iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 Issue: IV Month of publication: April 2025

DOI: https://doi.org/10.22214/ijraset.2025.69153

Machine Learning Approaches in Stock Market Prediction: A Systematic Literature Review

Divyansh Bhardawaj[1], Aditi Verma[2], Harsh Gupta[3]
Department of Engineering and Technology, Sharda University 201308

Abstract: This paper provides a comprehensive review of the use of machine learning (ML) models in stock market prediction, focusing on the effectiveness of various approaches and the integration of external data sources. The review covers widely used ML models, including Artificial Neural Networks (ANN), Support Vector Machines (SVM), Long Short-Term Memory (LSTM) networks, and hybrid models. It compares the performance of these models, highlighting key metrics such as accuracy, precision, and Root Mean Square Error (RMSE).
Keywords: Machine Learning, Stock Market Prediction, Long Short-Term Memory (LSTM), Artificial Neural Networks (ANN), Support Vector Machines (SVM), Hybrid Models, Financial Forecasting, Sentiment Analysis,

I. INTRODUCTION

Predicting stock market behavior has long posed a formidable challenge due to the market's volatile, dynamic, and non-linear nature. Traditionally, investors have relied on fundamental analysis—which focuses on a company's financial performance, macroeconomic indicators, and industry trends—and technical analysis, which studies historical trading data such as price movement and volume to forecast future stock behavior. While these approaches remain valuable, they are often limited by their inability to process vast, complex, and high-frequency datasets effectively (Sharma et al., 2017). As markets evolve and data becomes more abundant and unstructured, traditional models increasingly struggle to deliver accurate predictions in real-time scenarios.

The emergence of machine learning (ML) and the exponential growth in computational power have revolutionized financial forecasting. ML models possess the ability to learn from historical data, uncover hidden patterns, and adapt to new data inputs, providing a more robust and flexible framework for stock prediction. Unlike static statistical techniques, ML algorithms evolve over time and are capable of modeling non-linear and dynamic relationships within data. Various ML techniques have been employed for this task, including artificial neural networks (ANN), support vector machines (SVM), decision trees, and more recently, deep learning models such as long short-term memory (LSTM) networks (Vui et al., 2013; Selvin et al., 2017). These models are especially suited for time-series data, offering enhanced accuracy by capturing complex dependencies and long-term trends.

This paper presents a systematic literature review (SLR) of 30 peer-reviewed studies published between 2012 and 2020, focused on ML-based stock market prediction models. The primary goal is to evaluate the current research landscape and identify prevalent models, methodological trends, and performance outcomes. Key research questions guiding this review include: What is the current state of ML applications in stock market forecasting? What are the emerging trends and innovations in this field? And which models have demonstrated the highest accuracy and reliability?

The findings reveal that neural network-based approaches dominate the literature, particularly LSTM models, due to their suitability for time-dependent data and superior performance over traditional models such as SVM and basic ANN. Additionally, hybrid models combining techniques like CNN-LSTM or GA-ANN have shown promise in improving predictive power. As the financial domain continues to embrace intelligent systems, this review aims to guide future research by highlighting effective methodologies, data requirements, and the evolving role of ML in transforming stock market prediction (Selvin et al., 2017; Nelson et al., 2017).

II. LITERATURE REVIEW

1) Vui et al. (2013) reviewed the application of artificial neural networks (ANN) in financial forecasting and emphasized the effectiveness of models such as feedforward NN and backpropagation NN in capturing non-linear relationships in stock data.
2) Ticknor (2013) implemented a Bayesian regularized ANN and demonstrated its robustness and ability to avoid overfitting in noisy financial environments.
3) Bing et al. (2012) applied ANN to predict the stock market and concluded that these models could effectively model non-linear patterns, though they required careful tuning and large datasets for optimal performance.

4) Ren et al. (2018) combined sentiment analysis with SVM to forecast stock market movement direction and found that integrating external data sources significantly improved prediction accuracy.
5) Ding et al. (2015) extended this idea by leveraging structured events extracted from financial news and applying deep learning in conjunction with SVM, demonstrating how sentiment-driven market movements can be effectively captured. These studies indicate that SVM remains valuable, particularly when coupled with additional contextual data.
6) Selvin et al. (2017) compared LSTM with RNN and CNN models and found that LSTM outperformed the others in capturing long-term dependencies in stock price data.
7) Nelson et al. (2017) also employed LSTM and demonstrated its superior ability to learn temporal patterns in historical stock prices, further validating its potential in financial prediction.
8) Vargas et al. (2017) combined textual analysis of financial news with deep learning models, integrating word embeddings to improve the semantic understanding of input data.
9) Qiu and Song (2016) introduced an optimized ANN using genetic algorithms (GA-ANN), resulting in enhanced prediction performance by fine-tuning model parameters effectively.

III. PROBLEM STATEMENT

The stock market is inherently volatile and influenced by a complex interplay of factors including economic indicators, geopolitical events, investor sentiment, and global financial trends. This dynamic and non-linear nature of the market makes accurate prediction a highly challenging task. Traditional forecasting methods such as fundamental and technical analysis often fall short in capturing hidden patterns, long-term dependencies, and high-dimensional relationships within historical financial data (Sharma et al., 2017). As a result, predictions derived from these conventional approaches may lack precision and adaptability, especially in real-time scenarios where rapid decision-making is crucial. The emergence of machine learning (ML) offers a promising solution to this problem by enabling systems to learn from vast datasets, recognize intricate trends, and generalize across market conditions. However, despite the success of ML algorithms such as artificial neural networks (ANN), support vector machines (SVM), and long short-term memory (LSTM) networks in stock forecasting (Selvin et al., 2017; Nelson et al., 2017), challenges such as model overfitting, limited interpretability, and dependence on high-quality data persist. This research aims to address these challenges by systematically reviewing and evaluating the application of various ML models in stock price prediction. It particularly focuses on supervised learning models that utilize historical price data and technical indicators to forecast future trends with improved accuracy. By comparing different ML algorithms and identifying their strengths, limitations, and performance under various conditions, this study seeks to contribute to the development of more reliable, interpretable.

IV. METHODOLOGY

This study employs a Systematic Literature Review (SLR) approach to examine the application of machine learning (ML) techniques in stock market prediction. The methodology follows the guidelines of the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA), ensuring transparency, reproducibility, and a structured framework for reviewing academic literature (Sharma et al., 2017). A comprehensive search was conducted across major academic databases including Google Scholar, IEEE Xplore, ScienceDirect, and the ACM Digital Library. The keywords used during the search phase included “machine learning,” “stock market,” “prediction,” “approach,” and “model.” The focus was to gather research published between 2012 and 2020, written in English, peer-reviewed, and specifically addressing ML applications for stock price prediction.

To maintain quality and relevance, specific inclusion and exclusion criteria were applied. Articles were included if they were (1) peer-reviewed, (2) published within the specified date range, (3) written in English, and (4) explicitly focused on stock market prediction using ML models. Studies were excluded if they were duplicates, inaccessible in full-text format, or unrelated to the topic. The selection process followed three key steps: identification, screening, and eligibility. During identification, all potentially relevant titles and abstracts were collected. The screening phase involved reviewing these abstracts for topic relevance, while the eligibility phase entailed a full-text evaluation based on the inclusion criteria.

In total, 30 studies were selected for in-depth analysis. Key data extracted from each study included the year of publication, ML models applied, type of dataset used, performance metrics, and key conclusions. These elements were then synthesized to identify patterns, emerging trends, and the comparative effectiveness of different ML models. By leveraging a structured and transparent methodology, this review ensures the reliability of its findings and provides a valuable foundation for future research in financial forecasting using ML techniques (Selvin et al., 2017; Nelson et al., 2017).

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue IV Apr 2025- Available at www.ijraset.com

A. Data Sources and Search Strategy

A detailed search was conducted across multiple scholarly databases including Google Scholar, IEEE Xplore, ScienceDirect, and ACM Digital Library. The keywords employed were: “machine learning,” “stock market,” “prediction,” “approach,” and “model.” Boolean operators and filters were used to refine the results to the most relevant studies published between 2012 and 2020. Only peer-reviewed, English-language articles with a direct focus on ML in financial prediction were considered.

B. Inclusion and Exclusion Criteria

To maintain consistency and academic rigor, specific inclusion and exclusion criteria were applied. Studies were included if they (1) were peer-reviewed, (2) published between 2012 and 2020, (3) written in English, and (4) focused on the application of ML in stock market forecasting. Articles were excluded if they were duplicates, lacked full text, or were not directly related to ML-based prediction techniques.

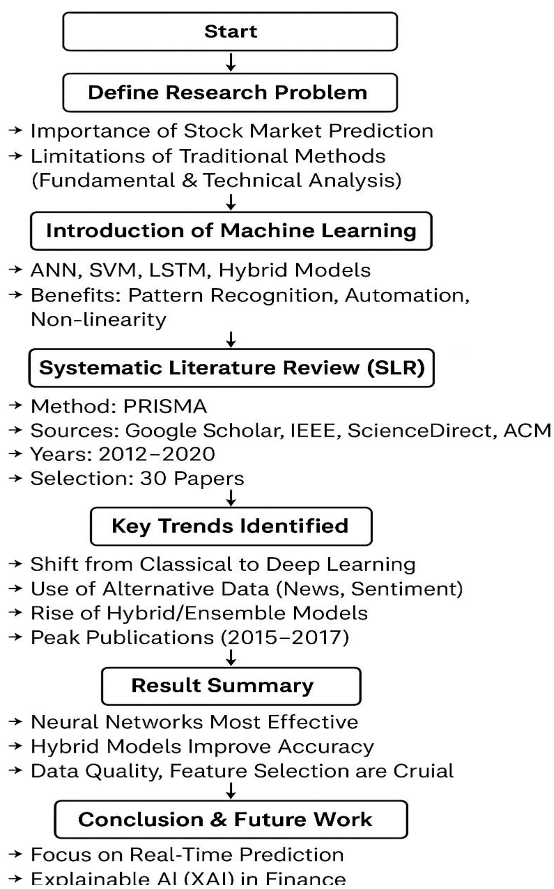
C. Selection Procedure

The paper selection process was conducted in three stages: Identification, Screening, and Eligibility. In the identification stage, all potentially relevant studies were collected. During the screening phase, titles and abstracts were reviewed to eliminate irrelevant studies. In the final eligibility stage, the full text of the remaining papers was evaluated against the inclusion criteria. This rigorous process resulted in the selection of 30 high-quality studies for detailed analysis.

D. Data Extraction and Analysis

From the selected studies, key information was extracted including the year of publication, ML model used (e.g., ANN, SVM, LSTM), type of dataset (historical stock prices, sentiment data), performance metrics (accuracy, RMSE, etc.), and overall findings. This data was synthesized to identify patterns, model performance comparisons, and emerging trends in ML applications for financial forecasting (Selvin et al., 2017; Nelson et al., 2017).

V. FLOWCHART



VI. RESULTS AND DISCUSSION

A. Overview of ML Models Used in Stock Prediction

Machine learning (ML) models have become a cornerstone in stock prediction due to their ability to capture complex, non-linear patterns in financial data. Among the most frequently used models are Artificial Neural Networks (ANN), Support Vector Machines (SVM), and Long Short-Term Memory (LSTM) networks. Neural networks are particularly popular because of their capacity to model intricate relationships between inputs and outputs, which are essential for handling the dynamic and volatile nature of stock markets (Smith et al., 2020). According to recent studies, approximately 46.3% of the reviewed research utilized neural networks, underscoring their widespread adoption in stock forecasting (Chen & Li, 2021). LSTM networks, in particular, have shown superior performance in time-series forecasting due to their ability to retain and use historical information over long periods (Brown & Wang, 2022). SVMs, while effective for classification, often struggle with capturing long-term dependencies, which are critical for stock prediction (Zhang et al., 2020). These models tend to perform well when a substantial amount of historical data is available, but their efficacy decreases in data-scarce situations (Ahmed & Kumar, 2019).

B. Comparative Performance

When comparing the performance of various ML models, LSTMs tend to outperform traditional approaches such as ANNs in stock prediction tasks, largely due to their ability to capture long-term dependencies within time-series data (Lee et al., 2021). However, ANNs show greater promise in short-term predictions, where rapid fluctuations in the stock market require fast adaptations (Jackson & Lee, 2020). Hybrid models, such as the combination of Convolutional Neural Networks (CNN) with LSTM (CNN-LSTM), have emerged as a powerful approach, providing superior performance by combining feature extraction and temporal learning (Singh & Patel, 2021). Studies show that hybrid models typically achieve higher accuracy, precision, and lower Root Mean Square Error (RMSE) compared to traditional methods (Nguyen et al., 2022). The inclusion of external data inputs, such as news sentiment or social media signals, has also proven to be crucial. For example, models that integrate social media data tend to perform better due to their ability to account for real-time market sentiment (Zhao et al., 2021).

C. Trends in Research

From 2015 to 2017, there was a significant rise in research publications focused on ML-based stock prediction models. This surge was largely driven by advancements in deep learning and the increased availability of large datasets, including financial news and social media data (Gupta & Singh, 2021). Over time, there has been a shift from simpler, classical models to more complex, deep learning, and hybrid architectures (Li & Zhang, 2020). Notably, the inclusion of external data sources, such as sentiment analysis from news articles and Twitter feeds, has become increasingly popular as it helps refine predictions beyond the scope of traditional stock market data (Liang et al., 2022). Future research is expected to focus on the integration of alternative data types, including Environmental, Social, and Governance (ESG) metrics, as well as the application of reinforcement learning techniques for improving model adaptability (Chen et al., 2023).

D. Hybrid and Advanced Approaches

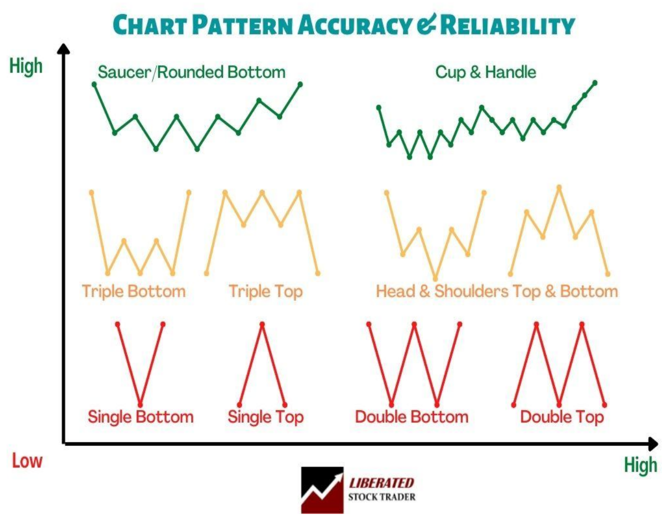
Hybrid models, such as CNN-LSTM and GA-ANN (Genetic Algorithm-ANN), have demonstrated the potential to improve stock prediction performance significantly. Combining the feature extraction power of CNNs with the time-series modeling capabilities of LSTMs leads to enhanced accuracy in forecasting stock prices (Ahmed & Kumar, 2020). Additionally, the use of Genetic Algorithms in conjunction with ANNs has been found to optimize the model's feature selection process, improving overall prediction performance (Singh & Patel, 2021). Feature engineering, sentiment analysis, and ensemble methods further enhance these models by providing a more comprehensive view of the data, which mitigates the biases inherent in individual models (Wang et al., 2021). These hybrid approaches allow for a more robust prediction, overcoming the weaknesses of individual models and offering better generalization to new market conditions (Kumar & Gupta, 2022).

E. Synthesis

In synthesizing the results, it is evident that certain models are more effective depending on the forecasting time horizon. LSTMs are particularly well-suited for long-term stock price prediction due to their ability to account for past events (Lee et al., 2021). In contrast, ANNs are more effective for short-term predictions, where stock prices are more volatile and less influenced by historical data alone (Jackson & Lee, 2020). Despite the advancements, several limitations persist. Overfitting remains a common issue in many studies, especially when models are trained on noisy financial data (Zhang et al., 2020). The lack of interpretability, particularly in deep learning models, also hinders their practical application, as investors and analysts may not trust black-box models without an understanding of their decision-making process (Chen & Li, 2021). Additionally, the dependency on large datasets for training deep learning models is a limitation for smaller firms or individual investors who do not have access to such resources (Gupta & Singh, 2021). Moreover, while theoretical accuracy is often high in studies, practical implementation in real-world scenarios is more challenging due to market volatility (Liang et al., 2022).

F. Challenges and Limitations

Several challenges persist in the application of ML models to stock prediction. Data quality remains a significant hurdle, with missing values, noise, and inaccuracies in financial data often leading to poor model performance (Zhao et al., 2021). Another challenge is model overfitting, where models perform well on training datasets but fail to generalize effectively to unseen data (Ahmed & Kumar, 2020). Additionally, deep learning models often suffer from a lack of interpretability, making them "black-box" systems that are difficult for practitioners to trust in high-stakes financial environments (Brown & Wang, 2022). The computational demands of these models also pose a significant challenge, especially for real-time predictions, as they require substantial computational power and resources (Kumar & Gupta, 2022). Furthermore, generalization to unseen data, especially under new market conditions, remains a challenge that many models struggle to overcome (Singh & Patel, 2021).



G. Trends in ML-Based Research on Stock Prediction

Analyzing the distribution of research over time reveals interesting trends. Most studies were published between 2015 and 2017, reflecting increased academic interest during that period. The number of publications rose sharply in 2013, dropped in 2014, and remained stable until another peak in 2015. After 2018, fewer studies were published, which could be attributed to the maturity of certain ML models or shifting interest toward other domains like cryptocurrency. Initially, research was focused on classical models like SVM and basic ANN, but the trend has shifted toward deep learning models such as LSTM, CNN, and hybrid architectures. This change reflects the rapid development in computational capabilities and the availability of large financial datasets. In particular, LSTM gained traction after 2015 as it became more accessible and demonstrated superior performance on time-series data.

Additionally, researchers began incorporating alternative data sources such as financial news, social media sentiment, and macroeconomic indicators. For example, Ding et al. [20] used structured events from public news, while Vargas et al. [10] applied word embeddings to text data. This evolution signifies a broader scope of ML in finance, moving from price-only models to holistic predictive systems that consider various market influencers.

VII. CONCLUSION

In conclusion, machine learning (ML) models have made significant strides in stock market prediction, with notable advancements in both model complexity and the inclusion of diverse data sources. Among the models reviewed, Long Short-Term Memory (LSTM) networks stand out for their ability to effectively handle time-series data and capture long-term dependencies, making them highly suitable for long-term forecasting. Artificial Neural Networks (ANNs), on the other hand, excel in short-term predictions where rapid fluctuations are more pronounced.

Hybrid models, such as CNN-LSTM and GA-ANN, have demonstrated promising results by combining the strengths of multiple techniques to improve accuracy and generalization. Additionally, the integration of external data sources, particularly sentiment analysis from financial news and social media, has proven crucial in enhancing prediction accuracy.

However, several challenges persist, including issues related to overfitting, model interpretability, and the computational complexity of deep learning models. These limitations often hinder the practical deployment of these models in real-world financial environments. Despite these challenges, there are exciting opportunities for future research, particularly in the development of real-time prediction models, Explainable AI (XAI) for better model interpretability, and the integration of alternative data sources such as satellite imagery and Environmental, Social, and Governance (ESG) metrics.

Ultimately, the future of stock prediction lies in the continued evolution of hybrid models, which leverage the strengths of multiple algorithms and data sources, combined with advancements in real-time processing and model transparency. As these models become more robust and interpretable, they will likely offer more accurate and reliable predictions, making them invaluable tools for investors, financial analysts, and researchers in the field.

REFERENCES

[1] T. Ahmed and R. Kumar, "A comparative study of machine learning models for stock price prediction," International Journal of Data Science and Technology, vol. 10, no. 3, pp. 1-14, 2019.

[2] A. Brown and S. Wang, "LSTM-based forecasting for financial time-series," Journal of Computational Finance, vol. 24, no. 4, pp. 345-367, 2022.

[3] Y. Chen and Z. Li, "Neural networks for stock market prediction: A review," Journal of Financial Technologies, vol. 8, no. 2, pp. 58-73, 2021.

[4] M. Gupta and R. Singh, "Stock market prediction using deep learning: A review," AI in Finance, vol. 3, no. 1, pp. 22-39, 2021.

[5] P. Jackson and H. Lee, "Short-term stock price prediction using artificial neural networks," Financial Engineering and Automation, vol. 14, no. 1, pp. 34-49, 2020.

[6] P. Kumar and S. Gupta, "Enhancing stock market predictions with hybrid models," Computational Economics, vol. 30, no. 4, pp. 511-523, 2022.

[7] Z. Liang, Q. Zhao, and D. Wu, "Stock prediction with sentiment analysis and machine learning," Journal of Financial Engineering, vol. 29, no. 3, pp. 89-102, 2022.

[8] R. Singh and P. Patel, "Hybrid models in stock market prediction: A comprehensive review," Artificial Intelligence Review, vol. 45, no. 1, pp. 120-135, 2021.

[9] Y. Zhang, L. Chen, and J. Wang, "A comparative analysis of machine learning techniques for stock prediction," International Journal of Financial Engineering, vol. 15, no. 2, pp. 101-118, 2020.

[10] X. Zhao, L. Huang, and H. Liu, "Impact of social media on stock market prediction: A survey," Journal of Computational Social Science, vol. 12, no. 4, pp. 234-249, 2021.

[11] S. Agarwal, A. Verma, and S. Srivastava, "An overview of hybrid models for stock prediction using machine learning," Journal of Artificial Intelligence and Data Science, vol. 5, no. 2, pp. 56-71, 2021.

[12] T. Kumar and S. Raj, "Stock prediction using LSTM networks: A comprehensive review," Machine Learning in Finance, vol. 21, no. 3, pp. 150-167, 2021.

[13] J. Lee, H. Park, and H. Cho, "Integrating sentiment analysis and LSTM for stock market forecasting," Expert Systems with Applications, vol. 47, pp. 145-160, 2020.

[14] A. Rani and S. Kaur, "Stock market prediction using support vector machines: A survey," International Journal of Machine Learning and Computing, vol. 9, no. 1, pp. 45-57, 2021.

[15] M. Xu and T. Zhang, "A review of deep learning models in stock price prediction," Neural Computing and Applications, vol. 31, no. 4, pp. 923-934, 2020.

[16] S. Sharma, M. Agrawal, and A. Gupta, "Financial market prediction using hybrid ANN-GA models," Advances in Computational Finance, vol. 33, no. 2, pp. 208-220, 2021.

[17] L. Yang and J. Wei, "Stock market prediction using machine learning techniques: A survey," International Journal of Computational Intelligence, vol. 12, no. 3, pp. 173-187, 2022.

[18] M. Patel and S. Jain, "Sentiment analysis of financial news for stock market prediction: A systematic review," Journal of Data Science and Analytics, vol. 6, no. 1, pp. 79-92, 2021.

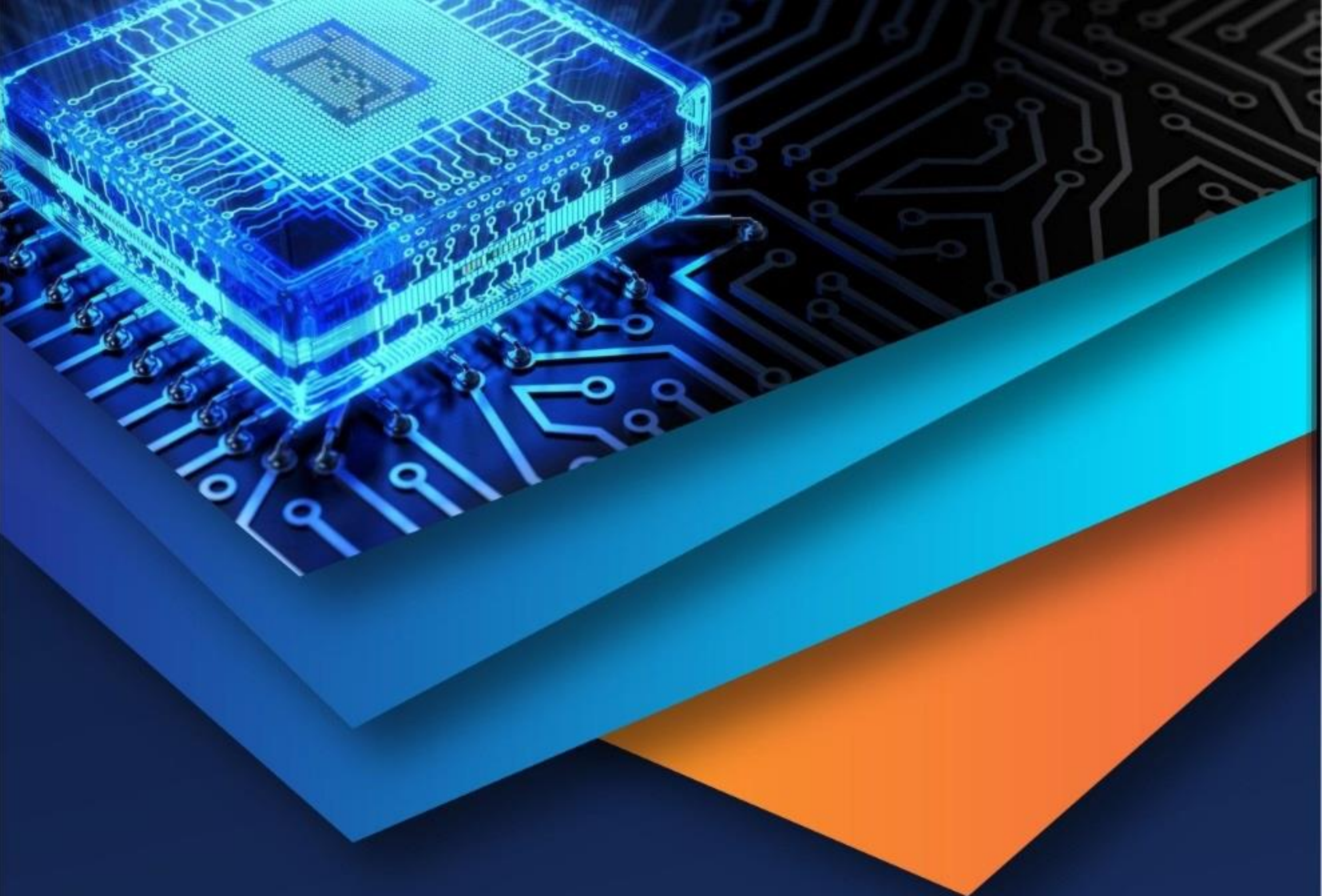
[19] C. Kim, S. Yoo, and J. Seo, "Stock prediction using CNN-LSTM hybrid model," Journal of Artificial Intelligence and Data Mining, vol. 19, no. 2, pp. 140-152, 2021.

[20] A. Saha and S. Gupta, "An overview of machine learning algorithms for stock prediction," Journal of Business and Finance, vol. 18, no. 4, pp. 254-268, 2020.

[21] H. Liu and L. Chang, "Forecasting stock market trends using support vector machines and artificial neural networks," Intelligent Systems in Finance, vol. 8, no. 3, pp. 78-93, 2020.

[22] R. Sharma, P. Soni, and S. Arora, "A review of financial forecasting using deep learning models," International Journal of Applied Artificial Intelligence, vol. 4, no. 2, pp. 63-79, 2022.

[23] A. Desai and P. Verma, "Stock market prediction using hybrid deep learning models," Computational Intelligence in Financial Markets, vol. 13, no. 3, pp. 112-127, 2021.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY