Journal of Electrical Engineering, Vol. 75, No. 3, 2024, pp. 237-248

sciendo

PAPERS

Forecasting material quantity using machine learning and times series techniques

Hanane Zermane[*1], Hassina Madjour[1], Ahcene Ziar[1], Abderrahim Zermane[2]

The current research is dedicated to harnessing cutting-edge technologies within the paradigm of Industry 5.0. The objective is to capitalize on advancements in Machine and Deep Learning techniques. This research endeavors to construct robust predictive models, utilizing historical data, for precise real-time predictions in estimating material quantities within a cement workshop. Machine Learning regressors evaluated based on several metrics, SVR (R-squared 0.9739, MAE 0.0403), Random Forest (R-squared 0.9990, MAE 0.0026), MLP (R-squared 0.9890, MAE 0.0255), Gradient Boosting (R-squared 0.9989, MAE 0.0042). The time series models LSTM and GRU yielded R-squared 0.9978, MAE 0.0100, and R-squared 0.9980, MAE 0.0099, respectively. The ultimate outcomes include improved and efficient production, optimization of production processes, streamlined operations, reduced downtime, mitigation of potential disruptions, and the facilitation of the factory's evolution towards intelligent manufacturing processes embedded within the framework of Industry 5.0. These achievements underscore the potential impact of leveraging advanced machine learning techniques for enhancing the operational dynamics and overall efficiency of manufacturing facilities

Keywords: advanced technologies, intelligent manufacturing, smart manufacturing, forecasting, machine learning, time series

1 Introduction

In every production system, human operators play a crucial role in operations, adjustments, and maintenance. However, the advancement of technology has paved the way for automated production systems, aiming to minimize human intervention and optimize resource utilization. Rapid technological changes have presented numerous challenges for factories; however, they have also brought about various emerging technologies such as the Internet of Things (IoT), big data, cloud computing, artificial intelligence, and its associated techniques. The concept of Industry 4.0 has emerged as a new paradigm for industrialization, leveraging these technologies to address these challenges effectively. At the heart of Industry 5.0 lies the concept of the smart factory is used for process optimization [1], production monitoring, control, and predictive maintenance [2]. Smart factories utilize these cutting-edge technologies to enhance the performance, quality, control, and transparency of the manufacturing process.

This research contributes to the advancement of Industry 5.0 and smart factories. By integrating IoT, big data analytics, cloud computing, artificial intelligence, and other intelligent techniques including machine learning and deep learning, smart factories create context-aware systems that enable both humans and machines to perform tasks based on information and data obtained from the physical and virtual realms [3-5]. The ultimate goal is to drive the factory towards intelligent and optimized operations within the industry 5.0 paradigm. The main focus of this work is the development of a predictive model of the quantity of raw material needed for the production of the clinker in a cement mill workshop in a cement factory (SCIMAT - Batna, Algeria) using a suite of machine learning algorithms including Support Vector Machine (SVM), Random Forest (RF), Multi-Layer Perceptron (MLP), and Gradient Boosting. Moreover, the powerful Deep Learning techniques used in time series include Long Short-Term Memory (LSTM), and the Gated Recurrent Unit (GRU).

The paper presents a comprehensive study that encompasses the following key aspects. It begins with motivation and contribution, a thorough literature review, categorizing previous studies, and highlighting the evolution of machine learning methodologies for regression tasks, in this domain. We then illustrate the data collection process, addressing challenges related to data quality and availability. Extensive feature engineering is explored to harness the potential of various data attributes, enhancing the accuracy and reliability of our predictive models. The results and discussion section illustrate the findings that underscore

[1] Laboratory of Automation and Manufacturing, Industrial Engineering Department, Batna 2 University, Batna, Algeria
[2] Safety Engineering Interest Group, Department of Chemical and Environmental Engineering, Faculty of Engineering, University of Putra Malaysia, 43400 Serdang, Selangor, Malaysia
[*]Corresponding author: h.zermane@univ-batna2.dz, https://orcid.org/0000-0003-4167-2578

https://doi.org/10.2478/jee-2024-0029, Print (till 2015) ISSN 1335-3632, On-line ISSN 1339-309X

the effectiveness of ensemble models and suggest the potential benefits of developing specialized models for each outcome. These insights could significantly impact production measures and contribute to the ongoing efforts to reduce downtimes and their associated consequences by enabling more accurate predictions of material fed to the workshop, enhancing the operational dynamics and overall efficiency of manufacturing facilities. The paper concludes with final suggestions, limitations, and feature studies.

2 Motivation and contribution

This study makes a valuable contribution to the industrial field such as Cement Production, because of the several advantages that the work highlights. Details about the data, the industrial process, and the deep learning techniques applied in the study are highlighted, in addition to the focus on findings and their economic impact on the real system. The developed model proved its efficiency in terms of constructing intelligent models that allow optimized participation more than human operators.

Accordingly, we believe that the following research directions are required for the next generation of prognostic and health management systems (PHM), especially in complicated industrial processes with enormous real-time alarms and faults. The final objective is to obtain a predictive system that can forecast in real time.

3 Literature review

Forecasting is a critical aspect of supply chain management and production planning, as accurate predictions of material quantities can help organizations optimize inventory, reduce costs, and enhance overall efficiency. Firstly, Machine Learning (ML) methods have been proposed in the academic literature as alternatives to statistical ones for time series forecasting. Authors in the review article provide insights into various forecasting methods and discuss their applications [6].

One of the most powerful machine learning techniques is Random Forest (RF) [7]. RF is essentially a collection of decision trees whose outcomes are aggregated based on voting [8]. The research of Han et al. harnessed the power of the random forest (RF) model to predict the compressive strength of LC3 [9]. Ma et al. used the support vector machine (SVM), decision tree (DT), and random forest (RF) models were developed to estimate the compressive strength of cement-based materials with mining waste. The RF algorithm obtained the highest value of R and the lowest value of RMSE, demonstrating the highest accuracy than SVM and DT [10]. The results of Atasham et al. demonstrated that SVM, RF, and ANN can predict the deteriorated compressive strength of concrete and align closely with the experimental results. In this study, the ANN model demonstrated the highest prediction accuracy with an R^2 of 0.924, exhibiting a higher prediction accuracy than RF and SVM models.

Despite the power of machine learning, deep learning techniques proved their efficiency in the industrial field. In recent years, Deep Learning time series techniques such as LSTM and GRU have gained prominence in time series forecasting due to their ability to capture complex temporal dependencies. Several survey papers provided an overview of various deep learning techniques, including LSTM and GRU, for forecasting [11-13].

The LSTM technique has been widely used in various fields of research involving sequential data, such as natural language processing [14], speech recognition [15], time series analysis, and machine translation. The GRU was developed as an effective alternative to the LSTM architecture [16, 17].

The present literature review explores many applications of deep learning techniques, including LSTM and GRU in forecasting since their introduction [18, 19]. LSTM is utilized in predicting equity price with corporate action events, and student performance prediction. In addition, their application for mining public opinion forecasting, as well as in the public health field for State of health estimation of Lithium-Ion batteries [20-23]. In its term, the GRU is applied for several uses; including, the prediction of reservoir parameters through well-logging data [24]. In most studies, LSTM and GRU are both utilized in forecasting situations combined with the Recurrent Neural Networks (RNN) for forecasting the electrical load in a power system. The models were tested, and the GRU model achieved the best performance in terms of accuracy and the lowest error. Results of Abumohsen et al. show that the GRU model achieved an R-squared of 90.228%, a Mean Square Error (MSE) of 0.00215, and a Mean Absolute Error (MAE) of 0.03266 [25]. The literature study presents a variety of models and methods for prediction problems. Each method has its strengths and weaknesses, and the choice of model depends on the data available and the study's specific context.

4 Methods

The paper discusses the intricacies of data collection, addressing challenges related to data quality and availability. Rigorous preprocessing techniques are employed to ensure the reliability and relevance of the data used in model development. This research employs a suite of machine learning algorithms for the prediction of the quantity of material needed for the production of the clinker, including Support Vector Machine (SVM), Random Forest (RF), Multi-Layer Perceptron (MLP), and Gradient Boosting. Moreover, time series techniques including LSTM and GRU are also selected for this task. SVM are supervised learning algorithms used for classification (SVC) and regression (SVR) tasks, particularly effective when dealing with non-linearly separable data. SVM identifies the optimal hyperplane that maximizes the margin between classes in the feature space [26]. Especially useful for scenarios where linear separation is not feasible. SVM employs the kernel trick to handle non-linear relationships in data by mapping it into a higher-dimensional space. Enables SVM to capture complex patterns and make accurate predictions. Support vectors are the data points closest to the hyperplane. They play a pivotal role in determining the position and orientation of the optimal hyperplane [27]. A Decision Tree Classifier is a versatile supervised learning algorithm used for both classification and regression tasks. It makes decisions by recursively splitting the dataset based on feature conditions until a stopping criterion is met, forming a tree-like structure of decisions. The algorithm selects the most informative feature to split the data at each node. The goal is to maximize information gain (for classification) or variance reduction (for regression). Nodes in the tree represent decisions based on feature conditions. Each decision node splits the data into subsets, guiding the traversal of the tree. Leaf nodes contain the final predicted output or class label. The algorithm assigns the majority class for classification tasks or the mean value for regression tasks. For classification, Decision Trees use entropy to measure impurity. Information gain is the reduction in entropy achieved by a split and guides the tree construction. Random Forest is an ensemble learning method based on Decision Trees [28]. It constructs a multitude of Decision Trees during training and outputs the mode of the classes (classification) or the mean prediction (regression) of the individual trees. Each tree is trained on a random subset of the data, and features are randomly selected for each split. For classification tasks, the mode (most frequent class) of the predictions from individual trees is taken as the final output. However, for regression, the mean prediction from all trees is used. Random Forest employs bagging, a technique where each tree is trained on a bootstrap sample (randomly sampled with replacement) from the original dataset. At each split, a random subset of features is considered, preventing individual trees from dominating the ensemble. RF Reduces overfitting and increases robustness. Because it is relevant to such a wide range of use cases, deep learning is generating a lot of interest. Choosing an algorithm is a key stage in the deep learning process, so ensure it genuinely matches the problem's use case [29, 30]. Recurrent Neural Networks (RNNs) are among the best models applied to sequential data. They allow both forward propagation and backward propagation, which is well-suited for time series data. Long Short-Term Memory (LSTM) was introduced by Sepp Hochreiter and Jürgen Schmidhuber in 1997 [18]. The authors presented the LSTM architecture as a solution to address the vanishing gradient problem in traditional recurrent neural networks. The LSTM technique is designed to overcome the limitations of traditional RNNs in capturing long-term dependencies in sequential data (see Fig. 1). LSTMs introduce a memory cell that can store information over long periods, selectively retaining or discarding information as needed. This memory cell is controlled by specialized gating mechanisms, including the input gate, forget gate, and output gate. The gates regulate the flow of information into, out of, and within the memory cell, allowing LSTMs to effectively handle sequences with long time lags and complex patterns [31].

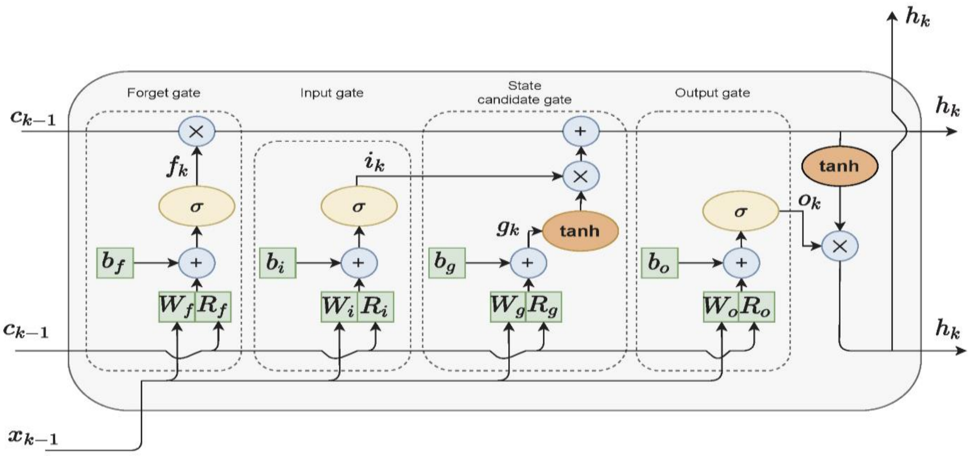


Fig. 1. The LSTM architecture [32]

LSTM networks are specifically designed to capture long-term dependencies in sequential data, making them well-suited for modeling historical trends and patterns. They achieve this through the use of memory cells and gating mechanisms that selectively retain or forget information over time [33]. GRU networks, on the other hand, simplify the architecture by combining the forget and input gates, resulting in a more streamlined model and a concise architecture (see Fig. 2). This reduction in the number of gates can lead to faster training and inference times, making GRU more computationally efficient. Moreover, the GRU can adaptively update the memory cells and control the flow of information. GRU achieves this through its reset gate and update gate mechanisms, which allow for selective retention or modification of information from previous time steps. This adaptability enables GRU to handle sequences with varying lengths and evolving patterns effectively. Furthermore, GRU has been shown to perform comparably to LSTM in many sequence modeling tasks while requiring fewer parameters. This simplicity makes GRU a popular choice in scenarios where computational resources are limited or when the model needs to be deployed on devices with constrained memory or processing capabilities [34]. It is important to note that the choice between LSTM and GRU depends on the specific task and dataset at hand. While LSTM is often preferred for tasks that require modeling very long-term dependencies, GRU can be a more efficient option when dealing with less complex sequential data or when computational efficiency is a priority [35, 36].

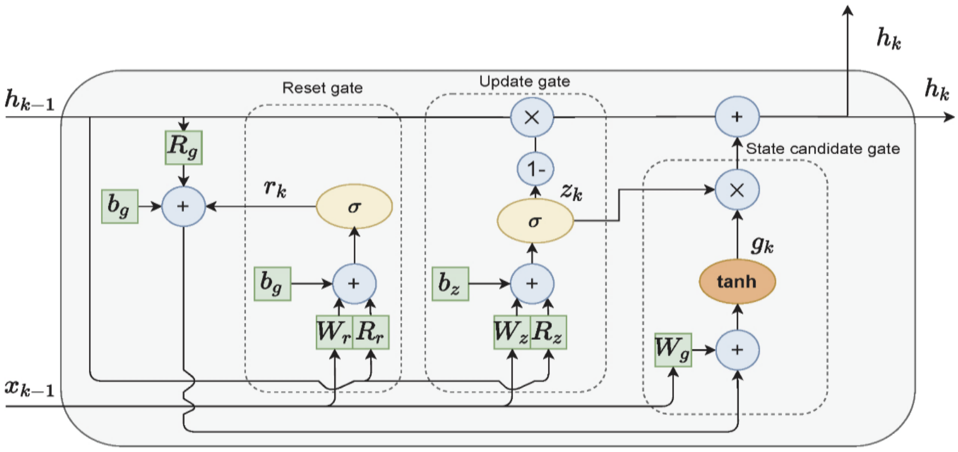


Fig. 2. The GRU architecture [32]

Both LSTM and GRU architectures effectively address the vanishing and exploding gradient problems that can hinder training in traditional RNNs. By regulating the flow of gradients, these architectures ensure stable learning and accurate predictions. Furthermore, the adaptive memory management of LSTM and GRU networks allows them to track dynamic changes in material quantity by updating their internal memory cells. This adaptability enables the models to capture evolving patterns and adjust their predictions accordingly. Moreover, LSTM and GRU networks demonstrate robustness to noisy or incomplete data, enabling them to learn from and generalize patterns even in challenging scenarios. This resilience makes them valuable in the context of real-world data from the cement workshop.

Finally, the flexibility and ease of implementation of LSTM and GRU networks make them widely accessible in popular deep learning frameworks, facilitating their integration into existing systems. Their versatility allows for experimentation, fine-tuning, and optimization, ensuring the predictive model for material quantity is tailored to specific requirements. Ultimately, by leveraging the strengths of LSTM and GRU networks, the cement workshop can optimize production processes, streamline operations, and minimize disruptions, driving efficiency and productivity within the intelligent automation and Industry 5.0 framework. Each developed model is trained and evaluated separately for each target variable. The script uses a pipeline that includes the preprocessing steps and the model itself. The model is trained on the training data, and then predictions are made on the test data. Any negative predictions (which may not make sense in the context of the problem, as the numbers cannot be negative) are replaced with 0. The performance of each model is evaluated using several metrics, including the Mean Squared Error (MSE), R-squared, and Mean Absolute Error (MAE). Each algorithm performed in this paper is selected based on its suitability for handling the unique characteristics of accident prediction data. The paper details the model development process, encompassing feature engineering, hyperparameter tuning, and model selection.

5 Materials

The study methodology, feature engineering approaches, model development, and extensive discussions on the implications of our findings are based on the rich dataset that allows the application of various machine-learning models to predict the amount of the materials. It provides a solid basis for examining the many factors that can contribute to the variation of material quantity. The models' performance can be evaluated and compared based on their accuracy in predicting the crash severity. In this study, the selected workshop for investigation is the cement mill workshop in the Ain Touta cement factory (SCIMAT) located in the East of Algeria. The cement mill workshop plays a crucial role in the cement production line. It involves a series of interconnected processes where the clinker materials undergo grinding, blending, and other necessary treatments to produce the desired product, which is the cement. The flowchart of the cement production plant of SCIMAT is illustrated in Fig. 3.

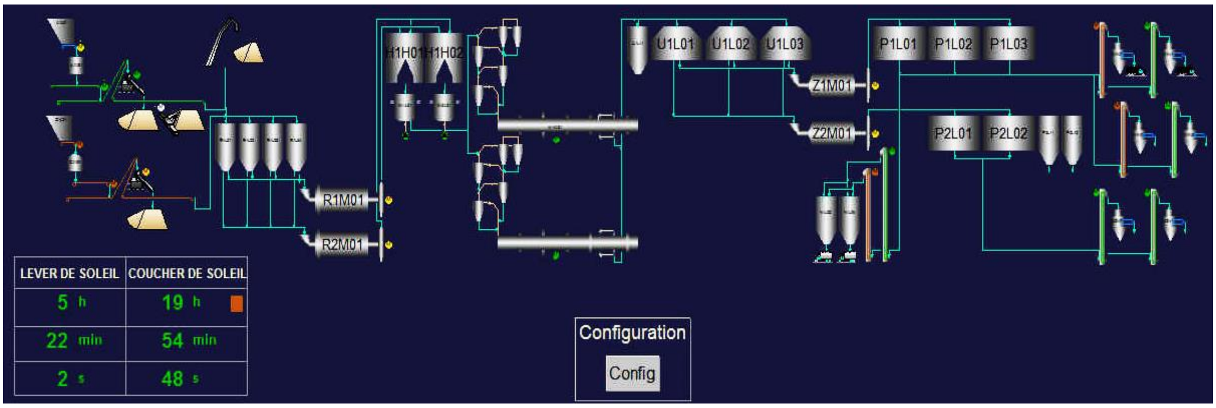


Fig. 3. Flowchart of the cement production plant of SCIMAT factory (FLS/ECS view)

To ensure smooth operation and continuous functionality, the raw mill workshop relies on a complex network of electrical, mechanical, and automated equipment. These include motors, conveyors, crushers, separators, and various control systems. Each component performs specific tasks to process and transform the raw materials into the required intermediate or final products. First, the pre-homogenized raw materials are dried and milled in a two-chamber raw ball mill. The milling process transforms the raw materials into a powder, which is essential for promoting chemical reactions during the burning phase. The resulting powder is conveyed by an elevator to a separator. The separator is responsible for separating the finished product from larger particles. The larger particles that are separated by the separator are sent back to the raw mill for further milling. The finished product, which is the cement, is transferred to the raw mill silos. The raw mill silos store the finished product until it is ready for further processing to produce the clinker in the rotary kiln under a temperature of 1450 °C. The produced clinker will be transferred to a cement ball mill to be milled into cement. The cement processing procedure executed in the cement workshop is illustrated in Fig. 4.

The workshop is supported by a range of auxiliary devices and systems. These include sensors, monitoring instruments, and maintenance equipment, which are employed to monitor operational parameters, detect anomalies, and conduct regular maintenance activities (Table 1). Their collective purpose is to ensure optimal performance, prevent downtime, and sustain the overall efficiency of the production process.

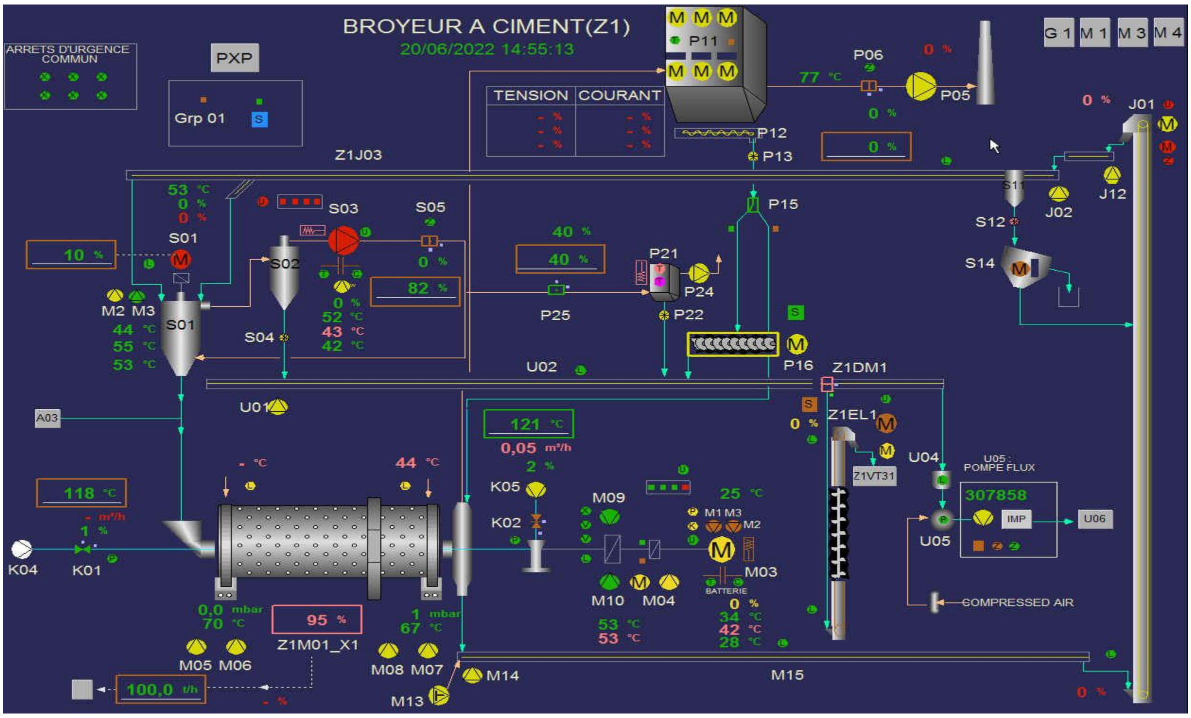


Fig. 4. Process of the cement ball mill workshop (FLS/ECS view)

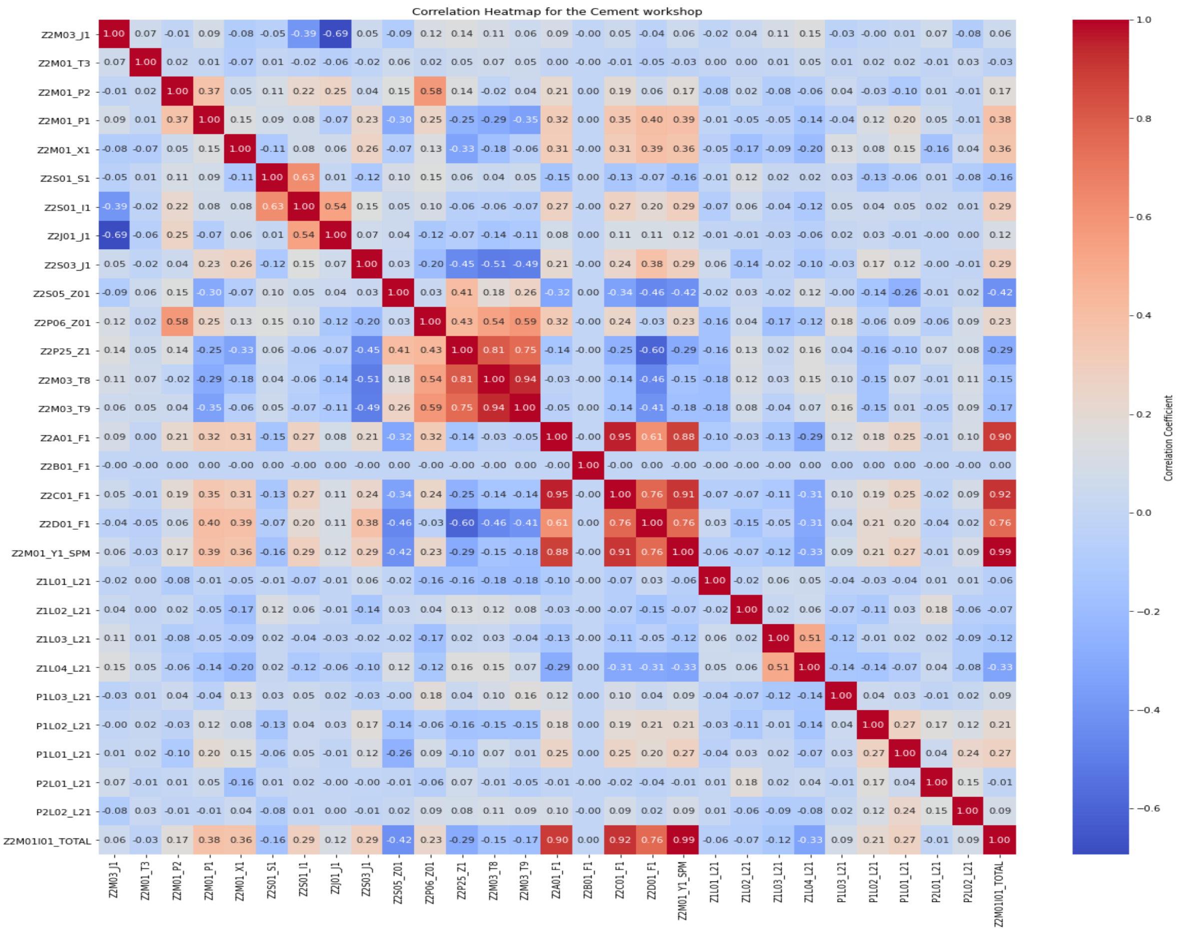


Fig. 5. The heatmap of all parameters

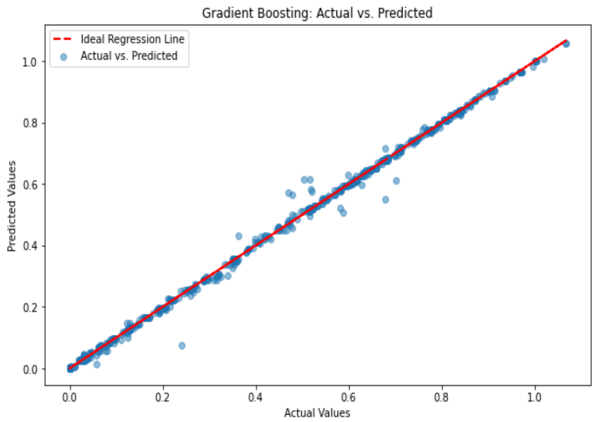
Table 1. Description of the cement mill workshop parameters

Parameters	Interval	Units	Designation
P1L01_L21	[0-100]	%	Silo Cement Level
P1L02_L21	[0-100]	%	Silo Cement Level
P1L03_L21	[0-100]	%	Silo Cement Level
P2L01_L21	[0-100]	%	Silo Cement Level
P2L02_L21	[0-100]	%	Silo Cement Level
Z1L01_L21	[0-100]	%	Clinker Hopper Level
Z1L02_L21	[0-100]	%	Clinker Hopper Level
Z1L03_L21	[80-100]	%	Silo Cement Level
Z1L04_L21	[0-100]	%	Silo Cement Level
Z2A01_F1	[0-140]	t/h	Transp.Tape Flow
Z2B01_F1	[0-140]	t/h	Transp.Tape Flow
Z2C01_F1	[0-8]	t/h	Transp.Tape Flow
Z2D01_F1	[0-40]	t/h	Transp.Tape Flow
Z2J01_J1	[0-120]	%	Elevator Power
Z2M01_P1	[0-4]	mbar	Crusher Pressure Input
Z2M01_P2	[0-40]	mbar	Crusher Pressure Output
Z2M01_T2	[0-150]	°C	Crusher Temperature Input
Z2M01_T3	[0-150]	°C	Crusher Temperature Output
Z2M01_X1	[0-100]	%	Crusher acoustic equipment
Z2M01_Y1_SPM	[0-140]	t/h	Total Feed
Z2M01I01_TOTAL	[0-150]	t/h	Total Feed Rate
Z2M03_J1	[0-120]	%	Engine Crusher Power
Z2M03_T8	[0-150]	°C	Crusher Bearing Temperature
Z2M03_T9	[0-150]	°C	Crusher Bearing Temperature
Z2P06_Z01	[0-100]	%	Butterfly Register Position
Z2P25_Z1	[0-100]	%	Butterfly Register Position
Z2S01_I1	[0-120]	%	Separator Current
Z2S01_S1	[0-100]	%	Separator Speed
Z2S03_J1	[0-120]	%	Fan Power
Z2S05_Z01	[0-100]	%	Butterfly Register Position

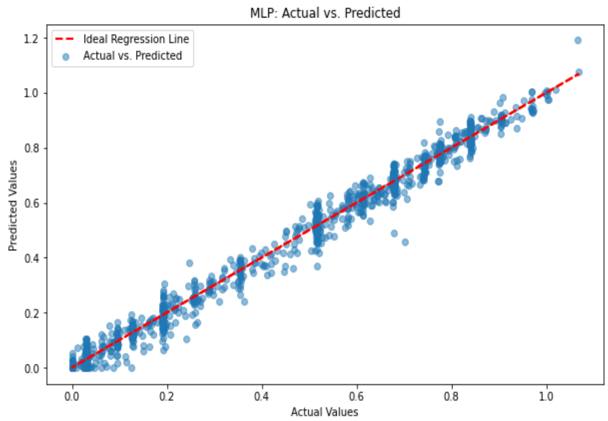
As a result, it was decided to include all the characteristics to capture their contributions and conduct a thorough analysis. By including all the characteristics, the analysis can take into account their potential combined effects on the production line. Typically, in cases where there is a high correlation between two or more characteristics, we may choose to eliminate some of them to avoid redundancy. However, in this particular dataset, since the correlations are generally low, it is necessary to include all the characteristics to capture their contributions and ensure a comprehensive analysis.

6 Results and discussion

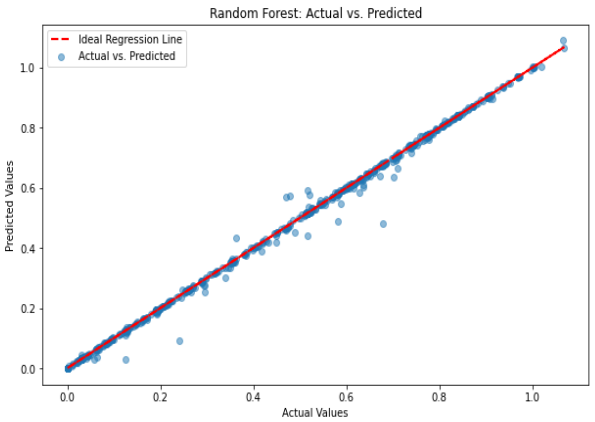
In our experiments, the data set is split into two parts, respectively the train set (60%), test set (30%), and validation set (10%). The training set is used to train the prediction model while the testing set is used to validate the performance of the trained model. More specifically, the accuracy of predictions on the testing set, the core, and key of further applications, plays an essential part in the validation and directly affects whether it could be used. During the first stage, the algorithms were applied to a training dataset and the performance was evaluated. Later, the algorithms were applied to a testing dataset to make predictions. The results of the evaluation process using the Machine Learning algorithm depicted in Fig. 6 (a-d), provide valuable insights into the predictive capabilities of the machine learning models and their ability to accurately forecast the state of the production line.



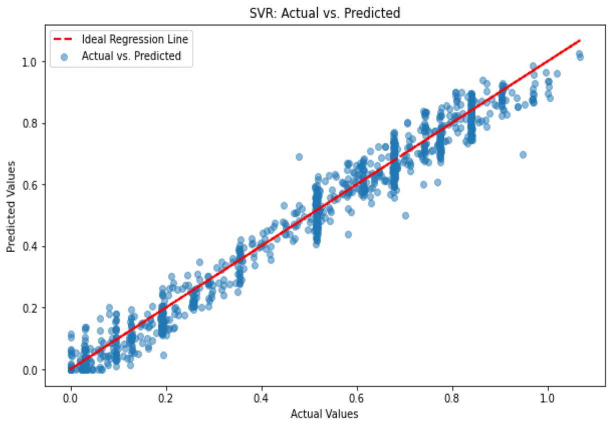
(a) Gradient boosting predictions



(b) Multi-layer perceptron predictions



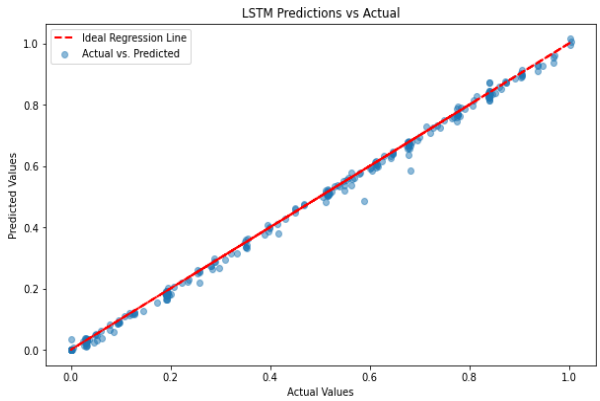
(c) Random forest predictions



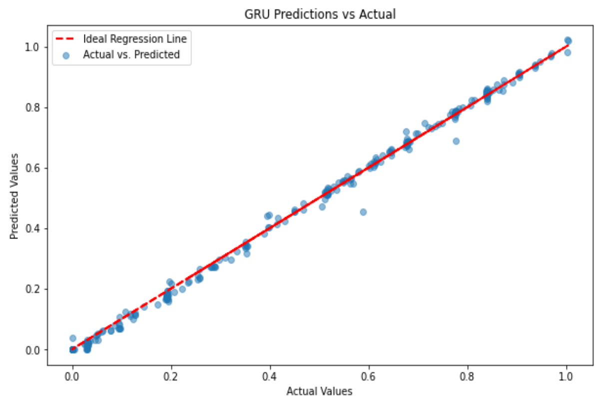
(d) Support vector regressor predictions

Fig. 6. Obtained results of first material forecasting of machine learning models

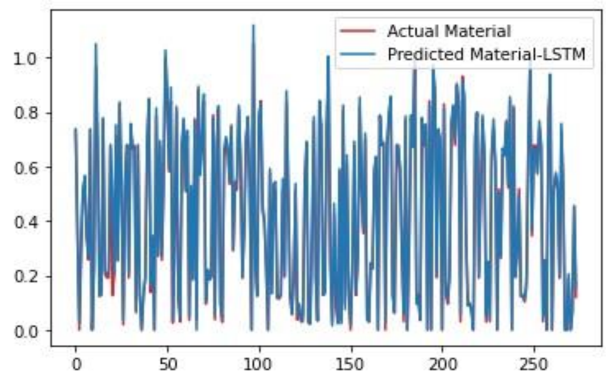
In the initial stage, the LSTM and the GRU algorithms are applied to the training dataset, and their performance is evaluated. Subsequently, the trained model is used to make predictions on the testing dataset, and the accuracy of these predictions is assessed. The results of the evaluation process using time series algorithms represented in Fig. 7 (a-d), provide a more valuable insights into the predictive capabilities of these models and their ability to truthfully predict the state of the feeders of the production line.



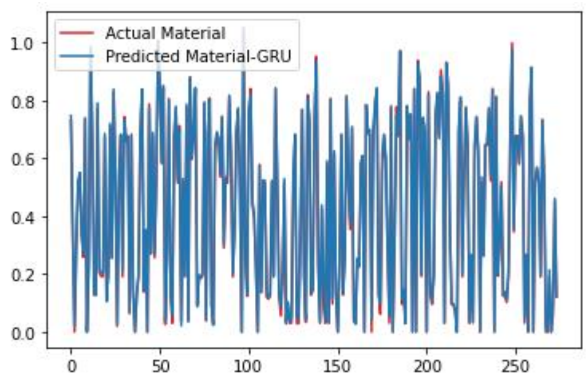
(a) Long short-term memory predictions



(b) Gated recurrent unit predictions



(c) Actual vs predicted using LSTM

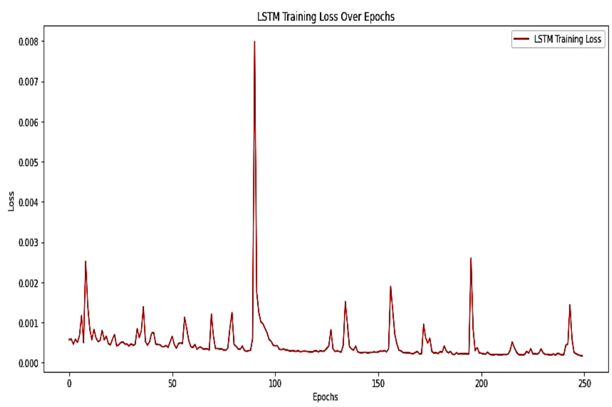


(d) Actual vs predicted using GRU

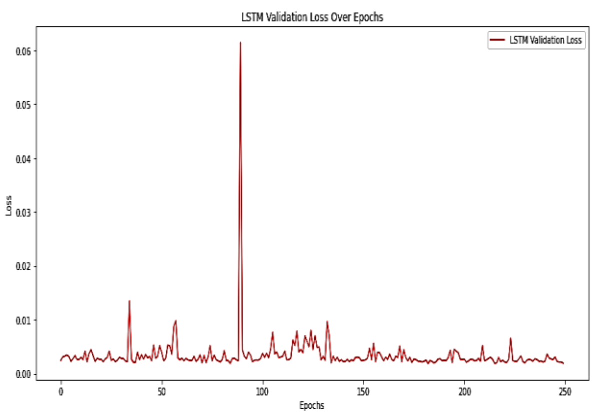
Fig. 7. Actual vs predicted first material forecasting using LSTM and GRU

The evaluating metrics of the predictions on the testing set is a crucial metric for evaluating the model's performance. It serves as a core aspect of further applications and directly influences the model's usability.

For time series prediction using LSTM or GRU, common loss functions include Mean Squared Error (MSE) and Mean Absolute Error (MAE). These loss functions are suitable for regression tasks when predicting a continuous value. The MSE measures the average squared difference between the predicted values (outputs) and the true values (labels). Mathematically, it is calculated as the mean of the squared differences between each corresponding pair of predicted and true values. However, the MAE measures the average absolute difference between the predicted values (outputs) and the true values (labels). It is calculated as the mean of the absolute differences between each corresponding pair of predicted and true values, see Fig. 8 (a, b).

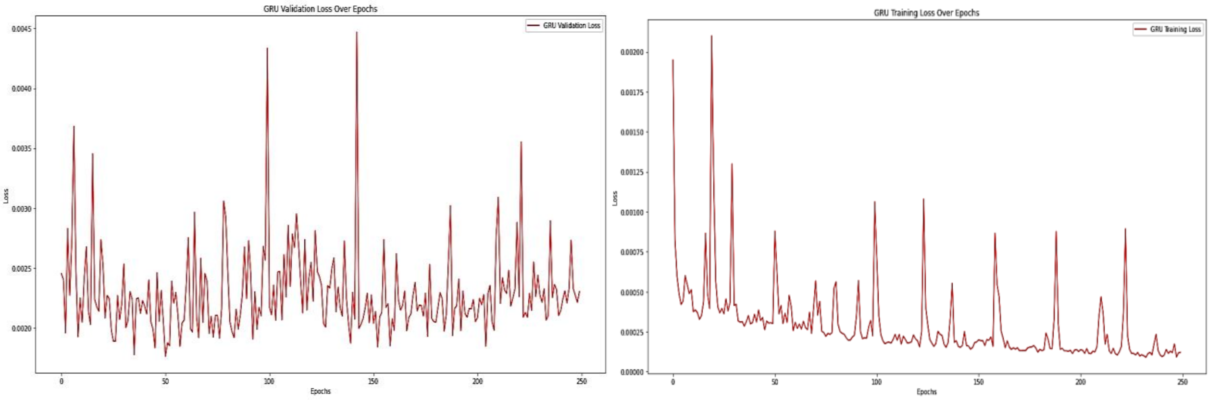


(a) LSTM training loss function



(b) LSTM validation loss function

Fig. 8. The LSTM model training and validation loss functions



(a) GRU training loss function (b) GRU validation loss function

Fig. 9. The GRU model training and validation loss functions

The results of the GRU model training and validation loss functions process are depicted in Fig. 9 (a, b), providing insights into the predictive capabilities of the model.

These metrics provide insights into the effectiveness of the models in accurately predicting the workshop-fed materials and their overall performance in the context of the specific application. By analyzing these metrics, it becomes possible to compare the performance of different models and identify the most suitable approach for predicting the workshop-fed materials in the given industrial setting. To assess the performance of various machine learning models on unseen data, industry-standard metrics such as loss function, r-squared Error, and Mean Absolute Error are utilized. The Mean Square Error of the predictions on the testing set serves as a crucial metric for evaluating the model's performance. It indicates how well the trained model generalizes to unseen data and directly influences its usability for practical applications. The goal is to achieve a high level of prediction on the testing set, demonstrating the model's ability to effectively predict the state of the production line.

Results demonstrate the overall system performance enhancement in predicting bearing failure when modeled data are included with SCADA data. Based on data from the cement plant, the performances of different machine-learning and deep-learning models (LSTM and GRU) on unseen data are then evaluated using industry-standard metrics. The evaluation results of these metrics for different machine learning models, based on the data obtained from the cement plant, are summarized, and presented in Table 2.

Table 2. The evaluation metrics of the predictive models

Metrics	R-squared	RMSE	MAE
SVR	0.9739	0.0530	0.0403
Random forest	0.9990	0.0096	0.0026
MLP	0.9890	0.0338	0.0255
Gradient boosting	0.9989	0.0106	0.0042
LSTM	0.9978	0.0149	0.0100
GRU	0.9980	0.0135	0.0099

The results and discussion presented in this paper underscore the complexity of industrial supervision systems prediction, a problem that requires an integrated, multi-method approach and algorithms. Thus, our work underlines the potential of machine learning algorithms and time series techniques for the prediction of the first materials. The findings inform the selection of appropriate algorithms and feature sets, with the ultimate goal of enhancing production and preventing downtimes.

7 Conclusion

The presented learning models and architectures showcase significant advancements that bring notable improvements in control flexibility, data handling capabilities, and the ability to process large amounts of information in complex industrial processes. These advancements have several advantages and implications for industrial operations. The increased control flexibility allows for better adaptability and responsiveness in managing complex systems. The learning model can dynamically adjust its parameters and decision-making based on real-time data, enabling more precise and efficient control of industrial processes. This flexibility enhances the ability to handle unexpected situations, adapt to changing conditions, and optimize performance in real time.

The improved data handling capabilities of the learning model enable the processing and analysis of large volumes of information. In complex industrial processes, there is a wealth of data generated from various sensors, machines, and control systems. The learning model can effectively extract meaningful insights from this data, facilitating informed decision-making, predictive maintenance, and fault detection. Machine Learning regressors achieved improved results, SVR (R-squared 0.9739, MAE 0.0403), Random Forest (R-squared 0.9990, MAE 0.0026), MLP (R-squared 0.9890, MAE 0.0255), Gradient Boosting (R-squared 0.9989, MAE 0.0042). However, noticeable results were obtained by the time series models LSTM and GRU. They yielded R-squared 0.9978, MAE 0.0100, R-squared 0.9980, MAE 0.0099, respectively. These achievements underscore the potential impact of leveraging advanced machine learning techniques for enhancing the operational dynamics and overall efficiency of manufacturing facilities.

The ability to handle large-scale data empowers industrial systems to leverage the benefits of big data analytics and effectively utilize information for optimizing operations. Furthermore, the advancements in the presented architecture contribute to enhanced productivity and reduced maintenance costs. By accurately predicting faults and alarms, the learning model enables proactive maintenance and minimizes unplanned downtime. This proactive approach helps prevent costly equipment failures and enables efficient scheduling of maintenance activities, ultimately improving overall productivity and reducing maintenance expenses. However, despite these advancements, there are still areas for further improvement and exploration.

The ultimate outcomes include improved and efficient production, optimization of production processes, streamlined operations, reduced downtime, mitigation of potential disruptions, and the facilitation of the factory's evolution towards intelligent manufacturing processes embedded within the framework of Industry 5.0. Consequently, Future research should focus on testing the presented dataset with a wider range of advanced deep-learning algorithms to identify the most suitable approach for specific industrial processes. Exploring alternative algorithms can potentially lead to even better efficiency, accuracy, and adaptability in predicting and managing faults and alarms. Moreover, it is crucial to address the challenges associated with developing prognostic and health management systems for complex industrial processes with numerous real-time alarms and faults. Research efforts should be directed toward developing more sophisticated algorithms and methodologies that can effectively handle the complexities and dynamics of such systems. The aim is to create autonomous systems that can supervise factories in real time, make timely decisions based on alarms and faults, and take appropriate actions to ensure optimal operation and performance. By pursuing these research directions, the field of prognostic and health management systems can advance further, leading to the development of intelligent and autonomous industrial processes that are capable of efficiently managing complex operations, minimizing disruptions, and maximizing productivity.

References

[1] Y. Li, S. Carabelli, E. Fadda, D. Manerba, R. Tadei, and O. Terzo, "Machine learning and optimization for production rescheduling in Industry 4.0," *Int. J. Adv. Manuf. Technol.*, vol. 110, no. 9–10, pp. 2445–2463, 2020.

[2] L. Wang, Z. Liu, A. Liu, and F. Tao, "Artificial intelligence in product lifecycle management," *Int. J. Adv. Manuf. Technol.*, vol. 114, no. 3–4, pp. 771–796, 2021.

[3] A. Sharma, Z. Zhang, and R. Rai, "The interpretive model of manufacturing: a theoretical framework and research agenda for machine learning in manufacturing," *Int. J. Prod. Res.*, vol. 59, no. 16, pp. 4960–4994, 2021.

[4] A. Kusiak, "Smart manufacturing," *Int. J. Prod. Res.*, vol. 56, no. 1–2, pp. 508–517, 2018.

[5] I. Van Heerden and A. Bas, "Viewpoint: Ai as author - bridging the gap between machine learning and literary theory," *J. Artif. Intell. Res.*, vol. 71, pp. 175–189, 2021.

[6] V. A. Spyros Makridakis, Evangelos Spiliotis, "Statistical and Machine Learning forecasting methods: Concerns and ways forward Spyros," *PLoS One*, vol. 13, no. 3, pp. 1–26, 2018.

[7] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.

[8] A. Zermane, M. Z. Mohd Tohir, H. Zermane, M. R. Baharudin, and H. Mohamed Yusoff, "Predicting fatal fall from heights accidents using random forest classification machine learning model," *Saf. Sci.*, vol. 159, no. November 2022, p. 106023, 2023.

[9] T. Han, B. K. Aylas-Paredes, J. Huang, A. Goel, N. Neithalath, and A. Kumar, "On the Prediction of the Mechanical Properties of Limestone Calcined Clay Cement: A Random Forest Approach Tailored to Cement Chemistry," *Minerals*, vol. 13, no. 10, pp. 1–19, 2023.

[10] H. Ma, J. Liu, J. Zhang, and J. Huang, "Estimating the Compressive Strength of Cement-Based Materials with Mining Waste Using Support Vector Machine, Decision Tree, and Random Forest Models," *Adv. Civ. Eng.*, vol. 2021, 2021.

[11] B. Lim and S. Zohren, "Time-series forecasting with deep learning: A survey," *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.*, vol. 379, no. 2194, 2021.

[12] H. Bousnguar, A. Battou, and L. Najdi, "Gated Recurrent units (GRU) for Time Series Forecasting in Higher Education," *Int. J. Eng. Res. Technol.*, vol. 12, no. 03, pp. 152–154, 2023.

[13] R. Dey and F. M. Salemt, "Gate-variants of Gated Recurrent Unit (GRU) neural networks," in *Midwest Symposium on Circuits and Systems*, 2017, pp. 1597–1600.

[14] Y. Wu *et al.*, "Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation," *arXiv*, pp. 1–23, 2016.

[15] H. Sak, A. Senior, and F. Beaufays, "Long Short-Term Memory Based Recurrent Neural Network Architectures for Large Vocabulary Speech Recognition," *arXiv*, 2014.

[16] K. Cho *et al.*, "Learning phrase representations using RNN encoder-decoder for statistical machine translation," in *EMNLP 2014 - 2014 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference*, 2014, pp. 1724–1734.

[17] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling," pp. 1–9, 2014.

[18] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Comput.*, vol. 9, pp. 1735–1780, 1997.

[19] R. W. S. Makridakis, A. Andersen, R. Carbone, R. Fildes, M. Hibon, R. Lewandowski, J. Newton, E. Parzen, "The Forecasting Accuracy of Major Time Series Methods," *J. Am. Stat. Assoc.*, vol. 81, no. 393, pp. 262–263, 1986.

[20] S. Minami, "Predicting Equity Price with Corporate Action Events Using LSTM-RNN," *J. Math. Financ.*, vol. 08, no. 01, pp. 58–63, 2018.

[21] Y. Xie, "Student Performance Prediction via Attention-Based Multi-Layer Long-Short Term Memory," *J. Comput. Commun.*, vol. 09, no. 08, pp. 61–79, 2021.

[22] G. M. S. Hossain, M. H. O. Rashid, M. R. Islam, A. Sarker, and M. A. Yasmin, "Towards Mining Public Opinion: An Attention-Based Long Short Term Memory Network Using Transfer Learning," *J. Comput. Commun.*, vol. 10, no. 06, pp. 112–131, 2022.

[23] I. Obisakin and C. V. Ekeanyanwu, "State of Health Estimation of Lithium-Ion Batteries Using Support Vector Regression and Long Short-Term Memory," *Open J. Appl. Sci.*, vol. 12, no. 08, pp. 1366–1382, 2022.

[24] Z. Yu, Y. Sun, J. Zhang, Y. Zhang, and Z. Liu, "Gated recurrent unit neural network (GRU) based on quantile regression (QR) predicts reservoir parameters through well logging data," *Front. Earth Sci.*, vol. 11, no. January, pp. 1–8, 2023.

[25] M. Abumohsen, A. Y. Owda, and M. Owda, "Electrical Load Forecasting Using LSTM, GRU, and RNN Algorithms," *Energies*, vol. 16, no. 5, pp. 1–31, 2023.

[26] S. H. Ahmadi and M. J. Khosrowjerdi, "Fault detection Automation in Distributed Control Systems using Data-driven methods : SVM and KNN," *TechRxiv. Prepr.*, pp. 0–7, 2021.

[27] C. Corinna and V. Vapnik, "Support-Vector Networks," *Mach. Leaming*, vol. 20, pp. 273–297, 1995.

[28] H. Zermane and A. Drardja, "Development of an efficient cement production monitoring system based on the improved random forest algorithm," *Int. J. Adv. Manuf. Technol.*, vol. 120, no. 3–4, pp. 1853–1866, 2022.

[29] J. P. Usuga Cadavid, S. Lamouri, B. Grabot, R. Pellerin, and A. Fortin, "Machine learning applied in production planning and control: a state-of-the-art in the era of industry 4.0," *J. Intell. Manuf.*, vol. 31, no. 6, pp. 1531–1558, 2020.

[30] T. Mohana-Priya, M. Punithavall, and R. Rajesh-Kanna, "Conceptual Review on Machine Learning Algorithms for Classification Techniques," *Int. J. Sci. Res. Comput. Sci. Eng. Inf. Technol.*, vol. 7, no. 1, pp. 215–222, 2021.

[31] K. Greff, R. K. Srivastava, J. Koutnik, B. R. Steunebrink, and J. Schmidhuber, "LSTM: A Search Space Odyssey," *IEEE Trans. Neural Networks Learn. Syst.*, vol. 28, no. 10, pp. 2222–2232, 2017.

[32] K. Zarzycki and M. Ławryńczuk, "LSTM and GRU neural networks as models of dynamical processes used in predictive control: A comparison of models developed for two chemical reactors," *Sensors*, vol. 21, no. 16, 2021.

[33] G. Van Houdt, C. Mosquera, and G. Nápoles, "A review on the long short-term memory model," *Artif. Intell. Rev.*, vol. 53, no. 8, pp. 5929–5955, 2020.

[34] N. Zafar, I. U. Haq, J. U. R. Chughtai, and O. Shafiq, "Applying Hybrid Lstm-Gru Model Based on Heterogeneous Data Sources for Traffic Speed Prediction in Urban Areas," *Sensors*, vol. 22, no. 9, pp. 1–20, 2022.

[35] T. B. Shahi, A. Shrestha, A. Neupane, and W. Guo, "Stock price forecasting with deep learning: A comparative study," *Mathematics*, vol. 8, no. 9, pp. 1–15, 2020.

[36] R. Fu, Z. Zhang, and L. Li, "Using LSTM and GRU neural network methods for traffic flow prediction," in *Proceedings - 2016 31st Youth Academic Annual Conference of Chinese Association of Automation, YAC 2016*, 2017, no. November 2016, pp. 324–328.

Received 12 April 2024