Chapter 9

Application of deep learning techniques for forecasting iron ore prices: A comparative study of long short-term memory neural network and convolutional neural network

Hoang Nguyen[a,b], Yoochan (Eugene) Kim[c], and Erkan Topal[c]
[a]*Department of Surface Mining, Mining Faculty, Hanoi University of Mining and Geology, Hanoi, Vietnam,* [b]*Innovations for Sustainable and Responsible Mining (ISRM) Research Group, Hanoi University of Mining and Geology, Hanoi, Vietnam,* [c]*Department of Mining Engineering and Metallurgical Engineering, WA School of Mines, Curtin University, Bentley, Perth, WA, Australia*

1 Introduction

Iron ore is a fundamental raw material in the steel industry and plays a significant role in the global economy as a strategic resource. With varying iron endowments and demand levels around the world, the global supply and demand for iron ore exhibit significant geographical differences, and it has a significant effect on the chaos of iron ore prices. Moreover, the spatial scale and scale effect of iron ore flowing worldwide is undergoing significant changes due to the ongoing process of global economic integration.

Besides, iron ore prices have been experiencing significant fluctuations in recent years due to a variety of political and economic factors. Such fluctuations have impacted companies' profits and investment plans, particularly those in downstream enterprises associated with metallurgy, construction, and automotive manufacturing. As a result, accurate and effective forecasting of iron ore prices has become increasingly vital in managing project risk and developing profitable investment strategies. With the potential for severe and frequent price

Applications of Artificial Intelligence in Mining and Geotechnical Engineering
https://doi.org/10.1016/B978-0-443-18764-3.00017-5

fluctuations, stakeholders must stay informed and up-to-date on the latest developments in the iron ore market.

To deal with this problem, previous studies have primarily focused on theoretical aspects, such as the mechanisms for trading and pricing iron ore and the factors that influence its price fluctuations. However, limited attention has been paid to predicting iron ore market prices accurately although several models have been introduced to forecast iron ore prices [1,2]. Kim et al. [3] demonstrated that iron ore price has lagged behind the co-integrating and causal relationship between 12 different commodity prices integrated, including oil, copper, gold, liquefied natural gas (LNG), aluminum, nickel, silver, copper, etc. Pustov et al. [4] modeled long-term iron ore prices using marginal costs against marginal incentive prices and forecasted \$85/t and \$124/t, respectively, considering the depletion of iron ore deposits. In another study, Wang et al. [1] suggested a combined model EMD (empirical mode decomposition)—NARNN (non-linear autoregressive neural network)—ARIMA (autoregressive integrated moving average) models for time-series iron ore import price to China which performed better than NARNN or seasonal autoregressive integrated moving average (SARIMA) models. Chen et al. [5] also suggested the Copula (time-varying Copula function), VAR (vector autoregression), BEKK (Baba, Engle, Kraft, and Kroner), GARCH (Generalized AutoRegressive Conditional Heteroskedasticity) models to explore the dependence, correlation, and volatility spillovers between the Baltic Dry Index (BDI), iron ore price and Brent crude oil price, concluding that iron ore is beginning to emerge its predictive indicator post-COVID-19. Ewees et al. [6] proposed a chaotic grasshopper optimization algorithm for the neural network (CGOA-NN) model for monthly iron ore price, which demonstrated an improvement in the forecasting accuracy obtained from the classic neural network (NN), Genetic Algorithm for neural network (GA-NN), particle swarm optimization for neural network (PSO-NN), and grasshopper optimization algorithm for neural network GOA-NN models by 60.82%, 32.18%, 16.49%, and 38.71% decrease in mean square error, respectively. In another study, Kim et al. [7] suggested a prediction model for iron ore price using the linear (Purelin) model of the Levenberg-Marquardt Artificial Neural Network, which proved to exhibit the better forecast result with an average accuracy of 5.92% for 1 month ahead, 9.48% for 2 months ahead, 11.21% for 3 months ahead during the time of price fluctuation in 2020–21, when compared against five other different models Bivariate Non-Linear Regression (BNLR), Multiple Linear Regression (MLR), Multiple Non-Linear Regression (MNLR) as well as logsig, tansig, and purelin model of Levenberg-Marquardt Artificial Neural Network modeling. Lv et al. [8] proposed an improved optimized neural network that had better performance for indicators like Mean Absolute Error (MAE), relative standard deviation (RSD), Mean Square Error (MSE), Root Mean Square Error (RMSE), and Average Absolute Relative Deviation (AARD) when compared with basic neural network, PSO-based, Intelligent Integrated Optimizer, Genetic Neural Network

(GNN). Tuo and Zhang [9] proposed a hybrid Ensemble Empirical Mode Decomposition-Gated Orthogonal Recurrent Unit (EEMD-GORU) model and a novel data reconstruction method to forecast the price index series of China's and international iron ore spot markets from the futures market.

In addition to existing neural networks for iron ore price prediction and development of decision-making support and project risk management references for related enterprises, use of deep learning techniques has not been considered for forecasting iron ore prices. As the iron ore market continues to evolve, further research in these areas will be crucial to ensure that stakeholders can make informed decisions and mitigate risks effectively. In recent years, deep learning techniques have been gaining popularity in the field of commodity price forecasting [10,11]. Specifically, the long and short-term memory neural network (LSTM) and convolutional neural network (CNN) have emerged as promising approaches for predicting metal prices [12–14]. In this book chapter, the feasibility and performance of these techniques are examined for forecasting iron ore prices, providing new insights into commodity price forecasting. With the potential for significant impacts on the mining and manufacturing industries, accurate and effective forecasting of iron ore prices is crucial for stakeholders. The insights gained from this research may pave the way for more advanced and reliable methods of commodity price forecasting in the future.

2 Methodology

2.1 Long short-term memory neural network (LSTM)

Recurrent Neural Network (RNN) is a deep learning technique that can be used to process sequence data, and it allows for the handling of sequential changes in the data [15]. Unlike general neural networks, RNNs are skilled at solving problems in which a word can have different meanings depending on the context in which it is used.

Early versions of RNNs faced a challenge known as the vanishing gradient problem, which caused gradient-based methods to take an excessively long time to train RNNs. This was because the error gradient, which gradient-based methods rely on, diminishes as it is transmitted back through the network [16]. As a result, the first layers in RNNs stop learning, making it difficult for the network to transfer information from earlier time steps to later ones when dealing with lengthy sequences. This results in RNNs having poor short-term memory.

Cell one in a typical RNN architecture can encounter a vanishing error gradient issue. This phenomenon results in an insufficient gradient signal for updating the weights, leading to problems in propagating accurate information through the network. As a consequence, incorrect or diluted information may be transmitted to subsequent cells, including cell four. This challenge hampers the RNN's ability to capture and retain long-range dependencies in sequential data. To mitigate this problem, Sepp Hochreiter and Jürgen Schmidhuber developed a new advanced architecture of RNN called long short-term memory (LSTM) [17], which are designed to handle and prevent such gradient-related issues, enabling more effective learning and information retention in sequential tasks.

The LSTM architecture has evolved over time, but the most common structure consists of a cell with three gates that regulate the flow of information within the cell and control the cell state. The three gates are the input gate, the output gate, and the forget gate [18]. These cells are then linked together, with each one serving as a memory module. Fig. 1 shows the structure of an LSTM model. The architecture of an LSTM cell involves several components, where X_t represents the input time step, ht is the output, C_t is the cell state, f_t represents the forget gate, i_t represents the input gate, and o_t represents the output gate. Additionally, there is an internal cell state represented by $\hat{C}_t$. The operations that occur within the light red circle are pointwise.

In the diagram above, the forget, input and output gates are represented by the symbols f_t, i_t, and o_t, respectively. The purpose of each gate is as follows:

- The forget gate decides which information in the internal cell state should be discarded.
- The input gate determines which new information should be stored in the internal cell state.
- The output gate produces the filtered version of the internal cell state, which is the output of the cell.

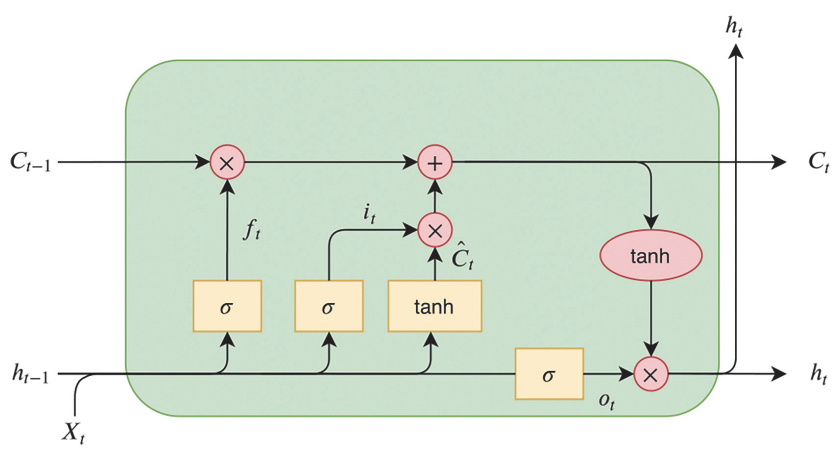


FIG. 1 Structure of LSTM model.

$$
\begin{aligned}
f_t &= \sigma\left(W_f \cdot [h_{t-1}, X_t] + b_f\right) \\
i_t &= \sigma(W_i \cdot [h_{t-1}, X_t] + b_i) \\
o_t &= \sigma(W_o \cdot [h_{t-1}, X_t] + b_o) \\
\widehat{C}_t &= \tanh(W_C \cdot [h_{t-1}, X_t] + b_C)
\end{aligned}
\tag{1}
$$

Then, the internal cell state is computed using Eq. (2) provided.

$$C_t = i_t \cdot \widehat{C}_t + f_t \cdot C_{t-1} \tag{2}$$

The final output of the cell, or h_t, is filtered with the internal cell state as described in the output Eq. (3).

$$h_t = o_t \times \tanh(C_t) \tag{3}$$

where each gate is associated with weights and biases, which are represented by the weight matrices $W_f, b_f, W_i, b_i, W_o, b_o$, and W_C, b_C, respectively. These weight matrices are used in conjunction with gradient-based optimization to enable the LSTM cell to learn.

Multiple LSTM cells are linked together, as shown in the diagram below, to enable the RNN-LSTM network to retain information from previous time steps and make predictions for time-series data [19]. The use of LSTM cell architecture addresses the issue of vanishing gradient, which was a limitation in earlier RNN architectures and prevented accurate time-series predictions, like forecasting iron ore prices.

Several different approaches have been used to forecast commodity prices (especially non-ferrous metals) using apply deep learning. Liu et al. [20] presented a novel hybrid deep learning model for forecasting non-ferrous metals (Zinc, Copper, and Aluminum) by combining the VMD (variational mode decomposition) method and the LSTM (long short-term memory) network. Ozdemir et al. [21] also proposed a nickel forecasting model using recurrent neural networks (RNN) based on long short-term memory (LSTM) and gated recurrent unit (GRU) networks and concluded that LSTM and GRU networks are very useful and successful in forecasting the nickel price variations owing to having average mean absolute percentage error (MAPE) values of 7.060% and 6.986%, respectively.

2.2 Convolutional neural network (CNN)

Convolutional neural networks (CNN) is one of the artificial neural network architectures widely used in computer vision and image processing related to artificial intelligence (AI). The initial concept started as simple and complex cells in 1959 [22]. A simple cell recognizes the orientation of an image and is followed by a response, as shown by the image in Fig. 2A.

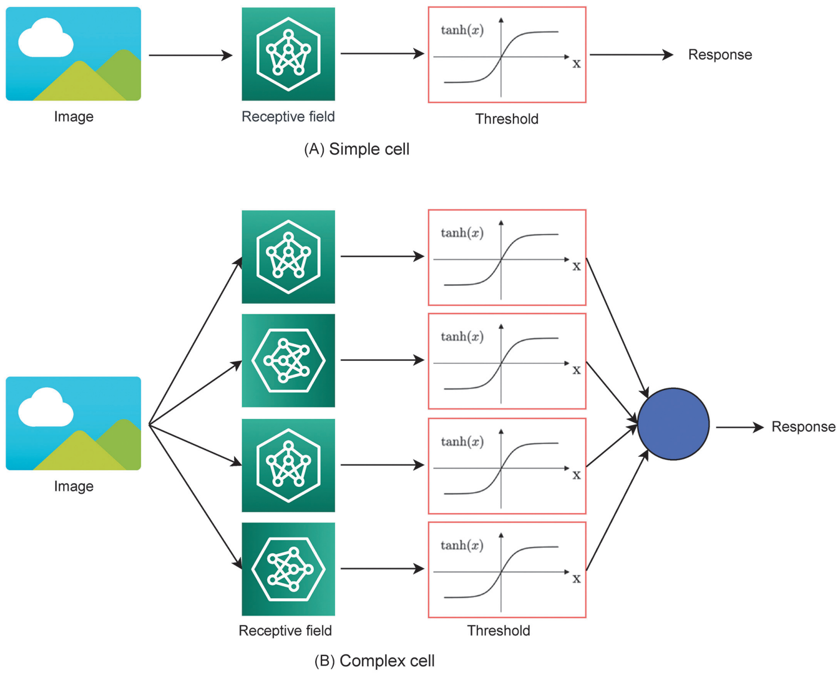


FIG. 2 (A) A simple cell responds to an image by recognizing an image and (B) a complex cell responds to an image by recognize images.

However, a complex cell also responds to orientations of images as well, it also has the capability to complex cells respond to edges and bars at any location in the scene, this capability is by summing information from multiple simple cells as shown in Fig. 2B.

Following the work of Hubel D and Wiesel T [22], Fukushima [23] designed a neural network that mimics simple and complex cell functions, which consists of two types of cells: S-cells, which operate as simple cells, and C-cells that operate as complex cells. These cells mimic the algorithmic structure of both simple and complex cells. These Neocognitron capture complex patterns by applying complex cells that gather their information from other lower-level complex cells or simple cells that detect simpler patterns. As one of the primary algorithms for deep learning, CNN has found numerous applications across a range of fields. Typically, a CNN model consists of an input layer, multiple convolution layers, several pool layers, several fully connected layers, and an output layer. The evaluation of a CNN model is achieved through the use of a loss function. With their ability to process complex data, CNNs have become increasingly important in areas such as time series, computer vision, and natural language processing, however, it has never been used to forecast commodity price. Song et al. [24] presented the structure of CNN and the role of the layers, which are presented in Fig. 3.

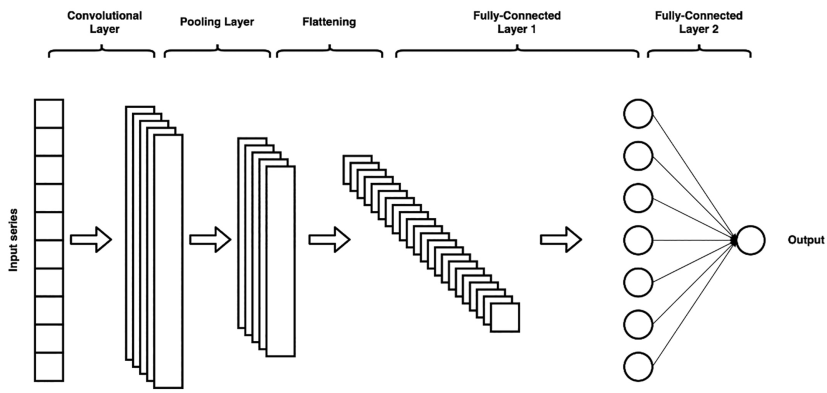


FIG. 3 Structure of a CNN model for the time-series problems.

3 Dataset used

The intricate dance of economic variables has always been of the utmost importance to both businesses and policymakers. In this chapter, a comprehensive dataset spanning from August 1991 to August 2021 has been carefully gathered and analyzed. This dataset includes a rich tapestry of six different variables, including the exchange rates between USD and CNY, and AUD, along with the prices of aluminum, crude oil, gold, and iron ore. By closely examining the complex interactions between these economic factors over the last 3 decades, we can gain valuable insights into their patterns and behaviors. Armed with this knowledge, businesses and policymakers can make informed decisions to navigate the uncertain and ever-changing landscape of the global market. The datasets are shown in Fig. 4. Herein, 70% of the dataset is used to train the model, 20% is used for the validation model during training, and the remaining 10% is used for testing the developed LSTM and CNN model for forecasting iron ore prices.

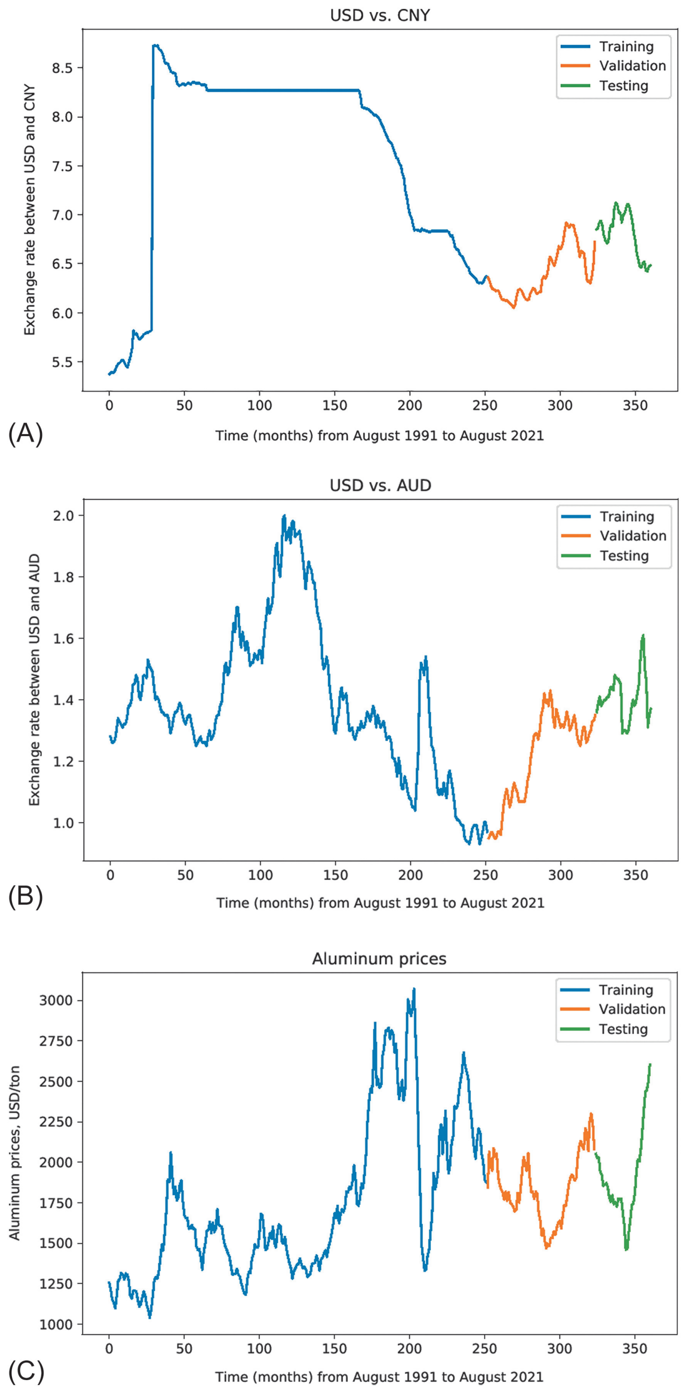
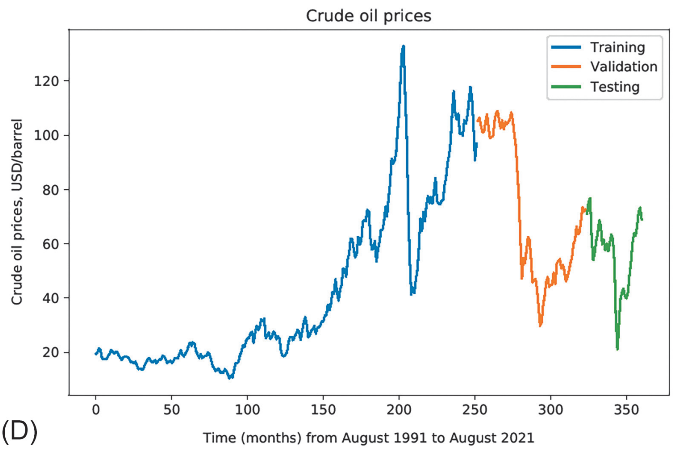
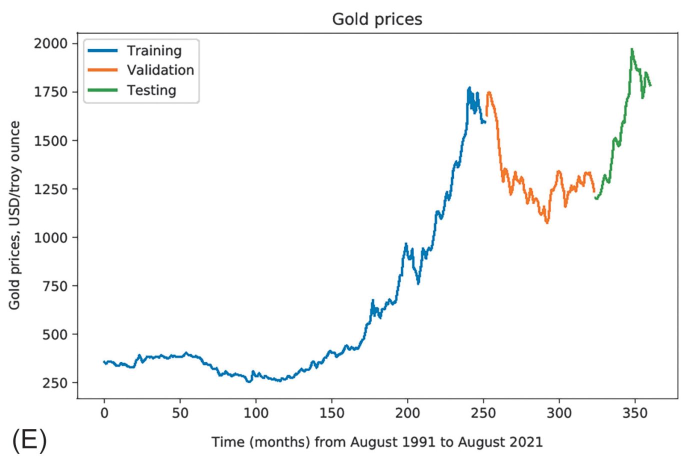


FIG. 4 Time-series dataset of the (A) USD vs CNY exchange rate, (B) USD vs AUD exchange rate, (C) aluminum price, (D) oil price, (E) gold price, and (F) iron ore price.

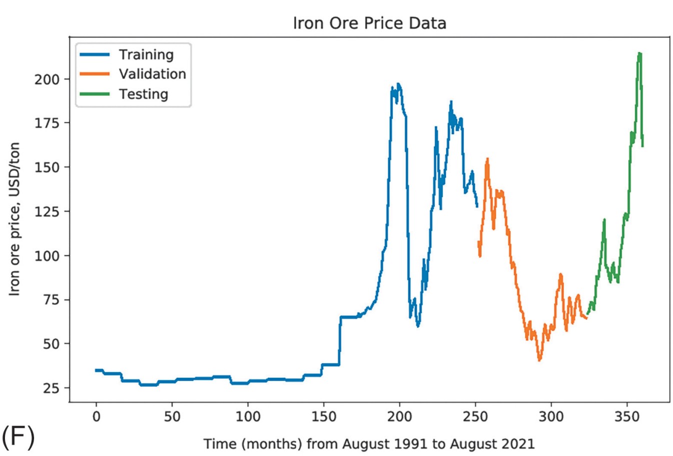
(Continued)



(D)



(E)



(F)

FIG. 4, CONT'D.

4 Results and discussion

In this book chapter, the main goal is to develop deep learning models for forecasting iron ore prices based on multivariate time-series data. To accomplish this, the first step was to normalize the dataset using the Min-Max scaler method. As the problem at hand is a time-series problem, the input and output data for the models were defined using five timesteps in each input sequence. With the dataset preprocessed, the structures of the LSTM and CNN models were established. The LSTM model was designed with a single hidden layer containing 16 neurons. Meanwhile, the CNN model had 32 filters and a kernel size of three for the Conv1D layer, a pool size of two for the max pooling 1D layer, and 10 neurons in the hidden layer. The "ReLU" activation function was applied to the CNN model, and both models were trained using the "Adams" algorithm.

To prevent overfitting, the early stopping technique was implemented with the best weights restored during the training of the CNN and LSTM models. The models were trained with 500 epochs and a batch size of 32, using the mean squared error (MSE) loss function. The optimization results are shown in Fig. 5.

Fig. 5 shows that both the LSTM and CNN models were well-trained and that the early stopping technique was effective in preventing overfitting. The training and validation curves exhibit high convergence during training, indicating that the models can accurately predict the target variable. To further validate the performance of the developed models, the testing dataset was imported to perform a peer test. The results of this test are displayed in Fig. 6, which show that both the LSTM and CNN models are capable of accurately forecasting iron ore prices based on multivariate time-series data.

The results obtained from testing the LSTM and CNN models on the testing dataset indicate that these models are capable of accurately forecasting iron ore prices with high accuracy. Despite the large fluctuations in iron ore prices over the 35 months from September 2018 to August 2021, caused by the COVID-19 pandemic and other economic and policy-related factors worldwide, the developed deep learning models demonstrated robust performance, as shown in Fig. 6. However, as evidenced by the last sample in Fig. 5, these models may encounter difficulties in accurately predicting sudden market changes or disruptions, resulting in decreased accuracy during such periods. To better understand the accuracy of the models, it is important to evaluate them using various performance metrics. Therefore, we calculated Root-Mean-Squared Error (RMSE), Mean Absolute Percentage Error (MAPE), and Variance Accounted For (VAF), as shown in Table 1. These metrics provide a comprehensive view of the performance of the models and their ability to predict iron ore prices accurately.

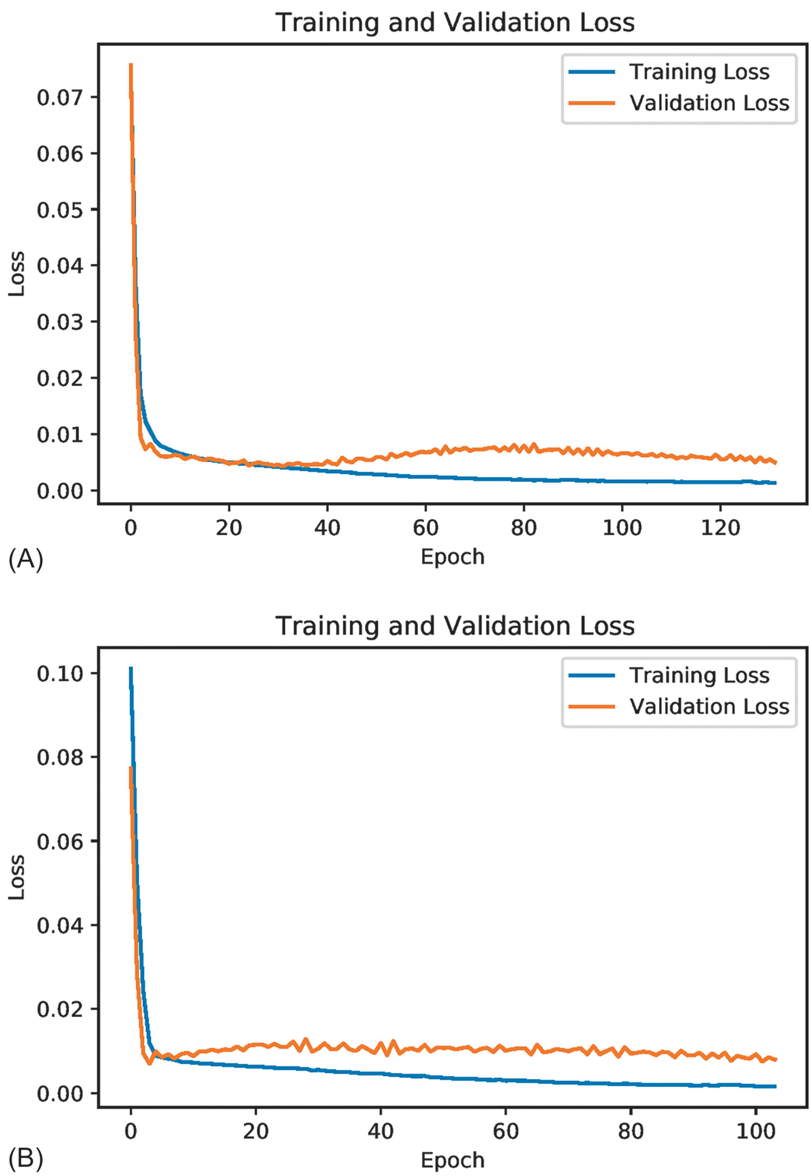


FIG. 5 (A) Training performance of the LSTM model for forecasting iron ore prices and (B) training performance of the CNN model for forecasting iron ore prices.

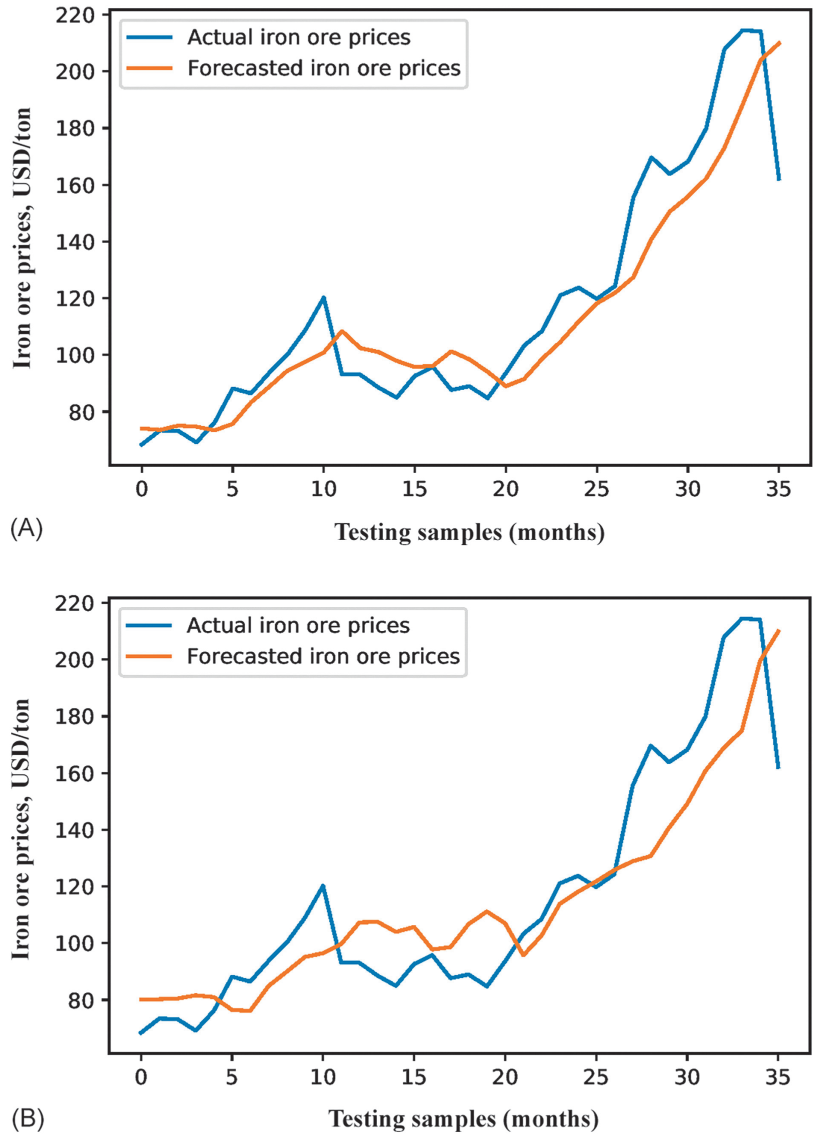


FIG. 6 (A) The forecasting performance of iron ore prices in the LSTM model and (B) the forecasting performance of iron ore prices in the CNN model.

The results of the deep learning models developed in this study have demonstrated that these techniques are effective in forecasting iron ore prices with high accuracy, even in times of large fluctuations such as during the COVID-19 pandemic. The accuracy of the models was evaluated using RMSE, MAPE, and VAF, which are presented in Table 1.

TABLE 1 Performance metrics of the LSTM and CNN models developed.

Model	RMSE	MAPE	VAF
LSTM	15.826	0.100	86.499
CNN	19.207	0.131	79.136

According to the results in Table 1, the LSTM model outperformed as compared to the CNN model in forecasting iron ore prices, achieving an accuracy of approximately 86.5% and a MAPE of 10%. In contrast, the CNN model achieved an accuracy of only 79.136% and a MAPE of 13.1%. These findings suggest that the deep learning techniques used in this study could be a reliable solution for forecasting iron ore prices, despite the accuracy of the models depending on other factors such as the parameters of the topology network and the structure of data.

It should be noted that the accuracy of the models can also be affected by the design of the data structure, such as the number of timesteps. While this study used several timesteps of five, utilizing different numbers of timesteps could potentially result in even higher accuracy. More investigation into these design choices could lead to more precise models for predicting iron ore prices. Furthermore, other types of variables can be included in the model to increase the accuracy of predictions. Nevertheless, this study has established a fundamental basis and uncovered new insights into the possibilities of deep learning techniques for forecasting iron ore prices.

5 Conclusion

The performance metrics calculated, and testing results of the LSTM and CNN models indicate that these models can be practical solutions for forecasting iron ore prices in the multivariate time-series problem. These models were able to capture the complex temporal patterns and non-linear relationships within the data, leading to more accurate predictions than traditional methods. Accurate predictions of iron ore prices can provide valuable insights for businesses and policymakers to make informed decisions and manage risks in the global market. Therefore, deep learning techniques such as the LSTM and CNN models can be considered useful tools for predicting the prices of iron ore.

However, there are limitations to these models that should be considered, such as the requirement for large amounts of data to train the models effectively and the models may struggle with predicting the sudden changes or disruptions in the market.

Based on the obtained results of this work, future works in this area could focus on improving the models' ability to handle sudden changes in the market by incorporating real-time data and external factors such as global pandemics, and political conflicts in the world. Additionally, the use of hybrid models combining LSTM and CNN with other forecasting techniques, such as linear regression or support vector machines, may provide even more accurate predictions.

Acknowledgments

The authors extend heartfelt gratitude to Hanoi University of Mining and Geology (Vietnam) and Curtin University (Australia) for their invaluable cooperation and support.

References

[1] Z.X. Wang, Y.F. Zhao, L.Y. He, Forecasting the monthly iron ore import of China using a model combining empirical mode decomposition, non-linear autoregressive neural network, and autoregressive integrated moving average, Appl. Soft Comput. 94 (2020) 106475, https://doi.org/10.1016/j.asoc.2020.106475 (accessed 03.04.23).

[2] T. Eisele, S. Kawatra, A review of binders in iron ore pelletization, Miner. Process. Extr. Metall. Rev. 24 (1) (2003) 1–90. In this issue.

[3] Y. Kim, A. Ghosh, E. Topal, P. Chang, Relationship of iron ore price with other major commodity prices, Miner. Econ. 35 (2022) 295–307, https://doi.org/10.1007/s13563-022-00301-x (accessed 04.02.23).

[4] A. Pustov, A. Malanichev, I. Khobotilov, Long-term iron ore price modelling: marginal costs vs incentive price, Resour. Policy 38 (4) (2013) 558–567, https://doi.org/10.1016/j.resourpol.2013.09.003 (accessed 15.12.22).

[5] Y. Chen, J. Xu, J. Miao, Dynamic volatility contagion across the Baltic dry index, iron ore price and crude oil price under the COVID-19: a copula-VAR-BEKK-GARCH-X approach, Resour. Policy 81 (2023) 103296, https://doi.org/10.1016/j.resourpol.2023.103296 (accessed 25.03.23).

[6] A.A. Ewees, M.A. Elaziz, Z. Alameer, H. Ye, J. Zhang, Improving multilayer perceptron neural network using chaotic grasshopper optimization algorithm to forecast iron ore price volatility, Resour. Policy 65 (2020) 101555, https://doi.org/10.1016/j.resourpol.2019.101555 (accessed 23.03.23).

[7] Y. Kim, A. Ghosh, E. Topal, C. Ping, Performance of different models in iron ore price prediction during the time of commodity price spike, Resour. Policy 80 (2023) 103237, https://doi.org/10.1016/j.resourpol.2022.103237 (accessed 27.03.23).

[8] J. Lv, W. Tang, H. Hosseinzadeh, Developed multiple-layer perceptron neural network based on developed search and rescue optimizer to predict iron ore price volatility: a case study, ISA Trans. 130 (2022) 420–432, https://doi.org/10.1016/j.isatra.2022.04.025 (accessed 31.03.23).

[9] J. Tuo, F. Zhang, Modelling the iron ore price index: a new perspective from a hybrid data reconstructed EEMD-GORU model, J. Manage. Sci. Eng. 5 (3) (2020) 212–225, https://doi.org/10.1016/j.jmse.2020.08.003 (accessed 02.04.23).

[10] X. Li, W. Shang, S. Wang, Text-based crude oil price forecasting: a deep learning approach, Int. J. Forecast. 35 (4) (2019) 1548–1560, https://doi.org/10.1016/j.ijforecast.2018.07.006. In this issue.

[11] H.B. Ameur, S. Boubaker, Z. Ftiti, W. Louhichi, K. Tissaoui, Forecasting commodity prices: empirical evidence using deep learning tools, Ann. Oper. Res. (2023), https://doi.org/10.1007/s10479-022-05076-6.

[12] Y. Hu, J. Ni, L. Wen, A hybrid deep learning approach by integrating LSTM-ANN networks with GARCH model for copper price volatility prediction, Phys. A Stat. Mech. Appl. 557 (2020) 124907, https://doi.org/10.1016/j.physa.2020.124907 (accessed 15.03.23).

[13] Y. Liang, Y. Lin, Q. Lu, Forecasting gold price using a novel hybrid model with ICEEMDAN and LSTM-CNN-CBAM, Expert Syst. Appl. 206 (2022) 117847, https://doi.org/10.1016/j.eswa.2022.117847 (accessed 21.03.23).

[14] A. Vidal, W. Kristjanpoller, Gold volatility prediction using a CNN-LSTM approach, Expert Syst. Appl. 157 (2020) 113481, https://doi.org/10.1016/j.eswa.2020.113481. In this issue.

[15] J. Ni, Y. Xu, Z. Li, J. Zhao, Copper price movement prediction using recurrent neural networks and ensemble averaging, Soft. Comput. 26 (17) (2022) 8145–8161, https://doi.org/10.1007/s00500-022-07201-w (accessed 19.03.23).

[16] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition, in: Procedure of the IEEE, 1998. http://yann.lecun.com/exdb/publis/pdf/lecun-01a.pdf. (accessed 04.04.23).

[17] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural Comput. 9 (8) (1997) 1735–1780, https://doi.org/10.1162/neco.1997.9.8.1735 (accessed 04.04.23).

[18] R. Chandra, A. Jain, D.S. Chauhan, Deep learning via LSTM models for COVID-19 infection forecasting in India, PLoS One 17 (1) (2022) e0262708, https://doi.org/10.1371/journal.pone.0262708 (accessed 04.04.23).

[19] T. Liu, T. Wu, M. Wang, M. Fu, J. Kang, H. Zhang, Recurrent neural networks based on LSTM for predicting geomagnetic field, in: 2018 IEEE International Conference on Aerospace Electronics and Remote Sensing Technology (ICARES), IEEE, 2018, https://doi.org/10.1109/ICARES.2018.8547087 (accessed 04.04.23).

[20] Y. Liu, C. Yang, K. Huang, W. Gui, Non-ferrous metals price forecasting based on variational mode decomposition and LSTM network, Knowl. Based Syst. 188 (2020) 105006, https://doi.org/10.1016/j.knosys.2019.105006 (accessed 04.04.23).

[21] A.C. Ozdemir, K. Buluş, K. Zor, Medium-to long-term nickel price forecasting using LSTM and GRU networks, Resour. Policy 78 (2022) 102906, https://doi.org/10.1016/j.resourpol.2022.102906 (accessed 04.04.23).

[22] T.N. Wiesel, D.H. Hubel, Receptive fields of single neurones in the cat's striate cortex, J. Physiol. 148 (3) (1959) 574–591. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1363130/. (accessed 03.04.23).

[23] K. Fukushima, Neocognitron: a self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position, Biol. Cybern. 36 (1980) 193–202. https://www.rctn.org/bruno/public/papers/Fukushima1980.pdf. (accessed 04.04.23).

[24] D. Song, X. Yuan, Q. Li, J. Zhang, M. Sun, X. Fu, L. Yang, Intrusion detection model using gene expression programming to optimize parameters of convolutional neural network for energy internet, Appl. Soft Comput. 134 (2023) 109960, https://doi.org/10.1016/j.asoc.2022.109960 (accessed 22.03.23).