayed effects in following the upward or downward trend.

Table 2 Comparison of results

Metrics	FFNN	CNN-LSTM	AE	Phased LSTM	Avg
R^2	0.6676	0.6842	0.6919	0.7100	0.7179
MAE	0.1310	0.1277	0.1260	0.1228	0.1214
$RMSPE$	0.2675	0.26123	0.2552	0.2495	0.2474

Avg stands for the results of taking the mean of the predictions of the three models

The bold font shows the best performing model for each performance metrics

The Phased LSTM model is considerable consistent at predicting the *sign* of the variation of the silicon content, that is, whether it will increase or decrease. Figure 4 shows a confusion matrix which determines the ability of the model to predict increase or decreases of the silicon content. It can be noted that the model is skillful at predicting the increase of silicon content. Such information can be used by the operators to take early corrective actions to keep the silicon content values within the upper bound limits. A second possible use for the models is as *soft sensors*. This implies the ability to virtually test scenarios where certain modifications to the control variables are performed and assess the impact on the target through the models' predictions.

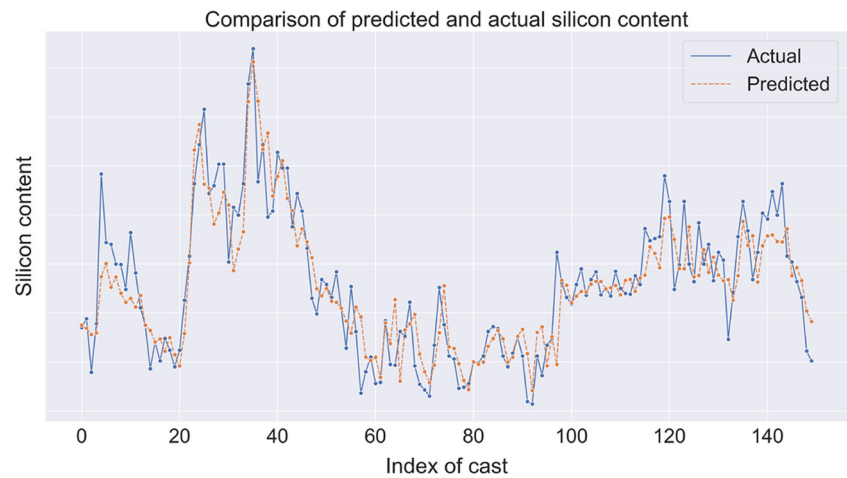
The main difficulties that hinder the performance of all DL models are the presence of similar scenarios (i.e. the set of measurements of the exogenous variables for the previous 6 h) with heterogeneous outcomes and the sampling rate of the target variable, which is at the same time irregular and sparse. Due to the nature of the process, it is impossible to regularly sample the hot metal, while increasing the frequency of the measurements is beyond the current technological capabilities. This is one of the main reasons why the Phased LSTM is particularly useful in this scenario as it tackles the event-based nature of the sampling.

5.1 Discussion and comparison with existing literature

Table 3 shows a summary of the best performing data-driven models for one-step ahead forecasting found in the literature against our proposed model. Reference [7] is not included as it studies different time horizons. Although it is not possible to make a direct fair comparison between the models studied in the literature due to different validation methodologies, datasets and metrics used, it can be seen that the proposed approach shows a good level of performance over a long time period, achieving $R^2 = 0.71$ and $MAE = 0.1228$. It must be noted that all the methods in Table 3, apart from Phased LSTM, have been tested on a shorter timescale and using hold-out validation strategies, hence potentially providing a less robust estimation of the model errors, especially if the test set is selected during stable process operations.

When comparing to [21], it can be seen that the Phased LSTM model outperforms the FFNN model ($MAE = 0.13549$) but not the corrected FFNN model ($MAE = 0.05009$), which relies on a correlation analysis to set variable weights. In such situations, the model would typically need to be periodically regenerated, hence making it impractical for real-world deployments. Similarly, when compared to [2], the Phased LSTM model performance is higher than the NARX FFNN ($R^2 = 0.517$) but lower than the Cluster NARX FFNN ($R^2 = 0.996$). Since the Cluster NARX FFNN uses information about temperature measurements

Fig. 3 Actual and predicted silicon content for the last 150 casts of the dataset for the Phased LSTM model. The scaled of the y-axis was intentionally hidden due to confidentiality constraints



and silicon content at the time of the prediction, it would not be a feasible approach for practical implementation and the high R^2 is achieved due to information being provided about the silicon content. While method [19] shows high accuracy, it is trained and tested on a simulated dataset.

The Phased LSTM is the only approach that can deal with event-based sequences with irregular sampling. Compared to methods [2, 21] and [22] the Phased LSTM does not rely on complex data pre-processing methods and extensive interpolation, making it more appropriate for real-world deployments. In order to maintain the predictive performance of the model, it would suffice to fine-tune it once a week (as it is demonstrated by the walk-forward validation approach). This poses an advantage with respect to models proposed in the literature.

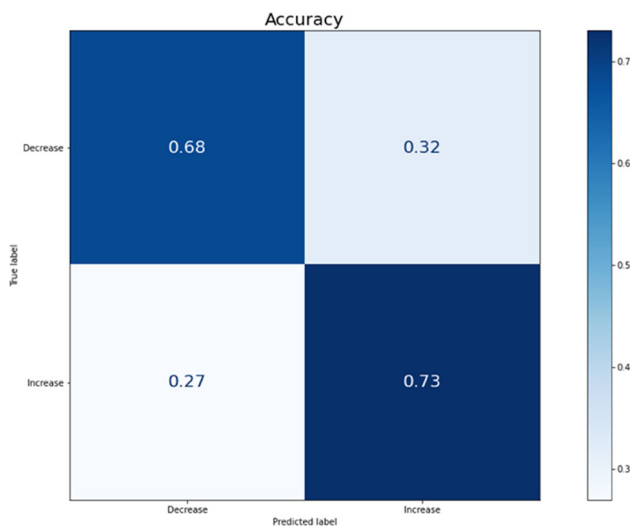


Fig. 4 Accuracy in predicting the sign of the variation for the next cast silicon content

6 Conclusions

In this paper, three DL models were applied to the prediction of silicon content in the next cast for a blast furnace. In order to handle the unprecedented scale and complexity of the available dataset, the forecasting problem was formulated in a novel event-based way, avoiding the need of interpolating the target and the explaining variables to a common regular frequency. From the comparison it emerges that the Phased LSTM model is the best performing model and it can be applied with high confidence to forecast the sign of the variation for the silicon content of the next cast and to test hypothetical scenarios. The methods have been validated using a robust walk-forward validation method in a period of about 1 year, which derives a robust estimation of the errors to provide confidence for industrial deployments. Comparing to existing literature, this is also the first study when such extensive validation has been carried out.

Additionally, the use of DL models removes the need of running frequent and often costly feature pre-processing routines and/or correlation analyses, which would be required by other alternatives in the literature to keep the predictive power up-to-date. Finally, the performance of the Phased LSTM model at dealing with irregularly sampled multivariate time series is verified in a challenging dataset. The results suggest that a more widespread adoption of predictive models able to learn from irregularly sampled data could benefit process industry, where event-driven signals from multiple sensors are common. Therefore this work paves the way towards the industrial deployment of DL forecasting methods that can deal with irregularly sampled data.

Due to the size and the nature of the dataset and, similarly to other works in the literature, this study focused on the next step ahead prediction to prove the ability of Phased LSTM to deal with event-based data. While prediction of next cast is

Table 3 Summary of results of similar approaches for prediction of silicon content of next cast

Paper	Data	Pre-processing	Main techniques	Results
Cui et al. [21]	1000 train samples 500 test samples 60 min frequency	Cubic spline interpolation Outlier detection (3σ method) Weighted average of last 6 h	FFNN optimised with GA plus a correction boosting step.	FFNN: RMSE = 0.18010 MAE = 0.13549 Corrected FFNN: RMSE = 0.06749 MAE = 0.05009
Hua et al. [22]	400 train samples 100 test samples (60 min frequency)	Lagged variables for past 4h (including present time)	A priori clustering (fuzzy C-means) plus SVR on each cluster.	MAPE = 0.1449 RMSE = 0.2988
Fontes et al. [2]	4550 train samples 975 validation samples 975 test samples (30 min frequency)	Fill missing values with rolling mean Replace outliers with PCHIP Lagged variables past 4 steps (2 h, includes present time)	A priori clustering (fuzzy C-means) plus a NARX FFNN on each cluster	NARX FFNN: RMSE = 0.073 $R^2 = 0.517$ Cluster+NARX FFNN: RMSE = 0.007 $R^2 = 0.996$
Ding et al. [19]	648 train samples 72 test samples The data is simulated.	Gaussian noise and Normalisation	LSTM network	RMSE = 0.0338 $R^2 = 0.9246$
Phased LSTM	15723 train samples (i.e. casts over 3-year). 4000 test samples (walk-forward validation) irregular sampling of target variable	Forward filling, rolling averages for missing value 6 h lag	Phased LSTM	MAE=0.1228 RMSPE=0.2495, $R^2 = 0.7100$

useful to overcome the problem of latency in cast measurements, from an industrial perspective, prediction of silicon content for longer time horizon would be desirable, given the lagged nature of the blast furnace operation. Future studies will investigate the proposed approach for longer time horizons.

As a future line, further improvement could be achieved by developing a hybrid model for the silicon content in the blast furnace process as the most promising step. For instance, a physical-chemical model could be used to fill in the missing target values and be later calibrated through a data-driven model.

Acknowledgements Cinzia Giannetti and Eugenio Borghini acknowledge the support of the IMPACT, Supercomputing Wales and Accelerate AI projects, which are part-funded by the European Regional Development Fund (ERDF) via Welsh Government. Cinzia Giannetti and Eugenio Borghini would like to thank Tata Steel UK for data access and industrial discussion.

Author Contributions C.G.: conceptualisation, methodology, analysis of results, validation, writing—original draft preparation, writing—reviewing and editing; E.B.: conceptualisation; methodology, software, analysis of results, validation, writing—original draft preparation, writing—reviewing and editing; A.C.: conceptualisation, analysis of results, writing—original draft preparation, writing—reviewing and

editing; J.R.: validation, writing—reviewing and editing; B. R.: validation, writing—reviewing and editing.

Funding This work was supported by the UK Engineering and Physical Sciences Research Council (EPSRC) project EP/S001387/1 and EP/V061798/1.

Declarations

Ethics approval All authors have approved the review and agree with its submission.

Conflict of interest The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Geerdes M, Toxopeus H, Vliet C (2019) Modern blast furnace ironmaking. IOS Press, Amsterdam, The Netherlands
- Fontes DOL, Vasconcelos LGS, Brito RP (2020) Blast furnace hot metal temperature and silicon content prediction using soft sensor based on fuzzy C-means and exogenous nonlinear autoregressive models. *Comput Chem Eng* 141
- Saxén H, Gao C, Gao Z (2013) Data-driven time discrete models for dynamic prediction of the hot metal silicon content in the blast furnace—a review. *IEEE Trans Ind Inform* 9(4):2213–2225
- Gao C, Jian L, Luo S (2012) Modeling of the thermal state change of blast furnace hearth with support vector machines. *IEEE Trans Ind Electron* 59(2):1134–1145
- Radhakrishnan VR, Mohamed AR (2000) Neural networks for the identification and control of blast furnace hot metal quality. *J Process Control* 10(6):509–524
- Chen W, Wang BX, Han HL (2010) Prediction and control for silicon content in pig iron of blast furnace by integrating artificial neural network with genetic algorithm. *Ironmak Steelmak* 37(6):458–463
- Diniz APM, Côco KF, Gomes FSV, and Salles JLF (2021) Forecasting model of silicon content in molten iron using wavelet decomposition and artificial neural networks. *Metals* 11(7)
- Jiang K, Jiang Z, Xie Y, Chen Z, Pan D, Gui W (2020) Classification of silicon content variation trend based on fusion of multilevel features in blast furnace ironmaking. *Inf Sci* 521:32–45
- Hua C, Wu J, Li J, Guan X (2017) Silicon content prediction and industrial analysis on blast furnace using support vector regression combined with clustering algorithms. *Neural Comput & Applic* 28(12):4111–4121
- Zhang X, Kano M, Matsuzaki S (2019) Ensemble pattern trees for predicting hot metal temperature in blast furnace. *Comput Chem Eng* 121:442–449
- Kaasra I, Boyd M (1996) Designing a neural network for forecasting financial and economic time series. *Neurocomputing* 10(3):215–236
- Borghini E, Giannetti C, Flynn J, Todeschini G (2021) Data-driven energy storage scheduling to minimise peak demand on distribution systems with pv generation. *Energies* 14(12)
- Chen L, Qin N, Dai X, Huang D (2020) Fault Diagnosis of High-Speed Train Bogie Based on Capsule Network. *IEEE Trans Instrum Meas* 69(9):6203–6211
- Wu C, Jiang P, Ding C, Feng F, Chen T (2019) Intelligent fault diagnosis of rotating machinery based on one-dimensional convolutional neural network. *Comput Ind* 108:53–61
- Todeschini G, Kheta K, Giannetti C (2022) An image-based deep transfer learning approach to classify power quality disturbances. *Electr Power Syst Res* 213:108795
- Essien A, Giannetti C (2020) A Deep Learning Model for Smart Manufacturing Using Convolutional LSTM Neural Network Autoencoders. *IEEE Trans Ind Informat* 16(9):6069–6078
- Giannetti C, Essien A (2021) Towards scalable and reusable predictive models for cyber twins in manufacturing systems. *J Intell Manuf*
- Giannetti C, Essien A, Pang YO (2019) A novel deep learning approach for event detection in smart manufacturing. *CIE49 proceedings, Beijing*
- Ding S, Yang H, Wang Z, Song G, Peng Y, Peng X (2018) Dynamic prediction of the silicon content in the blast furnace using lstm-rnn-based models. In *2018 International computers, signals and systems conference (ICOMSSC)*, pp 491–495
- Neil D, Pfeiffer M, Liu S-C (2016) Phased LSTM: accelerating recurrent network training for long or event-based sequences. In *Proceedings of the 30th international conference on neural information processing systems*, pp 3889–3897
- Cui Z, Han Y, Lu C, Wu Y, Chu M (2021) Prediction model of hot metal silicon content based on improved GA-BPNN. *Computational Intelligence and Neuroscience*, 2021
- Hua C, Wu J, Li J, Guan X (2017) Silicon content prediction and industrial analysis on blast furnace using support vector regression combined with clustering algorithms. *Neural Comput Appl* 28(12):4111–4121
- Liu X, Liu X, Jiang M et al (2002) Application of intelligent control and expert system on BF at Laiwu Iron and Steel Group Co [J]. *Metall Ind Autom* 26(6):15–18
- Zhao M, Liu X-G, Luo S-H (2006) Prediction of silicon content in hot metal based on genetic algorithms. *2006 6th World Congress on Intelligent Control and Automation* 2:7771–7774
- Li Y, Zou L, Jiang L, Zhou X (2019) Fault diagnosis of rotating machinery based on combination of deep belief network and one-dimensional convolutional neural network. *IEEE Access* 7:165710–165723
- Smith LN (2017) Cyclical learning rates for training neural networks. In *2017 IEEE Winter conference on applications of computer vision (WACV)*, pp 464–472

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Cinzia Giannetti¹  · Eugenio Borghini¹ · Alex Carr² · James Raleigh² · Ben Rackham²

Eugenio Borghini
eugenusb@gmail.com

Alex Carr
alex.carr@tatasteeleurope.com

James Raleigh
james.raleigh@tatasteeleurope.com

Ben Rackham
ben.rackham@tatasteeleurope.com

¹ Department of Mechanical Engineering, Swansea University Bay Campus, Fabian Way, Swansea SA1 8EN, UK

² Tata Steel Europe, Abbey Works, Port Talbot SA13 2NG, UK