Comparing Vision-Instruct LLMs, Vision-Based Deep Learning, and Numeric Models for Stock Movement Prediction

Qizhao Chen
Graduate School of Information Science, University of Hyogo, Kobe, Japan

***Abstract*—This research conducts a comparative study of several stock movement prediction approaches, evaluating large language models (LLMs) and vision-based deep learning models with stock image as input, as well as models that utilize numerical data. Specifically, the study investigates a prompt-based LLM framework that processes candlestick charts, comparing its performance with image-based models such as MobileNetV2, Vision Transformer, and Convolutional Neural Network (CNN), as well as models with numerical inputs including Support Vector Machine (SVM), Random Forest, LSTM, and CNN-LSTM. Although LLMs have demonstrated promising results in stock prediction, directly applying them to stock images poses challenges compared to numerical approaches. To address this, this study further improves LLM performance with post-hoc calibration, reducing prediction biases. Experimental results demonstrate that post-hoc calibrated LLMs with visual input achieve competitive performance compared to other models, highlighting their potential as a viable alternative to traditional stock prediction methods while simplifying the prediction process.**

***Keywords*—Convolutional Neural Network (CNN); Large Language Model (LLM); MobileNetV2; stock price prediction; time series forecasting; vision transformer**

I. Introduction

Forecasting stock price movements has long been a crucial area of research in financial markets. Recent developments in machine learning and deep learning have led to significant improvements in predictive models, allowing better identification of market patterns. Machine learning models, such as decision tree [1], [2] and logistic regression [3], [4], and deep learning frameworks, such as LSTM [5], CNN [6], and transformers [7], have shown great promise in improving the accuracy of stock price forecasts. By leveraging vast amounts of historical price data, technical indicators, and external factors such as news sentiment, these models can capture the non-linear and complex nature of financial markets.

More recently, Large Language Models (LLMs), originally developed for natural language processing (NLP), have been explored in financial forecasting due to their ability to analyze large volume of unstructured text data, such as news articles, financial reports, and social media sentiment [8]. Although LLMs have shown promise in understanding market sentiment and extracting insights from textual data [9], their direct application to visual financial data, such as candlestick charts, remains a challenge. Unlike numerical or text-based input, stock charts require an understanding of spatial and temporal patterns, which LLMs are not inherently designed to process.

Candlestick charts are important tools in technical analysis and they show price changes and movements over time for a stock. Traditionally, their interpretation has to rely on human expertise such as traders or investment analysts. Later, image-based deep learning models such as Convolutional Neural Network (CNN) are used to process and analyze the charts. Nowadays, with the rapid development of LLMs, the interpretation of the candlestick charts can be performed by LLMs. However, integrating prompt-based LLMs for candlestick chart analysis is still an emerging field, and their effectiveness is limited by biases in prediction confidence. Moreover, imbalanced input data poses another challenge. Sometimes, the stock time-series data includes more upward movements than downward movements. However, using traditional resampling techniques might disrupt the temporal dependencies inherent in the data. Therefore, without proper adjustments, LLMs applied to stock images may generate inconsistent or unreliable forecasts, reducing their practical utility in financial decision making [10].

To overcome these limitations, this research proposes a post-hoc calibration framework to improve the reliability of LLM-driven stock movement predictions. Post-hoc calibration techniques, such as Platt Scaling, Isotonic Regression, and Temperature Scaling, are commonly used to refine probabilistic outputs in machine learning models, particularly in classification tasks. By applying these calibration methods to LLM-generated predictions, the LLM performance can be further improved.

Furthermore, if the performance of prompt-based LLMs with post-hoc calibration proves effective, the stock price prediction process can be further simplified. Compared to traditional deep learning models such as CNNs or ViTs, which require extensive feature engineering, LLMs can automatically process raw data with minimal preprocessing. In addition, LLMs not only analyze visual input, but also use their pre-trained knowledge to gain a deeper understanding of financial markets.

Another key objective of this paper is to compare an image-based approach, which uses candlestick charts as input, with a numerical value-based approach for stock price prediction. This comparison is particularly important, as prior research has primarily focused on either visual or numerical data separately, without directly evaluating their relative effectiveness. This analysis offers new insights into the advantages and limitations of each method, contributing to a more comprehensive understanding of their respective impacts on stock prediction accuracy.

The main contributions of this paper are listed below.

- This paper conducts a comparative analysis of LLMs with advanced image-based models, such as MobileNetV2, and numerical input-based models, such as LSTM, in stock movement prediction. The findings highlight that LLMs, when calibrated with visual inputs, can not only simplify the prediction process, but also achieve competitive or superior performance. This demonstrates the potential of prompt-based LLM approaches as an effective alternative to traditional deep learning models for stock market prediction.
- This paper introduces a post-hoc calibration framework designed to improve the accuracy of LLM-generated stock forecasts by using models such as LLaMA and Qwen. This approach refines raw predictions by adjusting for potential biases and inconsistencies, ultimately improving the reliability and robustness of LLM-driven market forecasting.
- This paper also examines the impact of data augmentation techniques on stock image data and finds that proper application of techniques such as rotation and zooming can enhance the performance of deep learning models such as MobileNetV3.

The remainder of this paper is structured as follows. Firstly, related work is listed in Section II. The methodology is then described in Section III. In Section IV, the experimental results are presented. Section V provides a discussion of the results. Finally, Section VI concludes the paper.

II. Related Work

A. Stock Prediction Using Technical Indicators and Sentiment Analysis

Technical indicators, derived from historical price and volume data, are widely used to forecast stock movements. For example, Agrawal et al. [11] apply optimal LSTM together with several technical indicators such as the relative strength index (RSI) and moving average (MA) to predict the price of the stock. Moodi et al. [12] investigate various feature selection techniques to identify the most relevant indicators to predict stock prices. Julian et al. [13] apply Multilayer Perceptron with technical indicators and day-shifting method to predict the stock price.

Beyond numerical indicators, financial sentiment plays a crucial role in stock forecasting. NLP techniques have been widely used to analyze financial news, social media discussions, earnings reports, and analyst opinions. Studies have shown that combining technical indicators with sentiment analysis improves prediction accuracy, as stock prices are influenced by both market trends and investor sentiment [9], [14], [15]. For example, Vargas et al. [16] use deep learning models, including CNN and LSTM, to predict stock market movements by integrating technical indicators with financial news headlines. Khairi et al. [17] utilize a combined approach that integrates technical indicators, fundamental data, and news sentiment to predict stock prices.

B. Post-Hoc Calibration for Prediction Models

Deep learning models, particularly neural networks, often produce overconfident predictions, which can be problematic in high-stakes applications such as financial forecasting. Post-hoc calibration techniques address this issue by adjusting probability scores after model training, ensuring that confidence levels align more accurately with real-world probabilities. Rahimi et al. [18] present a method for post-hoc calibration of neural networks using a novel approach called g-Layers and also provides a theoretical support of post-hoc calibration methods. Furthermore, inspired by the concept of post-hoc calibration, Chen [19] applies Proximal Policy Optimization (PPO) to adjust the LLM-predicted output to improve the model performance.

C. Image-Based Stock Prediction

Compared to numerical features, stock images are less commonly used in stock price prediction. However, some researchers have explored using images as inputs. For example, Steinbacher [20] approaches stock price movement prediction as an image classification problem using CNN model. The study converts financial time series data into images and applies image classification techniques to predict stock price movements. Bang and Ryu [21] apply CNN to predict stock price using stock images but the prediction accuracy is only around 50%. Zhou et al. [22] propose a hybrid framework that integrates an LLM, a Linear Transformer (LT), and a CNN model to forecast stock price. CNN is used to extract features from stock image data. Jin and Kwon [23] study how stock chart characteristics impact the stock price prediction via CNN and they find that the prediction accuracy is improved when using solid lines, color, and a single image without axis marks.

D. Prompt-Based LLM Approach

The prompt-based approach has several advantages. For example, the prompts can be rapidly adjusted to meet the specific needs of the research, allowing for fine-tuning of the model's output to align with the desired results. This flexibility makes it easier to tailor the model's responses for different tasks, ensuring that the outputs are both relevant and precise for the research objectives. Some researchers have applied the prompt-based LLM approach in financial tasks. For example, Chen and Kawashima [9] use a prompt-based approach to compare the performances of several LLMs in financial sentiment analysis. Yang et al. [8] propose a FinGPT model, which is trained on financial data, including news, reports, and market data. Different prompts can be used to perform various tasks in finance.

As mentioned earlier, most existing studies focus exclusively on either image-based approaches or methods that utilize only numerical data for stock price prediction. The relative effectiveness of these two approaches remains unclear. In addition, strategies to enhance the vision-instruct LLM framework for stock forecasting have not been explored in previous research. This study aims to address these gaps in the current literature.

III. METHODOLOGY

Fig. 1 shows the whole picture of this study. For LLMs (LLaMA and Qwen) and image-based deep learning models such as MobileNetV3, candlestick charts generated using the historical stock data are used as input. For other models such as SVM, numerical stock data are directly used as input. Stock movement (Up or Down) prediction made by each model will be used for comparison.

A. Data

Daily stock price data for Apple, Tencent, and Toyota, spanning from 01/05/2015 to 08/05/2024, is retrieved from Yahoo Finance using the Python library yfinance. The data for each stock include the columns "Open", "Close", "High", "Low", and "Volume".

Candlestick charts, with a sliding window size of 20 days, are generated from stock price data using the Python library mplfinance to predict the stock movement (Up or Down) over the next five days. Volume data are incorporated below the candlestick chart to provide insight into trading activity. A sample of the image input is shown in Fig. 2.

B. LLM Model Setup

In this study, two advanced LLMs are utilized: Llama-3.2-11B-Vision-Instruct and Qwen2-VL-7B-Instruct. Both of these models were released in 2024. Llama-3.2-11B-Vision-Instruct, developed by Meta, is designed for visual recognition, image reasoning, captioning, and answering general questions about images. According to Meta, this model outperforms many existing open-source and proprietary multimodal models on common industry benchmarks. Qwen2-VL-7B-Instruct, part of the Qwen series developed by Alibaba Cloud, is also an instruction-following model.

Both models are fine-tuned using ground-truth market movements derived from historical stock prices. For this process, 80% of the image data is allocated for fine-tuning, while the remaining 20% is reserved for validation. The Unsloth Python library is employed for fine-tuning, enabling faster training speeds with reduced memory consumption. The model input will be structured as follows (Table I).

TABLE I. LLM INPUT FORMAT

```
[
{"role": "user",
"content": ["type": "text", "text": prompt,
"type": "image", "image": candlestick chart ],
{"role": "assistant",
"content": ["type": "text", "text": answer ],
]
```

C. Baseline Models

All models are fed with input (images or numerical values) with a sliding window size of 20 days to predict the stock prices over the next five days. Image-based models include MobileNetV2, Vision Transformer (ViT) and CNN. Models with numerical inputs include SVM, Random Forest, LSTM, and CNN-LSTM.

MobileNetV2, ViT and CNN are three prominent deep learning architectures that have shown effectiveness in various computer vision tasks, including image classification, object detection, and stock market prediction based on candlestick charts. Each of these models has distinct characteristics that make them suitable for different aspects of financial time-series forecasting through visual representations.

- MobileNetV2: MobileNetV2, proposed by study [24], is a CNN model designed to optimize performance on mobile platforms. It uses an inverted residual structure, where the residual connections are placed between the bottleneck layers. The intermediate expansion layer employs efficient depthwise convolutions to filter features, introducing non-linearity. In general, the MobileNetV2 architecture includes an initial convolutional layer with 32 filters, followed by 19 residual bottleneck layers. The MobileNet model used in this study is a pre-trained model available in the torchvision library.
- Vision Transformer: ViT, proposed by study [25], represents a paradigm shift in deep learning for image analysis by leveraging self-attention mechanisms instead of convolutions. Unlike CNNs, which rely on local receptive fields, ViT processes input images as sequences of non-overlapping patches and applies a transformer-based architecture to capture long-range dependencies. The ViT model used in this study is based on the pre-trained "google/vit-base-patch16-224" architecture from Hugging Face. The model utilizes 16×16 image patches and processes them through a transformer encoder to capture spatial dependencies.
- Convolutional Neural Networks: CNNs remain a fundamental approach in image-based analysis due to their hierarchical feature extraction capabilities. CNNs utilize convolutional layers to learn spatial hierarchies of features, ranging from low-level edges to high-level structures. This characteristic allows CNNs to effectively capture essential visual elements within candlestick charts, such as trend lines, support and resistance levels, and reversal patterns. The CNN used in this study consists of two convolutional layers: the first with 32 filters of size 3×3, followed by ReLU activation and max pooling, and the second with 64 filters. The feature maps are flattened and passed through a fully connected (FC) layer with 128 neurons, activated by ReLU. The final FC layer outputs a single neuron with a sigmoid activation.

Furthermore, models with numerical inputs such as SVM [26], Random Forest [27], and LSTM [28], [29] are widely used in stock price prediction. Hybrid models such as CNN-LSTM are also popular candidates [30].

- Support Vector Machine (SVM): SVM is a supervised learning algorithm used for classification and regression tasks. The main logic behind SVM is to find the optimal hyperplane that maximizes the margin between different classes in the feature space. The kernel used in SVM in this paper is Radial Basis Function (RBF).

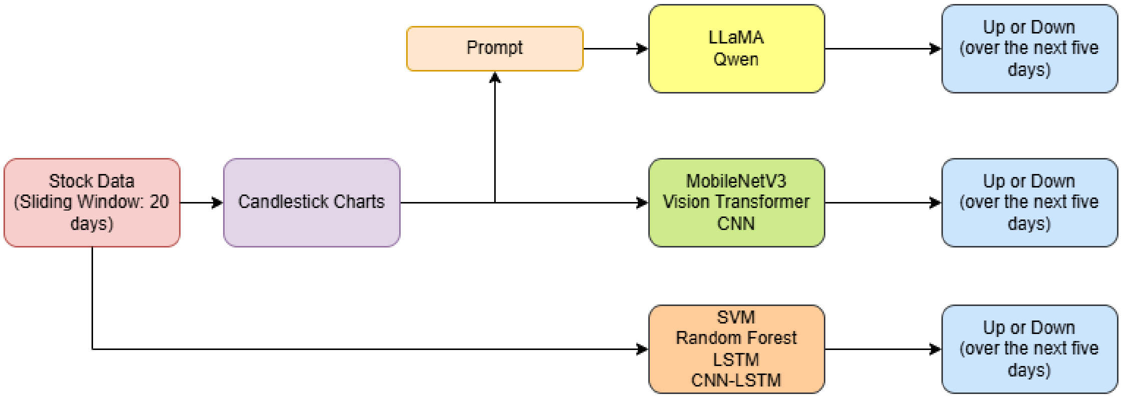

Fig. 1. Whole picture of this study.

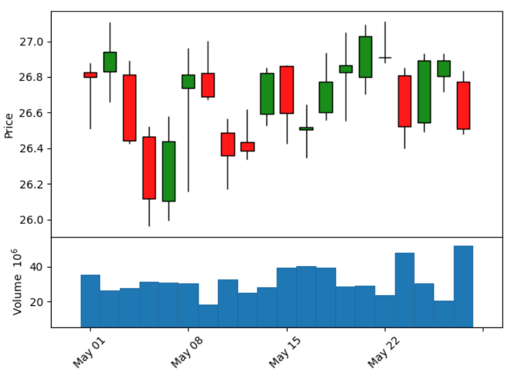

Fig. 2. Candlestick chart input sample.

- Random Forest: Random Forest is an ensemble method, which combines several decision trees to improve prediction accuracy. Each tree is trained on a subset of the data and makes independent predictions, which are then averaged (for regression) or voted on (for classification) to form the final output.
- LSTM (Long Short-Term Memory): LSTM is one type of recurrent neural network (RNN) and is originally designed to handle sequential data. It is particularly effective in capturing long-term dependencies due to its unique architecture, which includes memory cells that can retain information over time.
- CNN-LSTM: CNN-LSTM is a hybrid deep learning model that combines CNN and LSTM to capture both spatial and temporal dependencies in data. CNN excels at extracting spatial features from input sequences, such as images or time series data, by identifying local patterns through convolutional operations. These extracted features are then passed to an LSTM network, which is effective in modeling long-term dependencies and sequential relationships. This combination allows CNN-LSTM models to leverage both feature extraction and sequence learning. In this study, CNN-LSTM is fed with numerical values only.

D. Prompt Design

A structured prompt is created to guide the LLMs in forecasting the stock movement. Table II shows an example of the prompt. This prompt instructs LLMs to identify market trends by analyzing technical indicators and market sentiment, and to provide the predicted market trend along with its probability, which can then be used for post-hoc calibration.

TABLE II. PROMPT FOR STOCK PRICE PREDICTION

Analyze the provided 20-day candlestick chart of company_name and predict the stock price movement within the next five days.

Key Analysis Points:

- **Technical Indicators:** Assess trends using SMA, EMA, RSI, MACD, Bollinger Bands, and volume changes. Highlight any crossovers, divergences, or extreme values.
- **Candlestick Patterns:** Identify bullish or bearish patterns (e.g., engulfing, doji, hammer, shooting star) and explain their significance.
- **Support/Resistance Levels:** Pinpoint recent highs, lows, and key price levels acting as support or resistance.
- **Market Sentiment:** Consider overall sentiment from recent news, earnings reports, or macroeconomic events that could influence the stock price.

Output Format (JSON):

```
{
  "prediction": "Up" | "Down" | "Same",
  "probabilities": {
    "Up": 0.XX,
    "Down": 0.XX,
    "Same": 0.XX
  },
  "justification": "Technical indicators
  (e.g., RSI at 70 indicates overbought),
  patterns (bullish engulfing),
  support/resistance levels,
  and market sentiment
  (positive due to strong
  earnings report)."
}
```

E. Post-Hoc Calibration Techniques

Post-hoc calibration is a technique used to adjust the confidence scores of a machine learning model after training to better reflect the true likelihood of predictions. Many deep learning models, particularly neural networks, tend to be overconfident or underconfident in their outputs. Calibration methods help align the predicted probabilities with actual observed frequencies, making the model's confidence scores more reliable for decision-making.

To mitigate prediction biases and improve confidence estimation, one of the post-hoc calibration methods, Platt Scaling is used in this paper. Platt Scaling uses a logistic regression model to train on the model's logits to recalibrate probability outputs.

F. Evaluation Metrics

The models are assessed based on accuracy, precision, recall, and F1 score. Accuracy represents the proportion of correct predictions made by the models. Precision quantifies the percentage of predicted positive instances that are truly positive, emphasizing the reliability of positive predictions. Recall, in contrast, measures the number of actual positive cases that the model correctly classifies. The F1 score is the harmonic mean of precision and recall. This metric balances the importance of precision and recall.

IV. Results

The experimental results (Table III) present two LLMs, LLaMA and Qwen, with image-based deep learning models and models with numerical inputs in the context of stock price prediction using candlestick charts. The evaluation is conducted across three different stocks, Apple, Tencent, and Toyota, while also examining the impact of Platt Scaling calibration on the performance of LLMs.

In general, LLMs demonstrate strong predictive capabilities, particularly when calibration techniques are applied. LLaMA and Qwen show noticeable improvements in accuracy, precision, recall, and F1 score when Platt Scaling is used, suggesting that post-hoc methods can enhance their reliability. For example, in the case of Apple, LLaMA's accuracy increases from 0.79 to 0.86 after applying Platt Scaling, while its recall reaches 0.98, indicating a significant improvement in sensitivity to positive cases. Similarly, Tencent and Toyota also exhibit improved results for LLMs with calibration, reinforcing the effectiveness of adjusting confidence scores to refine predictions.

Among the image-based deep learning models, CNN consistently achieves strong performance across all stocks and often surpasses MobileNetV2 and Vision Transformer. Its ability to capture patterns in candlestick charts is evident, particularly in recall and F1 score, where it frequently outperforms other models. For example, CNN achieves an F1 score of 0.90 for Tencent and 0.88 for Toyota, demonstrating its robustness in financial time series analysis. MobileNetV2, although achieving competitive accuracy and recall, lags slightly behind CNN in precision. Vision Transformer, on the other hand, struggles in certain scenarios, particularly in recall, which indicates potential difficulties in recognizing crucial patterns in candlestick charts.

For models utilizing numerical inputs, CNN-LSTM demonstrates consistently strong performance, particularly in predicting stock movements for Apple and Tencent. It achieves high precision and recall, resulting in impressive F1 scores of 0.96 and 0.92, respectively. These results highlight the model's ability to effectively capture both spatial and temporal dependencies in financial data. However, its performance on Toyota is significantly weaker, with a recall of just 0.65 and an F1 score of 0.74, suggesting that the model may struggle with certain datasets or exhibit sensitivity to stock-specific characteristics. Furthermore, SVM, Random Forest, and LSTM perform similarly across stocks, with accuracy ranging from 0.73 to 0.77.

For an overall comparison, when predicting stock movement over the next five days, image-based models generally outperform those using numerical input, particularly CNN and LLaMA with post-hoc calibration, which achieve higher accuracy, precision, and recall.

For LLMs, the impact of calibration is particularly obvious. Without calibration, LLMs can generate competitive results, but their recall tends to be lower, which could lead to misclassification of important stock movements. The application of Platt Scaling addresses this issue by refining the decision boundaries, ultimately leading to a more balanced performance.

V. Discussion

During the training of deep learning models such as MobileNetV2, high training accuracy is observed alongside significantly lower testing accuracy, which indicates a potential overfitting issue. To mitigate this issue, this study tries to apply some data augmentation techniques and examine how the model performances will change by using different methods.

Data augmentation is a technique to improve the diversity of training data and reducing the risk of overfitting in deep learning models. Among the commonly used techniques of image data augmentation, **rotation** helps the model adapt to slight variations in chart orientation. **Flipping**, particularly horizontal flipping, allows the model to recognize patterns in different orientations while preserving the underlying price movement structure. **Zooming**, on the other hand, modifies the scale of the candlestick chart, enabling the model to capture different levels of detail within the image and potentially identify subtle but important trading signals. The effectiveness of these augmentation strategies in improving model performance is reflected in Table IV, which presents the results of MobileNetV2 in Toyota's stock movement prediction under different augmentation scenarios.

The results demonstrate interesting insights into how different augmentation strategies impact the model's accuracy, precision, recall, and F1 score.

When rotation is applied, the model's accuracy improves to 0.80, the highest among all tested configurations. The recall also increases significantly to 0.95, which suggests that model becomes better at identifying positive cases, which is crucial for stock price prediction. The F1 score rises to 0.89, indicating a better balance between precision and recall. This improvement highlights the effectiveness of rotation in making the model more robust to variations in the input data.

Flipping augmentation does not improve accuracy, which is still at 0.73, similar to the no-augmentation case. However, there is a slight increase in recall to 0.87 and an F1 score of 0.85, suggesting that flipping might help the model capture more positive instances without significantly affecting precision.

Zooming improves accuracy to 0.78, with a recall of 0.94 and an F1 score of 0.88. This means zooming helps the model

TABLE III. EXPERIMENTAL RESULTS FOR DIFFERENT MODELS ACROSS THREE STOCKS

Stock	Model	Accuracy	Precision	Recall	F1 Score
Apple	LLaMA	0.79	0.85	0.92	0.88
	LLaMA with Calibration	0.86	0.86	0.98	0.92
	Qwen	0.72	0.76	0.80	0.78
	Qwen with Calibration	0.78	0.81	0.84	0.82
	MobileNetV2	0.77	0.83	0.91	0.87
	Vision Transformer	0.78	0.49	0.50	0.49
	CNN	0.80	0.80	0.95	0.87
	SVM	0.77	0.77	1.00	0.87
	Random Forest	0.77	0.78	0.97	0.86
	LSTM	0.77	0.77	0.95	0.85
	CNN-LSTM	0.86	0.96	0.97	0.96
Tencent	LLaMA	0.75	0.80	0.78	0.79
	LLaMA with Calibration	0.82	0.85	0.88	0.86
	Qwen	0.70	0.75	0.74	0.74
	Qwen with Calibration	0.74	0.76	0.79	0.77
	MobileNetV2	0.79	0.83	0.94	0.88
	Vision Transformer	0.79	0.46	0.47	0.46
	CNN	0.85	0.85	0.96	0.90
	SVM	0.73	0.73	1.00	0.84
	Random Forest	0.71	0.74	0.93	0.82
	LSTM	0.73	0.73	0.98	0.84
	CNN-LSTM	0.86	0.86	0.98	0.92
Toyota	LLaMA	0.81	0.83	0.86	0.84
	LLaMA with Calibration	0.87	0.90	0.92	0.91
	Qwen	0.74	0.77	0.79	0.78
	Qwen with Calibration	0.79	0.81	0.83	0.82
	MobileNetV2	0.73	0.84	0.85	0.84
	Vision Transformer	0.77	0.49	0.49	0.49
	CNN	0.84	0.81	0.96	0.88
	SVM	0.74	0.74	0.97	0.84
	Random Forest	0.74	0.76	0.96	0.85
	LSTM	0.74	0.74	0.97	0.84
	CNN-LSTM	0.60	0.85	0.65	0.74

TABLE IV. PERFORMANCE OF MOBILENETV2 WITH DIFFERENT DATA AUGMENTATION TECHNIQUES

Augmentation	Accuracy	Precision	Recall	F1 Score
No Augmentation	0.73	0.84	0.85	0.84
Rotation	0.80	0.83	0.95	0.89
Flipping	0.73	0.82	0.87	0.85
Zooming	0.78	0.82	0.94	0.88
Rotation, Flipping	0.57	0.81	0.63	0.71
Rotation, Flipping, Zooming	0.77	0.82	0.92	0.87

perform better. It likely works by showing candlestick patterns at different scales, which helps with stock price prediction.

Furthermore, combining rotation and flipping result in a significant drop in performance, with accuracy plummeting to 0.57 and recall decreasing to 0.63. This suggests that these two augmentations, when used together, might introduce too much variability or noise, making it harder for the model to learn effectively. However, when rotation, flipping, and zooming are combined, the model's performance improves, achieving an accuracy of 0.77 and an F1 score of 0.87. This indicates that while combining augmentations can be risky, a balanced approach with multiple techniques can still yield positive results.

In summary, rotation augmentation alone provides the best performance, significantly improving accuracy and recall. Combining multiple augmentations can be beneficial but requires careful consideration to avoid introducing excessive noise. These findings suggest that data augmentation, when applied thoughtfully, can enhance the predictive power of models in stock price prediction.

VI. CONCLUSION

The experimental results show that LLMs such as LLaMA and Qwen can be effective in stock price prediction using candlestick charts, especially when calibrated with techniques such as Platt Scaling. Calibration improves the accuracy and reliability of LLMs, making them more suitable for financial forecasting. Among image-based deep learning models, CNN outperforms MobileNetV2 and Vision Transformer in recall and F1 score, demonstrating its effectiveness in candlestick chart analysis. While MobileNetV2 offers computational efficiency, Vision Transformer struggles with recall. Among models with numerical inputs, hybrid model CNN-LSTM achieves the best performance and other models such as SVM, Random Forest and LSTM achieve competitive performance. Last but not least, the impact of data augmentation techniques in model performance is also studied in this paper. The results indicate that rotation augmentation alone delivers the best performance, notably boosting accuracy and recall. Although combining multiple augmentations can be advantageous, it must be done cautiously to prevent excessive noise. These results indicate that thoughtful data augmentation can enhance model predictive power in stock price forecasting.

Future research can focus on assessing the robustness of LLMs against adversarial attacks. In practical financial applications, models may be vulnerable to minor changes in input data, such as subtle distortions in candlestick chart images, which could result in significantly different predictions. Investigating adversarial attacks on the visual components of the framework would offer deeper insights into the model's reliability and security. Implementing adversarial training could further improve the stability and robustness of the prediction system under noisy input conditions.

REFERENCES

[1] R. A. Kamble, "Short and long term stock trend prediction using decision tree," in *2017 International Conference on Intelligent Computing and Control Systems (ICICCS)*, 2017, pp. 1371–1375.

[2] S.-H. Cheng, "Predicting stock returns by decision tree combining neural network," in *Intelligent Information and Database Systems*, N. T. Nguyen, B. Attachoo, B. Trawiński, and K. Somboonviwat, Eds. Cham: Springer International Publishing, 2014, pp. 352–360.

[3] U. Ananthakumar and R. Sarkar, "Application of logistic regression in assessing stock performances," in *2017 IEEE 15th Intl Conf on Dependable, Autonomic and Secure Computing, 15th Intl Conf on Pervasive Intelligence and Computing, 3rd Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress(DASC/PiCom/DataCom/CyberSciTech)*, 2017, pp. 1242–1247.

[4] O. Shobayo, S. Adeyemi-Longe, O. Popoola, and B. Ogunleye, "Innovative sentiment analysis and prediction of stock price using finbert, gpt-4 and logistic regression: A data-driven approach," *Big Data and Cognitive Computing*, vol. 8, no. 11, 2024. [Online]. Available: https://www.mdpi.com/2504-2289/8/11/143

[5] Y. Gao, R. Wang, and E. Zhou, "Stock prediction based on optimized lstm and gru models," *Scientific Programming*, vol. 2021, no. 1, p. 4055281, 2021. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1155/2021/4055281

[6] L. Sayavong, Z. Wu, and S. Chalita, "Research on stock price prediction method based on convolutional neural network," in *2019 International Conference on Virtual Reality and Intelligent Systems (ICVRIS)*, 2019, pp. 173–176.

[7] K. L. Szydlowski and J. A. Chudziak, "Hidformer: Transformer-style neural network in stock price forecasting," 2024. [Online]. Available: https://arxiv.org/abs/2412.19932

[8] H. Yang, X.-Y. Liu, and C. D. Wang, "Fingpt: Open-source financial large language models," 2023. [Online]. Available: https://arxiv.org/abs/2306.06031

[9] Q. Chen and H. Kawashima, "Stock price prediction using llm-based sentiment analysis," in *Proceedings of the IEEE BigData 2024*. Washington DC, USA: IEEE, 2024, pp. 4828–4835.

[10] Q. Chen, "Image-driven stock price prediction with llama: A prompt-based approach," *International Journal of Modeling and Optimization*, 2025, to be published.

[11] M. Agrawal, D. A. U. Khan, and D. P. K. Shukla, "Stock price prediction using technical indicators: A predictive model using optimal deep learning," *International Journal of Recent Technology and Engineering (IJRTE)*, vol. 8, no. 2, pp. 2297–2305, Jul. 30 2019.

[12] F. Moodi, A. Jahangard-Rafsanjani, and S. Zarifzadeh, "Feature selection and regression methods for stock price prediction using technical indicators," 2023. [Online]. Available: https://arxiv.org/abs/2310.09903

[13] T. Julian, T. Devrison, V. Anora, and K. M. Suryaningrum, "Stock price prediction model using deep learning optimization based on technical analysis indicators," *Procedia Computer Science*, vol. 227, pp. 939–947, 2023, 8th International Conference on Computer Science and Computational Intelligence (ICCSCI 2023). [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1877050923017684

[14] N. Darapaneni, A. R. Paduri, H. Sharma, M. Manjrekar, N. Hindlekar, P. Bhagat, U. Aiyer, and Y. Agarwal, "Stock price prediction using sentiment analysis and deep learning for indian markets," 2022.

[15] L. S. Parvatha, D. Naga Veera Tarun, M. Yeswanth, and J. S. Kiran, "Stock market prediction using sentiment analysis and incremental clustering approaches," in *2023 9th International Conference on Advanced Computing and Communication Systems (ICACCS)*, vol. 1, 2023, pp. 888–893.

[16] M. R. Vargas, C. E. M. dos Anjos, G. L. G. Bichara, and A. G. Evsukoff, "Deep leaming for stock market prediction using technical indicators and financial news articles," in *2018 International Joint Conference on Neural Networks (IJCNN)*, 2018, pp. 1–8.

[17] T. W. A. Khairi, R. M. Zaki, and W. A. Mahmood, "Stock price prediction using technical, fundamental and news based approach," in *2019 2nd Scientific Conference of Computer Sciences (SCCS)*, 2019, pp. 177–181.

[18] A. Rahimi, T. Mensink, K. Gupta, T. Ajanthan, C. Sminchisescu, and R. Hartley, "Post-hoc calibration of neural networks by g-layers," 2022. [Online]. Available: https://arxiv.org/abs/2006.12807

[19] Q. Chen, "Stock price change prediction using prompt-based llms with rl-enhanced post-hoc adjustments," in *Proceedings of the 4th International Conference on Bigdata Blockchain and Economy Management*, 2025, to be published.

[20] M. Steinbacher, "Predicting stock price movement as an image classification problem," 2023. [Online]. Available: https://arxiv.org/abs/2303.01111

[21] D. R. Jeongseok Bang, "Cnn-based stock price forecasting by stock chart images," *Romanian Journal of Economic Forecasting*, vol. 3, pp. 120–128, 2023. [Online]. Available: https://ipe.ro/new/rjef/rjef3_2023/rjef3_2023p120-128.pdf

[22] L. Zhou, Y. Zhang, J. Yu, G. Wang, Z. Liu, S. Yongchareon, and N. Wang, "Llm-augmented linear transformer–cnn for enhanced stock price prediction," *Mathematics*, vol. 13, no. 3, 2025. [Online]. Available: https://www.mdpi.com/2227-7390/13/3/487

[23] G. Jin and O. Kwon, "Impact of chart image characteristics on stock price prediction with a convolutional neural network," *PLoS ONE*, vol. 16, no. 6, p. e0253121, 2021.

[24] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," 2019. [Online]. Available: https://arxiv.org/abs/1801.04381

[25] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An image is worth 16x16 words: Transformers for image recognition at scale," 2021. [Online]. Available: https://arxiv.org/abs/2010.11929

[26] K. Liagkouras and K. Metaxiotis, *Stock Market Forecasting by Using Support Vector Machines*. Cham: Springer International Publishing, 2020, pp. 259–271. [Online]. Available: https://doi.org/10.1007/978-3-030-49724-8_11

[27] S. Du, D. Hao, and X. Li, "Research on stock forecasting based on random forest," in *2022 IEEE 2nd International Conference on Data Science and Computer Application (ICDSCA)*, 2022, pp. 301–305.

[28] D. Wei, "Prediction of stock price based on lstm neural network," in *2019 International Conference on Artificial Intelligence and Advanced Manufacturing (AIAM)*, 2019, pp. 544–547.

[29] M. A. Nadif, M. T. Rahman Samin, and T. Islam, "Stock market prediction using long short-term memory (lstm)," in *2022 Second International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT)*, 2022, pp. 1–6.

[30] A. A, R. R, V. R. S, and A. M. Bagde, "Predicting stock market time-series data using cnn-lstm neural network model," 2023. [Online]. Available: https://arxiv.org/abs/2305.14378