Volume 1 Issue 2 2023
Pages:104-121
ISSN: 2980-3152

CURRENT TRENDS IN COMPUTING

http://ctc.karabuk.edu.tr
SURVEY ARTICLE

TECHNIQUES USED TO EXTRACT FEATURES FROM CANDLESTICK CHARTS IN THE STOCK MARKET: A SYSTEMATIC REVIEW

EDREES RAMADAN MERSAL[1*], HAKAN KUTUCU[2]

[1]*Department of Computer Engineering, Karabük University, 78050, Karabük, Türkiye*
[2]*Department of Software Engineering, Karabük University, 78050, Karabük, Türkiye*

ABSTRACT: In the financial sector, accurately forecasting stock market trends is essential for guiding the investment and trading decisions of investors and traders. These professionals often rely on candlestick charts to analyze and predict stock price fluctuations. In recent times, various methods and algorithms have been applied to leverage candlestick charts for prediction purposes. This systematic review aims to examine the application of Japanese candlesticks and machine learning techniques, including artificial neural networks, in predicting stock market trends. It also delves into the effective feature engineering strategies for extracting pertinent information from raw data, encompassing technical indicators and candlestick charts. The review encompasses 30 studies published between 2019 and 2023, selected based on criteria that include the utilization of candlestick charts in stock market analysis. The findings reveal that numerous studies employing automatic encoders, convolutional neural networks, and Gramian Angular Field (GAF) for feature geometry extraction from candlestick charts also identify common patterns.

1. INTRODUCTION

According to Naik & Mohan (2020), "forecasting *the stock market is difficult since it is volatile and nonlinear"*. However, it is considered a significant objective because of its impact on the decision-making process every time investors, traders, and financial institutions purchase and/or sell shares. Over the past several decades, a significant number of studies have been conducted in the field of financial markets to forecast the direction of movement using technical analysis (Lin, Liu, Yang, & Wu, 2021). Technical analysis contains historical price and volume volatility data, from which patterns can be identified, extracted, and used to make forecasts. The analysis of technical indicators and charts is one of the methods utilized to identify patterns and forecast the direction of the markets (Lin, Liu, Yang, & Wu, 2021). The origins of the candlestick chart can be traced back to the 18th century when Japanese traders attempted to identify trading trends in rice harvests. Specific patterns in the candlestick series can provide clues regarding the direction of future stock price swings (Thammakesorn & Sor[1]nil, 2019). Western investors have widely adopted candlestick patterns since the 1990s. Owing to their widespread adoption, candlestick patterns can be found in a wide variety of technical analysis tools (Udagawa, 2019; Jearanaitanakij & Passaya, 2019; Pan et al., 2020). The candlestick chart is regarded as the most helpful

E-mail: edreesa2001@gmail.com (*)
Keywords and phrases: Japanese Candlestick, Stock Market, Machine Learning.

chart for technical analysis, with each candle representing trade prices for a particular timeframe (minutes, hours, days, months, or years). The components of a candlestick are the body, the upper shadow, and the lower shadow. The body represents the area between the opening and closing prices, whereas the upper and lower shadows represent the highest and lowest traded prices, respectively (Jearanaitanakij & Passaya, 2019) as shown in Figure 1.

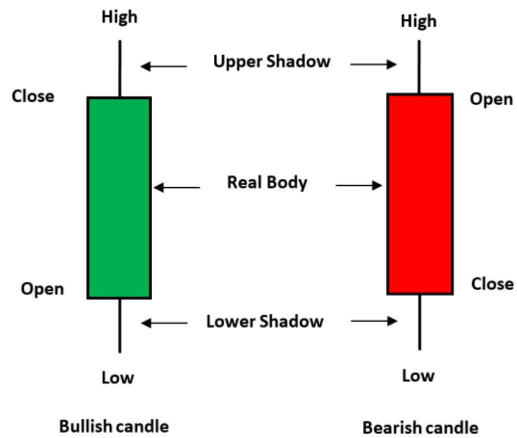

FIGURE 1. Bullish And Bearish Candles

The formation of several candlestick patterns between open and closed prices is a widely recognized phenomenon (Lin, Liu, Yang, Wu, et al., 2021). Figure 2 shows various candlestick patterns. However, extracting these patterns from historical data remains a challenging task. The primary reason is that the mining of candlestick patterns does not handle statistical combinations of candlestick patterns that exhibit high frequency, noise, a short timeframe, pattern distortion or other corresponding trends (Du et al., 2020). The complexity of candlestick patterns increases with the number of candlesticks on the chart, leading to an exponential increase in the number of possible geometries in the pattern.

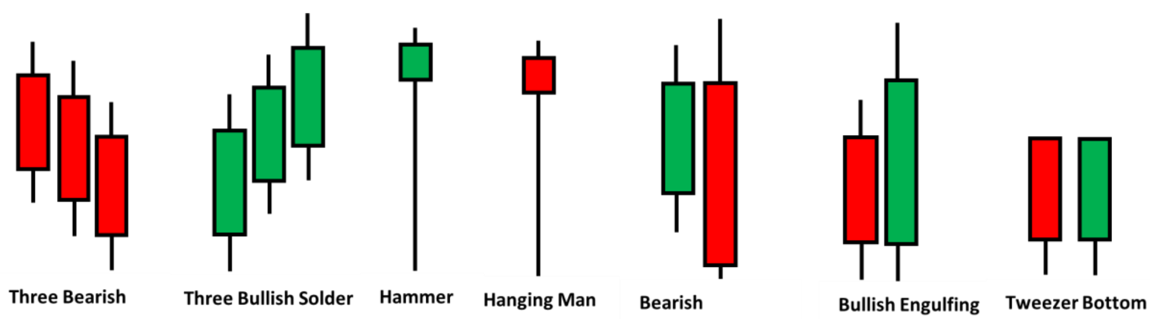

FIGURE 2. Some Candlesticks Patterns

Moreover, it is impossible to systematically match patterns retrieved from past data with those from more recent data. The results would show dramatic price fluctuations over time and new serial patterns would emerge (Liang et al., 2022). In contrast, the challenge of associating "the association of candlestick patterns" with the current trend of candlestick sequence occurrence has been resolved (Liang et al., 2022). Machine learning algorithms are extensively used to anticipate price variations from neural network methods, such as convolutional neural networks, recurrent neural networks, and modified neural networks (Lin, Liu, Yang, Wu, et al., 2021). Most studies of financial markets focus on technical indicators for candlestick charts, such as the color, length of the upper and lower shadows, and shape of the patterns that comprise a group of candlesticks. The standard candlestick chart in stock markets is expressed using tensors or vectors and disregards the financial properties of the chart as a graphical indicator. To extract the features or representations from the sub-chart, it is necessary to design pre-processing methods for a specific candlestick chart, which is a critical issue. Candlestick charts are used for trend prediction, and we locate literature that investigates this method.

This study provides a conceptual framework that classifies the significance of technical analysis using candlestick charts for forecasting after examining the existing literature on trend forecasting using candlestick charts in the stock market. Our work expands the existing literature by detailing the applications of candlestick charts for stock market forecasting. Therefore, we formulated the following Research Questions: Research Question 1: Which engineering methodologies, when implemented on candlestick charts, produce the most precise predictions of stock market trends? RQ2: When employing a Japanese candlestick chart to forecast the stock market, what specific timeframe should be employed?

2. METHODOLOGY

2.1 Research Methodology

The methodology section of this paper outlines the systematic literature review undertaken as part of our research. Systematic literature reviews are integral in synthesizing current knowledge, identifying research gaps, and laying the groundwork for further investigations. Our review was rigorously structured in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines. These guidelines are the benchmark for conducting and reporting systematic reviews.

The objective of our systematic review was to exhaustively collate and scrutinize literature relevant to our research queries. This method facilitated an impartial and comprehensive evaluation of the extant studies, enabling us to formulate conclusions grounded in evidence. This section delineates the procedural steps undertaken in our review. These steps include identifying pertinent databases and search terms, implementing specific inclusion and exclusion criteria, conducting the screening and selection of studies, and outlining the methods for data extraction and synthesis. Our systematic literature review process is depicted in Figure 3, following the PRISMA flow diagram format. A detailed exposition of each step is presented subsequently.

2.2 Identification of Studies via Databases and Registers

We initiated this review by searching three major electronic databases: Scopus, IEEE Xplore, and ProQuest. The objective was to comprehensively encompass potential studies, yielding 122 records (59 from Scopus, 21 from IEEE, and 42 from ProQuest).

2.3 Literature Search Strategy

To precisely target our search, we used a specific query with keywords "candlestick chart," "stock," and "market." The search string was as follows: TITLE-ABS-KEY-AUTH (candlestick chart AND stock AND market) AND (LIMIT-TO (PUBY YEAR, 2023) OR LIMIT-TO (PUBY YEAR, 2022) OR LIMIT-TO (PUBY YEAR, 2021) OR LIMIT-TO (PUBY YEAR, 2020) OR LIMIT-TO (PUBY YEAR, 2019)). This approach ensured a focused yet extensive search and capture of recent and relevant literature.

2.4 Inclusion and Exclusion Criteria

Before screening, the records were filtered out based on specific exclusion criteria. First, any record with less than one citation was removed under the assumption that higher citation counts may correlate with a study's influence and relevance. Second, records published before 2019 were excluded to ensure that the review focused on recent literature and reflected the latest research trends and findings. These exclusion criteria led to the removal of 40 records, with Scopus contributing to the majority of these exclusions (17 records), followed by ProQuest (13 records), and IEEE (10 records).

2.5 Screening and Selection Process

The next phase involved a screening process in which the titles and abstracts of the remaining 82 records were reviewed. During this phase, duplicates were identified and removed, resulting in 17 fewer records. This step was crucial for eliminating redundancy and ensuring that each study was unique, thereby maintaining the integrity of the review.

Subsequently, efforts were made to retrieve the full text of the 65 remaining reports. However, 15 of these were inaccessible, indicating potential issues such as subscription barriers or availability constraints.

The 50 reports that were successfully retrieved were assessed for eligibility based on the specific inclusion criteria relevant to the research question. This step resulted in further refinement, excluding 20 reports. The exclusions were due to two primary reasons: irrelevance to the research topic (5 reports) and the absence of candlestick charts in the methodology or analysis (15 reports), which implies that the use of candlestick charts was a significant factor for the review.

2.6 Data Extraction and Thematic Synthesis

The culmination of this meticulous process included 30 studies that met all established criteria. The literature review illustrates a rigorous and systematic approach to research synthesis. This process was essential to ensure that the conclusions drawn from the review were based on comprehensive high-quality evidence. The methodology adopted in this review reflects the best practices in research and provides a reproducible model for future studies.

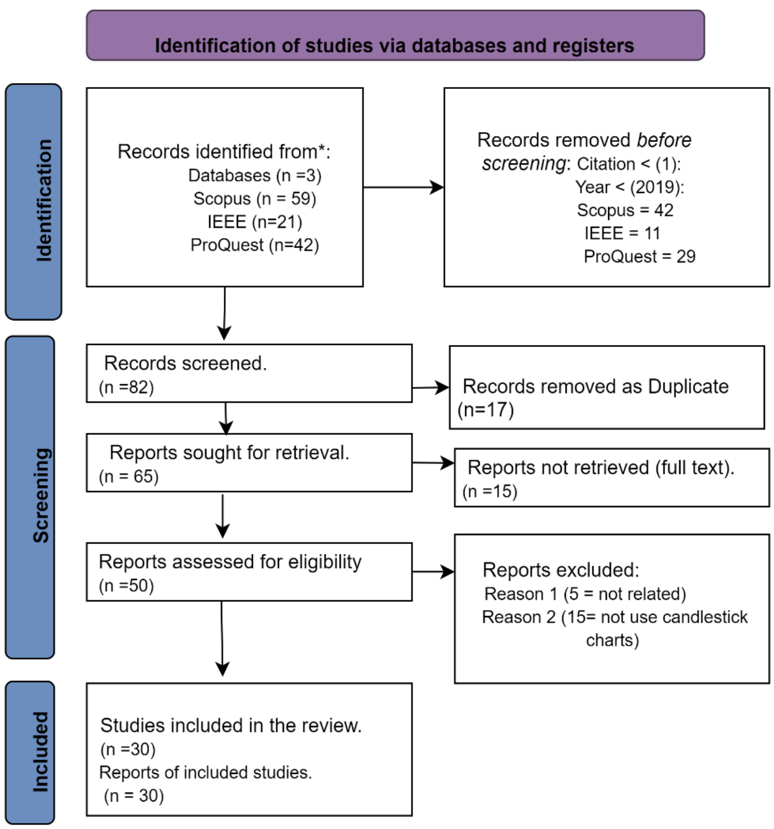

FIGURE 3. Prism flow framework

3. ANALYSIS OF RESULTS

The last step involved deconstructing the chosen articles into prediction techniques, including pattern analysis and prediction methods using deep learning and trading strategies with machine learning.

3.1 Analysis Based on Predictions Using Candlestick Patterns

In this section, we examine how candlestick chart pattern prediction can be used in the stock market. The analysis focuses on the techniques of the candlestick patterns used with different models; it covers the patterns that were applied, methods of dealing with the candlestick patterns, etc.

Andriyanto (2020) proposed predicting the direction of movement for trading using candlestick chart patterns and a convolutional neural network algorithm. The timing frame of the candlestick chart in which the experiments were conducted was the daily timing frame, such that each chart contained three days of candles, and the title of each chart was in the direction of an upward or downward movement. The label is determined by comparing the closing price of the third candlestick in the current chart with the closing price of the first candlestick in the upcoming data. If the current closing price is larger than the closing price in the upcoming data, the chart is labeled as downward trade, and if the opposite is lower, it is labeled as upward trade.

Jearanaitanakij & Passaya, 2019 proposed a method for predicting short-term trading using convolutional networks and candlestick patterns. They collected 1800 instances of Japanese candlestick pattern images, classified as bullish, bearish, or sideways. Subsequently, they employed convolutional neural networks. Better results were achieved with respect to training timing and accuracy compared with using ResNet-18.

According to Liang et al. (2022), prediction models such as a post-driven K-line prediction model and a similar sequence-driven K-line prediction model have been proposed. The predictive procedure was divided into two parts. First, sequential pattern mining was used to extract candlestick patterns from multidimensional candlestick data, and then the association between the different patterns and the corresponding future trends was calculated. Second, a new sequence similarity is proposed to compare various candlestick sequences with existing patterns.

To enhance the operational performance of algorithmic trading, Nakayama et al. (2019) propose a method for predicting short-term price trends using high-frequency trading series with deep learning, including all order types (market, limit, and cancellation). However, the images used in this study contained time and cost data. The encoding technique can utilize these features. They began by feeding a neural network with a variety of quality-of-order indices. Logistic regression and convolutional neural networks were employed for stock price forecasting.

Lin, Liu, Yang, Wu, et al. (2021) propose a pattern recognition model (PRML) that employs ML to improve financial choices. The empirical results show that when looking ahead when it comes to forecasting, two-day candlestick patterns are more effective than the two other popular machine learning models used for testing the dependence of the prediction model. The results prove that the investment strategy developed using the PRML is profitable. This study adds four new insights to the existing body of literature by (a) proposing a machine learning-based candlestick pattern recognition model, (b) applying four machine learning methods to all possible combinations of daily patterns in the pattern recognition process, (c) splitting the dataset in half, and (d) using the predicted parameters to run prediction tests on new, unknown data. This study applies PRML to five distinct strategy pools to examine how accurately pool predictions hold over time. The results of a one-day time period using the TOP10 strategy demonstrated that two-day candlestick patterns had the most predictive power, with an average annual return of 36.73%, a Sharpe ratio of 0.81, and an information ratio of 2.37.

JuHyok et al. (2020) proposed a novel method for predicting stock price reversal points by integrating up- and down-reversal point feature sets. The open price, high price, low price, closing price, and trading volume form the foundation upon which candlestick indications and technical indicators are built. After creating 27 sets of feature candidates, they identified the feature sets that were most useful for predicting upward and downward movements for each stock individually. Combining the outputs of the two LSTM-based upward and downward predictors increased prediction accuracy.

In Udagawa's (2019) study, a model with six parameters was proposed to retrieve similar candlestick patterns to enhance the accuracy of stock price predictions. The simulations conducted in this study demonstrate that higher risks are associated with higher returns. The algorithm leverages the slopes of the 5-day and 25-day moving averages to identify trends and determine whether the price falls within high- or low-price zones. Reverse trade criteria incorporate cumulative negative stock price movements and cumulative negative differences in the 5-day averages. The simulations further reveal that the likelihood of successful stock trades outweighs that of failed ones, and that there is a significant correlation between the number of days to hold stocks before executing a reverse stock trade and the gains generated.

Thammakesorn and Sornil (2019) proposed a strategy for generating stock trading strategies based on characteristics that have been shown to detect patterns in candlestick charts reliably. The Chi-square Automatic Interaction Detector is then used to incorporate the features into a trading strategy that resembles a tree. The results demonstrate that the developed strategies outperform some of the most well-liked trading strategies in terms of profitability, such as moving average convergence divergence, exponential moving average, relative strength index, stochastic oscillator, and average directional index.

Wang M. and Wang Y. (2019) utilized a recent data from twenty firms located in the United States to assess the effectiveness of various widely recognized candlestick patterns in predicting stock price movements. They computed the percentage of accurate predictions and expected earnings, ultimately leading to the conclusion that three of the four examined patterns were not profitable. However, the remaining pattern offers valuable insights into market trend reversals, prompting the conclusion that this specific pattern was indeed profitable.

Orte et al. (2023) presented a model designed to forecast the prices of crypto assets traded on futures markets. They employed a random forest model to explore three potential outcomes based on input variables, including technical indicators, candlestick patterns, and a combination of both. Subsequently, they investigated the model parameters, optimal time intervals, and investment horizons. In addition to presenting the model's output, they conducted a one-year out-of-sample prediction simulation, selecting the entirety of 2020 as the observation period because of the three conceivable outcomes in the stock market during that year: a sideways market, a bear market due to the global pandemic, and a bull market towards the end of the year. To emulate a real-world scenario, researchers retrained the model after each new data collection to ensure that it consistently had access to the most up-to-date information. In summary, substituting candlestick patterns for technical indicators has emerged as a significant strategy to enhance the effectiveness of the results.

Ahmad et al., (2020) proposed the amalgamation of technical analysis, candlestick analysis, and fuzzy algorithms. This combination aims to aid traders and investors in gaining a deeper understanding of the financial market and making more profitable decisions while also offering the means to avoid minimal risks.

Naranjo and Santos, (2019) introduced an intelligent decision tool that considers the influence of currency devaluation on forecasting. This technology employs fuzzy Japanese candlesticks.

In Santur's (2022) proposed model, there were four phases. The first phase involved the development of a system that employs object-oriented programming and a factory design pattern to recognize patterns in candlestick charts, encompassing 24 distinct patterns. In the second step, the “One Hot Encoder” procedure is utilized to determine the type of daily candlestick generated by each dataset. Data classification is based on the daily closing value, categorizing it as 'bearish' or 'bullish'. In the last stage, the community-learning algorithm XGBoost was applied.

3.2 Analysis Based on Prediction Using Deep Learning

In this section, the analysis focuses on forecasting methods using candlestick charts with deep learning convolutional neural networks, reinforcement learning, and other hybrid algorithms.

Hung and Chen (2021) developed a model for advanced prediction of stock prices using candlestick charts. Their method involved a three-step process: first, subdividing the candlestick chart; second, employing a convolutional autoencoder for optimal subschema representation using 48-pixel grayscale images with dimensions of $48 \times 48 \times 1$; and third, utilizing recurrent neural networks, specifically recurrent gated units, for the final prediction. The encoding process included a 2D convolutional layer, a 2D max pooling layer with ReLu activation, and a 3D convolutional layer.

Kusuma et al. (2019) discussed the predictive capabilities of candlestick charts derived from Taiwanese and Indonesian stock markets. The methodology involved using these charts to forecast stock price movements and the analysis considered three distinct trading periods to examine their relationships with market trends. This study utilized candlestick charts that encapsulated time-series data, with and without the inclusion of daily stock volume information. Experiments were conducted using candlestick charts of two dimensions, 50 and 20, to uncover hidden patterns and their correlations with market movements across various image scales. The authors used datasets to perform the analyses using a range of learning algorithms, including traditional machine learning approaches such as random forest and k-nearest neighbors as well as modern machine learning techniques such as convolutional neural networks (CNNs), residual networks, and visual geometry group (VGG) networks. The primary objective of this study was to explore how different parameters, specifically period times, image sizes, and feature sets, influence stock market trends. The authors aimed to ascertain the predictive accuracy of these models in determining whether the stock market will rise or fall the following day.

Naik and Mohan (2020) employed deep neural networks to forecast intraday stock prices by incorporating candlestick data with technical indicators. They utilized 10 different technical indicators to evaluate their values and then constructed daily candlestick patterns. This method entailed comparing the indicator values against price data from the previous ten days to determine whether the indicators exhibited an upward or downward trend. Furthermore, candlestick patterns were analyzed in relation to average market prices to determine the trend direction. These extensive data were subsequently utilized for training deep neural networks for stock market prediction, showcasing an innovative analytical approach to financial modeling.

J.-H. Chen and Tsai (2022) proposed candlestick patterns and Gramian Angular Field (GAF) time series coding as methods for the identification of objects. The proposed model uses deep neural networks and a novel architectural design, both of which contribute to the performance of the model in candlestick categorization as well as location recognition, using the GAF coding approach and YOLO-v1 model to recognize patterns with varying dynamic durations. These pieces of information are organized into eight distinct categories of candlestick patterns, each of which represents a primary candlestick signal. As a result, the best 10 results in each category were chosen as targets. Then, the top 10 targets were used, and the Dynamic Time Warping (DTW) approach was applied to aggregate the full dataset using a variety of window sizes. The final processed dataset included eight different classes of candlestick patterns, ranging in window sizes from 5 to 16. The YOLO-v1 model included six convolutional layers, batch normalization layers, and leak activation. The suggested model performs quite well in both the classification of candlesticks and the identification of their locations.

To beat the returns of the S&P 500 Index and to probe the workings of the trading system, Brim & Flann (2022) proposed a convolutional neural network (CNN) integrated within a Double Deep Q-Network (DDQN). The 30 most widely held stocks in the S&P 500 are used to train and test the DDQN. After CNN was trained, feature map representations were created to reveal the region of interest in the candlestick charts. This research indicates that the DDQN can generate superior returns compared to the S&P 500 Index. The results also show that CNN can shift its focus from all candles in a candlestick image to the most recent candles in the image.

Lee et al. (2019) suggested a convolutional neural network function approximator with a deep Q-network to approximate stock chart images as input and then they trained the model on the US market and tested it in 31 different countries over a period of 12 years. The portfolios yield approximately 0.1 to 1.0 percent return per transaction prior to transaction costs.

Barra et al. (2020) utilized convolutional neural networks (CNNs) in their research to predict the future trajectory of the Standard & Poor's 500 Index through the utilization of images generated from time series data by GAF. Each CNN was trained on a distinct multi-resolution image, allowing the examination of varying time intervals for a single observation. The efficacy of the proposed methodology was assessed by implementing a straightforward trading system based on an ensemble forecaster.

Pan et al., (2020) developed a portfolio learning model based on deep learning and applied it to the Chinese stock market. After collecting stock data, a set of images containing candlestick charts was formed. The candlestick charts are used to extract the features using the CAE encoder; subsequently, the K-means algorithm is applied to group the high-dimensional features. Finally, one stock from each category was selected according to the Sharpe ratio, and a high-return, low-risk portfolio was obtained.

To reduce noise, Fengqian and Chao (2020) suggested using real-time financial data to test the K-line theory, which generalizes price movements over time using Japanese candlesticks. Cluster analysis then extracts the learning characteristics of the candlesticks. The K-line learning features are also combined, and a high-frequency transactional technique is used with online adaptive control to learn the parameters through deep reinforcement in an unknown environment. This approach creates a high-frequency transaction strategy. To confirm its efficiency, the proposed strategy was tested against prediction-based approaches. They consist of both fuzzy neural networks and recurrent neural networks. The experimental results validate the robustness and prediction accuracy of the proposed method.

3.3 Analysis Based on Prediction Using Strategies

A trading strategy is a methodical approach employed to execute transactions within financial securities markets. This strategy is grounded in a set of predefined norms and criteria that guide decision-making processes in trading activities. The complexity of a trading strategy can range from basic to intricate and encompass various factors. These factors include the investor’s investment approach, market capitalization, application of technical indicators, utilization of fundamental analysis, focus on specific industry sectors, degree of diversification within a portfolio, intended duration of investment or holding period, level of acceptable risk, use of financial leverage, and considerations related to taxation. For trading purposes, strategy is defined as the rationale behind a set of proposed transactions involving an asset. The most popular strategy typically adheres to a simple rule of sorts. If the prediction issues a purchase signal, the algorithm will place a buy order and wait for a sell signal before calculating profit or loss.

Du et al. (2020) introduced a novel approach that integrates machine learning and image processing to predict market trends using intraday financial data. By applying a wavelet transform to the log return of stock prices, they generated images and reduced the noise. Subsequently, a convolutional neural network was utilized to detect patterns in these denoised images, enabling a binary classification of the daily time series into either "up" or "down" market states. This method effectively addresses the challenge of low signal-to-noise ratio in financial data and delivers competitive accuracy in predicting market states.

Lin, Liu, Yang, and Wu (2021) developed a machine-learning ensemble prediction model tailored to automatically select the optimal prediction methods for daily k-line patterns. By using artificial intelligence techniques with classic candlestick charts, this model advances stock market prediction studies. Rooted in Taoist cosmology, the eight-trigram classification, which draws insights from the high and low prices of two sequential trading days, is a fundamental component. To increase the prediction precision, this study introduces four types of technical indicators: overlap, momentum, volume, and volatility. This study proposes a methodological approach for identifying the most efficient machine prediction method across various feature modes by amalgamating several prediction models. This ensemble model streamlines the parameter optimization for the top six predictive models, namely Random Forest (RF), Gradient Boosting Decision Tree (GBDT), Logistic Regression (LR), K-Nearest Neighbors (KNN), Support Vector Machine (SVM), and Long Short-Term Memory (LSTM). The empirical findings indicate that RF and GBDT overshadow KNN and SVM in terms of short-term projection. Moreover, the proposed investment strategy, marked by a diminished maximum drawdown, elevated Sharpe ratio, and reduced Sortino index, surpasses the basic buy-and-hold approach. However, it is imperative to note that, in real-life scenarios, transaction costs profoundly influence trading decisions.

Rumpa et al., (2021) employed a support vector machine (SVM) to predict the next candle based on the characteristics of the previous one. Open, close, high, low, volume, and the previous candlestick color are the six features used in this study. The prediction accuracy rate of the model was 56%.

J. H. Chen and Tsai (2020) provided an approach for studying the eight candlestick patterns of the morning star, bullish engulfing, hammer, shooting star, evening star, bearish engulfing, hanging man, and inverted hammer. Time series are encoded as images using the Gramian angular summation field, and then a convolutional neural network is applied to those images. The accuracy of this study was 90.7% when evaluated using actual data.

H. Wang et al. (2021) proposed an approach for stock market prediction. They used the vector auto-regression (VAR) model and the vector error correction (VEC) model to model OHLC data methods in their study. The effectiveness and stability of the proposed method were examined through simulated studies.

Birogul et al. (2020) suggested a real-time object detection (YOLO) system to find “Buy-Sell” objects in 2D candlestick charts. The authors suggest that through these buying and selling elements, traders can make decisions about when to buy or sell an investment instrument. The study also discusses the importance of looking at charts from a different angle when making predictions in real-time after having seen the investment tool's visual trends in the past.

Hassen et al. (2020) proposed a cloud model for the qualitative prediction of stock market trends. They discuss the usefulness of fuzzy logic in decision-making, where there is a great deal of uncertainty and ambiguity. The proposed model uses hybrid weighting, which combines fuzzy time series and Japanese candlesticks. The cloud model helps in creating fuzzy membership criteria to effectively deal with uncertainty and ambiguity in the historical data of the stock with the intention of predicting the stock's open, high, low, and closing prices.

Poženel and Lavbič (2019) proposed a new method that can achieve returns. The proposed method used natural language processing (Word2Vec) with candlesticks. The proposed method was compared with three trading models: moving average convergence divergence (MACD), buy-and-hold, and moving average.

Table 1. Analyzed articles.

Author(S)	**Year**	**Market**	**Period**	**Timeframe**	**Model**	**Performance Metric**
Andriyanto, A.	2020	USA (IDX Mining Index) (JKMING)	01/01/2017-30/12/2019	Daily	CNN	Accuracy
Barra, S., Carta, S. M., Corriga, A., Podda, A. S., Recupero, D.R.	2020	USA (S&P 500)	01/02/2000-30/01/2015	1 hour 4 hours 8 hours 1 day	CNN	Accuracy

Brim, A., & Flann, N. S.	2022	USA (S&P 500)	2/01/2020-30/06/2020	Daily	Double Deep Q-Network	Statistical
Chen, J.-H., & Tsai, Y.-C.	2022	Foreign exchange (EUR/USD)	1/01/2000-1/01/2020	1-minute	CNN, YOLO	Accuracy
Du, B., Fernandez-Reyes, D., & Barucca, P.	2020	USA (S&P 500)	08/04/2009-10/04/2019	Daily	CNN	Accuracy Precision Recall, F1-Score
Fengqian, Di., & Chao, L.	2020	China's stock market (CSI 300 index)	2000-2016	1 minute	Deep reinforcement learning	Comparison
Hung, C. C., & Chen, Y. J.	2021	Taiwan Stock Exchange (TALEX, NIKKEI 225)	TALEX 21/07/1998-27/12/2016 NIKKI 5/01/2001-30/11/2020	Daily	CNN-Autoencoder RNN (Gated Recurrent Unit)	Accuracy Precision Recall, true positive rate, false positive rate
Jearanaitanakij, K., & Passaya, B.	2019	Stock exchange Thailand	-	2 weeks	CNN	Accuracy
Kusuma, R. M. I., Ho, T.-T., Kao, W.-C., Ou, Y.-Y., & Hua, K.-L.	2019	Stock market (Taiwan, Indonesia)	1/01/2000-31/12/2018	Daily	CNN	Accuracy, Precision, Recall, Specificity
Lee, J., Kim, R., Koh, Y., & Kang, J.	2019	USA (Russell 300 index)	2001-2018	Daily	Deep Q-network	Statistical
Liang, M., Wu, S., Wang, X., & Chen, Q.	2022	China's stock market (CSI 300-CSI 500)	1/01/2019-31/01/2021	Daily	K-line pattern mining	Accuracy
Lin, Y., Liu, S., Yang, H., & Wu, H.	2021	China's stock market	2000-20017	Daily	LR, SVM, KNN, RF, GBDT, and LSTM	Accuracy Precision Recall, F1-Score
Naik, N., & Mohan, B. R.	2020	India (National Stock Exchange)	2008-2018	Daily	ANN	Accuracy F1-Score
Nakayama, A., Matsushima, H., Izumi, K., Shimada, T., Sakaji, H., & Yamada, K.	2019	Tokyo stock exchange	7/2013-6/2014	-	CNN, LR	-
Orte, F., Mira, J., Sánchez, M. J., Solana, P.	2023	Cryptocurrency price (BTC/USD)	2013-2020	Daily	Random Forest	Accuracy
Pan, W., Li, J., & Li, X.	2020	China's stock market	1/01/2017-31/12/2018	Daily	CAE encoder K-mean	Sharpe ratio
Thammakesorn, S., & Sornil, O.	2019	Stock exchange Thailand	2008-2017	Daily	Chi-square Automatic Interaction Detector (CHAID)	Profit
U, J. H., Lu, P. Y., Kim, C. S., Ryu, U. S., & Pak, K. S.	2020	USA stock market, China's stock market	-	Daily	Long and short-term memory	Accuracy, Precision, Recall, F1-Score
Santur, Yunus	2022	Dow Jones (ABD), Nifty 50	(04/01/2007-11/12/2019) (04/01/2000-	Daily	xgboost	Accuracy

		(India), S&P 500 (ABD), Shanghai (China), Dax (Germany), Cac 40 (France), S&P TSX (Toronto), Russell 2000 (London), Ibex 35 (Madrid), Kospi (Korea), Bist 30 (Turkey)	11/12/2019) (04/01/2006-11/12/2019) (04/01/2000-11/12/2019) (03/01/2001-11/12/2019) (04/01/2000-23/07/2019) (03/01/2001-27/11/2019) (04/01/2000-27/12/2019) (03/01/2000-05/09/2019) (04/01/2000-27/12/2019) (04/01/2000-10/12/2019)			
Udagawa, Y.	2019	USA (NASDAQ)	25/07/2010-4/06/2019	Daily	Retrieval algorithm	Statistical
Wang, M., & Wang, Y.	2019	USA (S&P 500)	2010-2018	Daily	evaluation candlestick pattern	binomial test
Lin, Liu, Yang, Wu, et al.	2021	China's stock market	1/1/2000-30/10/2020	Daily	LR, KNN, RBM, and RF	Accuracy
Rumpa et al.	2021	Currency exchange (EUR/USD)	-	5-minutes	Support Vector Machine	Accuracy
Chen & Tsai	2020	Currency exchange (EUR/USD)	2010-2017	Daily	GAF-CNN	Accuracy, Precision, Recall,
Wang et al.	2021	China's stock market (Kweichow Moutai, CSI 100 index, 50 ETF)	27/08/2001-14/06/2019 30/12/2005-14/06/2019 23/02/2005-14/06/2019	Daily	Vector auto-regression Vector error correction	MAPE, RMSE, RMSEH, AR
Birogul et al.	2020	Turkey's stock market (Borsa Istanbul)	2000-2018	Daily	Convolutional neural network YOLO	Profit
Ahmad et al.	2020	Indonesian Stock Market Exchange (LQ45 stocks)	1/2020 - 7/2020	Daily	Fuzzy model	Profit
Hassen et al.	2020	USA (Boeing Company, Bank of America, DuPont, Ford Motor Co. General Electric, Hewlett–Packard, Microsoft, Monsanto, Toyota Motor, Wells Fargo,	(02/01/1962-27/06/2018) (03/01/2000-12/12/2014) (03/01/2000-12/12/2014) (03/01/2000-12/12/2014) (03/01/2000-12/12/2014) (03/01/2000-12/12/2014) (03/01/2000-12/12/2014)	Daily	cloud model	The mean squared error (MSE), mean absolute percentage error (MAPE)

		Yahoo, Exxon Mobil, Walt Disney)	(18/10/2000-12/12/2014) (03/01/2000-12/12/2014) (01/06/1972-27/06/2018) (03/01/2005-12/12/2014) (02/01/1970-21/05/2018) (02/01/1962-27/06/2018)			
Naranjo and Santos,	2019	USA (Nasdq100), Spanish (lbex35)	1/01/2011-30/04/2015 1/01/2011-20/09/2015	Daily	Fuzzy model	Sharpe Ratio
Poženel & Lavbič,	2019	Russell's top 50 index	-	Daily	NLP (Word2Vec)	Returns

4. DISCUSSIONS

The data obtained from the analysis conducted in the previous section are compiled and presented in Table 1. The table effectively illustrates the range of methodologies examined in this comprehensive analysis as well as the evolution of research approaches over time. One noteworthy finding is that despite the absence of intrinsic restrictions on the publishing year in the systematic investigation, the resulting collection of papers, following a meticulous application of the inclusion and exclusion criteria, encompassed the period from 2019 to 2023. This statement suggests that the primary focus of this research is to investigate and analyze the effectiveness of Japanese candlestick charts in predicting stock market trends and movements. The observed decrease in the number of publications focusing on stock market prediction using candlestick charts as images can be attributed to the complexity and limitations of image-based analysis. Analyzing candlestick charts as images using deep learning involves significant complexity. The process requires high computational resources and expertise in both finance and advanced machine-learning techniques. Additionally, candlestick charts primarily provide price and volume information, which may need to be improved for a comprehensive market analysis. Difficulty in Capturing Market Dynamics: Candlestick charts capture static snapshots of market conditions. However, stock markets are also influenced by a myriad of dynamic factors. For a better visual understanding, readers may consult Figure 4, which depicts the temporal evolution of the volumes of important publications. Simultaneously, Figure 5 provides a more comprehensive analysis of the various forecasting approaches that utilize the potential of the Japanese candlestick charts.

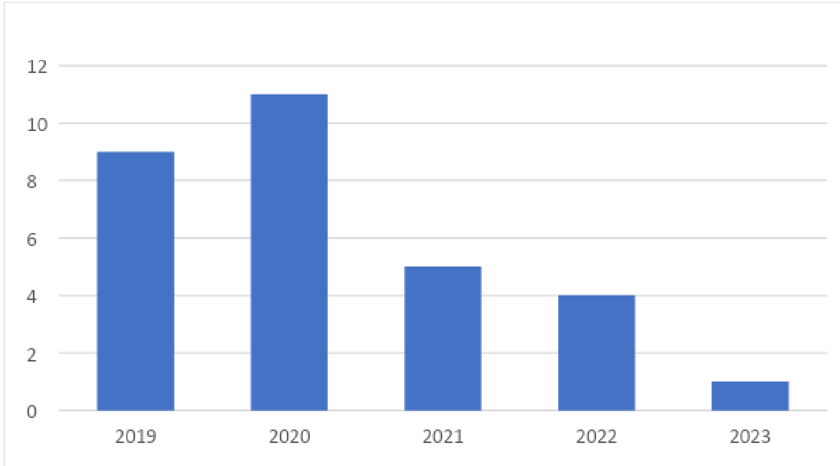

FIGURE 4. number of publications

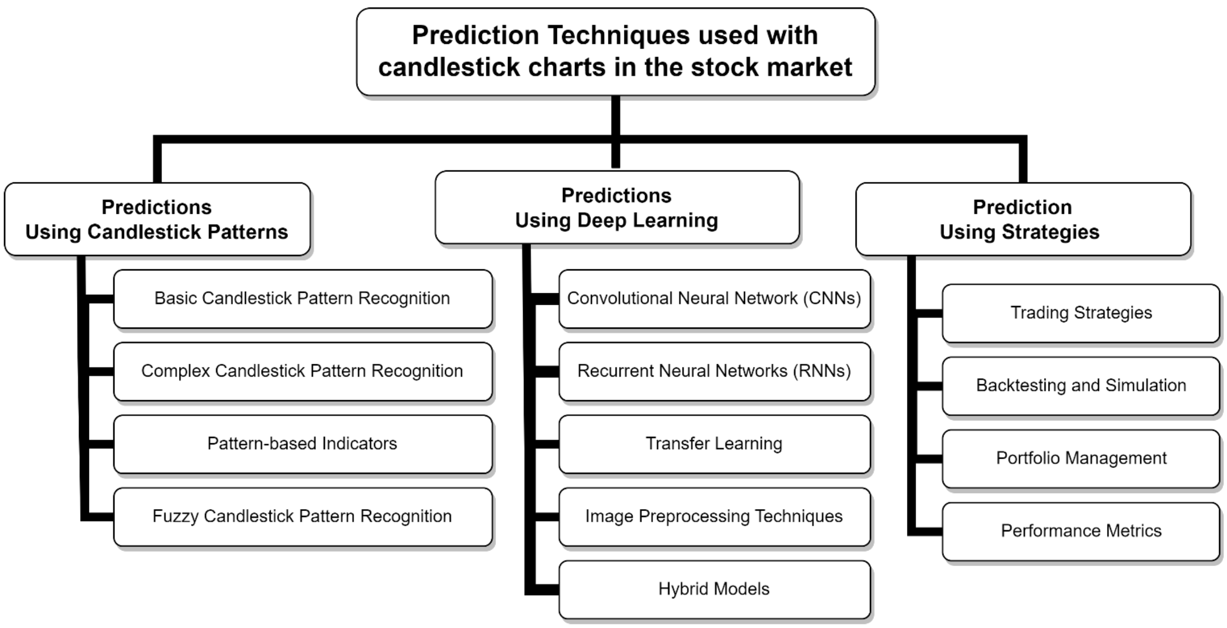

FIGURE 5. Analysis Of Forecasting Methods Using Japanese Candlestick Charts

Candlestick charts are a well-liked tool traders use to analyze and forecast stock market movements. Prior to addressing these inquiries, it is imperative to obtain a clear understanding of the methodology employed for image prediction. In the context of forecasting, the utilization of images does not enable accurate prediction of profits and returns associated with candle representations. While it is feasible to ascertain the forthcoming candle's direction, whether it will be bullish or bearish, this determination is adequate for assessing the potential for engaging in transactions involving the purchase or sale of shares, provided that the prediction proves correct.

The time series is divided into multiple stages, denoted as T_0 to T_n , where each stage is represented by a candle. The direction of the trend, whether upward or downward, is determined by comparing the closing price of the T_n candle with the closing price of the T_{n+1} candle. Thus, a collection of visual representations was generated. Figure 6 shows the selected candles as objects from T_0 to T_n and the next T_{n+1} is a prediction candle.

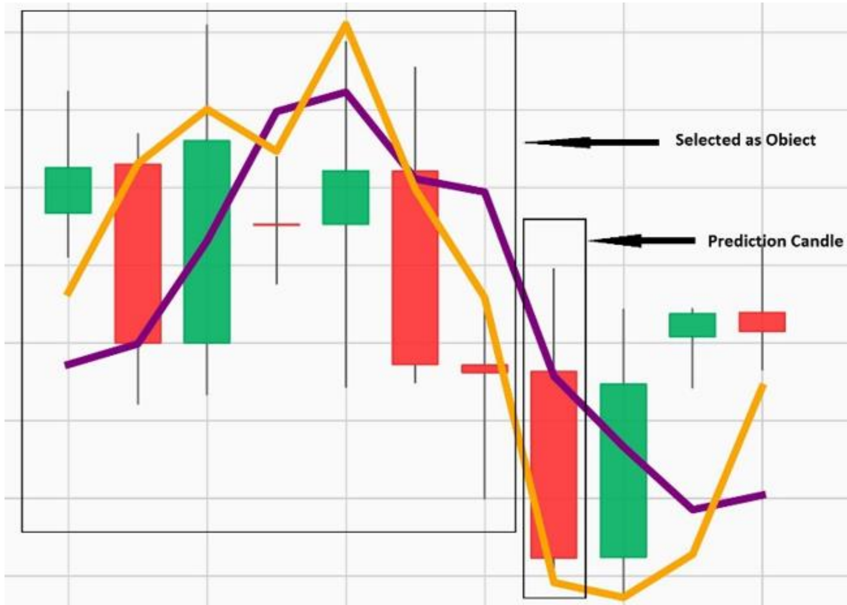

FIGURE 6**.** Object And Prediction Candle

Feature engineering is the process of selecting relevant and transformed data features to improve the performance of machine-learning models. Based on the research results, the answer to the first question is: What are the best engineering methods used with candlestick charts to predict the direction of the stock market? Feature engineering is the data-processing stage before building the model. Hung and Chen (2021) and Pan et al. (2020) used convolutional autoencoders to extract deep features from subcharts. In (Barra et al., 2020; Chen & Tsai, 2022), the authors converted time-series datasets to GAF and fed them to CNN. (Birogul et al., 2020); (Andriyanto, 2020); (Jearanaitanakij & Passaya, 2019); (Lee et al., 2019); (Kusuma et al., 2019) used a convolutional neural network for feature extraction. (Naik & Mohan, 2020) classified up or down movement using candlestick pattern and technical indicator features, the feature extraction created by condition for each technical indicator with N days, and for the candlestick pattern with the mean of stock price. (Du et al., 2020) The method used for feature extraction is the wavelet transform for denoising, and for feature selection, the information coefficient method selects the top five training indicators as the selected input features. (Thammakesorn & Sornil, 2019) The features used are twenty-six, and they are configured based on each of the colors, relative positions, and sizes of the candlesticks (Liang et al., 2022). A new sequence similarity approach was developed to match the sequences of various candlesticks to existing patterns and to apply the new sequences as predictive features. (Fengqian & Chao, 2020) To extract the features of high robustness, the K-line is decomposed, and the results of each sub-part of the K-line are clustered. The corresponding cluster centers were obtained using the K-means and fuzzy c-means clustering methods to cluster each part of the K-line and were used as the input state of the model. (Udagawa, 2019) A feature of candlestick patterns is that they use only the closing price and length of real bodies. U et al. (2020) used a combination of technical indicators and candlestick patterns. The type of candlestick pattern used was a one-day candlestick pattern created by calculation formulas. (Lin et al., 2021), and in different papers they used nine technical indicators with two of the candlestick charts: the feature of shape and the feature of candlestick relative position. In summary, to address the first research question, it is concluded that a comprehensive approach employing sophisticated neural networks, particularly Convolutional Neural Networks (CNNs), along with deep learning, reinforcement learning, and hybrid models, represents the most effective strategy for predicting stock market trends through the use of candlestick charts. The integration of these methodologies provides a robust and flexible framework for financial forecasting, adept at managing the intricacies and volatility inherent in stock markets. Figure 7 depicts the taxonomy employed in this study to classify the various models analyzed.

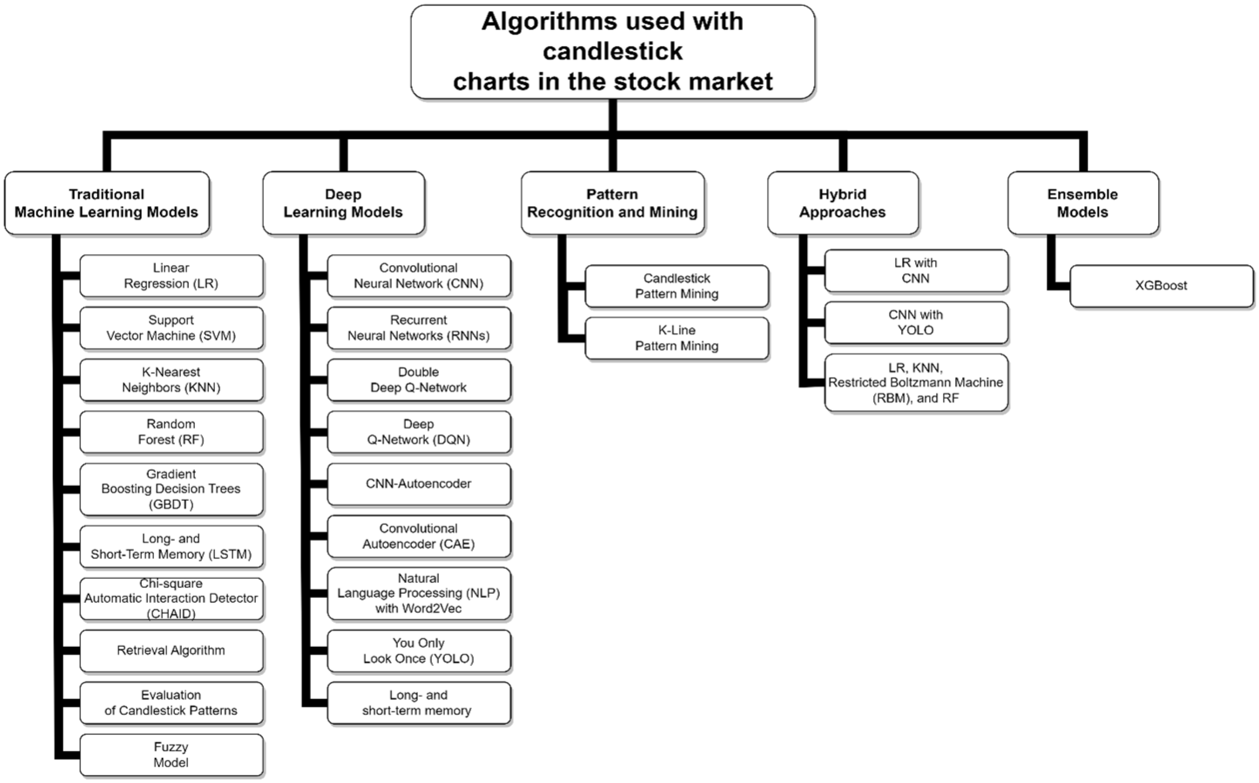

FIGURE 7. Taxonomy Of Stock Market Prediction Algorithms

Stock market data usually consists of previous price and volume information for a particular stock or index. The stock markets studied in the reviewed research include the US, China, Taiwan, Thailand, Indonesia, India, Japan, Turkey, Spain, France, Korea, the United Kingdom, Canada, and Germany. Figure 8 shows the groups of stock markets for which the forecasting studies were conducted. The answer to the second question is (RQ2): When utilizing a Japanese candlestick chart to predict the stock market, what timeframe should be used?

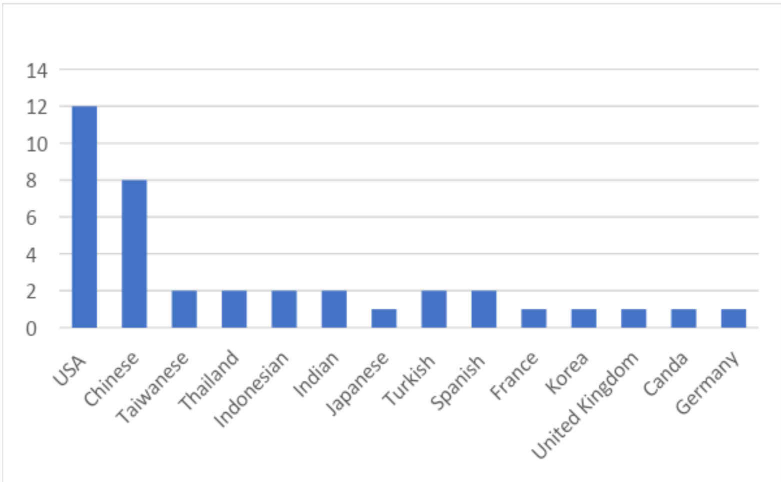

FIGURE 8. Stock Market

The timeframe used in the analysis was a daily timeframe, a 1-minute timeframe, a weekly timeframe, and a (1-4-8) hour timeframe. The most common timeframe in these studies is the daily frame. Daily data might be the most suitable and close substitute for short-term trading in situations where a trading strategy determines the time period. Moreover, historical data may be the most appropriate if the goal is to study long-term market trends or evaluate the performance of a long-term investment strategy. It is important to consider the timeframe of the appropriate forecast, as well as the availability and quality of data when selecting the time period for a dataset covering the stock market.

To evaluate the efficacy of stock market prediction models to determine their reliability, it is necessary to conduct a thorough performance study. The accuracy of a model's prediction can be evaluated by comparing its predicted values with the actual values of stock prices or returns. The number of studies that used various tools to evaluate the models is shown in Figure 9. These tools include accuracy, precision, recall, specificity, F1, Sharpe ratio, statistical and binomial tests, MAPE, RMSE, RMSEH, AR, and profitability.

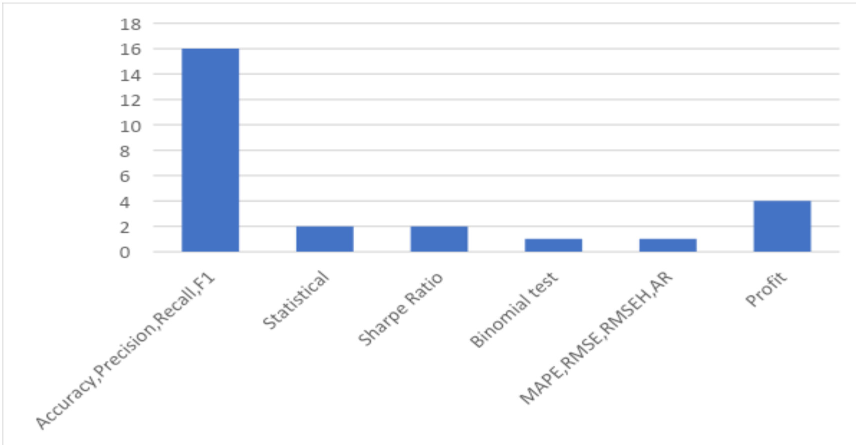

FIGURE 9. Evaluating Model Performance

5. CONCLUSION

This study conducts a comprehensive review of academic literature on candlestick forecasting in stock markets. The search and retrieval process yielded 30 articles that were scrutinized in our methodology. These articles specifically focus on the application of Japanese candlesticks in stock market analysis. They encompass various studies conducted to predict market trends using candlestick patterns, employing diverse algorithms and deep learning techniques, particularly the use of images.

A notable aspect of this study is its exploration of the use of convolutional neural network (CNN) models in conjunction with images for stock market prediction. This approach reflects the integration of computer vision techniques with Japanese candlestick charts, underscoring their significance in forecasting stock market directions.

The study's contribution is significant as it illuminates the application of computer vision and Japanese candlestick charts in stock market prediction. It also highlights potential areas for future research and model development, including the design of portfolio management strategies and prediction of market events through sentiment analysis, automated pattern mining, and analysis of percentage changes in stock price.

Conflict of Interest

The authors have no affiliation with any organization with a direct or indirect financial interest in the subject matter discussed in the manuscript.

REFERENCES

[1] Ahmad, M., Soeparno, H., & Napitupulu, T. A. (2020). Stock trading alert: fuzzy knowledge-based systems and technical analysis. 2020 International Conference on Information Technology Systems and Innovation, ICITSI 2020 - Proceedings, 155–160. https://doi.org/10.1109/ICITSI50517.2020.9264914

[2] Andriyanto, A. (2020). Sectoral stock prediction using Convolutional Neural Networks with candlestick patterns as input images. International Journal of Emerging Trends in Engineering Research, 8(6), 2249–2252. https://doi.org/10.30534/ijeter/2020/07862020

[3] Barra, S., Carta, S. M., Corriga, A., Podda, A. S., Recupero, D.R. (2020). Deep learning and time series-to-image encoding for financial forecasting. IEEE/CAA Journal of Automatica Sinica, 7(3), 683–692. https://doi.org/10.1109/JAS.2020.1003132

[4] Birogul, S., Temur, G., & Kose, U. (2020). YOLO Object Recognition Algorithm and "Buy-Sell Decision" Model over 2D Candlestick Charts. IEEE Access, 8, 91894–91915. https://doi.org/10.1109/ACCESS.2020.2994282

[5] Brim, A., & Flann, N. S. (2022). Deep reinforcement learning stock market trading utilizing a CNN with candlestick images. PLoS ONE, 17(2 February). https://doi.org/10.1371/journal.pone.0263181

[6] Chen, J. H., & Tsai, Y. C. (2020). Encoding candlesticks as images for pattern classification using convolutional neural networks. Financial Innovation, 6(1). https://doi.org/10.1186/s40854-020-00187-0

[7] Chen, J.-H., & Tsai, Y.-C. (2022). Dynamic Deep Convolutional Candlestick Learner. http://arxiv.org/abs/2201.08669

[8] Du, B., Fernandez-Reyes, D., & Barucca, P. (2020). Image Processing Tools for Financial Time Series Classification. http://arxiv.org/abs/2008.06042

[9] Fengqian, Di., & Chao, L. (2020). An Adaptive Financial Trading System Using Deep Reinforcement Learning with Candlestick Decomposing Features. IEEE Access, 8, 63666–63678. https://doi.org/10.1109/ACCESS.2020.2982662

[10] Hassen, O. A., Darwish, S. M., Abu, N. A., & Abidin, Z. Z. (2020). Application of cloud model in qualitative forecasting for stock market trends. Entropy, 22(9). https://doi.org/10.3390/e22090991

[11] Hung, C. C., & Chen, Y. J. (2021). DPP: Deep predictor for price movement from candlestick charts. PLoS ONE, 16(6 June 2021). https://doi.org/10.1371/journal.pone.0252404

[12] Jearanaitanakij, K., & Passaya, B. (2019). Predicting Short Trend of Stocks by Using Convolutional Neural Network and Candlestick Patterns. 2019 4th International Conference on Information Technology (InCIT2019).

[13] Kusuma, R. M. I., Ho, T.-T., Kao, W.-C., Ou, Y.-Y., & Hua, K.-L. (2019). Using Deep Learning Neural Networks and Candlestick Chart Representation to Predict Stock Market. http://arxiv.org/abs/1903.12258

[14] Lee, J., Kim, R., Koh, Y., & Kang, J. (2019). Global Stock Market Prediction Based on Stock Chart Images Using Deep Q-Network. https://doi.org/10.1109/ACCESS.2019.2953542

[15] Liang, M., Wu, S., Wang, X., & Chen, Q. (2022). A stock time series forecasting approach incorporating candlestick patterns and sequence similarity. Expert Systems with Applications, 205. https://doi.org/10.1016/j.eswa.2022.117595

[16] Lin, Y., Liu, S., Yang, H., & Wu, H. (2021). Stock Trend Prediction Using Candlestick Charting and Ensemble Machine Learning Techniques with a Novelty Feature Engineering Scheme. IEEE Access, 9, 101433–101446. https://doi.org/10.1109/ACCESS.2021.3096825

[17] Lin, Y., Liu, S., Yang, H., Wu, H., & Jiang, B. (2021). Improving stock trading decisions based on pattern recognition using machine learning technology. PLoS ONE, 16(8 August). https://doi.org/10.1371/journal.pone.0255558

[18] Naik, N., & Mohan, B. R. (2020). Intraday Stock Prediction Based on Deep Neural Network. National Academy Science Letters, 43(3), 241–246. https://doi.org/10.1007/s40009-019-00859-1

[19] Nakayama, A., Matsushima, H., Izumi, K., Shimada, T., Sakaji, H., & Yamada, K. (2019). Short-term Stock Price Prediction by Analysis of Order Pattern Images; Short-term Stock Price Prediction by Analysis of Order Pattern Images. In 2019 IEEE Conference on Computational Intelligence for Financial Engineering & Economics (CIFEr).

[20] Naranjo, R., & Santos, M. (2019). A fuzzy decision system for money investment in stock markets based on fuzzy candlesticks pattern recognition. Expert Systems with Applications, 133, 34–48. https://doi.org/10.1016/j.eswa.2019.05.012

[21] Orte, F., Mira, J., Sánchez, M. J., Solana, P. (2023). A random forest-based model for crypto asset forecasts in futures markets with out-of-sample prediction. Research in International Business and Finance, 64. https://doi.org/10.1016/j.ribaf.2022.101829

[22] Pan, W., Li, J., & Li, X. (2020). Portfolio learning based on deep learning. Future Internet, 12(11), 1–13. https://doi.org/10.3390/fi12110202
[23] Poženel, M., & Lavbič, D. (2019). Discovering the Language of Stocks. https://doi.org/10.3233/978-1-61499-941-6-243
[24] Rumpa, L. D., Limbongan, M. E., Biringkanae, A., & Tammu, R. G. (2021). Binary options trading: candlestick prediction using Support Vector Machine (SVM) on M5 time period. IOP Conference Series: Materials Science and Engineering, 1088(1), 012107. https://doi.org/10.1088/1757-899x/1088/1/012107
[25] Santur, Y. (2022). Candlestick chart-based trading system using ensemble learning for financial assets. Sigma Journal of Engineering and Natural Sciences – Sigma Mühendislik ve Fen Bilimleri Dergisi. https://doi.org/10.14744/sigma.2022.00039
[26] Thammakesorn, S., & Sornil, O. (2019). Generating Trading Strategies Based on Candlestick Chart Pattern Characteristics. Journal of Physics: Conference Series, 1195(1). https://doi.org/10.1088/1742-6596/1195/1/012008
[27] JuHyok U., PengYu L., ChungSong K., UnSok R., KyongSok P. (2020). A new LSTM based reversal point prediction method using upward/downward reversal point feature sets. Chaos, Solitons and Fractals, 132. https://doi.org/10.1016/j.chaos.2019.109559
[28] Udagawa, Y. (2019, December 2). Mining stock price changes for profitable trade using candlestick chart patterns. ACM International Conference Proceeding Series. https://doi.org/10.1145/3366030.3366053
[29] Wang, H., Huang, W., & Wang, S. (2021). Forecasting open-high-low-close data contained in candlestick chart. http://arxiv.org/abs/2104.00581
[30] Wang, M., & Wang, Y. (2019). Evaluating the effectiveness of candlestick analysis in forecasting U.S. stock market. ACM International Conference Proceeding Series, 98–101. https://doi.org/10.1145/3314545.3314555