INTERNATIONAL JOURNAL OF ACADEMIC RESEARCH IN ECONOMICS & MANAGEMENT SCIENCES

OPEN ACCESS JOURNAL

Vol 15, Issue 4, (2026) E-ISSN: 2226-3624

A Hybrid AR and Leading Technical Indicator Framework for TSMC Stock Price Prediction

Liang Ying Wei*

Graduate Institute of Digital Technology Innovation and Management Yuanpei University of Medical Technology, Hsin-Chu, Taiwan

*Corresponding Author Email: lywei@mail.ypu.edu.tw

DOI Link: http://dx.doi.org/10.6007/IJAREMS/v15-i4/29117

Published Online: 18 November 2026

Abstract

Accurately predicting stock price movements remains a challenging task due to the inherent volatility of financial markets and the complex temporal dependencies of financial time series. This study uses the stock price data of Taiwan Semiconductor Manufacturing Company (TSMC), one of the world's leading semiconductor manufacturers, as the basis for stock price prediction. Daily closing prices from 2016 to 2025 were collected and divided into training and testing datasets on a yearly basis. To improve prediction accuracy, an autoregressive model of order one (AR(1)) and technical indicators were combined with three machine learning algorithms: Random Forest (RF), Neural Network (NN), and Support Vector Regression (SVR). The performance of each prediction model was evaluated using Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE). The prediction errors generated by the different models were analyzed and compared using the testing dataset for each year. The experimental results indicate that TSMC's stock price exhibited significant upward trends and increased volatility during the period from 2020 to 2025, mainly due to the impacts of the COVID-19 pandemic, the U.S.–China trade war, and the rapid growth of the artificial intelligence (AI) industry. Among the six prediction models, the proposed AR(1)-TI-NN model, which integrates AR(1), technical indicators, and a neural network, achieved the best overall prediction performance. It produced the lowest average MAE (6.49995) and RMSE (10.06181) and achieved the best prediction accuracy in 18 out of 20 experimental comparisons. These findings demonstrate that incorporating technical indicators into the AR(1)-NN framework significantly improves stock price prediction accuracy and provides a reliable reference for investors and financial analysts in investment decision-making.

Keywords: Semiconductor Industry, Stock Price, Data Mining

Introduction

Stock market prediction has become an important research topic in finance, with applications in investment strategies, risk management, and economic forecasting(Eunsuk et al., (2017), Thomas et al., (2018)). Accurate prediction of stock prices enables investors to make more informed investment decisions and improve portfolio performance. However, the high volatility and uncertainty of global financial markets present significant challenges for stock

price forecasting (Rama (2001)), making accurate prediction difficult and increasing exposure to financial risks (Burton 2023). Consequently, developing more effective forecasting models capable of capturing complex market dynamics has become an important research objective.

Traditional financial forecasting models, such as the Autoregressive (AR), Autoregressive Moving Average (ARMA), and Autoregressive Integrated Moving Average (ARIMA) models (George et al., (2015)), have been widely used for time-series forecasting. However, as reported by Zhang (2003), these statistical models are primarily designed to capture linear relationships and often fail to model the nonlinear characteristics of financial markets, thereby limiting their forecasting accuracy under highly volatile market conditions. In contrast, artificial intelligence (AI) and machine learning techniques can effectively model both linear and nonlinear relationships, making them promising approaches for improving stock price prediction.

Technical analysis has long been used to assist investors in identifying market trends and predicting future price movements. Among various technical analysis tools, momentum indicators measure the rate of change in asset prices and are commonly used to evaluate the strength of market trends and identify potential buying or selling opportunities. Momentum trading is based on the assumption that assets exhibiting strong upward or downward price movements are likely to continue moving in the same direction for a certain period. By evaluating the strength and sustainability of price momentum, traders attempt to enter the market during accelerating trends and exit before momentum weakens or reverses.

Momentum indicators can generally be classified into two categories: leading indicators and lagging indicators. Leading indicators are designed to identify potential trend reversals before they occur, while lagging indicators are primarily used to confirm existing trends. Many momentum indicators are oscillators that fluctuate within a fixed range, allowing traders to identify overbought and oversold market conditions.

Among the leading momentum indicators, the Relative Strength Index (RSI), Stochastic Oscillator (%K and %D), and Williams %R are widely used in financial markets. The Relative Strength Index (RSI) ranges from 0 to 100 and measures the magnitude of recent price movements. An RSI value above 70 generally indicates that an asset is overbought and may experience a price correction, whereas a value below 30 suggests that the asset is oversold and may rebound. The Stochastic Oscillator compares the closing price with the price range over a specified period, typically 14 trading days. Values above 80 indicate overbought conditions, while values below 20 indicate oversold conditions. Traders frequently use crossovers between the %K and %D lines to identify potential trend reversals. Williams %R measures the position of the closing price relative to the highest price over a specified period and ranges from –100 to 0. Values above –20 indicate overbought conditions, whereas values below –80 indicate oversold conditions.

Motivated by the strengths of autoregressive models, technical indicators, and artificial intelligence, this study proposes a hybrid stock price prediction model that integrates the AR(1) model with leading momentum indicators, including RSI, %K, %D, and Williams %R, and combines them with machine learning algorithms. The proposed framework employs Random Forest (RF), Neural Network (NN), and Support Vector Regression (SVR) models to

predict the stock price of Taiwan Semiconductor Manufacturing Company (TSMC). The prediction performance of the proposed model is evaluated using Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE), and its forecasting capability is compared with that of conventional prediction models.

Related works

Time Series Analysis

Because historical stock prices are sequential observations recorded over time, stock price data can be analyzed using time series analysis. A time series is a sequence of observations arranged in chronological order, and its primary objective is to identify temporal patterns and dependencies among observations at different time points. Understanding these characteristics provides a foundation for developing forecasting models.

Time series data generally exhibit four fundamental characteristics:
Stationarity: The statistical properties of the time series, such as the mean and variance, remain relatively constant over time.
Trend: The long-term movement of the data, indicating a gradual upward or downward direction over a period of time.
Cyclicality: Recurrent fluctuations that occur over relatively long periods, reflecting repeated patterns associated with economic or market cycles.
Randomness: Irregular and unpredictable fluctuations caused by random or unforeseen factors that cannot be explained by systematic patterns.
Time series exhibiting stationarity, trends, or cyclicality are generally more suitable for forecasting because these characteristics provide identifiable patterns that predictive models can learn. In contrast, highly random time series contain limited structural information, making accurate forecasting considerably more challenging.

Random Forest

The Random Forest (RF) algorithm was first introduced by Tin Kam Ho (1995), who proposed the concept of random decision forests, and was subsequently refined by Leo Breiman (2001), who established the Random Forest algorithm in its modern form. Random Forest is an ensemble learning method that consists of multiple independently constructed decision trees. By combining the predictions of these trees, Random Forest generally achieves higher prediction accuracy and better generalization performance than a single decision tree.

In this study, the dataset is first divided into training and testing datasets. Suppose the training dataset contains **N** samples, each with **M** input features. Random Forest employs the bootstrap sampling technique, in which **N** bootstrap datasets are generated by randomly sampling the original training data with replacement. Consequently, approximately 36.8% of the original samples are not selected in each bootstrap sample. These excluded samples are referred to as **out-of-bag (OOB)** samples and can be used to estimate the prediction error without requiring an additional validation dataset.

For each decision tree, a random subset of **m** features (**m < M**) is selected at each split instead of considering all input features. This random feature selection reduces the correlation among decision trees and improves the robustness of the ensemble model. After all decision trees have been constructed, the final prediction is obtained by aggregating the outputs of the

individual trees. For classification problems, the final prediction is determined by majority voting, whereas for regression problems, such as stock price prediction, the final output is calculated as the average of all decision tree predictions.

Advantages of Random Forest

1. Efficiently handles large-scale datasets with high-dimensional features.
2. Produces robust prediction performance and is less prone to overfitting than a single decision tree.
3. Can effectively handle missing values and noisy data.
4. Performs well on imbalanced datasets.
5. Uses out-of-bag (OOB) samples to estimate prediction error without requiring a separate validation dataset.
6. Provides estimates of feature importance, which help identify influential variables.

Disadvantages of Random Forest

1. Requires relatively high computational cost and memory when constructing a large number of decision trees.
2. The ensemble structure reduces model interpretability, making the decision-making process more difficult to explain than that of a single decision tree.
3. Model training may become time-consuming for very large datasets or when a large number of trees is used.

Artificial Neural Networks

The Artificial Neural Network (ANN) was first proposed by Warren McCulloch and Walter Pitts in 1943 as a computational model inspired by the structure and functioning of the human brain. An ANN is designed to simulate the learning and information-processing capabilities of biological neural networks. Because the human brain consists of billions of interconnected neurons that communicate through electrical and chemical signals, ANN models imitate this mechanism by using interconnected artificial neurons (also called nodes) to process information and learn from data.

A biological neuron receives signals from other neurons, processes the information, and transmits the output to connected neurons. Similarly, an ANN consists of multiple interconnected nodes organized into different layers. Each connection is assigned a weight, and the network learns by adjusting these weights during the training process. Through repeated learning, an ANN can automatically identify complex nonlinear relationships within data, making it well suited for prediction, classification, and pattern recognition tasks.

The development of neural networks accelerated with the introduction of the **backpropagation (BP)** learning algorithm. Although the concept of backpropagation was proposed earlier, it gained widespread attention following the work of David E. Rumelhart, Geoffrey E. Hinton, and Ronald J. Williams (1986), who demonstrated its effectiveness for training multilayer neural networks. Backpropagation updates the connection weights by propagating prediction errors backward through the network, enabling the model to minimize the prediction error iteratively and improve learning performance.

A typical backpropagation neural network consists of the following three layers:

Input Layer

The input layer receives the input variables and passes them to the hidden layer without performing complex computations. In this study, the input variables include historical stock prices and leading technical indicators.

Hidden Layer

The hidden layer is located between the input and output layers and performs nonlinear transformations of the input data. It is responsible for extracting meaningful features and learning complex relationships among the input variables. The number of hidden layers and neurons is determined according to the complexity of the prediction problem. While increasing the number of hidden neurons may improve the model's ability to learn complex patterns, an excessively large network may increase computational cost and lead to overfitting.

Output Layer

The output layer generates the final prediction of the neural network. Since this study focuses on stock price forecasting, the output layer produces the predicted closing price for the next trading day.

Advantages of Artificial Neural Networks

1. Capable of modeling complex nonlinear relationships between input variables and target outputs.
2. Learns automatically from historical data without requiring explicit mathematical equations.
3. Provides strong prediction performance for large and complex datasets.
4. Demonstrates good adaptability and generalization when sufficient training data are available.
5. Can effectively integrate multiple input features, such as historical prices and technical indicators.

Disadvantages of Artificial Neural Networks

1. Requires a relatively large amount of training data to achieve satisfactory prediction performance.
2. Training can be computationally expensive and time-consuming.
3. Model performance is sensitive to the selection of network architecture and hyperparameters.
4. The internal learning process is difficult to interpret, making ANN a typical "black-box" model.
5. The model may overfit the training data if appropriate regularization techniques are not applied.

Support Vector Regression

Support Vector Machines (SVMs) were first proposed by Boser, Guyon, and Vapnik (1992) as a supervised machine learning algorithm for classification. The fundamental idea of SVM is to identify an optimal hyperplane that maximizes the margin between different classes. The data points located closest to the hyperplane are called **support vectors**, and these points

determine the position of the optimal separating hyperplane. By maximizing the margin, SVM improves the model's generalization ability and reduces the risk of overfitting.

Suppose the data points are (x_1, y_1), …, (x_i, y_i), with x_i∈R^n and y_i∈{1,-1}. Find a hyperplane f(x)=w^T+b such that y_i=1 when f(x)>0 and y_i=-1 when f(x)<0. Using f(x), we can distinguish which set the data belongs to. Based on these conditions, the equations are: H1: w^T+b=1, H2: w^T+b=-1. The distance from H1 to the hyperplane is |1-b|/‖w‖, and the distance from H2 to the hyperplane is |-1-b|/‖w‖. The resulting margin is 2/‖w‖.

Support Vector Regression (SVR) is an extension of SVM for regression problems. Instead of maximizing the separation between classes, SVR aims to construct a regression function that minimizes the prediction error while maintaining the largest possible margin. SVR introduces an **ε-insensitive loss function**, in which an ε-tube is created around the regression function. Prediction errors falling within this ε-tube are ignored and do not contribute to the loss function, whereas only observations outside the ε-tube, known as **support vectors**, influence the optimization process. This mechanism enables SVR to achieve good generalization performance while reducing the influence of noise in the training data.

To capture nonlinear relationships, SVR maps the original input data into a higher-dimensional feature space through kernel functions. Commonly used kernel functions include the linear kernel, polynomial kernel, radial basis function (RBF) kernel, and sigmoid kernel. Among these, the RBF kernel is the most widely adopted for stock price prediction because of its ability to model complex nonlinear patterns.

Advantages of Support Vector Regression

1. Effectively models complex nonlinear relationships through kernel functions.
2. Exhibits strong generalization ability and is less prone to overfitting than many conventional regression methods.
3. Performs well on high-dimensional datasets.
4. Produces reliable prediction performance when the number of training samples is relatively limited.
5. Is robust to noisy data due to the ε-insensitive loss function.

Disadvantages of Support Vector Regression

1. The prediction performance is highly dependent on the selection of kernel functions and hyperparameters.
2. Model training becomes computationally expensive for large-scale datasets.
3. Parameter tuning, such as selecting the penalty parameter (C), kernel parameters, and ε, can be time-consuming.
4. The model is less interpretable than traditional statistical regression methods.

Research Methods

This study focuses on the stock price prediction of Taiwan Semiconductor Manufacturing Company (TSMC), one of the world's leading semiconductor manufacturers. The experimental data consist of TSMC's daily closing stock prices from 2016 to 2025. The dataset was organized on a yearly basis, and each observation was arranged according to four consecutive trading days, denoted as t−2, t−1, t, and t+1. The closing prices at t−2, t−1, and t were used as input

variables to predict the closing price at t+1. If the stock market was closed on a scheduled trading day due to weekends or public holidays, the closing price of the next available trading day was used. For each year, the data from January to October were used as the training dataset, while the data from November to December were reserved as the testing dataset.

To investigate the effectiveness of different prediction approaches, this study developed six hybrid forecasting models. The baseline models combine a first-order autoregressive [AR(1)] model with three machine learning algorithms: Random Forest (RF), Neural Network (NN), and Support Vector Regression (SVR), resulting in the AR(1)-RF, AR(1)-NN, and AR(1)-SVR models. To further improve prediction accuracy, four leading technical indicators—the Relative Strength Index (RSI), Stochastic Oscillator (%K and %D), and Williams %R—were incorporated into the AR(1) framework. These enhanced features were then combined with the same three machine learning algorithms to construct three additional models: AR(1)-LI-RF, AR(1)-LI-NN, and AR(1)-LI-SVR.

The predictive performance of the six forecasting models was evaluated using two widely adopted error metrics: Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE). Lower MAE and RMSE values indicate higher prediction accuracy. The forecasting results obtained from the six models were compared to determine the effectiveness of incorporating leading technical indicators into the AR(1)-based machine learning framework.

Mean absolute error
This is the absolute difference between the predicted value and the actual value, as shown below.

$$\text{MAE} = \frac{\sum_{i=1}^{n} |y_1 - x_1|}{n}$$

y_1: predicted value
x_1: actual value
n: number of actual values

Root mean squared error:
This is the square root of the difference between the predicted value and the actual value, as shown below.

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n} (y_1 - x_1)^2}{n}}$$

y_1: predicted value
x_1: actual value
n: number of actual values

Experiments and Comparisons

TSMC Stock Price Fluctuations (2016–2025)



Figure 4-1 TSMC's stock price fluctuations from 2016 to 2025

Between 2016 and 2025, the stock price of Taiwan Semiconductor Manufacturing Company (TSMC, TWSE: 2330) generally showed a long-term upward trend, but was accompanied by several significant fluctuations. The following is an overview of each year:

Table 4-1

TSMC stock price overview for each year

Year	Stock Price Trend	Key Drivers
2016	Gradual increase (approximately NT$150 → NT$180)	Strong demand for 16nm technology and Apple A10 chip orders
2017	Steady growth (NT$180 → NT$230)	iPhone sales and mass production of the 10nm process
2018	High-level fluctuation (NT$220–270)	Cryptocurrency market slowdown and the U.S.–China trade war
2019	Strong rebound (NT$220 → NT$330)	Expansion of 5G applications and increased shipments of the 7nm process
2020	Significant surge (NT$300 → NT$530)	COVID-19-driven demand for cloud computing and high-performance computing (HPC)
2021	High-level volatility (NT$530–680)	Global semiconductor shortage and supply chain disruptions
2022	Sharp correction (NT$688 → NT$450)	U.S. interest rate hikes, technology stock sell-off, and economic slowdown
2023	Moderate recovery (NT$450 → NT$590)	Recovery in AI demand and completion of inventory adjustments
2024	Strong rally (Above NT$600 → over NT$1,000)	Explosive AI chip demand, especially from NVIDIA, and strong demand for CoWoS advanced packaging
2025	Continued upward trend with higher volatility (approximately NT$900 → over NT$1,100)	Sustained AI growth, strong earnings performance, and increased sensitivity to valuation and geopolitical risks

Table 4-2
Overall Stock Price Volatility (2016–2025)

Period	Volatility Level	Overall Trend
2016–2018	Low to Moderate	Steady upward trend
2019–2021	Moderate to High	Rapid appreciation driven by 5G and semiconductor demand
2022	High	Significant market correction
2023–2025	High	AI-driven bull market with increased price fluctuations

Summary

- **2016–2019:** Stable growth supported by advanced process technologies (10nm and 7nm) and smartphone demand.
- **2020–2021:** Rapid appreciation driven by the COVID-19 pandemic, cloud computing, and the global semiconductor shortage.
- **2022:** A significant correction caused by aggressive interest rate hikes, weakening global demand, and declining technology stock valuations.
- **2023–2025:** A strong AI-driven rally fueled by demand for AI accelerators, high-performance computing (HPC), and advanced packaging technologies (CoWoS). Despite the long-term upward trend, stock price volatility increased due to market sensitivity to AI growth expectations, capital expenditures, and geopolitical developments.

Analysis of Experimental Results

Table 4-3
TSMC MAE values from 2016 to 2025

	2016	2017	2018	2019	2020
AR(1)- RF	2.1981	5.6031	3.1301	20.6369	39.4512
AR(1)- NN	1.869	2.1996	3.3266	14.8441	13.4213
AR(1)- SVR	11.0089	2.1518	2.8736	3.6284	5.8288
AR(1)-TI- RF	4.7117	4.1444	2.8538*	32.0372	41.8854
AR(1)-TI-NN	1.7247*	1.8946*	2.8612	3.5654*	5.7813*
AR(1)-TI-SVR	20.1564	23.9974	8.7457	65.7492	155.4148
	2021	2022	2023	2024	2025
AR(1)- RF	6.5254	55.7686	5.5438	19.3665	21.767
AR(1)- NN	6.5416	50.8876	4.2118	17.2837	19.2218
AR(1)-SVR	4.4551	50.6389	37.5998	187.6647	371.1084
AR(1)-TI- RF	7.884	10.1378	6.9186	36.2326	89.122
AR(1)-TI-NN	4.3115*	8.5021*	4.0404*	13.7694*	18.5489*
AR(1)-TI-SVR	13.6672	49.2115	40.8054	188.9844	372.5083

* Best prediction among 6 algorithms

Table 4-4
TSMC RMSE values from 2016 to 2025

	2016	2017	2018	2019	2020
AR(1)- RF	4.7082	6.7676	4.0686	24.5129	44.0723
AR(1)-NN	4.2704	2.9432	4.1988	17.6538	17.6644
AR(1)-SVR	11.3162	2.8978	3.6515	14.8467	17.5956
AR(1)-TI- RF	5.9179	4.8755	3.4562*	34.453	46.9442
AR(1) -TI-NN	4.1405*	2.421*	3.4609	11.0821*	17.5045*

AR(1) -TI-SVR	20.2685	24.5846	9.9027	66.8777	157.0335
	2021	2022	2023	2024	2025
AR(1)- RF	7.8237	63.4493	6.9616	23.4478	26.6394
AR(1)-NN	7.8709	59.205	5.1007	20.1554	23.8516
AR(1)-SVR	5.7870	58.9913	39.707	189.6873	372.8506
AR(1)-TI-RF	11.5788	14.7863	8.7381	41.1703	94.3919
AR(1) -TI-NN	5.7823*	12.102*	5.042*	16.8178*	22.265*
AR(1) -TI-SVR	15.6107	57.7707	42.5673	190.906	374.2538

* Best prediction among 6 algorithms

Table 4-5
Average MAE Values of the Six Prediction Models

	AR(1)- RF	AR(1)-NN	AR(1)-SVR	AR(1)-TI-RF	AR(1) -TI-NN	AR(1) -TI-SVR
TSMC'stock price	17.99907	13.38071	67.69584	23.59275	6.49995*	93.92403

* Best prediction among 6 algorithms

Table 4-6
Average RMSE Values of the Six Prediction Models

	AR(1)- RF	AR(1)-NN	AR(1)-SVR	AR(1)-TI-RF	AR(1) -TI-NN	AR(1) -TI-SVR
TSMC'stock price	21.24514	16.29142	71.7331	26.63122	10.06181*	95.97755

* Best prediction among 6 algorithms

Table 4-7
Number of Times Each Prediction Model Achieved the Best Performance (Based on MAE and RMSE)

	AR(1)- RF	AR(1)-NN	AR(1)-SVR	AR(1)-TI-RF	AR(1) -TI-NN	AR(1) -TI-SVR
TSMC'stock price	0	0	0	2	18	0

Tables 4-5 and 4-6 summarize the average MAE and RMSE values of the six prediction models over the period from 2016 to 2025. Since lower MAE and RMSE values indicate higher prediction accuracy, these averages provide an overall comparison of the long-term forecasting performance of each model.

The experimental results demonstrate that integrating the AR(1) model with leading technical indicators and machine learning algorithms effectively improves stock price prediction accuracy. Among the six forecasting models, the proposed **AR(1)-LI-NN** model achieved the best overall performance, with the lowest average MAE (**6.49995**) and RMSE (**10.06181**). The second-best performer was **AR(1)-NN**, followed by **AR(1)-RF** and **AR(1)-LI-RF**, whereas the SVR-based models produced considerably larger prediction errors, particularly **AR(1)-LI-SVR**, which recorded the highest average MAE and RMSE values.

To evaluate the consistency of each forecasting model, Table 4-7 reports the number of times each model achieved the best prediction performance across all experiments, including both MAE and RMSE over the ten-year period (20 comparisons). The proposed **AR(1)-LI-NN** model

ranked first in **18 out of 20** experiments, while **AR(1)-LI-RF** achieved the best performance in the remaining **2** experiments. None of the other four models achieved the best prediction results in any experiment. These findings demonstrate that incorporating leading technical indicators into the AR(1)-NN framework significantly enhances both prediction accuracy and forecasting stability.

Compared with the baseline **AR(1)-NN** model, the proposed **AR(1)-LI-NN** model substantially reduced both MAE and RMSE, indicating that leading technical indicators, including the Relative Strength Index (RSI), Stochastic Oscillator (%K and %D), and Williams %R, provide valuable market information beyond historical stock prices. These indicators effectively capture short-term market momentum and potential trend reversals, enabling the neural network to learn more representative market patterns and improve forecasting performance.

The Random Forest models exhibited relatively stable prediction performance throughout most experimental years. Although **AR(1)-LI-RF** achieved the best prediction accuracy in two experiments, its overall performance remained inferior to that of **AR(1)-LI-NN**. In contrast, the Support Vector Regression (SVR)-based models generally produced much larger prediction errors, particularly during the highly volatile market period from 2024 to 2025. These results suggest that SVR is less capable of capturing the complex nonlinear characteristics and rapid price movements associated with significant market events.

The experimental results also indicate that market conditions have a considerable influence on forecasting performance. During the relatively stable period from 2016 to 2019, all prediction models achieved comparatively low prediction errors. However, beginning in 2020, the COVID-19 pandemic, the U.S.–China trade war, and the rapid growth of artificial intelligence (AI)-related semiconductor demand substantially increased TSMC's stock price volatility. Although prediction errors increased for all models during this period, the proposed **AR(1)-LI-NN** model consistently maintained the highest prediction accuracy, demonstrating its robustness and adaptability under different market conditions.

Overall, the experimental findings confirm that integrating AR(1), leading technical indicators, and neural networks provides a more effective stock price prediction framework than conventional AR(1)-based machine learning models. The proposed model consistently achieved the best forecasting performance across different market conditions and can serve as a practical decision-support tool for investors and financial analysts seeking more accurate stock price predictions.

Conclusions

This study proposed a hybrid stock price prediction framework that integrates a first-order autoregressive [AR(1)] model, leading technical indicators, and machine learning algorithms to improve the prediction accuracy of TSMC stock prices. Daily closing price data from 2016 to 2025 were used to construct six prediction models, including AR(1)-RF, AR(1)-NN, AR(1)-SVR, AR(1)-LI-RF, AR(1)-LI-NN, and AR(1)-LI-SVR. The predictive performance of these models was evaluated using Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE).

The experimental results demonstrate that incorporating leading technical indicators into the AR(1)-based prediction framework can effectively improve forecasting performance. Among

the six models, the proposed AR(1)-LI-NN model consistently achieved the highest prediction accuracy. It produced the lowest average MAE (6.49995) and RMSE (10.06181) over the 2016–2025 period and achieved the best performance in 18 out of 20 experimental comparisons. These results indicate that combining autoregressive time-series information with leading technical indicators enables the neural network model to capture both historical price dependencies and market momentum more effectively than the other prediction models.

The analysis also revealed that TSMC's stock price experienced relatively stable growth between 2016 and 2019, followed by substantially higher volatility from 2020 onward. The increased fluctuations were primarily associated with major market events, including the COVID-19 pandemic, the U.S.–China trade war, and the rapid expansion of artificial intelligence (AI)-related semiconductor demand. Despite these changing market conditions, the proposed AR(1)-LI-NN model maintained superior forecasting performance, demonstrating its robustness and adaptability under different market environments.

Overall, the findings suggest that integrating AR(1), leading technical indicators, and machine learning provides a more effective approach for stock price prediction than conventional autoregressive or machine learning models alone. The proposed framework can serve as a useful decision-support tool for investors, financial analysts, and portfolio managers by providing more accurate stock price forecasts.

Although this study focuses exclusively on TSMC stock prices, the proposed methodology can be extended to other semiconductor companies or financial assets. Future research may incorporate additional macroeconomic variables, market sentiment indicators, news and social media data, or deep learning models such as Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), and Transformer-based architectures to further improve forecasting accuracy and enhance the generalizability of the proposed prediction framework.

Theoretical and Contextual Contributions

This study makes both theoretical and contextual contributions to the literature on stock price prediction and financial time-series modeling. From a theoretical perspective, the study extends the application of hybrid forecasting approaches by integrating an autoregressive model of order one (AR(1)), technical indicators, and machine learning algorithms. Previous studies have demonstrated the effectiveness of statistical time-series models and machine learning techniques in financial forecasting; however, these approaches often emphasize either temporal dependence or nonlinear predictive relationships. The proposed AR(1)-TI-NN framework combines these two perspectives by using AR(1) to capture the linear temporal dependence of stock prices, technical indicators to represent market-related information, and a neural network to model potentially nonlinear relationships. The empirical results therefore provide additional evidence that combining statistical time-series information with technical indicators and machine learning can enhance stock price prediction performance.

The study also contributes to the existing knowledge by providing a systematic comparison of different prediction models under a consistent experimental framework. Rather than evaluating a single forecasting technique, this study compares six prediction models using yearly testing datasets and evaluates their performance based on MAE and RMSE. The finding that the AR(1)-TI-NN model achieved the lowest prediction error in 18 out of 20 experimental

comparisons provides empirical evidence regarding the value of integrating technical indicators with autoregressive and neural-network-based approaches. This comparative perspective helps clarify how different modeling components contribute to predictive performance and provides a basis for further research on hybrid financial forecasting models.

From a contextual perspective, this study focuses on Taiwan Semiconductor Manufacturing Company (TSMC), a globally significant semiconductor manufacturer, and examines daily stock price data from 2016 to 2025. This period covers substantial changes in the global semiconductor and financial environment, including the COVID-19 pandemic, U.S.–China trade tensions, and the rapid expansion of artificial intelligence applications. Therefore, the study provides empirical evidence from a specific but economically important semiconductor context in which stock prices are subject to both market volatility and technology-driven structural changes. By examining TSMC over an extended period, this research complements studies based on broader market indices or diversified portfolios and demonstrates how hybrid prediction models can be applied to an individual technology-oriented firm.

Overall, the theoretical and contextual contributions of this study lie in connecting statistical time-series modeling, technical analysis, and machine learning within a unified stock price prediction framework, while providing empirical evidence from the Taiwanese semiconductor industry. The findings may serve as a reference for future research that investigates hybrid forecasting models, alternative technical indicators, different machine learning architectures, or other technology-intensive financial markets.

References

Bayer, J. (2015). Learning sequence representations [Doctoral dissertation, Technical University of Munich].

Bayer, F. M., Bayer, D. M., & Pumi, G. (2017). Kumaraswamy autoregressive moving average models for double bounded environmental data. Journal of Hydrology, 555, 385–396.

Berry, M. J. A., & Linoff, G. (1997). Data mining techniques: For marketing, sales, and customer support. John Wiley & Sons.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroscedasticity. Journal of Econometrics, 31(3), 307–327.

Boser, B. E., Guyon, I. M., & Vapnik, V. N. (1992). A training algorithm for optimal margin classifiers. In Proceedings of the fifth annual workshop on computational learning theory (pp. 144–152). Association for Computing Machinery.

Box, G. E. P., & Jenkins, G. M. (1976). Time series analysis: Forecasting and control. Holden-Day.

Malkiel, B. G. (2003). The efficient market hypothesis and its critics. Journal of Economic Perspectives, 17(1), 59–82.

Caruana, R., Niculescu-Mizil, A., Crew, G., & Ksikes, A. (2004). Ensemble selection from libraries of models. In Proceedings of the 21st International Conference on Machine Learning (pp. 18–25). Association for Computing Machinery.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. Econometrica, 50(4), 987–1008.

Chong, E., Han, C., & Park, F. C. (2017). Deep learning networks for stock market analysis and prediction: Methodology, data representations, and case studies. Expert Systems with Applications, 83, 187–205.

Fayyad, U. M., Shapiro, G. P., Smyth, P., & Uthurusamy, R. (Eds.). (1996). Advances in knowledge discovery and data mining. AAAI Press/MIT Press.

Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). Time series analysis: Forecasting and control (5th ed.). John Wiley & Sons.

Hippert, H. S., Pedreira, C. E., & Souza, R. C. (2001). Neural networks for short-term load forecasting: A review and evaluation. IEEE Transactions on Power Systems, 16(1), 44–51.

Huarng, K. H. (2001). Effective lengths of intervals to improve forecasting in fuzzy time series. Fuzzy Sets and Systems, 123(3), 387–394.

Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32.

Chang, P. L., & Tsai, C. T. (2000). Evolution of technology development strategies for Taiwan's semiconductor industry: Formation of research consortia. Industry and Innovation, 7(2), 185–197.

Parker, D. B. (1985). Learning-logic: Casting the cortex of the human brain in silicon (Technical Report No. TR-47). Center for Computational Research in Economics and Management Science, Massachusetts Institute of Technology.

Cont, R. (2001). Empirical properties of asset returns: Stylized facts and statistical issues. Quantitative Finance, 1(2), 223–236.

Sutskever, I. (2012). Training recurrent neural networks [Doctoral dissertation, University of Toronto].

Fischer, T., & Krauss, C. (2018). Deep learning with long short-term memory networks for financial market predictions. European Journal of Operational Research, 270(2), 654–669.

Ho, T. K. (1995). Random decision forests. In Proceedings of the Third International Conference on Document Analysis and Recognition (Vol. 1, pp. 278–282). IEEE.

Zhang, G. P. (2003). Time series forecasting using a hybrid ARIMA and neural network model. Neurocomputing, 50, 159–175.