

Archives available at [journals.mriindia.com](http://journals.mriindia.com)

## International Journal on Advanced Computer Theory and Engineering

ISSN: 2319 - 2526

Volume 15 Issue 01s, 2026

### Feature Selection in Stock Market Prediction: A Comprehensive Review

<sup>1</sup>Mahesh M. Mahajan, <sup>2</sup>Dr. Nilesh A. Suryawanshi

<sup>1</sup> Research Scholar, NES's Gangamai College of Engineering, Nagaon

<sup>2</sup> Assistant Professor, NES's Gangamai College of Engineering, Nagaon

Peer Review Information	Abstract
<p><i>Submission: 08 Dec 2025</i></p> <p><i>Revision: 25 Dec 2025</i></p> <p><i>Acceptance: 10 Jan 2026</i></p> <p><b>Keywords</b></p> <p><i>Feature selection, Stock market prediction, Machine learning.</i></p>	<p>The current world is referred to as the "data world," as, according to Google, 328.77 million terabytes of data are generated every day and are continually increasing. One of the cause contributing to the data growth is the stock market. Thus, it is now necessary to reduce data by removing unnecessary data and extracting only the data that is important. The feature selection procedure is crucial for focusing on key data and reducing the dimensionality of the data. As per my knowledge there aren't many published articles that review the feature extraction and selection techniques utilized in stock market prediction at this time. The same motive will be covered in this paper, where we will analyze feature extraction and selection techniques utilized in the stock market. It includes embedded, filter, wrapper, supervised, unsupervised, semi-supervised, and hybrid methods.</p>

#### Introduction

The high volatility, non-stationary, noise, and large-scale heterogeneity characteristics of data, stock market prediction is fundamentally difficult. Modern prediction algorithms rely on technical indicators, macroeconomic parameters, and sentiment data. However, because of overfitting and redundancy, adding a lot of information frequently results in worse predictive performance. As a result, feature selection has become a most important preprocessing stage in pipelines for stock market prediction.

This study reviews the literature in great detail on employing several algorithms for feature selection, with an emphasis on research from last ten years that is 2015 to 2025. The study provides thorough descriptions of these techniques, datasets used, algorithms used, outcomes measured, evaluation metrics used in feature selection problems and findings about them. The evaluation also notes difficulties encountered and problems with generating the best feature subsets with various techniques. Although most Feature Selection methods have

historically been applied to static data, the increasing prevalence of dynamic data creation from websites and web-based apps brings additional challenges, such as noisy data that may hinder the efficiency of algorithms. The scalability of feature selection method is threatened by growing data sets, particularly when dealing with dynamic data. The dimensionality reduction technologies available today are insufficient to address these dynamic data issues (Nischal K. Verma et al., 2020). Using Feature Selection techniques reduces data and helps keep models away from overfitting.

In machine learning and pattern recognition applications, feature selection techniques including Filter, Embedded, and Wrapper approaches can save computation time, enhance prediction performance, and offer a more thorough comprehension of the data (A. Jović et al., 2015, G. Chandrashekar et al., 2014). Filter method, Embedded method and Wrapper method (Htun et al., 2023) are the basic feature selection methods along with it supervised, unsupervised, and semi-supervised feature selection techniques (Jun Chin et al. 2015)

are frequently used in machine learning issues such as clustering and classification (Jie Cai et al., 2018). In order to increase classification accuracy, many feature extraction and selection techniques seek to eliminate superfluous and irrelevant features from high-dimensional microarray data (Zena M. Hira et al., 2015). Log Gabor filters, Gabor filters, local binary patterns, higher-order local autocorrelation, and HLAC-like features are among the several feature extraction techniques for automatic facial expression detection (S. Lajevardi et al., 2010).

### 1. Filter Method

It is the process of locating and removing duplicate or unnecessary features from a dataset in order to prepare it for model training. Filter approaches do not directly entail training a machine learning model, in contrast to other feature selection techniques. To figure out the importance of a characteristic, they instead rely on statistical tests or metrics like chi-square, Pearson's correlation, etc.

### 2. Wrapper Method

In order to assess the significance of features, wrapper techniques rely on the machine learning model itself. They basically "wrap" the model you plan to employ for prediction around the feature selection procedure. These techniques evaluate various feature subsets iteratively, choosing the one that performs the best on the selected machine learning model.

### 3. Embedded Method

Embedded feature selection methods combine elements of wrapper and filter techniques, providing a trade-off between model-specific selection and efficiency. These techniques incorporate feature selection into the machine learning model's training procedure. Unlike the filter and wrapper techniques, respectively, they do not rely on independent statistical testing or the training of multiple models. During the training phase, these algorithms contain built-in mechanisms to give priority scores to features. A smaller feature set is obtained by excluding features with low significance values.

Without the use of any learning algorithms, filter techniques choose features based on each feature's statistical characteristics. Conversely, wrapper and embedded approaches choose the best collection of features for the algorithm by combining feature selection with a predefined learning algorithm.

This paper explores the significance of feature selection in achieving accurate and generalizable stock price predictions, drawing upon recent research findings. Forecasting the future trajectory of the stock market remains a persistent challenge, even in the age of advanced

machine learning (ML) algorithms. While these algorithms offer promising avenues for navigating the complex web of factors influencing stock prices, their effectiveness hinges on their ability to extract meaningful insights from vast and often high-dimensional datasets. This is where feature selection appears as a critical step in optimizing ML-based stock prediction models. This manuscript investigates the profound impact of feature selection on the accuracy of stock price predictions. We commence by exploring the inherent challenges associated with high-dimensional data and demonstrate how judicious feature selection strategically mitigates them.

Throughout this manuscript, we explore various feature selection methods, delve into their applicability to the historical context of stock market prediction, and critically examine the challenges associated with this crucial step in the modeling process. We will propose strategies to address these challenges effectively, ultimately demonstrating how carefully chosen feature selection can significantly enhance the efficacy and interpretability of ML- models based stock forecasting.

The next section in this paper contains literature review and the later one has the conclusion after the review.

### Objectives and Contributions

By analyzing current approaches, assessing their success across various feature methods and predictive models, and analyzing reported performance improvements using research results from earlier studies, this systematic review attempts to thoroughly investigate feature selection techniques applied to stock market prediction. The review also identifies limitations, and emerging trends. The comparative analyses summarizing methods, performance gains and datasets highlighting current trends and outlining an organized strategy for future financial forecasting research.

### Literature review

A methodical review of the previous years' worth of research papers, articles, and thesis is done using keywords like feature selection, and feature extraction. Then the search is narrowed down to stock market prediction step by step as shown in figure 1. First filter applied on research papers is of prediction that is feature selection or extraction method's used for prediction purpose for any case, then again it is narrowed down to stock market prediction only, and then the research papers are analyses in detail. This section gives the same review.

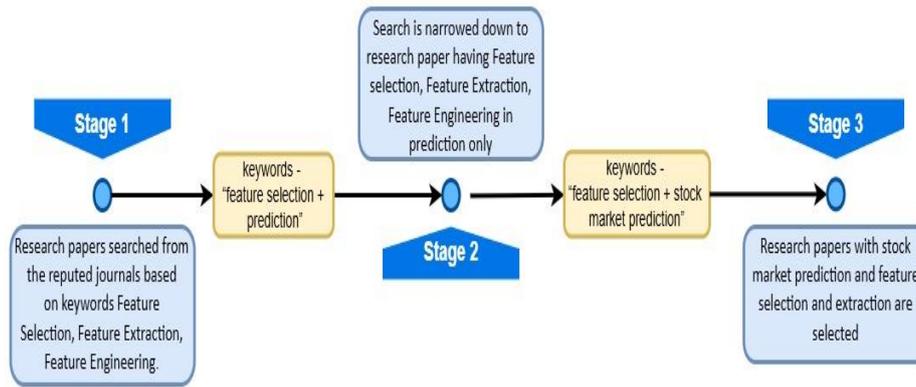


Fig. 1. Selection Criterion: Narrowing down the Research Paper selection

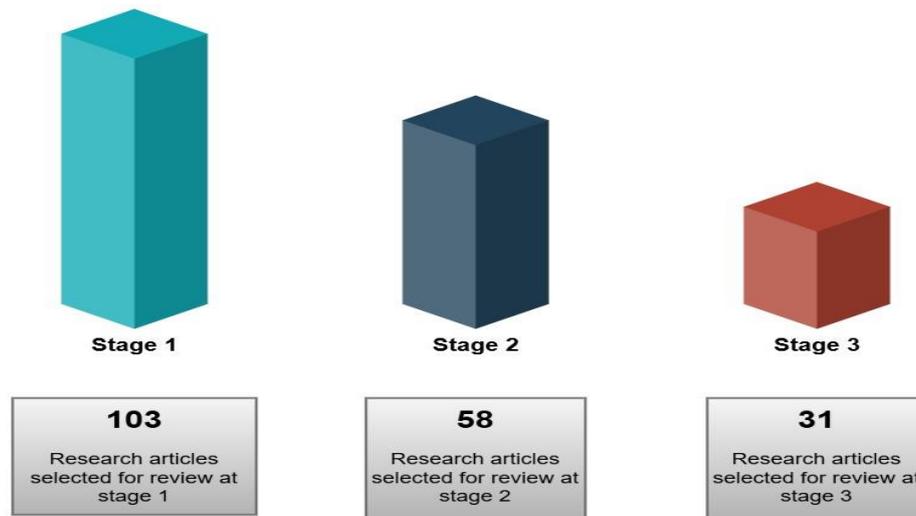


Fig. 2. Selection Criterion: Statistics

The figure 2 shows how research articles were gradually filtered through three stages of review. In the first stage, 103 articles were chosen for assessment. This number decreased to 58 in the second stage after further evaluation. By the third stage, only 31 articles remained, representing the most relevant studies. The steady reduction across the stages demonstrates a careful and structured screening process. Overall, the figure highlights the thorough approach used to select the most suitable research papers for final review.

Feature selection strategy can be used to decrease the algorithms complexity of any model-based on machine learning algorithm, which can also occasionally result in an enhancement of the performance of the simulated model (Zana O. Hamad et al., 2023). Using feature selection algorithms, Yuqinq He et al. conducted research on financial market theories and concepts as well as fundamental techniques for technical analysis of the stock market. Three algorithms—Feature Selection,

Sequential Forward Selection (SFS), Genetic Algorithm (GA) and Principal Component Analysis (PCA) are applied, and the outcomes are compared. According to it (Yuqinq He et al., 2023), volatility is the most important characteristic, whereas Williams %R and the Chaikin oscillator are the least important. It chooses 10 features from a total of 12, leaving out Williams and the Chaikin oscillator. Furthermore, because the Genetic Algorithm is random, the outcome may vary; yet, if the population and generation count are high enough, the outcome ought to approach a stable state and have an ideal solution. Therefore, the authors came to the conclusion that Genetic Algorithm would perform better in this scenario because it capitalizes on randomness, while Sequential Forward Selection (SFS) (Liu Boshu et al., 2007) might produce a local optimal solution but run the risk of "nesting problem" (Yuqinq He et al., 2023).

Current stock forecasting methods frequently employ a single feature selection procedure,

which may miss crucial presumptions on the correlation between input and output variables. Anwar Ul Haq et al. (2021) have observed that the accuracy and performance of prediction models can be enhanced by integrating features obtained by different feature selection approaches to create an ideal feature subset. The authors of this paper present a novel method for producing an ideal feature subset by combining characteristics chosen from several feature selection strategies. They use random forests, support vector machines, and logistic regression models as their three feature selection methods. Each method bases its ranking of the technical indicators' significance on its own set of presumptions regarding the fundamental link between the input and output variables. For instance, RF takes into account statistical interactions between variables, whereas L1-LR assumes a linear regression function (Haq et al., 2019). Features with low importance measures below a threshold are dropped from each ranked feature set. The remaining features are then clubbed together into clusters, and the most significant feature from each cluster is selected to produce the final subset. The process results in multiple feature subsets, including the ones opted by each individual method and the final set generated by combining them and evaluate the importance of each feature. The features are then grouped into clusters, and the most important feature from each cluster is selected to produce the final subset. In order to predict future price movements, the chosen features are subsequently fed into a deep generative model that also incorporates an attention mechanism and an extractor of market signals. To manage the stochastic character of stock market data, the market signal extractor decodes market movement from hidden variables, and the attention mechanism makes a distinction between different temporal auxiliary outputs' predictive relationships. The study's findings show that integrating features chosen from various feature selection techniques and providing them to the deep generative model surpasses cutting-edge methods for predicting market trends. Anwar Ul Haq et al. (2021) state that this method efficiently chooses a better ideal feature set, enhances prediction models performance and assisting in the development of more precise stock forecasts.

Mohammad Tubishat et al., uses one of the nature inspired algorithm salp swarm algorithm based on the salps swarming behaviour. The technique is specifically created for feature selection, with the goal of extracting the most significant characteristics from datasets. It is based on the swarm intelligence paradigm, which enables

efficient exploration of a huge search space of solutions (Mio). In enhanced SSA, salps use their swarming tendency in the deep seas to form a chain known as the salp chain. An optimization approach was used to represent this salp chain, and it involved classifying the population into two subcategories: leaders & followers. The chain's followers adhere to one another and to the leader's instructions. Mathematically, each salp's location in the SSA population is defined in a search space of  $n$  dimensions, where  $n$  represents the number of variables. Here, the authors present an advanced version of the SSA approach that addresses the feature selection issue using the Local Search approach (LSA) and the Opposition Based Learning (OBL) technique. Because ISSA chose fewer features overall across all 18 datasets, it performed better in feature selection than the SSA method (Mohammad Tubishat et al., 2019). Novel Improved SSA (Miodrag Zivkovic et al., 2022) is mostly utilized for sensor data while selecting features. With the goal of extracting the most significant features from datasets, the technique is especially made for feature selection. It adheres to the Learnheuristics (Miodrag Zivkovic et al., 2022) Approach, which combines metaheuristic and machine learning techniques (R. Sarkhel et al., 2017). While metaheuristic optimization algorithms are used to optimize difficult problems that may not have simple solutions, utilization of machine learning algorithms is done in learnheuristics to learn from data and generate predictions. Learnheuristics seeks to improve the efficacy and efficiency of resolving difficult optimization and decision-making problems by combining these two methods. Enhancing data processing, increasing accuracy, optimizing resource utilization, speeding up decision-making, and solving problems effectively are just a few advantages it offers. The utilization of deep neural networks for forecasting stock prices is discussed in the study, and it is noted that these networks can discover from the data abstract representations that less hidden layer networks would not be able to (Yaohao Peng et al., 2021). Additionally, they apply the three feature selection techniques—SFFS, TS, and LASSO—to the feature set of technical indicators, employing the filtered versions of these indicators as training process explanatory variables. Sequential Feature Selection (SFS), Sequential Backward Selection (SBS) and a forward wrapper selection technique, are combined in Sequential Forward Floating Selection. A heuristic search technique 'Tournament Screening' (Chen et al. 2009), that produces candidate features by selecting the most advantageous features from subsets that are mutually exclusive. Recursively,

the variables are divided into smaller batches, and within each subset, a "tournament" is held, with the characteristics that "survive" the competitions being categorized as the greatest. The Least Absolute Shrinkage and Selection Operator (LASSO) is a widely used minimizing technique for linear regression. They have used total 124 technical indicators (Yaohao Peng et al., 2021) out of which 51 they found in their literature survey and remaining 73 they got on few financial websites - TradingView, Fidelity Investments, StockCharts, and Trading Technologies.

In an effort to support investors in making wise trading decisions, the study addresses the difficulties related to predicting stock price and presents the PCC- Pearson correlation coefficient (A Aloraini, 2015) and Spearman rank correlation framework for this purpose. A multi-indicator method for feature selection based on PCC and BLS is used to predict stock prices, by selecting input features based on the PCC with a threshold of 0.5 for the four experimental stocks using PCC (Guanzhi Li et al., 2022) from the available 35 features and performance is compared with ten other ML methods. The authors considered the following goals: [1] gather an ample data set with a variety of parameters influencing the closing price; [2] BLS performance is improved by applying PCC for feature selection; and [3] evaluate the effectiveness of the suggested PCC-BLS model. The dataset used by authors consist of stock price data gathered from Sept 2010 to Sept 2021, technical and financial indicators, and the Shenzhen Securities Component Index (SZI) and Shanghai Securities Composite Index (SSEC) obtained from the NetEase website to calculate the prediction accuracy of the closing price of listed stocks from the Shanghai Stock Exchange using the proposed PCC-BLS framework (Guanzhi Li et al., 2022). While the proposed model lacks the practical application.

In order to quantify stock market uncertainty and ascertain the reaction of Black Swan occurrences in the stock market, Samit Bhanja and Abhishek Das make reference to the Black Swan Event Analysis algorithm (BSEA) and Black Swan Event's Impact (BSEI) algorithm put forth in the paper. These algorithms are made to examine and identify characteristics associated with Black Swan events and how they affect trends in the stock market. This study's technique include putting forth a hybrid stock trends prediction framework (HSTPF) that contrasts several machine learning classifiers, extracts technical information, deploys a CNN-based autoencoder (Esterabi and Dami 2021; Chong et al. 2017; Xie et al. 2021; Gunduz 2021), and uses a BSEA

algorithm. The methodology uses a tiny second dataset to anticipate stock patterns and is divided into three functional phases. Dataset of Nifty 50 and S&P BSE SENSEX is used to achieve objectives like capturing market uncertainty with a BSEA algorithm, extracting technical features for improved prediction accuracy, reducing dataset dimensionality with a CNN-based autoencoder, and comparing machine learning classifiers for optimal trend prediction. After applying above said methodology they have measured prediction accuracy of stock trend using HSTPF, especially during Black Swan events and the study include the confirmation of Gradient Boosting Machine (GBM) as the superior machine learning classifier for proposed Hybrid Stock Trends Prediction Framework (HSTPF), which exhibited stable performance and outperformed other prototypes even during Black Swan events.

Changjun Zhou and Shile Chen used a combination of optimal factors selected through Genetic Algorithm (S.Jadhav et al. 2018) and the LSTMNN- Long ShortTerm Memory neural network model with a couple of phases including GA for ranking factor significance and optimizing selection of features for the LSTM model. Past data from CSI 300 and China Construction Bank stock from the first of January 2010, to the first of April, 2020 is used to forecast stock price. The JoinQuant quantitative platform is utilized to gather historical data of the CSI 300 stock and CC Bank. The proposed model is then applied to this information. In this experiment, the splitting of data is 80% - 20% for training and testing respectively. The application of genetic algorithms for feature selection, the process of encoding potential solutions using genetic algorithms, fitness function calculations, chromosome crossing and mutation procedures, LSTM network construction, and mean square error model evaluation are all covered in this study.

The author Saud S. Alotaibi, 2021 emphasize the significance of feature selection, optimization techniques, and prediction accuracy in their summary of the main procedures and results of the suggested stock market forecasting model. The dataset used in the study includes market data from the companies DAR ALARKAN, ALINMA and SABIC. Statistical and technical indicator features are extracted by authors from above dataset along with optimal features are selected using the Red Deer Adopted Wolf Algorithm (RDAWA). Saudi stock market closing prices of the days - day 8th, day 16th, and day 21st are predicted using an ensemble technique with classifiers like RF1, optimized NN, RF2, and

SVM. The study presents a novel ensemble technique for stock market prediction that includes extraction of features, selection of optimal features, and prediction utilizing classifiers such as RF1, SVM, optimized NN, and RF2. RDAWA a hybrid optimization method is used in the suggested model for feature selection and to fine-tune the neural network's optimal weights in order to improve prediction accuracy and Mean, Standard Deviation, Variance, Skewness, Kurtosis, ATR, EMA, RSI, ROC are measured and results are compared with existing models. According to the findings, the RDAGW model performed comparatively better than previous methods in concern with of error metrics including MSE, RMSE, MAPE, and MAE, indicating that it is a quite useful method for accurate stock market prediction. The objective of Understanding the predictability of future returns and introducing an ensemble technique using conventional classifiers and a novel RDAGW algorithm for NN training was achieved by the authors.

The study employed the graph convolutional feature based convolutional neural network (GC-CNN) model as the specific algorithm to be introduced (Wei Chen et al. 2021), utilized and examined. To aim the stock trend prediction, this model combines an improved GCN (IGCN) with a Dual-CNN to concurrently catch stock market data and discrete stock features. The GC-CNN approach performs better than several different stock trading strategies and stock trend forecast techniques. Wei Chen et al. employ a methodology on the dataset of stock market data from six randomly selected Chinese stocks, including technical indicators and preprocessed data that proposes a novel approach to stock trend forecasting using a GC-CNN model, taking into account information from both stock market and discrete stock sources. The study chooses companies from the Chinese stock market, preprocesses data, assigns labels based on closing prices, implements TensorFlow and Sklearn, uses a sliding window method for training and testing, and assesses performance (Wei Chen, Manrui Jiang, 2021) using both computational and financial factors. Different predictions and popular stock trading techniques are the basis for simulated trading, which is carried out using a variety of assessment measures for comparison. A starting capital of 50,000 RMB is allocated for the financial evaluation.

A new methodology is implemented for prediction system having GA-XGBoost with an improved 3-stage feature selection process for stock price direction prediction. XGBoost is used as the primary predictor in, Daehan Won, Sang

Won Yoon, and Kyung Keun Yun's (2021) technique, which highlights the scalability, accuracy, and effectiveness of XGBoost in machine learning contests. A GA is also integrated for feature extraction. The machine learning techniques such as supervised learning algorithms (e.g., Linear Regression, Random Forests), unsupervised learning methods, binary classification tasks, feature engineering, development of an executable Jar Project (Rebwar M. Nabi et al. 2021) with a GUI using Java Programming Language, and the use of Waikato Environment for Knowledge Analysis (WEKA) as a machine learning platform. The goal was to determine the optimal classifier technique and increase classification accuracy by incorporating new features. In the end, after using many feature selection algorithm like SFS Search, Genetic Search, CFS Subset Evaluation and CFS Subset Evaluation, Ranker Search and CHI, Ranker Search and RFE Evaluation, Ranker Search and IG, Ranker Search and CCE, Ranker Search and REF, Ranker Search and PCA, it is concluded that feature selection and feature engineering worked together to achieve 100% accuracy in the majority of cases, specifically using PCA with Multilayer Perceptron and SVM. Yaping Hao et al. discusses highlights the significance of forecasting the stock market index trajectory, presents a hybrid neural network that combines signals from several time scales, and outperforms benchmark models to show the network's efficacy. The authors extract features on various time scales through a CNN, utilizes three LSTM networks to catch time reciprocities in the features, and use fully connected layers predicts the stock index price trends. The recommended method is constructing a hybrid neural network using feature learning across many different scales, which incorporates information from many time scales and produces predictions that are more accurate than those produced by models based on a one time scale feature. The hybrid neural network simplifies the model by using a single CNN for feature extraction and achieves better predictive performance, and providing relevant work, tests, findings, and discussions.

According to recent studies published, the limitations of individual statistical methods are further highlighted by the survey conducted by Htun et al. (2023), which brings together evidence from various markets and confirms that combining filter and wrapper strategies consistently yields moderate yet reliable performance gains. Beyond static selection, Kim et al. (2024) present a (DFSS) dynamic feature selection system that adjusts chosen features over time and across different industries,

highlighting the increasing demand for feature selection mechanisms that adapt to moving market conditions. Similarly, He et al. (2024) show that combination of effective optimization with precise feature selection boosts error-based metrics like RMSE and MAE when applied to stock indices, highlighting the efficacy of optimization-improved hybrid approaches. As shown by Pabuccu and Barbu (2024), annealing-based feature selection increases robustness in financial time-series forecasting, at the cost of higher computational cost, illustrating the

ongoing interest in meta-heuristic strategies. In addition to highlighting new trade-offs between scalability, system complexity and performance in real time financial forecasting applications, these recent contributions collectively support the trend toward optimization-aware and adaptive feature selection techniques.

The following Table 1 depicts summary of few research papers based on some points like Algorithm used, Dataset used by authors and outcomes and findings as per my knowledge and understandings.

**Table 1:** Summary of few latest research papers

Author	Technique used	Feature selection Category	Model for Prediction	Performance Gain	Key Limitations
Htun et al. (2023)	Filter + Wrapper	Hybrid	ML & DL models	Consistent accuracy improvement ( $\approx 3-6\%$ )	Review-based, no new model
Kim et al. (2024)	DFSS	Dynamic	LSTM, RF	Improved adaptability and stability	System complexity
He et al. (2024)	Comprehensive FS + Optimization	Hybrid	ML ensembles	Reduced error metrics (RMSE, MAE)	Dataset-dependent
Pabuccu & Barbu (2024)	Annealing-based FS	Meta-heuristic	ML regressors	Improved forecasting robustness	Computational overhead
Li et al. (2022)	PCC	Filter	BLS	+3-5% accuracy	Threshold sensitivity
Bhanja & Das (2022)	BSEA + Autoencoder	Hybrid	GBM	Improved trend stability	Market-specific
Chen & Zhou (2021)	GA	Wrapper	LSTM	Reduced MSE	High computational cost
Alotaibi (2021)	RDAWA	Hybrid	Ensemble	Lower RMSE, MAE	Limited generalization
Chen et al. (2021)	GC-CNN	Embedded	CNN	Improved returns	Long training time
Yun et al. (2021)	GA-XGBoost	Hybrid	XGBoost	Higher accuracy	Model complexity
Hao & Gao (2020)	CNN-based FS	Embedded	CNN-LSTM	Improved trend accuracy	Limited interpretability
Tubishat et al. (2020)	ISSA	Meta-heuristic	ML classifiers	Higher accuracy, fewer features	Not finance-specific

### Comparative Analysis of Literature

The representative feature selection methods for stock market prediction across different datasets and classifiers are compared in Table 3.1. The significance of feature selection in stock market forecasting is confirmed by the countable performance improvements, which are reported by all reviewed studies. The improvements range from roughly 3% to 7%. Mutual Information a filter-based technique offers low computational costs and moderate gains, but they are not able to catch interdependencies of features. Although their improvements are still limited, embedded techniques such as LASSO provide model-aware

selection and efficiency. Hybrid approaches gives higher gains that combine filtering techniques with genetic algorithms, especially for heterogeneous technical and sentiment data. PSO-based feature selection in combination with XGBoost produce the greatest accuracy improvement in meta-heuristic methods, which generates the best performance. Overall, the comparison shows that in stock market prediction tasks, hybrid and meta-heuristic feature selection techniques combined with ensemble classifiers consistently outperform standalone approaches.

**Table 2:** Comparison of Feature Selection Techniques

Dataset used	Feature Selection Method	Features used	Classifier	Reported Improvement
S&P 500	Mutual Information	Technical	SVM	+4.2% Accuracy
NSE	GA + Filter	Technical + Sentiment	Random Forest	+6.8% F1-score
NASDAQ	LASSO	Technical	ANN	+3.1% Accuracy
Nikkei 225	PSO-based FS	Hybrid	XGBoost	+7.4% Accuracy

### Conclusion

By reducing dimensionality and removing non required features, feature selection consistently improves performance of prediction, as this systematic review confirms. Gains in performance are typically range from 3% to 7%, according to a comparative analysis of the reviewed studies, with hybrid and meta-heuristic feature selection techniques outperforms standalone filter and embedded approaches. Specifically, the maximum gains are obtained when optimization-driven methods are coupled with ensemble classifiers. Feature selection significantly increases the robustness of machine learning models, even though perfect stock market prediction is still unattainable due to inherent market uncertainty. Future research should focus on hybrid and adaptive feature selection models to further increase predictive accuracy in stock market forecasting as the complexity of financial data continues to rise.

### References

- [1] Agrawal, P., Abutarboush, H.F., Ganesh, T., Mohamed, A.W.: Metaheuristic algorithms on feature selection: A survey of one decade of research (2009–2019). *IEEE Access* **9**, 1–1 (2021). <https://doi.org/10.1109/access.2021.3056407>
- [2] Aloraini, A.: Penalized ensemble feature selection methods for hidden associations in time series environments: Case study on equities companies in Saudi stock exchange market. *Evolving Systems* **6**(2), 93–100 (2014). <https://doi.org/10.1007/s12530-014-9124-y>
- [3] Alotaibi, S.S.: Ensemble technique with optimal feature selection for Saudi stock market prediction: A novel hybrid Red Deer–Grey algorithm. *IEEE Access* **9**, 64929–64944 (2021). <https://doi.org/10.1109/access.2021.3073507>
- [4] Ang, J.C., Mirzal, A., Haron, H., Abdull Hamed, H.N.: Supervised, unsupervised, and semi-supervised feature selection: A review on gene selection. *IEEE/ACM Transactions on Computational Biology and Bioinformatics* **13**(5), 971–989 (2016). <https://doi.org/10.1109/tcbb.2015.2478454>
- [5] Sujana Hiregundagal Gopal Rao. (2023). A Review of Intrusion Detection Methods for In-Vehicle Networks at the Semiconductor Level. *International Journal of Intelligent*

- Systems and Applications in Engineering, 11(10s), 1032–1036. Retrieved from <https://ijisae.org/index.php/IJISAE/article/view/8000>
- [6] Bouktif, S., Fiaz, A., Awad, M.: Augmented textual features-based stock market prediction. *IEEE Access* **8**, 40269–40282 (2020). <https://doi.org/10.1109/access.2020.2976725>
- [7] Chen, S., Zhou, C.: Stock prediction based on genetic algorithm feature selection and long short-term memory neural network. *IEEE Access* **9**, 9066–9072 (2021). <https://doi.org/10.1109/access.2020.3047109>
- [8] Chen, W., Jiang, M., Zhang, W.G., Chen, Z.: A novel graph convolutional feature-based convolutional neural network for stock trend prediction. *Information Sciences* **556**, 67–94 (2021). <https://doi.org/10.1016/j.ins.2020.12.068>
- [9] Chen, Z., Chen, J.: Tournament screening cum EBIC for feature selection with high-dimensional feature spaces. *Science in China Series A: Mathematics* **52**(6), 1327–1341 (2009). <https://doi.org/10.1007/s11425-009-0089-4>
- [10] Chong, E., Han, C., Park, F.C.: Deep learning networks for stock market analysis and prediction: Methodology, data representations, and case studies. *Expert Systems with Applications* **83**, 187–205 (2017). <https://doi.org/10.1016/j.eswa.2017.04.030>
- [11] Dami, S., Esterabi, M.: Predicting stock returns of Tehran Exchange using LSTM neural network and feature engineering technique. *Multimedia Tools and Applications* **80**(13), 19947–19970 (2021). <https://doi.org/10.1007/s11042-021-10778-3>
- [12] Gunduz, H.: An efficient stock market prediction model using hybrid feature reduction method based on variational autoencoders and recursive feature elimination. *Financial Innovation* **7**(1), 1–15 (2021). <https://doi.org/10.1186/s40854-021-00243-3>
- [13] Hao, Y., Gao, Q.: Predicting the trend of stock market index using the hybrid neural network based on multiple time scale feature learning. *Applied Sciences* **10**(11), 3961–3974 (2020). <https://doi.org/10.3390/app10113961>
- [14] Haq, A.U., Zeb, A., Lei, Z., Zhang, D.: Forecasting daily stock trend using multi-filter feature selection and deep learning. *Expert Systems with Applications* **168**, 114444 (2021). <https://doi.org/10.1016/j.eswa.2020.114444>
- [15] Htun, H.H., Biehl, M., Petkov, N.: Survey of feature selection and extraction techniques for stock market prediction. *Financial Innovation* **9**(1), 1–18 (2023). <https://doi.org/10.1186/s40854-022-00441-7>
- [16] Li, G., Zhang, A., Zhang, Q., Wu, D., Zhan, C.: Pearson correlation coefficient-based performance enhancement of broad learning system for stock price prediction. *IEEE Transactions on Circuits and Systems II: Express Briefs* **69**(5), 2413–2417 (2022). <https://doi.org/10.1109/tcsii.2022.3160266>
- [17] Liu, B., Li, S., Wang, Y., Lu, L., Li, Y., Cai, Y.: Predicting the protein SUMO modification sites based on properties sequential forward selection (PSFS). *Biochemical and Biophysical Research Communications* **358**(1), 136–139 (2007). <https://doi.org/10.1016/j.bbrc.2007.04.097>
- [18] Nabi, R.M., Saeed, S.A.M., Harron, H.B., Fujita, H.: Ultimate prediction of stock market price movement. *Journal of Computer Science* **15**(12), 1795–1808 (2019). <https://doi.org/10.3844/jcssp.2019.1795.1808>
- [19] Niu, T., Wang, J., Lu, H., Yang, W., Du, P.: Developing a deep learning framework with two-stage feature selection for multivariate financial time series forecasting. *Expert Systems with Applications* **148**, 113237 (2020). <https://doi.org/10.1016/j.eswa.2020.113237>
- [20] Peng, Y., Albuquerque, P.H.M., Kimura, H., Saavedra, C.A.P.B.: Feature selection and deep neural networks for stock price direction forecasting using technical analysis indicators. *Machine Learning with Applications* **5**, 100060 (2021). <https://doi.org/10.1016/j.mlwa.2021.100060>
- [21] Pudjihartono, N., Fadason, T., Kempa-Liehr, A.W., O’Sullivan, J.M.: A review of feature selection methods for machine learning-based disease risk prediction. *Frontiers in Bioinformatics* **2**, 1–25 (2022). <https://doi.org/10.3389/fbinf.2022.927312>
- [22] Sarkhel, R., Chowdhury, T.M., Das, M., Das, N., Nasipuri, M.: A novel harmony search algorithm embedded with metaheuristic opposition-based learning. In: Thampi, S.M., El-Alfy, E.S.M. (eds.) *Journal of Intelligent & Fuzzy Systems* **32**(4), 3189–3199 (2017). <https://doi.org/10.3233/jifs-169262>
- [23] Tubishat, M., Idris, N., Shuib, L., Abushariah, M.A.M., Mirjalili, S.: Improved Salp Swarm Algorithm based on opposition-based learning and novel local search algorithm for feature selection. *Expert Systems with Applications* **145**, 113122 (2020). <https://doi.org/10.1016/j.eswa.2019.113122>

- [24] Verma, N.K., Salour, A.: Feature selection. In: *Studies in Systems, Decision and Control*, pp. 175–200. Springer, Singapore (2020). [https://doi.org/10.1007/978-981-15-0512-6\\_5](https://doi.org/10.1007/978-981-15-0512-6_5)
- [25] Yuan, X., Yuan, J., Jiang, T., Ul Ain, Q.: Integrated long-term stock selection models based on feature selection and machine learning algorithms for China stock market. *IEEE Access* **8**, 22672–22685 (2020). <https://doi.org/10.1109/access.2020.2969293>
- [26] Yun, K.K., Yoon, S.W., Won, D.: Prediction of stock price direction using a hybrid GA-XGBoost algorithm with a three-stage feature engineering process. *Expert Systems with Applications* **186**, 115716 (2021). <https://doi.org/10.1016/j.eswa.2021.115716>
- [27] Zivkovic, M., Stoean, C., Chhabra, A., Budimirovic, N., Petrovic, A., Bacanin, N.: Novel improved Salp Swarm Algorithm: An application for feature selection. *Sensors* **22**(5), 1711 (2022). <https://doi.org/10.3390/s22051711>
- [28] Htun, H.H., Biehl, M., Petkov, N.: Survey of feature selection and extraction techniques for stock market prediction. *Financial Innovation* **9**(1), 1–18 (2023). <https://doi.org/10.1186/s40854-022-00441-7>
- [29] Kim, W., Lee, J., Park, S., et al.: Developing a dynamic feature selection system (DFSS) to predict stock prices across major industries. *Applied Sciences* **14**(16), 7314 (2024). <https://doi.org/10.3390/app14167314>
- [30] He, X., Wang, J., Xiao, L., Liu, J., Ma, X.: A hybrid forecasting system based on comprehensive feature selection and intelligent optimization for stock price index forecasting. *Mathematics* **12**(23), 3778 (2024). <https://doi.org/10.3390/math12233778>
- [31] Pabuccu, H., Barbu, A.: Feature selection with annealing for forecasting financial time series. *Evolving Systems* (2024). <https://doi.org/10.1186/s40854-024-00617-3>