

A Comparative Study of the Effectiveness of Machine Learning for Predicting Stock Prices in the New Energy Vehicle Industry

Binrong Yang *

Jinan University-University of Birmingham Joint Institute, Jinan University, Guangzhou, China

* Corresponding Author Email: bxy254@student.bham.ac.uk

Abstract. This paper investigates the effectiveness of machine learning models in predicting stock prices in the New Energy Vehicle (NEV) industry - a key part of the current transformation of the global transport system and electricity supply. This study applies five models, ARIMA, LSTM, GRU, CNN and TCN using the sample 20 NEV companies, to predict stock prices. And further use the risk metrics such as MSE, RMSE, R^2 , MAE and MAPE to measure the prediction performance. The results shows that the CNN model stands out for its superior accuracy in predicting stock movements, highlighting its potential as a valuable tool for investors and financial analysts in the fast-growing NEV market. The findings highlight the importance of utilizing advanced neural network algorithms to improve investment strategies and risk assessment in the face of market volatility and economic fluctuations.

Keywords: Deep learning model; CNN; LSTM; Stock prices; New Energy Vehicle Industry.

1. Introduction

In the 21st century, the new energy vehicle market has succumbed as a centre for global focus because global consensus on cutting carbon footprint and attaining a sustainable future is being strengthened. Above all, the electric cars have enjoyed maximum due to their key role of freeing from oil lines and air pollution reduction by developed states like the USA, China, and European Union. Those countries and regions are playing a major role basing on these by providing to the core which have expedited the establishment of new energy vehicle industry through means such as providing subsidies on vehicle purchase, setting up of charging infrastructure as well as having the sites' run on stringest emission standards on their products. The breakthrough technology and pricing consolidation have led the electric cars to the transition from niche markets to the mass markets thereby increasingly providing a better choice among the consumers for sustainable transportation.

Besides, it can be seen that the huge trend of energy for its clean and sustainable environment across the whole globe. In such environment, nothing else develops so fast if not new energy vehicle (NEV) market still in making. The investors and financial analysts have lately demonstrated a noticeable interest to know the exact Store prices of companies which are part of the NEV industry, but at the same time the high volatility and stability of that particular sector is adding more to the complexity of exactly predicting their stock performance. The previous research will acknowledge five models where ARIMA, long short-term memory networks (LSTM), gated recurrent units (GRU), convolutional neural networks (CNN), and Temporal Convolutional Networks (TCN) are exploiting the strengths and potentials of these recurrent architectures in matters related to non-linear patterns and long-term time series data.

This research will cater to the effect and correctness of these software/techniques in predicting stock prices in the new energy cars/vehicles sector. Because the energy vehicles industry has been and remains one of the swiftly expanding sectors of the global economy, the perpetuation of the task of sufficiently obtaining and publishing data for effective stock price forecasting has become particularly essential. When investors make the comparison of different forecasting models, there is a guarantee of gaining better knowledge of which models will be of help in a new energy vehicles industry environment and ultimately make the right investment decisions. Markets analysts can make the most out of the two studies in this sense. First, the study has enumerated the models' strengths and weaknesses. A very vital part is the contribution of the study in providing the market analysts

with a yardstick to compare models and hence, the improvements of forecasting strategies may be obtained for better and reliable forecasts. Last but not the least, accurate stock price prediction in new energy vehicle sector not only enhances investors' certainty but also flourishes consolidated and healthy market.

2. Literature review

Stock technology forecasting with machine learning models has been prevalent since the 1980s. There has been inexorable development in the science field within the 21st century, which is depicted by the integration of deep learning technologies, such as CNN, into these research areas. Cao and Wang have illustrated the superiority of deep learning algorithms over the traditional back-propagation technique as well as the convenient Support Vector Machine (SVM) model in stock price prediction [1]. In addition, the results have shown that exploiting a mix of traditional machine learning models with deep learning models has enabled the increasing adoption of the latter for intelligent systems development [1]. Tsai and Wang's study, which combined Artificial Neural Networks (ANN) and Decision Trees (DT), achieved an accuracy level of 77% in the electronics industry, then it showed better prediction performance compared to using ANN or DT alone [2]. This indicates the significance of the need to experiment with different machine learning models when it comes to stock price prediction. Furthermore, Shahi, et al. have verified that LSTM and GRU forecasts for stock prices are widely used and valuable neural network models in the field [3]. Sim, Kim, and Ahn have explicitly shown that CNN stand out by having the flexibility of a model that adds runs in every iteration to address the mismatches of traditional neural networks and SVM in terms of accuracy, indicating deep learning models as the soon-to-be dominant technology [4]. On the other hand, Fischer and Krauss have confirmed the practicality of LSTM in financial market prediction, stressing the possibility that deep learning technologies can provide solutions in the financial forecasting of complex tasks [5].

In the niche of stock price forecasting in the new energy auto industry, Chen Menglong, Fan, and Wu have employed machine learning methods, such as tree models and bagging techniques, and conducted an empirical study. It turns out that these methods have generated accurate stock price predictions [6]. To be more precise, they have proved that random forest models are better in terms of accuracy and reliability than other typical machine learning algorithms [6]. These discoveries suggest the assistance of machine learning methods in forecasting stock prices in the new energy vehicle field. Nevertheless, though the fast progression and successful use of deep learning technology in different areas, giving a closer look at the possibilities of this technology in modeling the stock price of the new energy vehicle industry would be reasonable. Future stock market predictions would greatly benefit when deep learning applications are expanded, as they thrive on sophisticated non-linear data. For instance, Wang et al. also used cloud computing when developing a forecasting model, which proves the diversity in modeling methods and indicates that multitudinous models can be used to predict changing stock prices of new energy vehicles [7]. Accordingly, a combination or comparison of these different models should be considered as a norm for future studies to take advantage of the best features of each technology in new energy vehicle stock price prediction.

This study focuses on the nuanced capabilities of machine learning models in the new energy vehicle stock market, further refining the bridge to the application of deep learning in stock prediction, especially in the rapidly developing new energy vehicle sector.

3. Methodology

3.1. Data sources & processing

3.1.1. Data sources

Noticeable things about this research: It refers to the data of 20 companies that have been meticulously chosen as to be the best representative of the global new energy vehicles (NEV) market. The selection has been made on the ground of the authoritative sources such as, the International Energy Agency's "Global EV Outlook 2023" report (IEA, 2023), and the analysis provided by one of the financial websites, specifically, Xueqiu (<https://xueqiu.com/5516314868/249621153>). These exact same websites and reports researched according to an extensive online search and literature review in order for the chosen companies to fully encompass the core trends and market performance of the NEV industry.

Table 1. Selected companies and their ticker symbols

Company	Ticker Symbol	Company	Ticker Symbol
BYD Company Limited	byddf	Renault S.A.	rno.pa
Volkswagen AG	vwagy	Nissan Motor Co., Ltd.	7201.t
General Motors Company	gm	Mitsubishi Motors Corporation	7211.t
Stellantis N.V.	stla	Toyota Motor Corporation	7203.t
Hyundai Motor Company	hymtf	SAIC Motor Corporation Limited	601127.ss
BMW AG	bmw.de	NIO Inc.	NIO
Daimler AG	mbg.de	XPeng Inc.	XPEV
Tesla, Inc.	tsla	Lucid Group, Inc.	LCID
Geely Automobile Holdings Limited	0175.hk	Rivian Automotive, Inc.	RIVN
Volvo Group	volv-b.st	Li Auto Inc.	LI

The data collected for this study include the daily adjusted closing prices from January 1, 2019, to February 29, 2024, of the aforementioned companies as recorded on the Yahoo Finance platform. The choice of adjusted closing prices as the data foundation for the study is due to its consideration of factors such as stock splits and dividends, which more accurately reflect changes in stock value, thereby providing a more accurate measure for analyzing long-term stock price trends. Through in-depth analysis of the data within this period, this study aims to reveal patterns in stock price movements within the NEV industry, providing a scientific basis for forecasting future price trends.

3.1.2. Data preprocessing

Through in-depth analysis of the data within this period, this study aims to reveal patterns in stock price movements within the NEV industry, providing a scientific basis for forecasting future price trend. To maintain a consistent timeline across the stock data, initial datasets retained missing values due to non-trading days. However, for independent stock analysis, these missing values were later removed to ensure complete datasets for reliable model training.

The creation of this preliminary dataset marks the initial phase of data preparation. Subsequent steps, including data normalization and feature scaling (which is often synonymous with normalization), are critical for preparing the data for specific models. These preprocessing steps are essential to adjust the scale of the data, thereby facilitating the convergence of algorithms and improving model performance. The whole step for data preparation, that is, data preprocessing for individual models, will be presented and exploitation in a way that meets the specific model need.

3.2. Model Selection

The objective of this research is to ascertain which machine learning model most effectively forecasts stock prices within the new energy vehicle sector. To this end, five models have been selected for comparison due to their distinct characteristics and proven capabilities in handling time-series data: ARIMA, LSTM, GRU, CNN, and TCN.

ARIMA (Autoregressive Integrated Moving Average) represents a conventional model for forecasting time series data, distinguished by its proficiency in identifying and leveraging linear trends and connections within the data. Given its robustness in dealing with stationary series, ARIMA serves as a foundational benchmark in our comparative analysis, especially for datasets exhibiting clear trends or seasonal patterns.

LSTM (Long Short-Term Memory) networks, which is a branch of RNN, perform best in modeling long-term dependencies in sequence data [8]. If compared to other memory cells, the ability of the LSTM to remember something over longer time distances makes them very well suitable for the stock price prediction of companies operating in the new energy vehicles sector, where long-term trends and cyclicity impact performance.

GRU (Gated Recurrent Unit), which is the next evolution of RNN, enhances the LSTM architecture, and the network is capable of memorizing long-term sequential structures [8]. The efficiency of GRU in training and prediction, alongside their effectiveness in sequence modeling, renders them a compelling choice for this study.

CNN (Convolutional Neural Networks), though predominantly associated with image processing, have demonstrated remarkable success in analyzing sequential data. By capturing local dependencies and recognizing patterns across time, CNN offer a novel approach to time-series forecasting, including stock price predictions.

TCN (Temporal Convolutional Networks) integrate the strengths of traditional convolutional networks with the specific demands of time-series analysis. Through dilated convolutions, TCN efficiently model long-range dependencies, making them particularly promising for financial time-series data where past events may have long-term effects [9].

Each of these models brings a unique perspective to stock price forecasting, with distinct advantages and potential limitations. The forthcoming analysis will critically assess their performance in the context of new energy vehicle industry data, focusing on their ability to handle the intricacies of financial market fluctuations and predict future trends with accuracy.

3.3. Evaluation standards

The performance of the selected machine learning models—ARIMA, LSTM, GRU, CNN, and TCN—in forecasting stock prices within the new energy vehicle sector will be evaluated using a comprehensive set of statistical metrics. These metrics, detailed below with their respective formulas, are chosen to assess accuracy and precision from multiple dimensions:

Mean Squared Mistake (MSE): This statistic determines the average made even deviations between predicted worths and actual values, showing the mean degree of settled mistakes.

Root Mean Squared Mistake (RMSE): By taking the square root of MSE, we obtain RMSE, which aims to use an error magnitude metric that is simpler to analyze and comprehend.

R-squared (R^2): Acting as an important scale of version fit, R^2 shows the proportion of the variance in the reliant variable that can be anticipated from the independent variables.

Mean Absolute Mistake (MAE): This statistic steps the average magnitude of errors in a set of forecasts, overlooking the instructions of mistakes, thus ignoring whether the mistakes are positive or unfavorable.

Mean Outright Portion Mistake (MAPE): This metric stands for the ordinary absolute deviation in between forecasted worths and real worths as a percent of the true values, giving a standardized indication of error dimension [10].

Each metric contributes unique insights into the forecasting capabilities of the models, allowing for a detailed and nuanced evaluation of their performance in predicting stock prices in the new energy vehicle sector.

4. Results & discussion

4.1. Preliminary Screening with ARIMA

The ARIMA model was employed as a baseline to establish a point of reference for subsequent comparisons. However, its performance was found to be substantially inadequate. The model's inability to capture the dynamic nature of stock price fluctuations was evident, as reflected in the graph portraying BYD Company's stock prices, where ARIMA's predictions resembled a horizontal line, failing to match the actual price movements.

The chart below, using BYD as an example, shows that ARIMA does not accurately capture price movements.

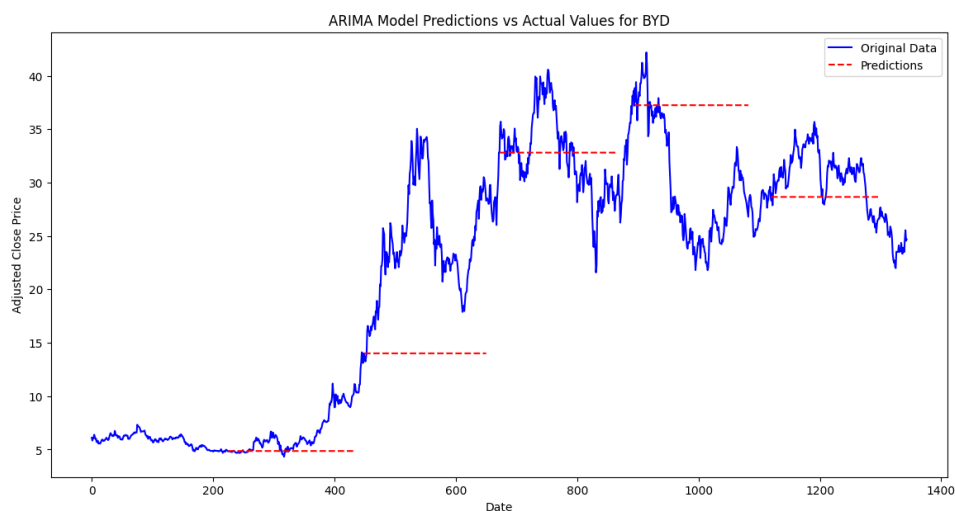


Fig 1. Predicted and actual values of the ARIMA model using BYD as example.

4.2. Deep Learning Model Cross-Validation

Upon excluding ARIMA from further analysis due to its poor performance, we proceeded with cross-validation of the deep learning models. The comparative results were visually represented in a radar chart, encapsulating multiple performance metrics such as RMSE, MSE, MAE, MAPE, and the coefficient of determination (R^2).

4.3. Model Performance & Prediction Accuracy

The data revealed that the CNN consistently outperformed the other models in terms of lower RMSE, MSE, and MAE values, indicating higher accuracy in forecasting stock prices. The reduced MAPE value suggested that the percentage error in predictions was minimal, reinforcing the model's efficacy.

Table 2. Performance of different models of BYD.

Model	MSE	RMSE	R^2	MAE	MAPE
LSTM	1.4033	1.0583	0.9142	0.8380	3.3449
GRU	0.0018	0.0364	0.8555	0.0277	81036.5416
CNN	2.3899	1.2838	0.8668	1.0367	4.3066
TCN	4.7985	2.0398	0.7291	1.5720	6.8067

Note: Due to space constraints, this article only shows data for two representative companies out of 20 companies

Table 3. Performance of different models of Tesla.

Model	MSE	RMSE	R ²	MAE	MAPE
LSTM	185.9795	12.1987	0.8991	9.9858	5.0666
GRU	0.0030	0.0469	0.7129	0.0371	9.4861
CNN	107.1216	10.1386	0.9354	7.8439	4.5010
TCN	570.9363	23.4882	0.6752	17.7648	10.7282

Counter to this result, though, the LSTM and GRU models were effective in capturing temporal sequences but they lagged behind the CNN model in terms of detecting temporal dependencies. The TCN, similar the CNN, had potential benefits but in the end, its performance was poorer.

CNN model which is the most efficient of the models has the largest R². Similarly R² values and error metrics at a low score determine the CNN as having accurate predictive power and ability to reflect the trend and volatility of stock exchange. This ability to have its cake and eat it too, by being both accurate and the proper fit, makes the CNN model the most intriguing, in any sort of stock price prediction.

A comparison of the different metrics between the models can be seen in the following figure,

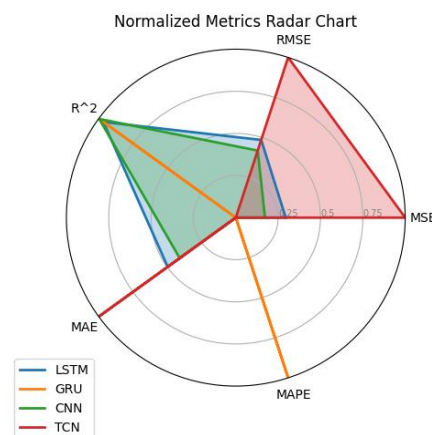


Fig 2. Radar chart comparing the performance of different models.

A comprehensive comparative analysis resulted in the conclusion that in terms of performance and irritation, the CNN model was mostly preferable. The predictive value of this model was confirmed by detailing the stock prices forecast of the simulation on BYD and Tesla. In accuracy of the stock prices predicted by our CNN model we found that they were highly correlated with the actual stock prices. This is a good sign of the great model's efficiency.

Below you can see that the model's predictions for BYD and Tesla's prices are very close to the true values.

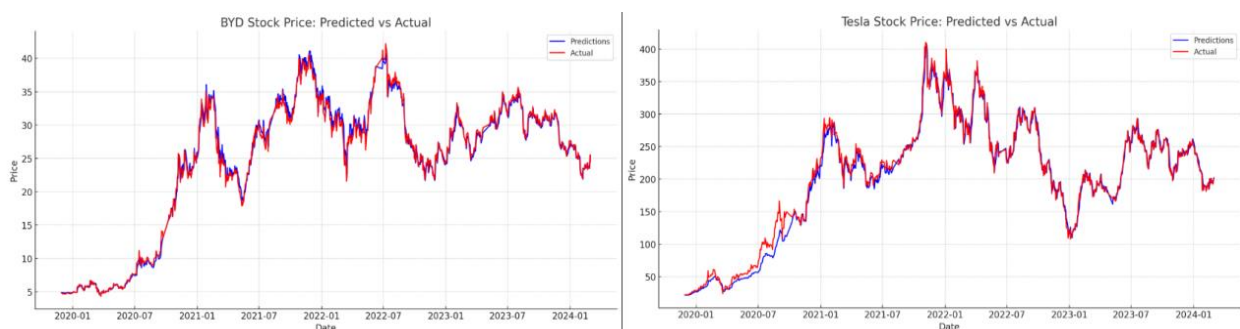


Fig 3. Predicted and actual values of the CNN model using BYD and Tesla as examples.

4.4. Generalization Capability & Variability Analysis

To address potential concerns regarding overfitting and to validate the CNN model's generalization capability, a variability analysis was conducted.

The box plots of the CNN model for three important metrics for different companies are plotted below:

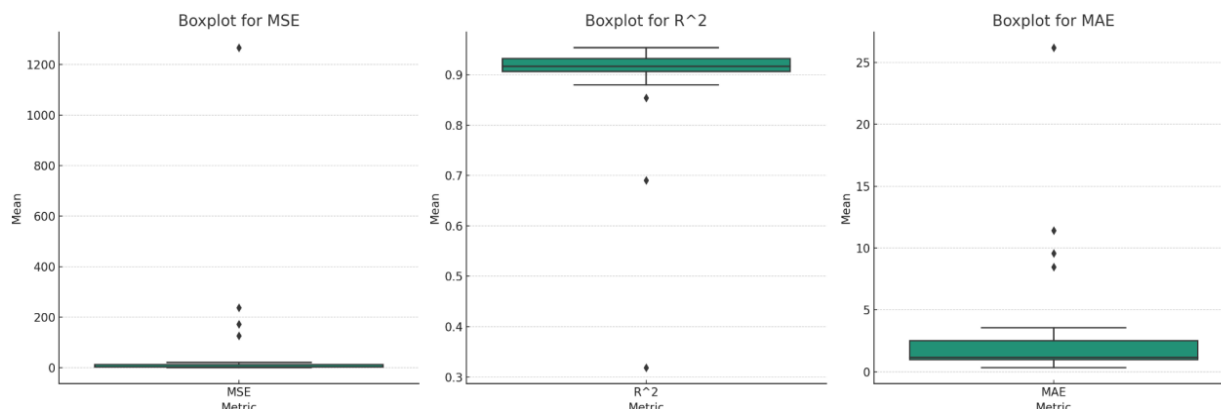


Fig 4. Box plot of metrics distribution for CNN model.

MSE (Mean Square Error): The box plot for this metric shows that the distribution of the model's MSE values across validation sets is relatively concentrated, but there are some outliers. This means that while the model's prediction error is small in most cases, there are still a few cases where the error is large. This may indicate that the model is not performing well with some specific types of data.

R² (Coefficient of Determination): The box plot of the R² values shows that the model explains the variation in the data well on most of the validation sets, with a more concentrated distribution of R² values and a relatively high median. This indicates that the model has good predictive and generalisation abilities. However, as the box plots also show some outliers, this may point to weaker explanatory power of the model in some specific cases.

MAE (Mean Absolute Error): the box plot of MAE shows that the model's prediction error is generally low across validation sets, which is a good sign. However, just like the MSE, there are some outliers in the distribution of the MAE, indicating that the model's prediction accuracy is lower in specific cases.

Boxplots for MSE, R², and MAE were produced, indicating that the model maintained robust performance across a range of scenarios. The CNN model's predictions were not only consistently accurate but also demonstrated a reliable predictive quality in the face of diverse data conditions [11].

5. Conclusion

This research has reached its end and provided fresh insights for financial technology that is here to stay with its advance in stock price prediction, specifically for new energy vehicle (NEV) industry that is emerging as a new dimension. The study outlines its comprehensive and detailed examination and comparison of five neural network algorithms, which gives the CNN model a priority in handling the time-series data. This can be generalised as an important academic research. The indirect reference of the role of machine learning in forecasting the share price volatility, and also the establishment of detection of new standard of classification algorithms, specifically CNN, was made.

The study enhances the comprehension of the factors which impact on the trends of the NEV market while enhancing the creation of reliable predictive models. The results are then used as a point of reference for the ongoing work that can be used to define the AI which has been suggested for bringing preciseness in the forecast and thus providing a way in which AI can be integrated as a part of financial analyses.

Nevertheless, the study is subject to limitation making it possible for future studies to feed from its work. While the model may appear more accurate by focusing only on period-adjusted prices, using trading volume, market sentiment, and other economic statistics can provide better results compared to using only adjusters.

References

- [1] Cao, J., & Wang, J. (2019). Stock price forecasting model based on modified convolution neural network and financial time series analysis. *International Journal of Communication Systems*, 32(12), e3987.
- [2] Tsai, C. F., & Wang, S. P. (2009, March). Stock price forecasting by hybrid machine learning techniques. In *Proceedings of the international multiconference of engineers and computer scientists* (Vol. 1, No. 755, p. 60).
- [3] Shahi, T. B., Shrestha, A., Neupane, A., & Guo, W. (2020). Stock price forecasting with deep learning: A comparative study. *Mathematics*, 8(9), 1441.
- [4] Sim, H. S., Kim, H. I., & Ahn, J. J. (2019). Is deep learning for image recognition applicable to stock market prediction? *Complexity*, 2019.
- [5] Fischer, T., & Krauss, C. (2018). Deep learning with long short-term memory networks for financial market predictions. *European journal of operational research*, 270(2), 654-669.
- [6] Chen, M. L., Fan, C., & Wu, Z. P. (2024). Research on stock price prediction of new energy vehicle industry based on machine learning algorithms. *Journal of Jilin Business and Technology College*, (1), 93-100.
- [7] Wang, M. X., Xiao, Z., Peng, H. G., Wang, X. K., & Wang, J. Q. (2022). Stock price prediction for new energy vehicle enterprises: An integrated method based on time series and cloud models. *Expert Systems with Applications*, 208, 118125.
- [8] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, 9(8), 1735-1780.
- [9] Bai, S., Kolter, J. Z., & Koltun, V. (2018). An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. *arXiv preprint arXiv:1803.01271*.
- [10] Makridakis, S. (1993). Accuracy measures: theoretical and practical concerns. *International journal of forecasting*, 9(4), 527-529.
- [11] Ismail Fawaz, H., Forestier, G., Weber, J., Idoumghar, L., & Muller, P. A. (2019). Deep learning for time series classification: a review. *Data mining and knowledge discovery*, 33(4), 917-963.