Cloud Computing and Data Science
https://ojs.wiserpub.com/index.php/CCDS/

UNIVERSAL WISER PUBLISHER

Research Article

Stock Price Prediction: A Machine Learning Approach

Moses Ashawa*, Aaron Young

Department of Cybersecurity and Networks, Glasgow Caledonian University, Glasgow, G4 0BA, UK
E-mail: Moses.Ashawa@gcu.ac.uk

Received: 25 September 2025; **Revised:** 24 November 2025; **Accepted:** 25 November 2025

Abstract: Accurately predicting stock prices remains a complex task due to market volatility influenced by economic indicators and geopolitical events. This study presents a Hybrid Stock Sequence Learner (HSSL) model that integrates Support Vector Machine (SVM), Support Vector Regression (SVR), and Linear Regression (LR) to improve forecasting performance in dynamic financial environments. The model employs an attention-gated mechanism and regularization to capture higher-order feature interactions while preventing overfitting. SVM captures non-linear patterns, LR enhances interpretability by calculating feature impacts, and SVR enables adaptive modelling through high-dimensional feature mapping. The HSSL model was empirically evaluated using historical stock data from five entities, namely Apple, Microsoft, Walt Disney, Alphabet, and the S&P 500 index. This data was sourced from Yahoo Finance. Results show that HSSL achieves a Mean Squared Error (MSE) of 4.414 and a Root Mean Squared Error (RMSE) of 29.843, outperforming baseline models. These results demonstrate that our model effectively reduces prediction errors and captures market trends. Our proposed approach offers a robust and interpretable forecasting framework suitable for short-term stock price prediction and decision-making in highly volatile markets, serving as a vital tool for stock investors to evaluate potential risks and adjust strategies to minimize losses.

***Keywords*:** Market microstructure, microeconomic factors, attention map, trading volume, volatile markets, stock market

1. Introduction

The stock market serves as a centralized platform for the trading and circulation of stocks and to enable companies to raise capital by issuing shares to investors. It is widely regarded as a barometer of a country's financial health and economic stability. The origins of stock trading and price prediction trace back to the early 17th century with the establishment of the Amsterdam Stock Exchange in 1602 [1]. Subsequently, in 1792, the New York Stock Exchange (NYSE) emerged from the union of stockbrokers and merchants under the Buttonwood Agreement (BA). BA introduced security and standardization to stock trading practices [2] The Chinese stock market was later established following the communist revolution in 1949 [3] which led to the advent of modern financial data platforms, such as Yahoo Finance [4]. Launched in 1996 as part of Yahoo's internet services, Yahoo Finance has become one of the leading sources for financial data and market information [5]. These advancements have profoundly reshaped global trade dynamics and investment strategies.

During the late 1990s and early 2000s, Yahoo Finance expanded its offerings to include both real-time and

DOI: https://doi.org/10.37256/ccds.7120268731

historical stock data. This expansion played a pivotal role in facilitating stock price prediction and financial analysis for major corporations, including Alphabet Inc., Apple Inc., Microsoft Corporation, Walt Disney Co., and the S&P 500 index. These entities monitor critical market metrics such as market penetration, enterprise adoption, and strategic acquisitions, which function as essential macroeconomic indicators for forecasting broader market trends. The stock market is influenced by a complex interplay of factors, including economic conditions, political and geopolitical events, market liquidity, natural disasters, pandemics (e.g., COVID-19), market cycles, and inflation, among others [6-8]. Due to the interconnected nature of global markets, financial activities and policies enacted in one region can have a significant impact on stock prices worldwide. Interventions by governments or central banks aim to stabilize markets. However, their effectiveness depends on the specific economic context, which, in most cases, may introduce either volatility or stability. Given this complexity, stock price movements are often non-linear and difficult to predict using simple linear models. This challenge requires an intelligent forecasting approach to optimize investment returns and manage risk effectively.

In this paper, we propose a hybrid Machine Learning (ML) predictive model combining Support Vector Machine (SVM), Support Vector Regression (SVR), and Linear Regression (LR) to forecast stock prices over a 28-day horizon. The SVM component captures the non-linear relationships between input features and stock prices through kernel functions that accommodate the complex patterns inherent in financial time series. LR interprets the influence of individual features on stock price movements. Earnings Per Share (EPS) [9] and the Price-to-Earnings (P/E) ratio [10] are used as independent variables to indicate stock performance. SVR maps the input features into a high-dimensional space, enabling effective regularization and dynamic adaptation to market fluctuations. The model's performance was quantitatively assessed using Mean Squared Error (MSE) and Root Mean Squared Error (RMSE) metrics.

In this paper, the origins of stock trading and price prediction are traced back to the 17th to 19th centuries, beginning with the establishment of the Amsterdam Stock Exchange (ASE) in 1602 [11]. Later, the New York Stock Exchange (NYSE) was founded in 1792 through the Buttonwood Agreement [12]. Early approaches to stock prediction primarily relied on fundamental analysis and market rumours. These foundational theories aimed to predict market movements by examining financial assets, earnings, and overall market positioning. Investment decisions were often influenced by insider information and prevailing public sentiment as conveyed through market reports [13]. A prominent example is Dow's Theory, proposed by Charles Dow. The theory posits that market prices reflect all known information and that historical price patterns can be analyzed to forecast future movements [14]. Despite their historical significance, these theories offered only a rudimentary understanding of stock market behaviour and exhibited notable limitations. The theory tended to depend excessively on generalized market assertions without adequately incorporating the complexities of broader economic factors. By examining historical price movements and identifying consistent patterns according to Dow's Theory, it becomes feasible to forecast future price changes. Dow Theory established the foundational framework for technical analysis [15]. It significantly influenced the emergence of the Efficient Market Hypothesis (EMH) [16]. Although Dow Theory does not account for unforeseen events or sudden shifts in market dynamics, it remains a fundamental component of contemporary financial analysis and forecasting. Building upon Dow Theory, technical analysis has gained widespread adoption as a method for predicting stock prices by focusing on the evaluation of price charts and trading volumes to detect market trends [17-18].

However, proponents of technical analysis argue that price movements are non-random and adhere to identifiable trends, thereby enabling predictive insights. This analytical approach contributed to the conceptual development of EMH, which formally emerged during the mid-20th century [19-20]. The EMH postulates that historical market behaviour exhibits recurring patterns [21]. However, the theory fails to adequately account for exogenous shocks and structural transformations within the market. Events such as geopolitical disruptions or technological innovations can significantly alter market dynamics. Such market alteration can diminish the reliability of historical patterns in forecasting future price movements. In highly volatile or rapidly evolving market environments, this limitation becomes particularly pronounced as price trajectories may deviate noticeably from historical trends. Although Dow Theory and the discounting principle suggest that market prices reflect all publicly available information, a temporal lag often exists between the release of new information and its full assimilation into asset prices. This latency introduces the potential for short-term inefficiencies and mispricing. Under the strong form of EMH, technical and fundamental analysis are rendered ineffective as price movements are presumed to be entirely unpredictable. While EMH has laid the theoretical foundation for neoclassical finance, its assumption of linear return distributions in financial markets has been increasingly challenged. This has led to the emergence of the Fractal Market Hypothesis (FMH), which states that asset prices follow a fractional Brownian motion and exhibit fractal characteristics [22]. Unlike the linear assumptions of EMH, FMH incorporates long-range dependence and self-similarity attributes that better capture the complex, multi-scale behaviour of financial markets [23-25]. Despite their conceptual appeal, FMH-based models exhibit limited empirical robustness. Moreover, their inherently backward-looking orientation especially within technical analysis frameworks constrains predictive accuracy under dynamic and volatile market conditions.

This study provides the following key technical contributions:

A systematic analysis has been conducted on stock trading history, dating back to the 17th to 19th centuries, highlighting its strengths and weaknesses. This analysis offers a deeper understanding of how early financial markets functioned, including the origins of stock exchanges, trading mechanisms, and market behaviours. This contribution bridges the gap between past and present financial practices by enriching economic history with a systematic approach.

A hybrid novel Hybrid Stock Sequence Learner (HSSL) model is proposed that predicts stock market prices for the next 28 days, with an emphasis on closing prices at the end of each month, using stock and price data from five prominent and reliable stock companies, Alphabet Inc., Apple Inc., Microsoft, Walt Disney Co., and S&P 500.

The model significantly improves the timeliness of stock price predictions by forecasting within a 28-trading-day window. The model shows a substantial reduction compared to the extended prediction horizons of 1,000 days or more found in existing models.

The HSSL-Predict model bridges the gap between long-term predictive models and the immediate decision-making needs of stock price investors.

The subsequent division of this paper is as follows: Section 2 provides an overview of the related research. Section 3 provides the methodology including the data source and preprocessing. Section 4 provides the proposed model. Section 5 presents and discusses the experimental results. Section 6 provides the conclusion and future work.

2. Related works

In this section, we review the history, evolution, theories, and platforms that gave rise to 21st-century trading methods, which are closely related to our proposed approach.

2.1 *Modern stock*

The introduction of electronic trading platforms in the 1980s marked a significant shift in financial markets, enabling faster and more efficient execution of trades [26]. This advancement laid the groundwork for the development of computer-based trading algorithms [27]. By the 1990s, algorithmic trading gained popularity by leveraging mathematical and statistical models to predict stock price movements based on arbitrage opportunities among correlated assets. Subsequent technological innovations led to the emergence of High-Frequency Trading (HFT) in the 2000s [28]. The trading framework leveraged High-Performance Computing (HPC) infrastructures and low-latency network architectures to detect and exploit microsecond-level price differentials across multiple financial markets. Such HFT strategies typically operate under the assumption of a stable market microstructure, where price continuity and liquidity provision remain consistent. Under these normal market conditions, algorithmic execution and arbitrage mechanisms contribute to efficient price discovery and minimal volatility propagation.

However, the dependence on these stability assumptions introduces a significant limitation: HFT algorithms often exhibit reduced adaptability during market turbulence, such as flash crashes, liquidity shocks, or abrupt structural shifts in trading behaviour. During such anomalous events, feedback loops and latency races between competing algorithms can amplify volatility rather than dampen it, potentially leading to transient liquidity vacuums. The profitability and operational resilience of HFT systems are increasingly constrained by evolving regulatory frameworks such as transaction speed restrictions, minimum resting time requirements, and enhanced market surveillance mechanisms designed to mitigate systemic risks. While these interventions aim to promote fairness and transparency, they inadvertently alter latency dynamics and execution efficiency, thereby challenging the sustainability of traditional HFT models and heightening their susceptibility to market instability during periods of stress.

Modern stock prediction is rooted in data science and Natural Language Processing (NLP) methodologies [29].

These methods use extensive datasets from historical prices and a wide range of supplementary financial indicators to detect underlying patterns that inform predictive models. State-of-the-art machine learning and artificial intelligence algorithms use market microstructure features [30] to model both linear and non-linear dependencies in financial time series data. Beyond traditional pattern recognition, these intelligent systems detect anomalies and forecast short- and long-term trends by harnessing the power of big data encompassing diverse sources such as market transactions, macroeconomic variables, satellite imagery, and consumer behaviour inferred from credit card usage. In a study [31], authors proposed a support vector regression-based framework for predicting stock prices across both large and small-cap equities using high-frequency intraday data. However, their model was static and lacked country-specific adaptation. The proposed model did not consider advanced regression techniques and did not demonstrate how well linear regression would perform in predicting stock prices in a dynamic market involving different stock companies.

Ding and Qin [32] have proposed a stock price prediction model using logistic regression. The study reported a prediction accuracy of 95%. Although performance metrics such as Multi-Input Multi-Output (MIMO)-loss and loss variation were employed in their evaluation, the study did not consider MSE. The absence of MSE in the evaluation restricts a comprehensive assessment of the model's robustness, particularly in financial forecasting, where even minor prediction errors can lead to significant monetary consequences. In contrast [33], they conducted a comprehensive survey of deep learning-based and linear correlation-based models for predicting stock prices. They incorporated multiple error metrics, such as RMSE, MAE, and MAPE, to enable a more comprehensive analysis of predictive performance. The findings highlight that deep learning and hybrid architectures generally outperform single-model approaches. While Ding and Qin's model demonstrated high classification accuracy, the lack of MSE limits its evaluative complexity. Compared to Hu et al.'s study, which underscores the importance of multifaceted performance assessment, the omission of such measures in Ding and Qin's work reduces confidence in the model's generalizability and resilience. Agrawal [34] proposed a regression-based approach using data from Tesla and the New York Stock Exchange for stock price prediction.

Although the method demonstrated high predictive accuracy, the dataset was limited to only Google and Apple. This limitation constrains the model's generalizability and robustness across heterogeneous stocks and varying market regimes. The restricted training scope inherently increases the likelihood of overfitting, as the model may capture idiosyncratic patterns unique to the selected companies rather than learning generalizable market representations. Consequently, such models are prone to systematic bias and deteriorated predictive accuracy when deployed on unseen financial instruments or under altered market dynamics. In a related study [35] , authors have proposed a stock price prediction framework utilizing LR and Decision Tree (DT) algorithms on the Amazon stock dataset. While their model achieved a relatively low prediction error, it was trained exclusively on the closing price feature, disregarding other critical and informative attributes such as high, low, and trading volume. This limited feature space not only restricts the model's ability to capture multi-dimensional market dependencies but also undermines its temporal adaptability in highly volatile financial environments. From a cybersecurity and data integrity perspective, such feature sparsity further reduces the model's resilience against adversarial perturbations and data manipulation risks, which are increasingly relevant in algorithmic trading systems.

2.2 *ML*

ML models have significantly transformed various domains, including financial market analysis and trading. ML models improve the accuracy and efficiency of predictive analytics. In [36], Jiang conducted a comprehensive survey of deep learning techniques used in market prediction. The survey detailed a wide array of data sources drawn from different stock exchanges. This diversity highlights the broad applicability and varying nature of financial datasets used in stock price forecasting. The study provides a structured overview of state-of-the-art deep learning architectures. Also, the survey discusses aspects of model reproducibility; however, it falls short of addressing the practical challenges associated with replication. These challenges include dependencies on specific computational resources, software configurations, and restricted access to proprietary and sensitive datasets. In [37], Ji et al. proposed an Long Short-Term Memory (LSTM) deep learning technique to predict the stock market using social media and traditional financial attributes as input data sources.

Using the Doc2Vec architecture, the method built text features vector attributes while balancing the text dimensions with an auto-encoder. Applying autoencoders enhanced the model by reducing feature complexity while retaining the most important information. The combination of deep learning techniques, such as Doc2Vec, autoencoders, and LSTM, requires significant computational power and resources, which may not be feasible for all users or organizations, especially in real-time trading scenarios. One of the observed limitations in this proposed model is that the integration of these advanced deep learning techniques, such as Doc2Vec, autoencoders, and LSTM, demand substantial computational power and resources. This requirement may pose challenges for many organizations during real-time trading, where speed and efficiency are critical. Following a comprehensive review of existing approaches and an evaluation of their respective strengths and limitations, it is necessary to have a hybrid model that can capture the complex, non-linear relationships in stock markets and will provide an edge over single-machine models.

This paper proposes a hybrid HSSL-Predict model designed to enhance predictive accuracy by integrating the interpretability of linear models with the non-linear learning capabilities of machine learning techniques across diverse market conditions. The proposed hybrid HSSL-Predict model is designed to leverage the complementary strengths of each algorithm, namely, the interpretability and simplicity of linear regression, the robustness of SVM in handling classification boundaries, and the superior non-linear function approximation capabilities of SVR. This integration improves generalization performance and predictive accuracy across varying market trends.

3. Methodology

This section presents the development process of the proposed HSSL model. The model integrates SVM, SVR, and LR modules. This section details the dataset sources and the tools used for implementing the proposed model.

3.1 *Tools and techniques*

The SVM-LR-SVR hybrid model was implemented on a 64-bit Windows-based AORUS system equipped with an Intel Core i9 processor running at 6.0 GHz and 32 GB of RAM. Model training, testing, cross-validation, hyperparameter optimization, and selection were conducted using the Scikit-learn library [38] within the Python environment [39]. Data preprocessing and statistical analysis were carried out using the Pandas and NumPy libraries. For data visualization, the Matplotlib library was employed.

3.2 *Dataset and preprocessing*

Five publicly available stock market datasets were collected from Yahoo Finance [40-44]. These datasets encompass historical information on stock prices, dividends, stock splits, and capital gains. The selection of stocks for this study was guided by three principal criteria: company reputation, price stability, and data accessibility. Emphasis was placed on companies with established credibility. This is a recognition of their potential impact on market performance. The selected stocks exhibited relatively stable price behaviour.

Table 1. Overview of selected companies and their corresponding ticker symbols

Company	Ticker code
Alphabet Inc.	GOOG
Apple Inc.	AAPL
Microsoft Corporation	MSFT
Walt Disney Co.	DIS
S&P 500	^GSPC

This criterion was considered for the reliability of predictive modelling. The data sets comprise historical records from Alphabet Inc., Apple Inc., Microsoft Corporation, The Walt Disney Company, and a representative index of S&P 500 companies. Table 1 presents a detailed description of the datasets. For model development, 92.33% of each dataset was allocated for training, corresponding to approximately 337 days out of a total 365-day span. The data collection period ranges from Jan. 31, 2024, to Jan. 31, 2025, to capture consistent market behaviour influenced by prevailing economic and political conditions. The data preprocessing workflow in this study is outlined in Algorithm 1.

The algorithm takes two inputs: Operating System (OS) (the operating system module used to handle file paths and operations) and Comma-Separated Values (CSV) (the module to handle CSV reading and writing), and outputs a preprocessed CSV file. The function Data_Preprocessing (stock_data.csv) accepts a stock data file as a string parameter representing the path to the stock data CSV file. The file is then opened in append mode ('a+') and read mode. Inside the file context, two variables, namely stock_data_csv_reader and stock_data_csv_writer, are instantiated to read from and write to the CSV file. The cleaned values are unpacked into five variables: date, open_price, high_price, low_price, and close_price, representing stock price attributes. The dataset was converted into a CSV format. During this process, a formatting inconsistency was identified: the date entries were structured in the mm/dd/yyyy format, whereas the implemented models required the dd/mm/yyyy format for accurate parsing. This discrepancy led to errors, as dates with day values exceeding 12 were misinterpreted as invalid months. To rectify this issue, the date fields were reformatted according to the CSV structure.

Algorithm 1 Data preprocessing
Input: *OS*, csv
Output: Preprocessing file

```
def Data_Preprocessing (stock_data.csv)
Stock_data_file (str)
If (stock_data_csv_file) exists:
    os.unlink(stock_data_csv_file)
With open(stock_csv_file, 'a+', newline='') as file do
        stock_data_csv_reader = stock_data_csv.reader(file)
        stock_data csv_writer = stock_data csv.writer(file)
For each row in the stock_data csv_reader:
If '@' in row:
row = [entry.replace('@', '') for entry in row]
date, open_price, high_price, low_price, close_price = row
csv_writer.writerow
return (f"Preprocessing complete for file: {csv_file}").
```

3.3 *Framework design*

3.3.1 *SVM model*

The initial phase of the model development process involved integrating essential libraries and tools necessary for implementing the SVM model. The pandas library (imported as pd) is used for efficient management of stock market data. For visualization and performance assessment of the model, matplotlib.pyplot (imported as plt) was employed. The SVR class was imported from the SVM module of scikit-learn to facilitate the deployment of SVM in regression-based tasks. Model performance was evaluated using the MSE metric. The dataset was partitioned into training and testing subsets, denoted as X_train, y_train, X_test, and y_test, respectively. This partitioning was achieved using the train_test_split function provided by scikit-learn.

The SVR model was instantiated using the Radial Basis Function (RBF) kernel [45] mathematically defined in Equations (1)-(2) as follows:

$$\gamma \ll \exp^{-}\left(\frac{\gamma}{\delta}\right)^{2} \tag{1}$$

$$K\left(x,\, x'\right) = \exp\left(-\gamma \,\|\, x - x' \,\|^{2}\right) \tag{2}$$

where x is the input vector in the feature set. $\| x - x' \|^2$ is the squared Euclidean distance between the two vectors. γ is a scaling parameter that determines the radius of influence of the support vectors. The parameter δ^2 represent the variance of the Gaussian distribution that governs the spread and smoothness of the RBF kernel in the proposed model. The RBF kernel then maps the input data into an infinite-dimensional space where a linear separator is constructed to achieve non-linear decision boundaries in the original input vector space. As the distance $\| x - x' \|^2$ increases, the similarity value $K(x, x')$ exponentially decays toward zero.

The RBF kernel is well-suited for capturing complex non-linear patterns, which are prevalent in stock price prediction tasks. This is essential in stock price forecasting, where data points may exhibit overlapping or non-linear patterns due to complex market dynamics, as discussed in sections 1 and 2. This indicates that the model distinguishes between upward and downward price movements with high generalization capability. In evaluating the SVM model post-training, the "predict()" function was utilized to generate predicted values (y_pred) from (X_test data categories. The predicted values were compared with the actual target values (y_test) using MSE. A lower MSE reflects improved model accuracy, whereas a higher MSE signals greater discrepancies between predictions and actual outcomes. Reporting the MSE is essential for assessing the SVM model's predictive accuracy, serving as a critical step in evaluating and optimizing the proposed hybrid model.

3.3.2 *SVR model*

SVR is a machine learning algorithm widely used for addressing regression problems [46] SVR is extensively applied in financial time series forecasting for stock price prediction due to its capacity to handle non-linear and high-dimensional data. By formulating the prediction task as a convex optimization problem, SVR identifies a function that accurately maps historical market indicators to future stock prices. Its ability to control model complexity while minimizing generalization errors makes it a powerful tool for capturing the subtle patterns and fluctuations inherent in stock market data. In our approach, SVR formulates the learning task as a convex optimization problem, where the objective is to determine a regression function that accurately maps financial input features to corresponding target values. One of the major strengths of SVR is its ability to maintain model generalizability and effectively balance the trade-off between model complexity and prediction accuracy. This attribute makes SVR effective in high-dimensional spaces.

Let $f(x)$ represents a function with ϵ deviation. Let y_i represent actual target values for training stock data. Given a set of training data $\{(x_i, y_i)\}_{i=1}^{n}$, SVR is defined in Equation (3) as:

$$\text{SVR} = \min_{\omega,\, b,\, \xi,\, \xi^*} \frac{1}{2} \| \omega \|^2 + K \sum_{i=1}^{n} \left(\xi_i + \xi_i^* \right)$$

$$\text{Subject to: } y_i - \left(\omega - \phi(x_i) + b \right) \le \epsilon + \xi_i$$

$$\left(\omega \cdot \phi(x_i) + b \right) - y_i \le \epsilon + \xi_i$$

$$\xi_i \xi_i^* \ge 0, \text{ for all } i \tag{3}$$

where ω is the model parameter for the weight vector that determines the orientation of the regression hyperplane. The bias term b adjusts the position of the hyperplane. To accommodate deviations from the exact predictions, slack variables ξ_i and ξ_i^* that allows the model to tolerate a predefined margin of error, thereby enhancing its ability to generalize in the presence of market noise and volatility. Let α_i, α_i^* denote the Lagrange multipliers associated with the optimization constraints [47] Equation (3) is transformed to incorporate these variables. The prediction function of SVR is mathematically expressed in Equation (4) as:

$$f(x)=\sum_{i=1}^{n}\left(\alpha_i-\alpha_i^*\right)+\frac{1}{2}\left(-\gamma\,\|\,x_i-x\,\|^2\right)+b \tag{4}$$

where $f(x)$ is the predicted stock price corresponding to the input feature vector x, which includes historical prices, market indicators, and other market-related features. x_i represent the ith training data point, where each vector contains feature values associated with known stock prices. Equation (4) captures the non-linear mapping from historical market indicators to future stock prices.

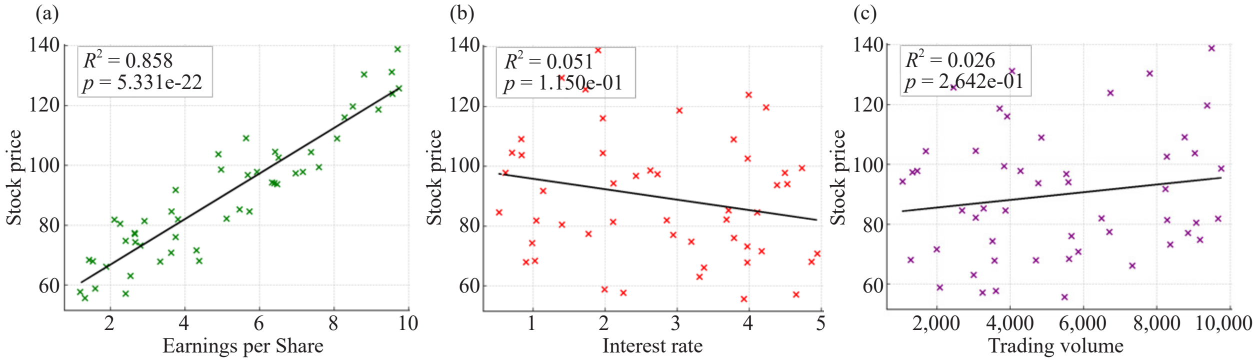

Figure 1. Visual comparison of the relationship between stock prices and key independent variables

Figure 1 illustrates the output behaviour of Equation (4) within the input domain $x \in [0.4]$. This demonstrates how the proposed SVR model approximates target stock prices over the specified interval. The resulting curve reflects the learned regression function whose shape is significantly influenced by the selected support vectors, the corresponding Lagrange multipliers, and the parameter γ, which controls the width of the kernel in the financial feature space. These components collectively determine our model's ability to capture the underlying patterns in stock price data and affect the smoothness of the predictive function. A systematic comparison between the proposed SVM and SVR was conducted using a structured procedure, as detailed in Algorithm 2.

3.3.3 *LR model*

In [48], the authors discussed LR as a technique applied in two contexts. Firstly, regression analysis is typically used for making forecasts an by *d* predictions. Secondly, regression analysis can also serve as a tool for identifying potential causal relationships between two sets of variables. By modelling the relationship between these variables, regression analysis estimates the value of the dependent variable '*y*' based on a given range of values for the independent variable '*x*'. In the context of stock price prediction, LR is used not only for forecasting future prices but also for examining potential associations between market variables [49]. In our approach, the LR model estimates future stock prices by analyzing how changes in explanatory variables, such as trading volume, historical price movements, moving averages, the relative strength index, and macroeconomic indicators, impact the dependent variable. For our LR model to produce reliable and unbiased estimates, the following five key assumptions must hold:

1. Linearity: The relationship between the dependent variable and each independent variable is assumed to be linear.
2. Independence of errors: Residuals are assumed to be uncorrelated with one another.
3. Homoscedasticity: The variance of residuals remains constant across all levels of the independent variables in this model.
4. Normality of errors: The residuals are normally distributed.
5. No multicollinearity: Predictor variables are not highly correlated with each other, ensuring the stability of our model.

Algorithm 2 SVM and SVR selection approach
Input: Historical data and historical prices with other parameters.
Output: Predicted stock prices $\hat{y}$
1: Input X
2: Compute γ
3: **for each** pair (x_i, x_j) in X:
4: Compute $K(x_i, x_j)$
5: Choose model type:
6: **If** classification (SVM):
7: optimise SVM (w, b, and α)
8: Else if regression (SVR):
9: optimise SVR (w, b, and α)
10: Train the model using X
11: **for** new x', predicting stock prices: $\hat{y}(x')$
12: Output predicted stock prices $\hat{y}$.

Consider a set of independent variables denoted as $X = \{X_1, X_2, ..., X_n\}$ where each X_i corresponds to the ith predictor variable. In the context of Multiple Linear Regression (MLR), our LR model includes an intercept term β_0 which represents the expected stock price when all predictors are zero. The model also includes a set of coefficients $\beta = \{\beta_1, \beta_2, ..., \beta_n\}$ where each β_i computes the corresponding predictor X_i to the predicted stock price with earnings per share, interest rate, and other variables. Equation (5) defines the LR model as follows:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ..., + \beta_n X_n. \quad (5)$$

The individual effects of key predictors on the estimated stock prices are illustrated in Figure 2, where EPS demonstrates a strong positive influence, interest rate exhibits a negative relationship, and trading volume shows a relatively mild positive association.

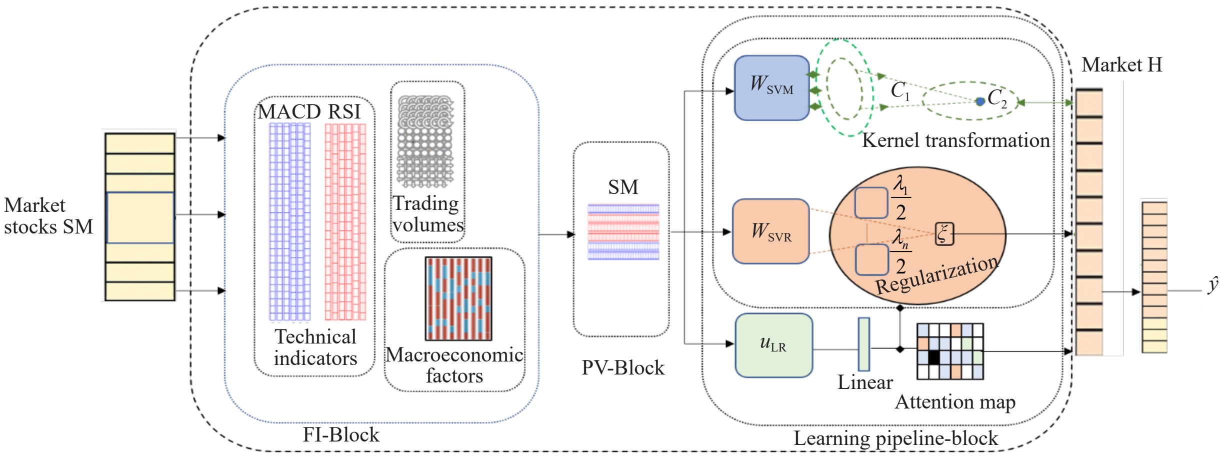

Figure 2. Architecture of the proposed hybrid stock prediction framework

In Figure 1a, the scatter plot illustrates a strong positive linear correlation between EPS and the predicted stock prices. This relationship is consistent with financial theory, which speculates that higher earnings often reflect greater profitability and financial health. This thereby attracts investor interest and leads to upward pressure on stock valuations. In Figure 1b, a negative linear association was observed between interest rates and stock prices. This implies that as interest rates increase, predicted stock prices tend to decline. This inverse relationship aligns with macroeconomic principles. This result supports the inclusion of interest rates as a meaningful macroeconomic predictor in stock valuation as proposed in this model. In Figure 1c, the relationship between trading volume and stock prices reveals a weak positive trend. This suggests that increased market activity may have a marginal influence on stock price appreciation.

This relationship may reflect heightened investor interest or market sentiment, though its explanatory power appears limited relative to EPS and interest rates. The flow diagram of the LR model is illustrated in Figure 3 where financial indicators serve as independent variables mapped to a target price through linear transformations and weighted contributions. Each cell in the matrix reflects the magnitude and direction of contribution derived from the product of the regression coefficient β_i and the corresponding predictor value X_i. The intercept term β_0 is held constant across all instances to contribute to the predicted output.

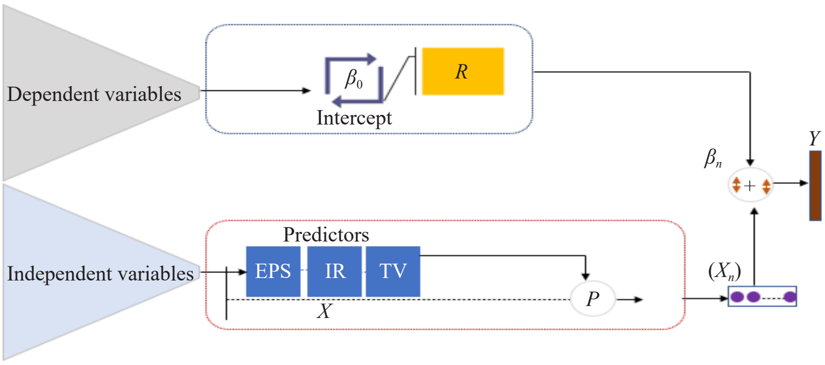

Figure 3. LR model showing the correlation between independent variables (EPS, interest rate, and trading volume) and the dependent variable (predicted stock price) through regression coefficients and intercepts

4. Proposed hybrid model

In this section, we present a proposed HSSL model. The hybrid model, which integrates SVM, SVR, and LR to predict stock prices over 28 days, is illustrated in Figure 2. The model is designed according to a structured procedure, as detailed in Algorithm 3. This model primarily consists of a Financial Indicator Block (FI-Block), a Predictor Variables Block (PV-Block), and a Learning Pipeline Block (LP-Block). The FI-Block extracts relevant stock market features derived from a range of financial indicators. These include technical indicators such as the Moving Average Convergence Divergence (MACD) and the Relative Strength Index (RSI). These indicators are used to capture price momentum and market trends. The FI-Block incorporates additional financial metrics, such as trading volumes and macroeconomic variables, to provide a comprehensive representation of market dynamics. The PV-Block encapsulates the derived predictor variables used for model training and inference. By capturing temporal and structural patterns within the data, the PV-Block enables the model for effective training. This supports robust inference for accurate stock price predictions. This block accentuates the temporal dependencies and latent structural patterns present within the time series data by encoding historical context and lagged values. Engineered features derived from financial technical indicators, such as MACD and RSI, represent momentum and mean-reversion behaviours in financial time series. The interaction between the FI-Block and the PV-Block provides a rich and information-dense input space for model training.

LP-Block extracts and integrates predictive signals from the processed feature space consisting of three key components: kernel transformation, regularised learning module, and linear attention mechanism, as illustrated in Figure 4. The feature set derived from technical indicators undergoes kernel transformation through SVM to project the input into a high-dimensional feature space $C_1 \rightarrow C_2$. This transformation enhances the non-linear separability of financial patterns. To control overfitting and improve the generalisation of the proposed model on unseen financial data, regularisation was optimised using λ. Working in parallel with the linear regression unit, contributing features modulated by an attention map were extracted from market data. To overcome the issue associated with stock price fluctuations, the attention output, which mapped features in the attention map, was then combined with a linear layer, which was subsequently integrated into a linear transformation layer for final stock price prediction.

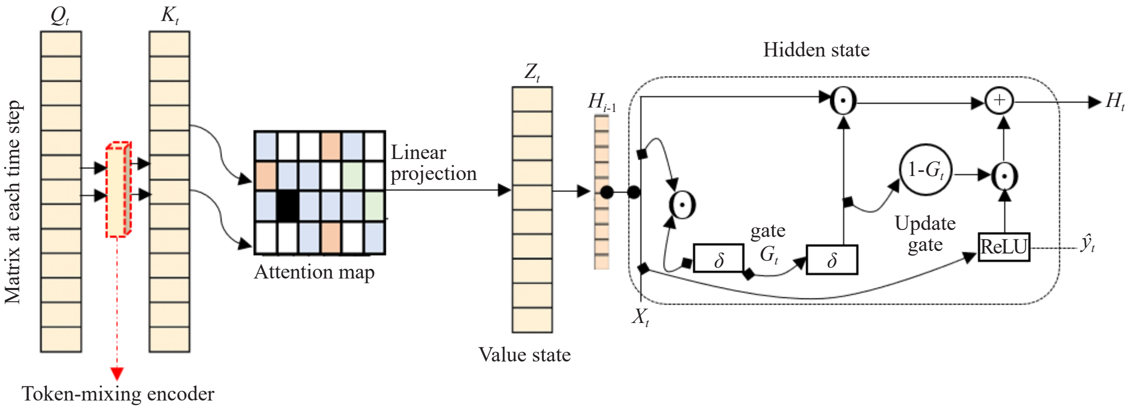

Figure 4. Self-attention map

4.1 *Problem expression*

For the convenience of expression, we define the historical time window and the current time as T and t. In this paper, $T = 28$. We present the stock historical data at time t instance as $X_t \in \mathbb{R}^{mxT} = [t - T + 1,\ t]$, where each $X^{t\text{-}T+1:t}$ is a sequence of feature values over time T steps. The complete feature set X_t which forms the input aggregation in FI-Block at time t to yield the mapping transposition $X_{K,\,t}$ in LP-Block for model training is expressed as:

$$X_t = \left[x_{\text{MACD}}^{t-T+1:t},\ x_{\text{RSI}}^{t-T+1:t},\ x_{\text{VOL}}^{t-T+1:t},\ x_{\text{MACRO}}^{t-T+1:t} \right]$$

$$\text{Subject to: } X_{K,\,t} = \phi\left(X_t\right),\ \phi : \mathbb{R}^{mxT} \to \mathbb{R}^{d'} \tag{6}$$

where d' is the dimensionality correlation of the kernel space for the trading day. In this paper, the transformation is designed to capture complex and non-linear financial patterns that cannot be effectively modeled in the original space. To selectively focus on the 28-day history to enable the model to assign dynamic importance to time, the linear feature matrix in LP-Block is mapped using an attention mechanism, expressed as:

$$A_t = \text{softmax}\left(\frac{Q_t K_t^T}{\sqrt{d_k}} \right),\ Z_t = A_t V_t$$

$$V_t \in \mathbb{R}^{28 \times d_k} \tag{7}$$

where A_t is the temporal attention weight matrix at every time step. Z_t is the output of the attention module, which encodes the refined feature correlations. d_k is the value vector in the attention mechanism, which aggregates the attention scores during mapping of the weighted values V_t of the feature matrix Q_t at time step t. In some market behaviors where macroeconomic factors play essential roles in determining the quality of stock prices, a correlation-based procedure is required to identify related stocks and their prices, adding complexity for investors, as it demands not only technical expertise in algorithmic methods but also a strong foundational understanding of market behavior. To solve this challenge, it is essential to reduce feature complexity through regularization. This optimization technique minimizes the structural risk in the stock matrix feature, helping to maintain the model's prediction flexibility. Let y_i denote the actual observed stock price at time step i. The regularized objective Support Vector Regression Objective (SVRO) over a stock multiview data is expressed in Equation (8) as:

$$\text{SVRO} = \min_{\omega\text{SVR}\xi\xi_i^*} \frac{1}{2} \| \omega\text{SVR} \|^2 + \lambda \sum_{i=1}^{n} \left(\xi_i + \xi_i^* \right)$$

$$\text{SVRO} = \min_{\omega\text{SVR}\xi\xi_i^*} \frac{1}{2} \| \omega\text{SVR} \|^2 + \lambda \sum_{i=t-27}^{n} \left(\xi_i + \xi_i^* \right)$$

$$\text{Subject to:} \begin{cases} y_i - \left\langle \omega^T,\ \phi\left(X_i\right) \right\rangle + b \le \epsilon + \xi_i \\ \left\langle \omega^T,\ \phi\left(X_i\right) \right\rangle + b - y_i \le \epsilon + \xi_i^* \\ \xi_i \xi_i^* \ge 0,\ \text{for all } i \end{cases} \tag{8}$$

where ωSVR is the support vector regression weight vector. ϵ, ξ_i and ξ_i^* are slack variables that allow the model to construct a regularized and noise-tolerant prediction surface, and also to allow for controlled deviations between the predicted stock price $\hat{y}_i$ and the actual price y_i. These slack variables capture the degree to which predictions fall outside the designated tolerance zone defined by the ϵ insensitive loss function. Instead of attempting to perfectly match every observed stock price, which can lead to overfitting in financial data, the model allows for a controlled error buffer. These variables together help the model capture dominant patterns in stock movements while ignoring non-informative variations. This design enhances both forecasting accuracy and generalizability to unseen financial data. λ is the regularization parameter controlling the propose model complexity for every input x_i at time x_t.

To improve performance, the model employs feature fusion by concatenating the attention output Z_t and the raw input features X_t, to capture both learned global dependencies and fine-grained historical information. This fusion is flattened and linearly transformed through a regression weight vector ω_{LR}^T to yield the prediction as expressed in Equation (9). To dynamically balance the influence of global and local features, we introduced a gating mechanism. The gate vector $G_t \in [0,\ 1]^d$ controls both Z_t and X_t during linear projection, expressed in Equation (10). To capture complex correlations between global and local features in volatile markets, we applied a non-linear transformation to the concatenated value representation before the final prediction as expressed in Equation (11).

$$\hat{y}_t = \omega_{LR}^T \left(vec\left(Z_t\right) + \lambda \cdot vec\left(X_t\right) \right) + b \tag{9}$$

$$\hat{y}_t = \omega_{LR}^T \left(G_t \odot vec\left(Z_t\right) + \left(1 - G_t\right) \odot vec\left(X_t\right) \right) + b \tag{10}$$

$$\hat{y}_t = \delta\left(\omega_f^T \left[vec\left(Z_t\right) \| vec\left(X_t\right) \right] + b_f \right) + b \tag{11}$$

where ω_f and b_f are intermediate fusion weights and bias, δ is the activation function to learn higher-order interactions between predictor variables during model prediction especially in volatile and dynamic market.

Algorithm 3 Proposed HSSL model for stock price prediction
Input: Multivariate stock features over a historical window $T = 28$ at current time t
Output: Sequence of predicted stock prices $\hat{y}_t$
1: **for** $t = 1$ to $X_{t\text{-}1}$ **do**
2: Combine multivariate stock data to be predicted
3: Transform the raw multiview stock data into high-dimensional kernel space $\phi : \mathbb{R}^{mxT} \rightarrow \mathbb{R}^{d'}$
4: **end for**
4: Initialize the attention encoding of the model
5: **for** $d = t\text{-}T$ to
6: Transform $X_{K,\,t} = \phi\left(X_t\right)$
7: Compute attention scores $Z_t = A_t V_t$
8: Define optimisation SVRO

9: $\langle y_i - \langle w^T, \ \phi(x_i) \rangle + b \rangle \leq \varepsilon + \xi_i, \ w^T, \ \phi(x_i) + b) - y_i \leq \varepsilon + \xi_i^*$

10: Compute the intermediate prediction from the kernel-projected support vectors

11: $\alpha_i \phi(Z_{t,\,i}) + b$

12: Concatenate raw input and attention-encoded features

13: **end for**

14: Update model through 1 – *Gt*

15: Return predicted price $\hat{y}_t$.

4.2 *Model evaluation*

The performance of the proposed model was comprehensively evaluated using the MSE, RMSE, and coefficient of determination (R^2) metrics. The MSE quantifies the average squared difference between the observed values y_i and the predicted values $\hat{y}_i$, with larger values indicating greater prediction error and potential areas for model refinement. The RMSE, derived as the square root of the MSE, provides an interpretable measure of error in the same units as the response variable, thereby facilitating a practical understanding of model accuracy. The MSE, RMSE, and R^2 are computed in Equations (12)-(14) as follows:

$$\text{RMSE} = \sqrt{\sum_{i=1}^{n} \frac{(\hat{y}_i - y_i)^2}{n}} \tag{12}$$

$$\text{MSE} = \frac{1}{n} \Sigma (y_i - \hat{y}_i)^2 \tag{13}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (\hat{y}_i - y_i)^2}{\sum_{i=1}^{n} (y_i - \overline{y_i})^2}. \tag{14}$$

5. Results and discussion

This section presents the performance evaluation of the proposed model across different price segments, as outlined in Table 2. Table 3 reports the MSE, RMSE, and R^2 values obtained from each technique. Among the standalone models, LR yielded the lowest MSE of 22.034 and RMSE of 30.372, along with a high coefficient of determination R^2 = 0.9748, indicating strong predictive capability. SVR also demonstrated competitive performance, with a marginally higher MSE of 23.049 and RMSE of 31.004, although it exhibited a slightly reduced R^2 value of 0.9604 compared to LR. The SVM model yielded an MSE of 24.193, an RMSE of 32.393, and the highest R^2 among the individual models (R^2 = 0.9702). In contrast, the hybrid HSSL model achieved superior overall performance, attaining the lowest MSE (19.132) and RMSE (29.512). Along with the highest R^2 value (0.9802), thereby confirming its enhanced predictive accuracy and generalisation capability. These results suggest that the hybrid approach effectively leverages the complementary strengths of the individual models, resulting in improved predictive accuracy and reduced error metrics compared to the standalone methods.

Table 2. Classification of stock prices by range

Stock ticker	Stock price range
MSFT	260-400
DIS	80-110
AAPL	150-190
GOOG	90-150
S&P 500	4,000-4,800

Table 3. Comparative performance of the individual and the proposed hybrid model

Method	MSE	RMSE	R^2
SVM	24.193	32.393	0.9702
SVR	23.049	31.004	0.9604
LR	22.034	30.372	0.9748
HSSL	19.132	29.512	0.9802

5.1 *MSFT results*

As illustrated in Figure 5a, the SVM model demonstrated limited effectiveness in predicting this dataset, producing overly uniform outputs that failed to reflect the observed price fluctuations accurately. While the model succeeded in capturing the general directional trend, it lacked sensitivity to short-term variations. The linear regression model, in contrast, demonstrated substantially higher predictive accuracy, as shown in Figure 5b.

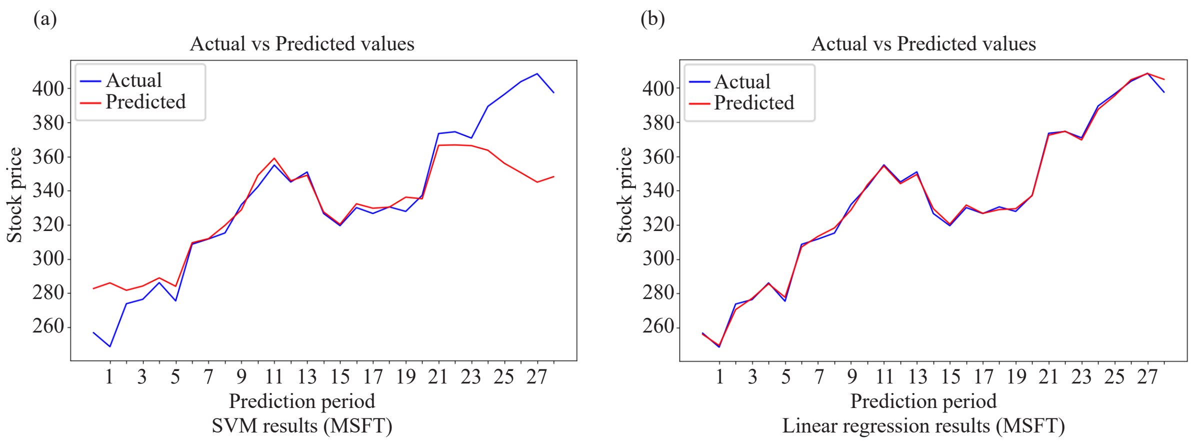

Figure 5. Comparative visualisation of SVM and LR model predictions on the MSFT stock price dataset

However, it faced challenges in capturing sharp or abrupt price movements. In general, the results demonstrate that the SVM model exhibits a high level of alignment with the ground truth data in the early and middle segments of the prediction period (days 1-20), effectively capturing both the upward trend and local fluctuations. This indicates that the SVM model is capable of learning time-based dependencies and local non-linear patterns inherent in financial time series data. However, minor deviations are observed in the latter part of the prediction window (days 22-28), where the predicted values tend to underestimate the actual price trajectory, particularly during steep upward movements. This discrepancy may be attributed to the lag in response of SVM to sudden market volatility, as kernel-based learning methods often assume local stationarity, which is violated in high-frequency financial environments. From a time series analysis perspective, the MSFT dataset likely exhibits a combination of trend and seasonal components, coupled with short-term market noise.

5.2 *DIS results*

Figure 6a illustrates that the SVM model demonstrated improved predictive performance on the DIS dataset relative to previous datasets. However, the model's predictions remained overly smoothed, resulting in limited responsiveness to abrupt price fluctuations, despite the SVM model accurately capturing the overall price trend. In contrast, Figure 6b shows that the LR model exhibited strong predictive accuracy on this dataset. The LR model effectively captured several pronounced price changes within the testing data, although it encountered challenges in precisely forecasting smaller and incremental price variations. These findings highlight that while the SVM model can successfully identify general price trends in this dataset, its smoothing tendency limits its effectiveness in capturing short-term volatility and financial indicator changes, which are crucial for high-frequency trading and risk management. The LR model's strength lies in detecting sharp price movements in this dataset, making it suitable for scenarios where significant trend shifts dominate. But its relative difficulty in modelling minor price increments suggests a limitation in addressing subtle market dynamics.

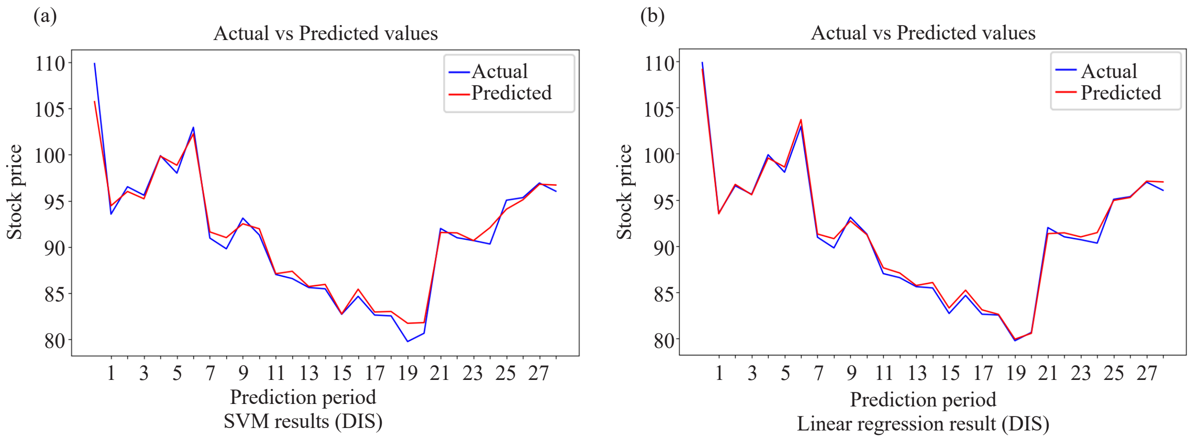

Figure 6. Comparative visualization of SVM and LR model predictions on the DIS stock price dataset

5.3 *AAPL results*

Figure 7a illustrates the performance of the SVM and LR models on the AAPL dataset. Compared to its performance on the MSFT dataset, the SVM model demonstrated moderately improved predictive capability. While it effectively captured the overall trend and approximated the general shape of the price trajectory, its predictions remained overly smoothed, failing to accurately reflect the finer-grained price fluctuations.

In Figure 7b, the linear regression model exhibited strong predictive performance on this dataset. It successfully captured the underlying trend and produced several accurate predictions on the test data. However, similar to SVM, the LR model tended to overly flat its forecasts, resulting in diminished sensitivity to abrupt changes between consecutive data points.

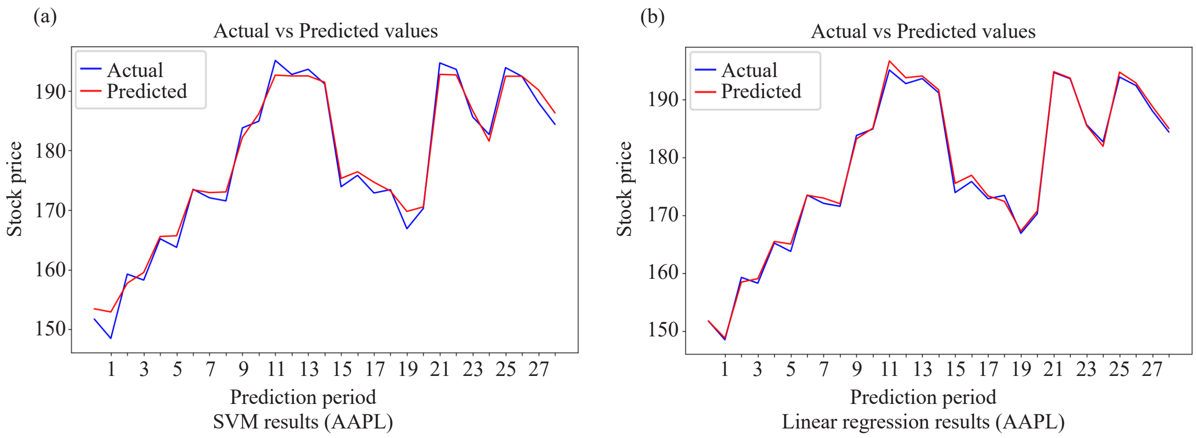

Figure 7. Comparative visualisation of SVM and LR model predictions on the AAPL stock price dataset

5.4 *GOOG results*

The graph in Figure 8a shows that the SVM model struggled to make predictions on this dataset compared to the previous datasets. However, it again demonstrates that the model made predictions that were too even and did not anticipate price changes well, although it was correct in predicting the general price trend. Figure 8b shows that the LR model was very successful in making predictions for this dataset. The graph shows that it made several accurate predictions on the testing data, especially with more sharp changes in prices; however, it struggled slightly.

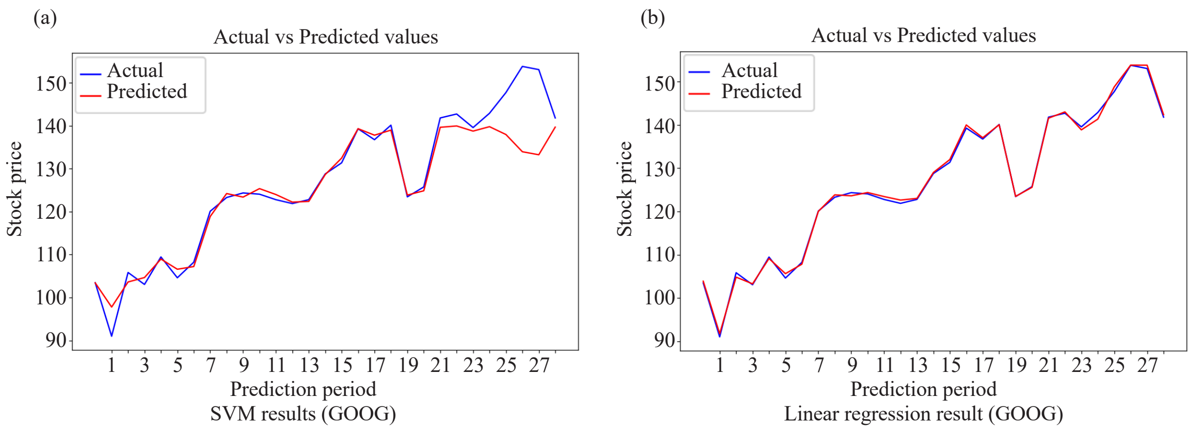

Figure 8. Comparative visualization of SVM and LR model predictions on the GOOG stock price dataset

The predicted values (red) closely follow the overall upward trend of the actual stock prices (blue), indicating that the model is capable of learning and generalizing the long-term price movement. While the model tracks the general price direction well, discrepancies between the actual and predicted values are visible, particularly around the following points: Periods 1-2, where a short-term dip in the actual data is not accurately reflected by the predicted values, which show a smoother response. Periods 17-18 and 24-28: The model underestimates sharp price increases or decreases, suggesting it may struggle with volatility or rapid market changes. This result shows that SVM is better at capturing non-linearities than LR, but is limited if the kernel choice is suboptimal. This result is suitable for complex and volatile financial data where capturing temporal dependencies is essential.

5.5 *S&P 500 results*

The predictive performance of the SVM model on the S&P 500 dataset is illustrated in Figure 9a. The results indicate that the model failed to capture the underlying dynamics of the actual stock price movements. Notably, the predicted values exhibit a pronounced smoothing effect, with the outputs remaining relatively flat across the prediction horizon. This behavior suggests that the SVM model was unable to adapt to the volatility and directional shifts present in the actual market data. During periods of sharp declines, such as between prediction periods 14 and 22, the model's predictions remain nearly constant. This significantly underrepresents the observed price drop; similarly, in the earlier periods between day 1 and 7, the actual prices show a steep downward trend. The model continues to produce values clustered around a narrow range during this period. This insensitivity to both short-term fluctuations and longer-term trends suggests that the model exhibits a high bias in this dataset. The SVM's inability to generalize on this dataset may be attributed to the complex, non-linear nature of financial time series data, such as the S&P 500. Figure 9b illustrates the performance of the LR model in predicting S&P 500 stock prices over a 28-period forecast horizon.

The blue line represents the actual stock prices, while the predicted values are shown in red. The proximity between the two curves suggests a relatively strong correlation between predicted and observed values across the majority of the time horizon. The model demonstrates good performance during periods of smooth trend transitions from periods 1-5 days and 13-15 days. The model accurately captured the general downward trajectory of the market. This indicates the model's effectiveness in modeling linear trends and its responsiveness to consistent price movements. However, during more volatile regions around Periods 7-11 and 20-24, the model experienced minor deviations between the predicted and actual values.These discrepancies are likely from the linear model's inability to capture non-linear dynamics and sudden market shocks, which are characteristic of real-world financial time series. Quantitative evaluation indicates that the proposed model demonstrates low bias and variance across the defined evaluation window, suggesting stable generalization performance. Within this context, the LR model applied to the dataset provides a robust baseline for short-term forecasting tasks. Its predictive reliability is particularly pronounced when combined with rigorous feature engineering techniques, which enhance the model's capacity to capture relevant patterns in the data. Furthermore, the LR model can serve as a critical component within hybrid modeling frameworks, where it complements more complex nonlinear models by providing interpretable and computationally efficient estimations, thereby improving overall forecasting accuracy and robustness under varying market conditions.

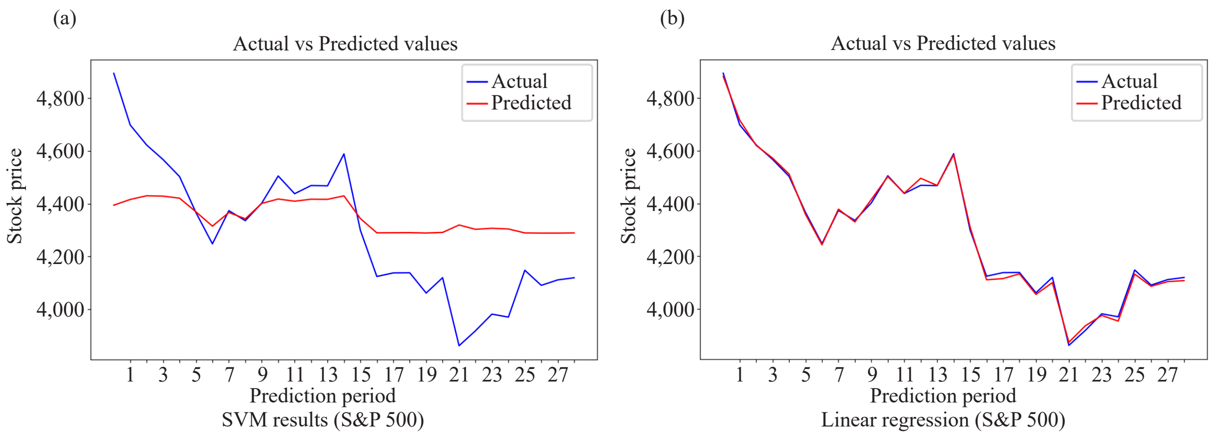

Figure 9. Comparative visualization of SVM and LR model predictions on the S&P 500 stock price dataset

5.6 *HSSL model results*

The HSSL model combines the strengths of SVM, SVR, and LR models to achieve higher prediction accuracy. Figure 10a illustrates the comparative performance between the actual and predicted stock prices generated using the proposed HSSL model across a prediction horizon of 28 periods. The graphical output demonstrates a strong alignment between the actual (blue curve) and predicted (red curve) stock price trajectories. This suggests a high-fidelity predictive capability of the HSSL model. The predicted curve closely tracks the actual price movements with minimal lag, particularly during periods of volatility with sharp declines around day 21 and recovery around period 22. This suggests that the model is robust at learning temporal dependencies and short-term patterns intrinsic to financial time series. Figure 10b presents a comparative visualization of the actual versus predicted stock prices generated by the proposed model with the stock price range bounded within 4,000 to 4,800 units to capture short-to mid-term market dynamics.

Despite the presence of micro-fluctuations in the stock values, the predicted values by the hybrid model exhibit smoothness while retaining critical inflection points. This reflects its capability to suppress high-frequency noise without compromising the accuracy. The near-overlap between actual and predicted sequences throughout the prediction period demonstrates excellent generalization to unseen data. The HSSL model maintains responsiveness to abrupt regime shifts, such as the sudden price drop and recovery between days 20-22. This reinforces the model's capacity to adapt to the structural discontinuities often present in stock market behavior, which is a crucial trait for short-term prediction systems.

The close alignment of both series throughout the prediction horizon strongly indicates that the HSSL model is capable of tracking non-linear stock price fluctuations with high stability. This observation is evident in days of high volatility, such as days 15-18 and days 21-23, where the model shows sharp upward and downward trends. The minimal phase lag and narrow deviation band between the actual and predicted curves confirm the robustness of the model's ability to handle complex market patterns. The negligible visual error margin between predicted and actual curves suggests low MAE and RMSE. This indicates the predictive precision of the proposed model, which is attributed to the complementary strengths of SVM, SVR, and LR. While SVM and SVR capture complex, non-linear structures and generalization boundaries, linear regression stabilizes the model with its interpretability and low variance. Additionally, the proposed model responds dynamically to sudden market shocks, such as price drops around day 15, and rebounds around day 17. This is crucial for investors to make real-time trading decisions and for a risk-aware portfolio management approach.

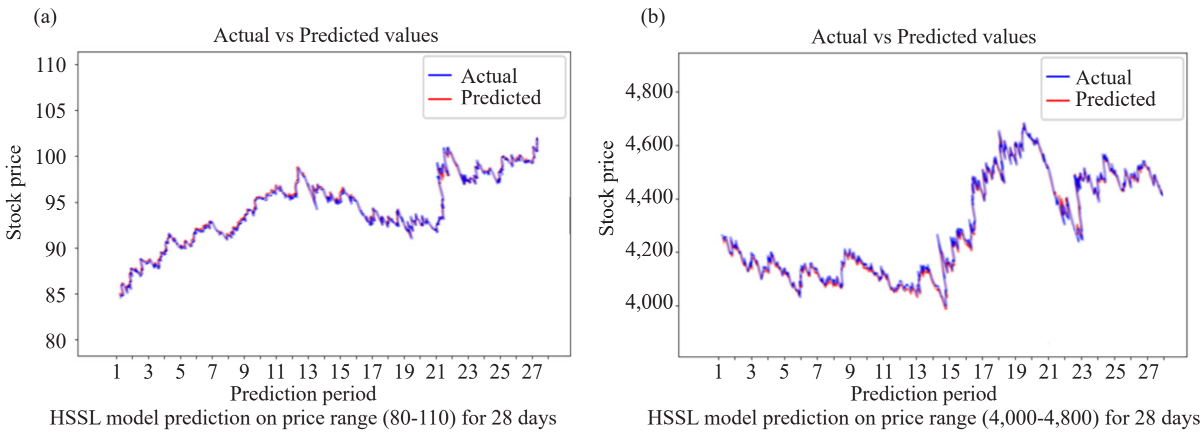

Figure 10. Comparison of actual and predicted stock prices using the proposed HSSL model over a 28-day prediction period within the price range of 80-110 and 4,000-4,800 units

5.7 *Results discussion*

The experimental evaluation demonstrates clear performance differentials across the standalone SVM, SVR, and LR models, as well as the proposed hybrid HSSL framework. These results offer important insights into the predictive behaviour of linear and non-linear learners when applied to heterogeneous financial time series characterised by regime shifts, short-term volatility, and company-specific market signals. Across all datasets, the hybrid HSSL model consistently achieved the lowest error metrics (MSE = 19.132; RMSE = 29.512) and the highest coefficient of determination (R^2 = 0.9802), outperforming each individual model. This improvement is attributed to the complementary strengths of the constituent modules: SVM and SVR contribute non-linear feature transformations capable of modelling localised fluctuations, while LR provides a stabilising linear component that mitigates overfitting. The combination of attention-gated feature modulation and regularised kernel projections further enhances stability when predicting under dynamic market conditions.

A more granular analysis across stocks reveals important behavioural patterns. For MSFT, both the SVM and LR models captured the global price direction, yet SVM exhibited a notable lag in periods of rapid upward movement (days 22-28). This behaviour aligns with the known sensitivity of RBF-kernel SVMs to local stationarity assumptions; abrupt deviations violate this assumption and reduce responsiveness. LR showed superior short-term alignment but struggled with sharp, transient fluctuations, consistent with its linear functional form. By contrast, the HSSL model significantly reduced phase lag and preserved local structure, suggesting that feature fusion between kernel-projected and raw linear components successfully captures both smooth and non-smooth transitions. DIS and AAPL exhibited similar patterns. SVM predictions were systematically smoothed and failed to reflect high-frequency variations, while LR produced greater fidelity during pronounced price swings but again underperformed on minor incremental changes. These characteristics reflect the distinct feature distributions of the DIS and AAPL datasets, where high-impact macroeconomic announcements cause abrupt movement superimposed on otherwise mild intra-day variability. The hybrid model effectively reconciled these behaviours, preserving trend structures while capturing a broader spectrum of volatility signatures.

GOOG and the S&P 500 presented more challenging cases due to their higher volatility and more complex temporal dependencies. GOOG demonstrated the greatest divergence between SVM predictions and ground truth, with consistent underestimation of rapid upward movements and over-smoothing across the sequence. LR, while more accurate, still experienced difficulty modelling regions exhibiting non-linear acceleration. The S&P 500 results further exposed the limitations of the standalone models: SVM predictions collapsed toward a flat, high-bias solution, indicating difficulties in handling the index's multiscale volatility and long-memory effects. LR showed better alignment but failed to adapt to short-term shocks, reflecting the inherent linearity of the mapping.

The HSSL model, however, maintained strong predictive fidelity across both datasets, particularly in regions containing sharp directional changes (days 20-22 for S&P 500). The performance gain is explained by the HSSL's ability to integrate global temporal structure (captured through the LR component), high-dimensional non-linear relations (captured by SVM/SVR kernel transformations), and selective attention weighting over the 28-day historical window. This fusion mitigates the weaknesses of each individual model and allows the hybrid framework to preserve inflection points while filtering noise, which is an essential requirement for short-term forecasting applications in financial markets. Taken together, the empirical findings indicate that linear models alone are insufficient in the presence of non-stationary temporal dynamics, while non-linear kernel learners alone are prone to smoothing or performance degradation when volatility accelerates. The hybridisation strategy used in the HSSL model demonstrably enhances robustness, reduces prediction error, and improves generalisation across stocks with differing volatility regimes, making it a more suitable architecture for short-horizon forecasting in practical investment settings.

6. Conclusion

This study proposes a HSSL model spanning a 28-period prediction horizon for high-frequency stock price prediction, supporting short-term investment strategies. The experimental evaluation was conducted on five major financial companies, namely MSFT, DIS, AAPL, GOOG, and the S&P 500. A comparative performance analysis was conducted involving standalone SVM, SVR, and LR models, with the hybrid HSSL model demonstrating superior predictive capabilities. Quantitative evaluation based on standard error metrics, including MSE, RSME, and R^2, revealed that combining multiple machine learning algorithms enhances the accuracy of the stock price model. These findings suggest that the hybridisation of non-linear and linear learners enhances the model's capacity to capture both abrupt and smooth variations in stock price behaviour. This makes the proposed model highly effective under dynamic market conditions.

Although the proposed model demonstrates strong predictive performance, it has certain limitations that require further investigation. Firstly, the evaluation was conducted solely on the selected dataset, which may limit the model's generalisability and robustness when applied to different financial contexts. Secondly, the model has not yet been tested on alternative asset classes, such as cryptocurrencies, which exhibit distinct volatility patterns and market behaviours compared to traditional financial instruments. Future work will aim to extend the model's evaluation to a wider range of financial assets, including highly volatile instruments like Bitcoin and other cryptocurrencies, to assess its adaptability and effectiveness across diverse market conditions.

Data availability statement

The dataset used for this research is publicly available at: https://github.com/kabokorikoma/Stock-market-dataset.

Conflict of interest

The authors declare no competing financial interest.

References

[1] Macaulay CR. Stock market development: Building a risk society since 1688. *International Review of Applied Economics*. 2024; 1-24.

[2] Zaitun S, Munene A, Wanjau JN. Moderating role of savings on the relationship between interest rates and the stock market performance. A case of Nairobi Securities Exchange. *African Journal of Education, Science and Technology*. 2023; 7(3): 654-661.

[3] Da Cunha CR. *Machine Learning for the Physical Sciences: Fundamentals and Prototyping with Julia*. London: CRC Press; 2023.

[4] Mathanprasad L, Gunasekaran M. Analysing the trend of stock marketand evaluate the performance of market prediction using machine learning approach. In: *2022 International Conference on Advances in Computing. Communication and Applied Informatics (ACCAI)*. India: IEEE; 2022. p.1-9.

[5] Liu HY, Wang Y, He D, Wang, C. Short term response of Chinese stock markets to the outbreak of COVID-19. *Applied Economics*. 2020; 52(53): 5859-5872.

[6] Li W, Chien F, Kamran HW, Aldeehani TM, Sadiq M, Nguyen VC, et al. The nexus between COVID-19 fear and stock market volatility. *Economic Research-Ekonomska Istraživanja*. 2022; 35(1): 1765-1785.

[7] Tiwari AK, Abakah EJA, Karikari NK, Gil-Alaña LA. The outbreak of COVID-19 and stock market liquidity: Evidence from emerging and developed equity markets. *The North American Journal of Economics and Finance.* 2022; 62: 101735.

[8] Boako G, Alagidede IP, Sjo B, Uddin GS. Commodities price cycles and their interdependence with equity markets. *Energy Economics*. 2020; 91: 104884.

[9] Lusiana H. The effect of return on equity (ROE) and earning per share (EPS) on stock prices in indonesia stock exchange 2015-2018. *Ilomata International Journal of Tax and Accounting*. 2020; 1(3): 132-138.

[10] Nafia A, Yousfi A, Echaoui A. Equity-market-neutral strategy portfolio construction using LSTM-based stock prediction and selection: An application to S&P 500 consumer staples stocks. *International Journal of Financial Studies*. 2023; 11(2): 57.

[11] Cuznetov A. The role of colonial trading companies in the institutionalization and formation of contemporary joint stock companies. In: *Proceedings of the Central and Eastern European eDem and eGov Days 2023*. New York: Association for Computing Machinery; 2023. p.204-208.

[12] Jimenez-Andrade JR. Origins of financial analysts in the United States. *Accounting History*. 2023; 28(2): 338-353.

[13] Nti IK, Adekoya AF, Weyori BA. A systematic review of fundamental and technical analysis of stock market predictions. *Artificial Intelligence Review*. 2020; 53(4): 3007-3057.

[14] Wagdi O, Salman E, Albanna H. Integration between technical indicators and artificial neural networks for the prediction of the exchange rate: Evidence from emerging economies. *Cogent Economics & Finance*. 2023; 11(2): 2255049.

[15] Jagirdar SS, Gupta PK. Charting the financial odyssey: A literature review on history and evolution of investment

strategies in the stock market (1900-2022). *China Accounting and Finance Review*. 2024; 26(3): 277-307.
[16] Syshchuk A, Panasiuk I. History and practice of technical methods of the international currency market. *Economic Journal of Lesya Ukrainka Volyn National University*. 2020; 1(21): 12-21.
[17] Sánchez-Granero MA, Balladares KA, Ramos-Requena JP, Trinidad-Segovia JE. Testing the efficient market hypothesis in Latin American stock markets. *Physica A: Statistical Mechanics and Its Applications*. 2020; 540: 123082.
[18] Li AW, Bastos GS. Stock market forecasting using deep learning and technical analysis: A systematic review. *IEEE Access*. 2020; 8: 185232-185242.
[19] Nyakurukwa K, Seetharam Y. Alternatives to the efficient market hypothesis: An overview. *Journal of Capital Markets Studies*. 2023; 7(2): 111-124.
[20] Liu G, Yu CP, Shiu SN, Shih IT. The efficient market hypothesis and the fractal market hypothesis: Interfluves, fusions, and evolutions. *Sage Open*. 2022; 12(1): 21582440221082137.
[21] Ehiedu VC, Obi KC. Efficient market hypothesis (EMH) and the Nigerian stock exchange in the midst of global financial crises. *International Journal of Academic Management Science Research (IJAMSR)*. 2022; 6(8): 263-273.
[22] Blackledge J, Lamphiere M. A review of the Fractal Market Hypothesis for trading and market price prediction. *Mathematics*. 2021; 10(1): 117.
[23] Brătian V, Acu AM, Oprean-Stan C, Dinga E, Ionescu GM. Efficient or fractal market hypothesis? A stock indexes modelling using geometric brownian motion and geometric fractional brownian motion. *Mathematics*. 2021; 9(22): 2983.
[24] Ghosh B, Bouri E. Long memory and fractality in the universe of volatility indices. *Complexity*. 2022; 2022(1): 6728432.
[25] Liu H, Li C, Shao Y, Zhang X, Zhai Z, Wang X, et al. Forecast of the trend in incidence of acute hemorrhagic conjunctivitis in China from 2011-2019 using the seasonal autoregressive integrated moving average (SARIMA) and exponential smoothing (ETS) models. *Journal of Infection and Public Health*. 2020; 13(2): 287-294.
[26] Yin T, Liu C, Ding F, Feng Z, Yuan B, Zhang N. Graph-based stock correlation and prediction for high-frequency trading systems. *Pattern Recognition.* 2022; 122: 108209.
[27] Lanbouri Z, Achchab S. Stock market prediction on high frequency data using Long-Short Term Memory. *Procedia Computer Science*. 2020; 175: 603-608.
[28] Sharma SK, Dwivedi YK, Metri B, Lal B, Elbanna A. Transfer, diffusion and adoption of next-generation digital technologies. In: *Proceedings of the IFIP WG. 8*. Cham: Springer; 2024.
[29] Khalil F, Pipa G. Is deep-learning and natural language processing transcending the financial forecasting? Investigation through lens of news analytic process. *Computational Economics*. 2022; 60(1): 147-171.
[30] Fang F, Chung W, Ventre C, Basios M, Kanthan L, Li L, et al. Ascertaining price formation in cryptocurrency markets with machine learning. *The European Journal of Finance*. 2024; 30(1): 78-100.
[31] Henrique BM, Sobreiro VA, Kimura H. Stock price prediction using support vector regression on daily and up to the minute prices. *The Journal of Finance and Data Science*. 2018; 4(3): 183-201.
[32] Ding G, Qin L. Study on the prediction of stock price based on the associated network model of LSTM. *International Journal of Machine Learning and Cybernetics*. 2020; 11(6): 1307-1317.
[33] Hu Z, Zhao Y, Khushi M. A survey of forex and stock price prediction using deep learning. *Applied System Innovation*. 2021; 4(1): 9.
[34] Agrawal SC. Deep learning based non-linear regression for Stock Prediction. In: *IOP Conference Series: Materials Science and Engineering*. Bristol, UK: IOP Publishing; 2021. p.012189.
[35] Karim R, Alam MK, Hossain MR. Stock market analysis using linear regression and decision tree regression. In: *2021 1st International Conference on Emerging Smart Technologies and Applications (eSmarTA)*. Sana'a, Yemen: IEEE; 2021. p.1-6.
[36] Jiang W. Applications of deep learning in stock market prediction: Recent progress. *Expert Systems with Applications*. 2021; 184: 115537.
[37] Ji X, Wang J, Yan Z. A stock price prediction method based on deep learning technology. *International Journal of Crowd Science*. 2021; 5(1): 55-72.
[38] Pölsterl S. Scikit-Survival: A library for time-to-event analysis built on top of scikit-learn. *Journal of Machine Learning Research*. 2020; 21(212): 1-6.
[39] Hao J, Ho TK. Machine learning made easy: A review of scikit-learn package in python programming language. *Journal of Educational and Behavioral Statistics.* 2019; 44(3): 348-361.
[40] Yahoo Finance. *Microsoft Corporation (MSFT)*. Available from: https://uk.finance.yahoo.com/quote/MSFT/history/

[Accessed 3rd August 2024].
[41] Yahoo Finance. *The Walt Disney Company (DIS)*. Available from: https://uk.finance.yahoo.com/quote/DIS/history/ [Accessed 3rd August 2024].
[42] Yahoo Finance. *Apple Inc. (AAPL)*. Available from: https://uk.finance.yahoo.com/quote/AAPL/history/ [Accessed 3rd August 2024].
[43] Yahoo Finance. *Alphabet Inc. (GOOG)*. Available from: https://uk.finance.yahoo.com/quote/GOOG/history/ [Accessed 3rd August 2024].
[44] Yahoo Finance. *S&P 500 (^GSPC)*. Available from: https://uk.finance.yahoo.com/quote/%5EGSPC/history/ [Accessed 3rd August 2024].
[45] Ding X, Liu J, Yang F, Cao J. Random radial basis function kernel-based support vector machine. *Journal of the Franklin Institute*. 2021; 358(18): 10121-10140.
[46] Zhang F, O'Donnell LJ. Support vector regression. In: *Machine Learning*. Amsterdam: Academic Press; 2020. p.123-140.
[47] Izonin I, Tkachenko R, Shakhovska N, Lotoshynska N. The additive input-doubling method based on the SVR with nonlinear kernels: Small data approach. *Symmetry*. 2021; 13(4): 612.
[48] Maulud D, Abdulazeez AM. A review on linear regression comprehensive in machine learning. *Journal of Applied Science and Technology Trends*. 2020; 1(2): 140-147.
[49] Alita D, Putra AD, Darwis D. Analysis of classic assumption test and multiple linear regression coefficient test for employee structural office recommendation. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*. 2021; 15(3): 295-306.