Financial Market Forecasting using RNN, LSTM, BiLSTM, GRU and Transformer-Based Deep Learning Algorithms

T.O. Kehinde
Department of Industrial and Systems Engineering,
The Hong Kong Polytechnic University, Kowloon, Hong Kong
temitope.kehinde@connect.polyu.hk

Waqar Ahmed Khan
Department of Industrial Engineering and Engineering Management, College of Engineering, University of Sharjah, P.O. Box 27272, Sharjah, United Arab Emirates
wakhan.dr@gmail.com

Sai-Ho Chung
Department of Industrial and Systems Engineering,
The Hong Kong Polytechnic University, Kowloon, Hong Kong
nick.sh.chung@polyu.edu.hk

Abstract

In recent years, there has been a notable surge of interest in deep learning techniques due to their potential application in predicting financial market movements. Their proficiency in effectively handling the complex, unpredictable, and dynamic nature of financial markets establishes them as valuable resources for both investors and scholars. The aim of this study is to conduct a comprehensive assessment of the predictive precision of five deep learning models, namely RNN, LSTM, BiLSTM, GRU, and Transformer, in forecasting the performance of prominent global stock indices such as the FTSE 100, S&P 500, and HSI. The study demonstrated that the Transformer model exhibited higher accuracy and more efficient convergence compared to other models across several datasets, as assessed by commonly used evaluation metrics like Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Huber loss, and Log-Cosh. On the other hand, the Recurrent Neural Network (RNN), despite its relatively straightforward architecture, frequently reached convergence within a comparable range of epochs as many sophisticated models. However, it significantly fell behind in terms of predicting performance. The Long Short-Term Memory (LSTM), Bidirectional LSTM (BiLSTM), and Gated Recurrent Unit (GRU) models demonstrated comparable performances, but with some dependency on the particular dataset. The Transformer model exhibited greater forecasting accuracy in comparison to its peers across all datasets and performance criteria. The results of our study underscore the effectiveness of the Transformer model in predicting future returns in financial markets. This suggests that incorporating this model into investment strategies can yield significant advantages, such as higher returns.

Keywords
Stock market prediction, deep learning, neural network, LSTM, and Transformer

1. Introduction
The prediction of financial market trends, specifically in relation to stock prices, is a subject that garners significant attention and holds great significance (kehinde et al. 2023a). This is mainly owing to the ever-changing and unpredictable nature of stock movements. Significant fluctuations can compromise the instability of global financial systems (Anagnostidis et al. 2016), as exemplified by the events that unfolded during the 2008 financial crisis (Apergis and Dastidar 2023). Throughout history, methodologies such as technical and fundamental analysis have been the basis for forecasting stock market trends (Krishnapriya and James 2023). Nevertheless, in light of the unpredictable and turbulent environment characterized by increased instability and the abundance of vast amounts of data

influencing market patterns, a growing demand arises for more advanced and intricate models. This phenomenon has resulted in a growing propensity towards the utilization of machine learning algorithms (Khan et al. 2020a; Khan et al. 2020b; Khan 2023), particularly deep learning models (Cavalcante et al. 2016). Stocks are financial instruments that symbolize a partial ownership stake in a corporation (Kehinde et al. 2023b), offering the opportunity for financial rewards when the company's value increases. Historically, corporations have employed the practice of issuing stocks as a mechanism to generate funds, whereas investors perceive it as a channel for the accumulation of wealth. The act of investing in stocks presents a potential for significant financial gains. Yet, it is essential to acknowledge that this endeavour is not without its share of risks, primarily stemming from the unpredictable nature of stock price fluctuations. The utilization of effective forecasting techniques can provide traders and market regulators with valuable solutions to minimize financial losses and optimize investment returns. Nevertheless, the task of forecasting stock price fluctuations is widely recognized as a formidable endeavour, primarily due to the intrinsic non-linear nature of the market, its inherent volatility, and its vulnerability to a multitude of domestic and international influences. The financial field has observed notable accomplishments in utilizing deep learning frameworks for the purpose of predicting stock market trends. Frameworks such as Artificial Neural Networks (ANNs), Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Long Short-Term Memory Neural Networks (LSTMs) have continuously demonstrated superior performance compared to older frameworks (Fischer and Krauss 2018). Among these options, the LSTM model stands out due to its ability to effectively retain and exploit sequential data patterns over extended periods (Selvin et al. 2017). For individuals interested in gaining insights into the most recent advancements in the utilization of deep learning techniques for financial prediction, it is recommended to consult the scholarly articles of Jiang (2021), Kumbure et al. (2022), Raj et al. (2022), and Nazareth and Reddy (2023). Though, it is essential to acknowledge that despite the considerable progress made in stock prediction through deep learning techniques, existing architectures such as ANN, CNN, RNN, and LSTM possess certain limitations. Persistent issues include the requirement for extensive data, vulnerability to overfitting, and difficulties in handling some typical time series data patterns. The difficulties raised have been effectively addressed by the introduction of the Transformer architecture by Vaswani et al. (2017), marking a significant milestone in the field of deep learning. The Transformer model, incorporating a self-attention mechanism, presents a notable advantage in terms of parallel training capabilities. This enables the model to efficiently capture global data patterns, surpassing the sequential structures of RNNs and LSTM networks. While the Transformer model has been widely adopted for analyzing unstructured data, its application to structured datasets, such as stock market technical indicators, is still in its early stage (Wang et al. 2022).

1.1 Objectives
This study aims to build prediction models for three important stock indices: the Financial Times Stock Exchange (FTSE) 100 Index, the Standard & Poor's 500 (S&P500) Index and the Hang Seng Index (HSI). This study is novel because it models the daily closing price and volume in its forecasts. The selection of these major market indices was deliberate, as they represent the economic strength of Europe, North America and Asia. Five advanced deep learning models, namely RNN, LSTM, BiLSTM, GRU, and the Transformer model, were utilized in this study. We evaluate their effectiveness by assessing their performance across five metrics: mean absolute error (MAE), mean squared error (MSE), root mean squared error (RMSE), Huber loss, and Log-Cosh loss. This analysis aims to determine the most efficient model among them. The subsequent sections of this paper will provide a more comprehensive analysis. Section 2 will review relevant literature, Section 3 will provide further information on the selected deep learning models, Section 4 will outline the methods used for data gathering and preprocessing, Section 5 will describe the outcomes of our research, and Section 6 will present the final remarks.

2. Literature Review
With the rapid advances in Artificial Intelligence (AI), machine learning and deep learning have emerged as principal methodologies for stock prediction. They have found usefulness in forecasting stock prices, indices, trends, and market upheavals. Academic research in the field of finance has extensively endeavoured to unravel the complex array of factors that impact the returns on financial investment, especially in the stock market domain. The predominant objective in the past has been focused on explanatory modelling; nevertheless, there has been a growing emphasis on predictive modelling since the 2000s. This shift in attention can be attributed to the availability of larger datasets, which have unveiled intricate linkages. The forecasting of financial time series is significantly challenging due to their inherent chaotic and complex nature (Tao et al. 2023). The initial theories suggested that stock prices adhered to a random walk pattern, rendering them unpredictable. This concept was summarized in the efficient market hypothesis, which refutes the possibility of anomalous returns. Nevertheless, this hypothesis has encountered both scrutiny and endorsement throughout its existence. The conventional approach to financial time series analysis relied on the

assumptions of linearity and stationarity, frequently employing the linear regression model. Nevertheless, the presence of non-linear connections and other complexities within financial time series has resulted in the inadequacy of traditional linear models. Deep learning models, specifically ANNs, have emerged as viable alternatives due to their ability to effectively process non-linear and intricate data. ANNs have played a pivotal role in the advancement of deep learning techniques. Kara et al. (2011) showcased that ANN offers more precise predictions when juxtaposed with SVM. Similarly, Lin et al. (2021) highlighted the shortcomings of SVM in predicting large-scale stock data. In their exploration of ensemble machine learning for daily market patterns, they surmised that incorporating technical indicators often boosts prediction accuracy. Nabipour et al. (2020) initiated a comparative study between deep learning and machine learning models on discrete and continuous data sets. Their study incorporated a gamut of machine learning models like decision trees, random forests, SVM, and more, while on the deep learning front, they analyzed LSTM and RNN. Interestingly, their results favoured models working with binary data over those with continuous data. While fundamental and technical analyses are traditional pillars for stock forecasting, the influx of unstructured textual data, such as financial news, social media comments, and earnings reports, has shifted the paradigm. Sentiment analysis, a key focus of natural language processing, has increasingly been applied to textual data, aiming to filter sentiments and predict market bearings accordingly (Rajput and Bobde 2016). Picasso et al. (2019) innovatively integrated sentiment and technical analysis, leveraging deep learning algorithms for market trend predictions. This sentiment-centric approach was further bolstered by Jin et al. (2020) and Köksal and Özgür (2021), who harnessed LSTMs with attention mechanisms and analyzed Twitter datasets, respectively. Building on the ethos of technical analysis that stock prices swiftly integrate new market information, deep learning models are primed to discern patterns from historical data for accurate future value predictions (Long et al. 2019). A seminal work by Selvin et al. (2017) tested the efficacy of CNNs, RNNs, and LSTMs for stock prediction, with LSTMs showcasing unparalleled prowess due to their inherent sequence memory capabilities. Technical analysis posits that stock market prices promptly assimilate all new information. Leveraging the historical patterns embedded in past data, deep learning models are primed to deliver precise future value forecasts than their machine learning counterparts (Long et al. 2019). RNNs have been specifically designed to handle temporal dependencies, while LSTM networks have further enhanced the memory capacity of RNNs to handle the challenge of vanishing gradients. However, it is important to acknowledge that LSTM models also possess specific limitations.

In recent times, the Transformer model, which was initially devised for the purpose of natural language processing, has exhibited promising capabilities in the domain of financial forecasting. The utilization of parallel processing and attention mechanisms in this system enables it to surpass conventional models in specific tasks. Ding et al. (2020) pioneered the application of a Transformer-based model for stock trend prediction, showcasing its superior performance over LSTM models. Building on this, the authors introduced a refined methodology that enhances the Transformer model through the integration of a multi-scale Gaussian prior, orthogonal regularization, and a trading gap splitter. This approach is lauded for its exceptional ability to discern long-term, short-term, and hierarchical intricacies within financial time series. Empirical tests on two real-world exchange markets further validate its superior performance against numerous benchmark methods. Mohammadi Farsani and Pazouki (2020) showcased the effectiveness of the Transformer architecture is grounded in its self-attention mechanism. Their research, using electricity consumption and traffic patterns datasets, illustrated that the Transformer model offers enhanced accuracy in time-series forecasting with reduced computational demands. In another study, Yoo et al. (2021) employed a methodology to forecast future stock price fluctuations by leveraging correlations between multiple stocks. The method employed by the researchers, DTML, utilizes a data-axis transformer that incorporates multi-level contexts. This approach aims to acquire knowledge regarding the dynamic and asymmetric connections among stocks by analyzing past prices and market index data. The authors assert that their approach attains accuracy levels that are considered the most advanced in the field, as well as generating financial gains across six distinct datasets originating from various countries. More recently, Muhammad et al. (2023) introduced a deep-learning model based on transformers for the purpose of predicting stock prices. The model is trained and evaluated using data from the Dhaka Stock Exchange (DSE), which is recognized as the central stock market in Bangladesh. The time series characteristics are encoded using the time2vec method, and afterwards, the transformer model is employed on eight individual stocks using their historical daily and weekly data. The authors demonstrate that their model attains favourable outcomes and satisfactory RMSE across a majority of the stocks, hence suggesting the potential effectiveness of transformer-based models in the domain of stock price prediction. Other existing studies, such as Köksal and Özgür (2021), predominantly pivot around the model's aptitude for sentiment analysis, gleaning insights from textual sources like financial news and social media commentary. This present research diverges from this norm by harnessing robust and sophisticated deep learning techniques, including RNN, LSTM, BiLSTM, GRU, and Transformer, to forecast major stock market indices, sidestepping its reliance on unstructured textual datasets. The analysis encompasses pivotal notable indices like FTSE100, S&P500 and HSI. While initial investigations indicate that Transformers hold potential as a new avenue for financial forecasting, their application in financial research remains relatively unexplored. The aim of this study is to further explore the capabilities and potential of the Transformer model, together with other deep learning models in the specific domain of financial market forecasting.

3. Methods

This section provides an overview of the chosen deep-learning models employed for stock market prediction in our model development process.

3.1 RNN

The most challenging part of the time series prediction problem is figuring out how to model all the interdependent data. RNN is one of the earliest attempts, and it solves this issue by inserting a memory cell, an internal state that stores historical data. Although RNN is able to accurately characterize the contextual relationship between sequential data, this relationship weakens as the gap distance between them grows. Back-propagation issues, including disappearing gradients and gradient explosions, have been linked to RNNs' propensity for long-term reliance (Huang et al. 2019). Figure 1 depicts a typical structure of RNN framework.

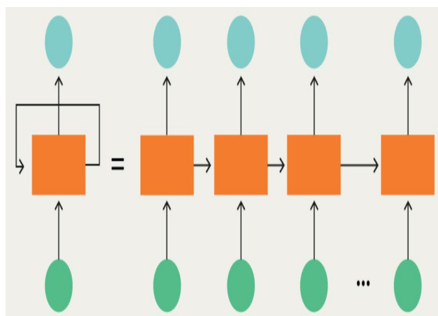

Figure 1. RNN Architecture.

3.2 LSTM

LSTM is a type of RNN that can enhance models with a specific gate structure. There are three gates involved in the interaction: the input gate, the forget gate, and the output gate. The relevant data will be retained and transmitted to the next neuron, while the irrelevant details can be ignored to make room. Figure 2 shows a typical structure of LSTM framework.

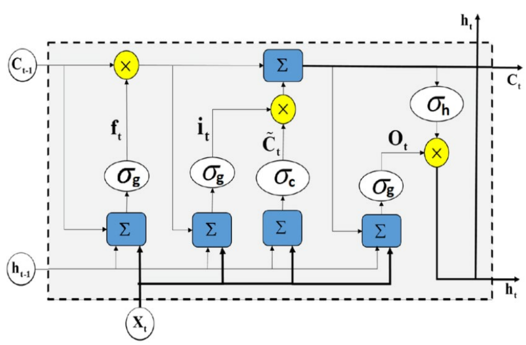

Figure 2. LSTM Architecture.

3.3 BiLSTM

Bidirectional long short-term memory (BiLSTM) is a type of RNN that may learn long-term dependencies between time steps in time series or sequence data in both directions. These types are helpful if you want the RNN to absorb information about the entire time series at each iteration.

3.4 GRU
In 2014, Cho et al. (2014) presented gated recurrent units (GRUs) as a gating technique for recurrent neural networks. Similar to a long short-term memory (LSTM) with a forget gate, the GRU lacks an output gate and hence has fewer parameters than an LSTM. There have been instances where GRU performed better than LSTM.

3.5 Transformer
The Transformer comprises embeddings, an encoder, a decoder, self, and multi-head attention. The Transformer's embedding layer is a simple linear one. The linear layer's output is sent to the transformer encoder module. Encoder transformer modules can have as many as N layers. Each transformer encoder module has a multi-head attention layer, followed by a feed-forward network layer. The output of each layer is combined and normalized in the encoder and decoder modules. The preprocessed data is sent into the decoder. After being processed by the linear layer, this data is sent to the decoder subsystem. The decoder component consists of a feed-forward network followed by two multi-head attention layers. The first multi-head attention layer receives the linear layer's output as its input. The output of the previous multi-head attention layer is fed into the next multi-head attention layer, as is the case with the final encoder module. The layer's output was sent to a feed-forward network. The final layer's output from the transformer decoder is sent into the linear layer to calculate the expected closing price. The framework of the proposed model is depicted in Figure 3.

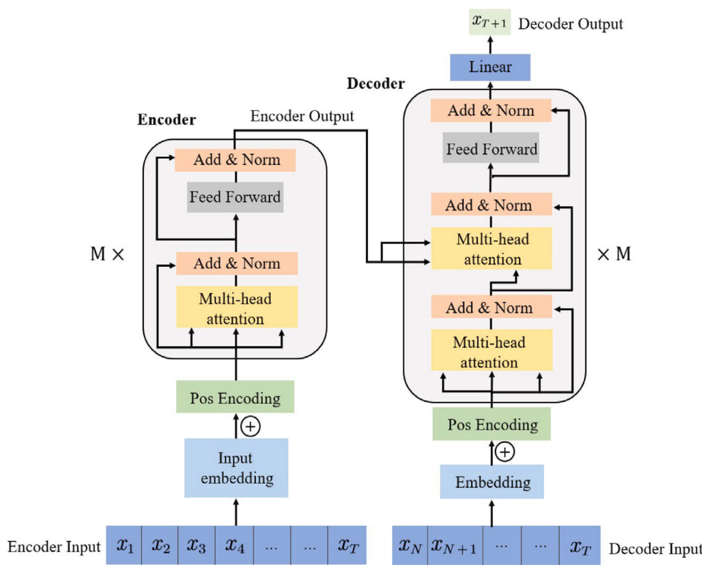

Figure 3. Transformer framework.

4. Data Collection
We build our models and run them 15 times to ensure that the results are consistent across the FTSE100, S&P500, and HSI, the three most widely followed stock market indices. The information was gathered from the publicly available Yahoo Finance database. All three stock indices, such as their high price, low price, adjusted close, and volume, were recorded, as well as their respective opening and closing prices. There are ten full years' worth of data collected, beginning on January 1, 2013, and ending on December 31, 2022. Google Colab is used as our hardware processor to execute all programs.

4.1 Preparing the Data
Only the Closing Price and Volume were used in the original set of datasets. Then, the datasets were examined to see whether any variables were missing. There were no blanks in the information we gathered. After that, we did a 70:30 split across the datasets. The performance indicators are tested on a subset (around 30%) of the data used for testing. The data volatility was reduced by using MinMax Standardization, which smoothed out the numbers and limited them to values between 0 and 1. The equation below further explains how standardization is achieved.

$$a = \frac{x - \max}{\max - \min}$$

Where a is the standardized price at particular time t.
For the output, we estimate the output of the present day at time t, using the closing value of the next day at time, t+1.

4.2 Hyperparameters tuning

We explore different hyperparameter values on the training set to determine an optimal solution. We use MAE as a loss function to compare predicted and actual outputs. Adam optimizer is used in the training phase, while a batch size of 128 is used in all the developed models. An epoch is a single iteration over the full training dataset when training a model using machine learning. The iterative optimization algorithms used to train models, such as gradient descent, rely on this principle heavily. The model runs through the training data in mini-batch sizes at the start of each epoch and adjusts its internal parameters based on the mistakes or gradients it calculates. The objective is to maximize the learning rate while minimizing the loss function. As a hyperparameter, the number of epochs controls how many times the model will iteratively examine all of the data in the training set. Selecting an appropriate epoch size helps to limit the model overfitting and underfitting. Reducing overfitting with an early stopping epoch model is an optimization strategy that does not sacrifice model accuracy. Overfitting can be avoided if training is terminated early enough, which is the basic principle behind early stopping. In this work, 1000 epochs were set, and an early stopping epoch was implemented to avoid wastage of energy and time.

5. Results and Discussion

This section analyses the performance metrics of each of the five models under consideration. The performance evaluation uses the loss function during the back-testing experiments.

5.1 Performance Analysis and Discussion

The results of our experiment are duly represented using Table 1-4, and Figure 4-10. The analysis of each Table and Figure goes thus; First, Table 1 and Figure 4 show the error analysis for the FTSE 100 index dataset. From Table 1, after conducting an analysis of the performance measures of the FTSE 100 index data, it becomes apparent that the RNN model, which is considered the most rudimentary model, lags behind in accuracy across all metrics. This observation implies that the RNN model may have certain difficulties in effectively reflecting the complexities of the stock index. On the contrary, the LSTM model exhibits significant enhancements; nonetheless, its accomplishments are overshadowed by the more sophisticated BiLSTM, GRU, and Transformer models. The Transformer and BiLSTM models are particularly notable in this context, with the former exhibiting slightly better performance in terms of metrics such as MAE, Huber Loss, and Log Cosh. This suggests that the Transformer model possesses an improved capacity to identify intricate patterns. Nevertheless, the comparable performance of the model with BiLSTM in terms of RMSE and MSE highlights the effectiveness of both models in forecasting the movements of the FTSE. This comparative analysis highlights the advantages of utilizing advanced deep learning models such as Transformer and BiLSTM in financial forecasting applications that involve complex datasets such as the FTSE 100 index.

Second, Table 2 and Figure 5 show the error analysis for the S&P 500 index dataset. The models' prediction accuracy on the S&P 500 data strongly resembles the patterns identified in the FTSE 100 index dataset but with relatively smaller magnitudes of error. The RNN consistently exhibits the most significant errors in all metrics, highlighting the limits of its basic architecture in effectively interpreting complex financial inputs. The LSTM model demonstrates enhanced performance compared to the RNN; however, it is surpassed by more sophisticated models. Both the Transformer and BiLSTM models are notable; however, the Transformer model demonstrates higher performance across all metrics. Specifically, the algorithm's superior performance in metrics such as MAE, Huber Loss, and Log Cosh suggests its enhanced ability to accurately capture complex patterns and relationships within the S&P 500 dataset. This comparison emphasizes the ongoing supremacy of advanced deep learning models, particularly the Transformer model, which stands out as the favoured option for financial forecasting tasks related to the S&P 500 index.

Third, as evident in Table 3 and Figure 6, the HSI index datasets analysis revealed notable insights dissimilar from the patterns observed in the FTSE 100 and S&P 500 datasets. The errors exhibited by the RNN and LSTM models are comparable in terms of the MAE, RMSE, and MSE metrics. This indicates that, in the case of HSI data, the increased complexity of the LSTM model does not yield a substantial benefit over the RNN model. Notably, the BiLSTM, which exhibited favourable outcomes in earlier datasets, demonstrates suboptimal performance with significantly elevated error rates across several metrics, deviating from the established patterns. The GRU model exhibits an intermediate level, as its error rates tend to be higher than those of the Transformer model but lower than those of the BiLSTM model. The Transformer model demonstrates superior accuracy, as seen by its consistent performance over previous datasets, consistently achieving the lowest error rates across all metrics. This highlights the proficiency of the model in modelling and predicting complex financial time-series data, as demonstrated by all three stock market indices.

Next, from Table 4 and Figure 7, considering the epoch usage, the Transformer model demonstrates continuous convergence at a rapid pace across the FTSE 100, S&P 500, and HSI datasets, highlighting its notable efficiency in training and performance. It is noteworthy that the RNN, despite its straightforward architecture, often necessitates a similar number of epochs as more sophisticated models such as the LSTM and GRU. This observation implies that although the RNN may iterate in a similar manner, it may not optimize as efficiently. There is a noticeable degree of variability in the stopping epochs, whereby the LSTM model requires the highest number of epochs for the HSI dataset but exhibits a closer alignment with other models for the FTSE 100 dataset. In contrast, the convergence speeds of GRU and BiLSTM frequently exhibit comparable performance. The disparities in convergence across datasets emphasize that while some models, like the Transformer, are universally efficient, the suitability of others might be more context-specific, reaffirming the importance of a dataset-specific approach in model selection.

Table 1: Error Analysis for FTSE 100 stock index dataset.

MODELS	**RNN**	**LSTM**	**BiLSTM**	**GRU**	**Transformer**
MAE	0.0319	0.0202	0.0153	0.0155	0.0149
RMSE	0.036	0.0249	0.0197	0.02	0.0197
MSE (10^{-2})	0.13	0.06183	0.038929	0.040151	0.038877
Huber Loss (10^{-2})	0.065111	0.031118	0.019508	0.020107	0.019357
Log Cosh (10^{-2})	0.065086	0.03111	0.019503	0.020102	0.019351

Table 2: Error Analysis for S&P 500 stock index dataset.

MODELS	**RNN**	**LSTM**	**BiLSTM**	**GRU**	**Transformer**
MAE	0.013	0.0089	0.0066	0.0062	0.0054
RMSE	0.0142	0.0104	0.0085	0.0082	0.0077
MSE (10^{-2})	0.020238	0.010847	0.007242	0.006669	0.005961
Huber Loss (10^{-2})	0.010417	0.005522	0.003636	0.003305	0.002917
Log Cosh (10^{-2})	0.010416	0.005522	0.003635	0.003305	0.002917

Table 3: Error Analysis for HSI stock index dataset.

MODELS	**RNN**	**LSTM**	**BiLSTM**	**GRU**	**Transformer**
MAE	0.0129	0.0127	0.016	0.0142	0.0123
RMSE	0.0166	0.0166	0.0197	0.0178	0.0165
MSE (10^{-2})	0.027703	0.02759	0.038794	0.031837	0.027196
Huber Loss (10^{-2})	0.013536	0.013595	0.019334	0.015703	0.013264
Log Cosh (10^{-2})	0.013535	0.013592	0.019331	0.0157	0.013262

Table 4: Stopping Epoch

INDEX	**RNN**	**LSTM**	**BiLSTM**	**GRU**	**Transformer**
FTSE100	27	35	29	31	16
S&P500	20	35	23	23	16
HSI	23	40	25	34	17

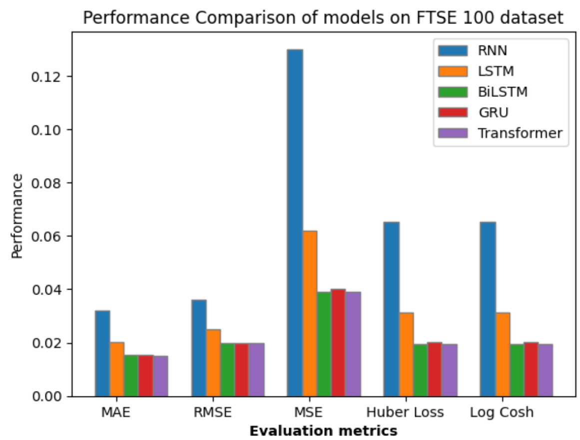

Figure 4: Model performance on FTSE100 index.

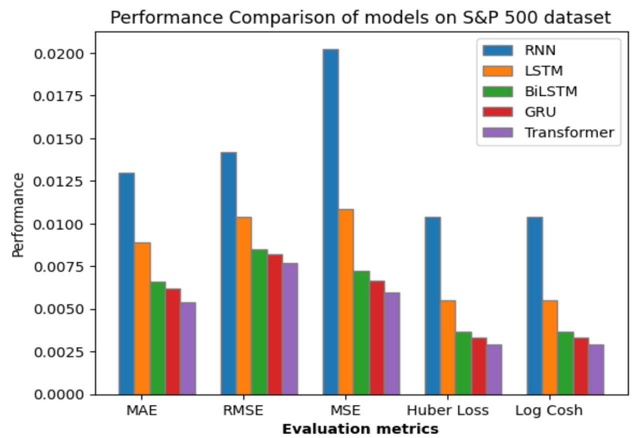

Figure 5: Model performance on S&P500 index.

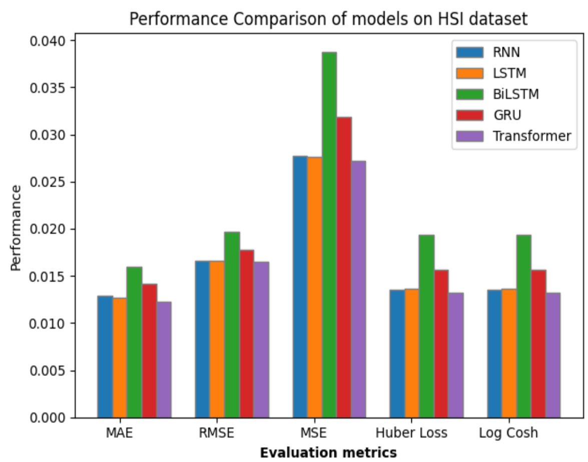

Figure 6: Model performance on HSI.

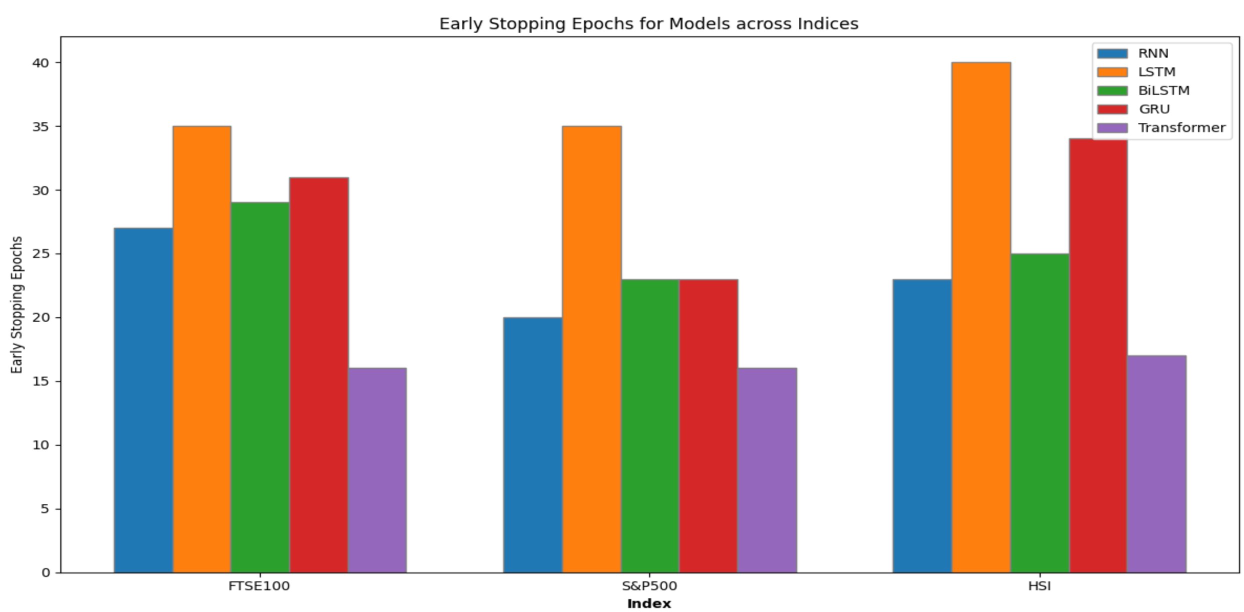

Figure 7: Early Stopping Epoch plot for all three indices.

Last, the loss functions of all the individual stock market index models were plotted as shown in Figure 8-10.

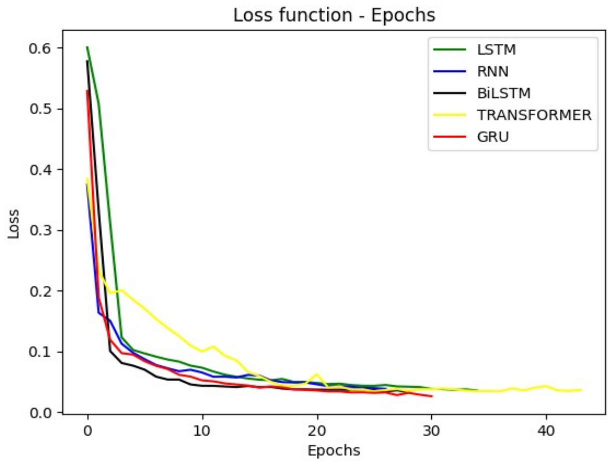

Figure 8: FTSE 100 index model loss function.

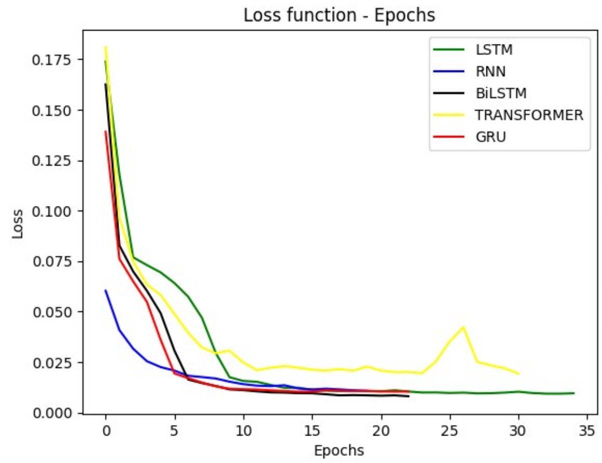

Figure 9: S&P500 index model loss function.

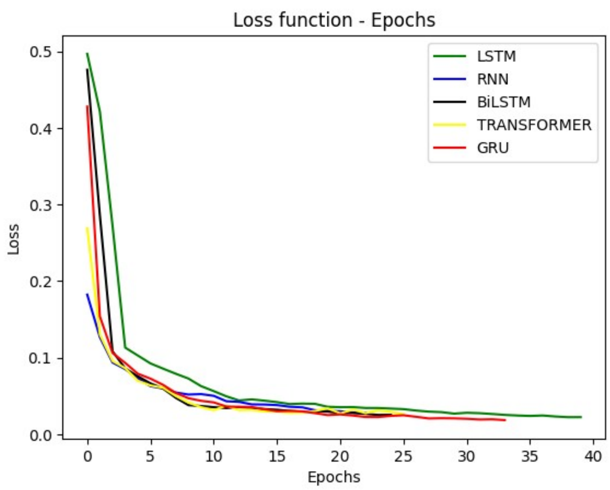

Figure 10: HSI model loss function.

6. Conclusion

In this work, we conducted a meticulous evaluation of five advanced deep learning models, namely RNN, LSTM, BiLSTM, GRU, and Transformer, to determine the most proficient model for predicting the performance of three global stock indexes. The evaluation was performed on the FTSE 100, S&P 500, and HSI indices. Based on all the five evaluation metrics considered, including MAE, RMSE, MSE, Huber Loss, and Log Cosh, the results clearly emphasized the exceptional capabilities of the Transformer model, showcasing its superior accuracy and efficient convergence. In contrast, the RNN exhibited a delay in performance, whereas the LSTM, BiLSTM, and GRU displayed varying levels of competence, depending on the specific dataset under consideration. Nevertheless, it is imperative to recognize the limitations of the study. The evaluation, albeit thorough, was limited to the predetermined designs and datasets, thus disregarding other influential models or external economic factors that may have further effects on forecasting accuracy. Furthermore, it should be noted that the evaluation criteria now employed, although widely accepted, may not fully encompass all aspects of predictive accuracy. This highlights the necessity for a more comprehensive approach to assessing performance. In anticipation of future studies, a promising landscape exists for further investigation. Future works can venture and perform similar experiments in other markets such as cryptocurrency, bond market, and forex market, exploring different iterations of the Transformer model, including external economic or geopolitical factors, or utilizing ensemble techniques to combine the advantages of many models present potential avenues worth considering. Moreover, the utilization of developing hybrid deep learning models has the potential to enhance the forecasting process, potentially achieving higher accuracy levels. In summary, even though the Transformer has emerged as a leading contender in this field of study, the ever-evolving domain of financial market prediction calls for periodic appraisal to understand and model the dynamic nature of a typical financial market system.

Acknowledgement

This research was funded by Hong Kong Polytechnic University using student account code RLN7. The authors acknowledge the financial and technical assistance provided by the Hong Kong Polytechnic University Research Committee.

References

Anagnostidis, P., Varsakelis, C. and Emmanouilides, C., Has the 2008 financial crisis affected stock market efficiency? The case of Eurozone, *Physica A: statistical mechanics and its applications*, vol. 447, pp. 116-128, 2016.

Apergis, N. and Dastidar, S., Local stock liquidity and local factors: fresh evidence from US firms across states. *Research in International Business and Finance*, 102112, 2023.

Cavalcante, R., Brasileiro, R., Souza, V., Nobrega, J. and Oliveira, A., Computational intelligence and financial markets: A survey and future directions, *Expert Systems with Applications*, vol. 55, pp. 194-211, 2016.

Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H. and Bengio, Y., Learning phrase representations using RNN encoder-decoder for statistical machine translation, *arXiv preprint arXiv:1406.1078*, 2014.

Ding, Q., Wu, S., Sun, H., Guo, J., and Guo, J., Hierarchical Multi-Scale Gaussian Transformer for Stock Movement Prediction, *In IJCA,* pp. 4640-4646, 2020.

Fischer, T. and Krauss, C., Deep learning with long short-term memory networks for financial market predictions, *European journal of operational research*, vol. 270, no. 2, pp. 654-669, 2018.

Huang, Y., Shen, L. and Liu, H., Grey relational analysis, principal component analysis and forecasting of carbon emissions based on long short-term memory in China, *Journal of Cleaner Production*, vol. 209, pp. 415-423, 2019.

Jiang, W., Applications of deep learning in stock market prediction: recent progress, *Expert Systems with Applications*, vol. 184, pp. 115537, 2021.

Jin, Z., Yang, Y. and Liu, Y., Stock closing price prediction based on sentiment analysis and LSTM, *Neural Computing and Applications*, vol. 32, pp. 9713-9729, 2020.

Kara, Y., Boyacioglu, M. and Baykan, Ö., Predicting direction of stock price index movement using artificial neural networks and support vector machines: The sample of the Istanbul Stock Exchange, *Expert systems with Applications*, vol. 38, no. 5, pp. 5311-5319, 2011.

Kehinde, T., Chan, F. and Chung, S., Scientometric review and analysis of recent approaches to stock market forecasting: Two decades survey, *Expert Systems with Applications*, vol. 213, pp. 119299, 2023.

Kehinde, T., Chung, S., and Chan, F., Benchmarking TPU and GPU for Stock Price Forecasting Using LSTM Model Development*, In Science and Information Conference Cham: Springer Nature Switzerland,* pp. 289-306*,* July 2023.

Khan, W., Chung, S., Awan, M. and Wen, X., Machine learning facilitated business intelligence (Part I) Neural networks learning algorithms and applications. *Industrial Management & Data Systems*, vol 120, no. 1, pp.164-195, 2020.

Khan, W., Chung, S., Awan, M. and Wen, X., Machine learning facilitated business intelligence (Part II) Neural networks optimization techniques and applications. *Industrial Management & Data Systems*, vol 120, no. 1, pp.128-163, 2020.

Khan, W., Balanced weighted extreme learning machine for imbalance learning of credit default risk and manufacturing productivity. *Annals of Operations Research*, pp.1-29, 2023.

Köksal, A. and Özgür, A., Twitter dataset and evaluation of transformers for Turkish sentiment analysis, *2021 29th Signal Processing and Communications Applications Conference (SIU)*, pp. 1-4, IEEE, June 2021.

Krishnapriya, C. and James, A., A Survey on Stock Market Prediction Techniques, In *2023 International Conference on Power, Instrumentation, Control and Computing (PICC), IEEE,* pp. 1-6, April 2023.

Kumbure, M., Lohrmann, C., Luukka, P. and Porras, J., Machine learning techniques and data for stock market forecasting: A literature review, *Expert Systems with Applications*, vol. 197, pp. 116659, 2022.

Lin, Y., Liu, S., Yang, H. and Wu, H., Stock trend prediction using candlestick charting and ensemble machine learning techniques with a novelty feature engineering scheme, *IEEE Access*, vol. 9, pp. 101433-101446, 2021.

Long, W., Lu, Z. and Cui, L., Deep learning-based feature engineering for stock price movement prediction, *Knowledge-Based Systems*, vol. 164, pp. 163-173, 2019.
Mohammadi Farsani, R. and Pazouki, E., A transformer self-attention model for time series forecasting, *Journal of Electrical and Computer Engineering Innovations (JECEI)*, vol. 9, no. 1, pp. 1-10, 2020.
Muhammad, T., Aftab, A., Ibrahim, M., Ahsan, M., Muhu, M., Khan, S., and Alam, M.., Transformer-based deep learning model for stock price prediction: A case study on Bangladesh stock market. *International Journal of Computational Intelligence and Applications*, 2350013, 2023.
Nabipour, M., Nayyeri, P., Jabani, H., Shahab, S. and Mosavi, A., Predicting stock market trends using machine learning and deep learning algorithms via continuous and binary data; a comparative analysis, *IEEE Access*, vol. 8, pp. 150199-150212, 2020.
Nazareth, N. and Reddy, Y., Financial applications of machine learning: A literature review, *Expert Systems with Applications*, pp. 119640, 2023.
Picasso, A., Merello, S., Ma, Y., Oneto, L. and Cambria, E., Technical analysis and sentiment embeddings for market trend prediction, *Expert Systems with Applications*, vol. 135, pp. 60-70, 2019.
Raj, P., Mehta, A. and Singh, B., Stock Market Prediction Using Deep Learning Algorithm: An Overview, *International Conference on Innovative Computing and Communications: Proceedings of ICICC 2022,* vol. 2, pp. 327-336, Singapore, September, 2022.
Rajput, V. and Bobde, S., Stock market forecasting techniques: literature survey, *International Journal of Computer Science and Mobile Computing*, vol. 5, no. 6, pp. 500-506, 2016.
Selvin, S., Vinayakumar, R., Gopalakrishnan, E., Menon, V. and Soman, K., Stock price prediction using LSTM, RNN and CNN-sliding window model, *2017 international conference on advances in computing, communications and informatics (icacci)*, pp. 1643-1647, September 2017.
Tao, Z., Wu, W., and Wang, J., Series decomposition Transformer with period-correlation for stock market index prediction. *Expert Systems with Applications*, 121424, 2023.
Vaswani, A., Shazeer, N., Parmar, N. and Uszkoreit, J., Attention is all you need in Advances in Neural Information Processing Systems, *Search PubMed*, pp. 5998-6008, 2017.
Wang, C., Chen, Y., Zhang, S. and Zhang, Q., Stock market index prediction using deep Transformer model, *Expert Systems with Applications*, vol. 208, pp. 118128, 2022.
Yoo, J., Soun, Y., Park, Y., and Kang, U., Accurate multivariate stock movement prediction via data-axis transformer with multi-level contexts. *In Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining,* pp. 2037-2045, 2021.

Biographies

Kehinde Temitope is a Ph.D. candidate at the Department of Industrial and Systems Engineering, The Hong Kong Polytechnic University, Hong Kong. He has published in journal and conference articles. His research interest includes portfolio optimization using MCDM techniques, Machine Learning applications for financial market prediction, Inverse Data Envelopment Analysis, Stochastic Optimization and many more. Temitope is a student member of many professional bodies, including the European Operations Management Association (EurOMA), Industrial Engineering and Operations Management (IEOM), Institute of Industrial and Systems Engineers (IISE), International Association of Engineers (IAENG), and Production and Operations Management Society (POMS).

Waqar Ahmed Khan received a Ph.D. in Industrial and Systems Engineering (ISE) from the Hong Kong Polytechnic University (PolyU) in 2020. He is currently an Assistant Professor with the Department of Industrial Engineering and Engineering Management, College of Engineering, University of Sharjah, P.O. Box 27272, Sharjah, United Arab Emirates. He has published in journals such as TRC, TRE, IJPR, IMDS, and ANOR. His research interests include deep learning, transportation, and Industry 4.0.

Sai-Ho Chung, Ph.D., is an Associate Head and Associate Professor of the ISE at the PolyU. He obtained his Ph.D. degree from the University of Hong Kong. His research interests are in the field of logistics, supply chain management, supply chain finance, production scheduling, distribution network, machine learning, container port terminal, aviation, etc. He has published over 100 SCI journal papers. His publications appear in POM, TR (Part B/C/E), DSJ, Risk Analysis, EJOR, IEEE (SMC, TIE, EM, SJ), DSS, IJPE, IJPR, COR, etc. He serves as an editorial board member in TRE and edited several special issues in SCI journals.