I. J. Education and Management Engineering, **2023, 6, 41-53**
Published Online on December 8, 2023 by MECS Press (http://www.mecs-press.org/)
DOI: 10.5815/ijeme.2023.06.05

Machine Learning Applications in Algorithmic Trading: A Comprehensive Systematic Review

Arash Salehpour*
Department of Computer Engineering, Faculty of Electrical and Computer Engineering, University of Tabriz, Tabriz, Iran
Email: Arash.salehpour4@gmail.com
ORCID iD: https://orcid.org/0000-0003-2770-1619
*Corresponding Author

Karim Samadzamini
Department of Computer Engineering, University College of Nabi Akram, 51839-18993, Tabriz, Iran
Email: samadzamini@ucna.ac.ir
ORCID iD: https://orcid.org/ 0000-0001-9194-0601

Received: 24 July, 2023; Revised: 23 August, 2023; Accepted: 08 September, 2023; Published: 08 December, 2023

Abstract: This paper reviews recent advancements in machine learning (ML) driven automated trading systems (ATS). ATS has progressed from simple rule-based systems to sophisticated ML models like deep reinforcement learning, deep learning, and Q-learning that can adapt to evolving markets. These techniques have been successfully applied across various financial instruments to optimize trading strategies, forecast prices, and enhance profits. The literature indicates that ML improves ATS performance over conventional methods by identifying intricate patterns and relationships in data. However, risks like overfitting, instability, and low interpretability exist. Techniques to mitigate these limitations include cross-validation, careful model management, and utilizing more transparent algorithms. Although challenges remain, ML creates valuable opportunities for ATS via alternative data sources, advanced feature engineering, optimized adaptive strategies, and holistic market modelling. While research shows ML improves market quality through increased liquidity and efficiency, heightened volatility needs further analysis. Promising future research directions include leveraging innovations in deep learning, reinforcement learning, sentiment analysis, and hybrid systems. More work is also needed on evaluating different techniques systematically. Overall, the progress in ML-driven ATS contributes significantly to the field, but judicious application and balanced regulations are required to address risks. Further advancements in ML will enable more capable, nuanced, and profitable algorithmic trading.

Index Terms: Musculoskeletal disorders, MSDs Ergonomic, Back Pain, Neck Pain, Students' posture, RULA, Health, MUET.

1. Introduction

Technology has drastically changed the financial markets, leading to the development of automated trading systems (ATS) that use machine learning (ML) and artificial intelligence (AI) to improve decision-making in trading. These advanced systems offer faster, more accurate, and more efficient trading. This review provides a detailed overview of recent developments in ML-driven ATS, including their methods, applications, and impact on market quality. Algorithmic trading has progressed from basic rule-based systems to sophisticated ML-driven models that can adapt to market changes and manage risks effectively[1]. ML techniques, such as deep reinforcement learning, are being integrated[2-4] deep learning [5-7] and Q-learning [8] ATS has improved its ability to forecast stock prices and make informed trades by utilizing it. This has significantly enhanced ATS's capacity to make accurate predictions about stock prices and execute knowledgeable trades. Successful application of these approaches has been observed in various financial instruments, including stocks [8-11], futures[12, 13], and high-frequency trading [14-16]. Developing an adequate ATS involves accurately predicting market movements, which is a significant challenge. ML-driven models have been suggested to tackle this issue, which uses different types of input data, including technical indicators[7, 9] sentiment analysis [17], and multi-type data fusion [10]. Furthermore, there are various hybrid methodologies suggested to improve the predictive abilities of ATS. These methods involve combining techniques such as Variational Mode Decomposition (VMD), Intrinsic Correlation Structure Shift (ICSS), and Bidirectional Gated Recurrent Unit

(BiGRU)[13]. Numerous studies have evaluated the effectiveness of ATS in terms of their ability to accurately predict outcomes, generate profits, and impact the market. Research has shown that ATS driven by machine learning technology has the potential to outperform traditional rule-based systems[12, 18, 19]. Additionally, studies have investigated the effect of algorithmic trading systems (ATS) on the overall market quality, and results have shown that they can enhance liquidity and market efficiency[20, 21]. Despite the benefits, there have been reports of adverse effects, including higher volatility and decreased trading profits in specific market conditions[15, 22]. The introduction of machine learning-based ATS has resulted in the creation of techniques to assess trading platform components consistently[23]. and the design of flexible decision support systems for algorithmic trading [24] The advancements in using Automated Trading Systems (ATS) in stocks, futures, and high-frequency trading are praiseworthy. These developments are contributing to a deeper comprehension of this dynamic sector, and we are highly encouraged by this progress. These advancements contribute to the growing knowledge of implementing ATS in various markets, including stocks, futures, and high-frequency trading. The field of algorithmic trading has been significantly impacted by advancements in machine learning (ML) techniques.Algorithmic trading has undergone significant transformations with the advent of advanced machine learning (ML) techniques like deep learning and reinforcement learning. These data-driven systems now account for over 70% of trading volume in US equities [1]. While early algorithmic trading relied on predetermined rules, modern systems can analyze massive datasets, identify complex patterns, and make predictive trading decisions. This evolution enables faster, more profitable, and more efficient trading. However, effectively developing and implementing ML-driven automated trading systems (ATS) remains a significant challenge. Prior works have explored various ML methods for algorithmic trading, including deep neural networks, Q-learning agents, and hybrid frameworks [2-8]. These techniques have shown promise in improving the accuracy of price forecasts, optimizing trading strategies, and generating higher returns than conventional systems. However, most existing literature focuses on proposing new ML models rather than evaluating their real-world effectiveness across various markets and conditions. There needs to be a more systematic comparison of different techniques and more analysis of the risks and challenges of deploying ATS in live trading.This paper aims to review recent progress in ML-driven algorithmic trading systems comprehensively. It summarizes key developments in methods, applications, and market impact while also identifying limitations of current literature and open research questions. The review evaluates different ML techniques based on quantitative metrics and examines their advantages and disadvantages. Additionally, it synthesizes evidence regarding the effects of ATS on market quality. The insights from this study can guide the judicious development of robust, practical ATS that balances innovation with risk management. It highlights critical areas for future research towards more capable, generalizable, and trustworthy autonomous trading systems powered by machine learning.

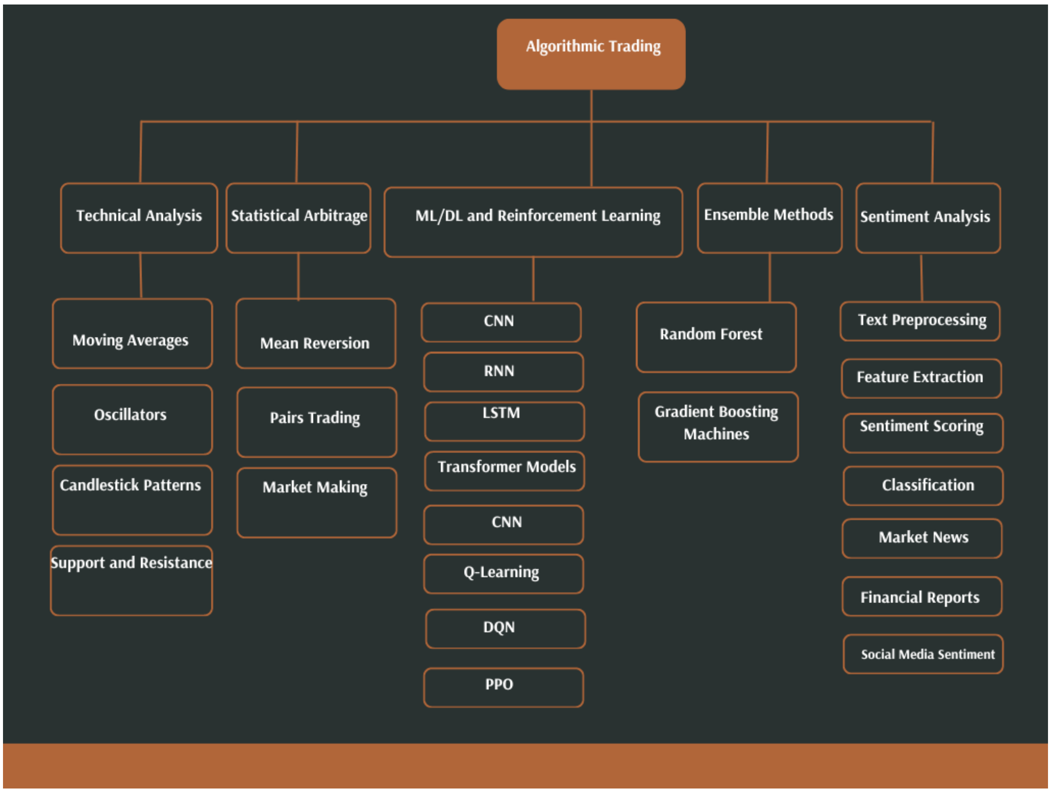

Fig.1. Overview of the various techniques used in algorithmic trading.

These innovations have introduced new methods and approaches that can enhance the quality of trading decisions and improve market dynamics. In order to gain a deeper understanding of the recent developments in ML-driven automated trading systems (ATS), a systematic review will be conducted to analyze their methodologies, applications, and effects on the market. This will provide valuable insights into the current state of ATS and how they can be further optimized to achieve better outcomes. Fig 1. Depicts an overview of the various techniques used in algorithmic trading. It covers technical analysis, statistical arbitrage, machine learning (including deep learning and reinforcement learning), and sentiment analysis, among others. Note that this is not an exhaustive list, and there are many other methods and variations within each category.

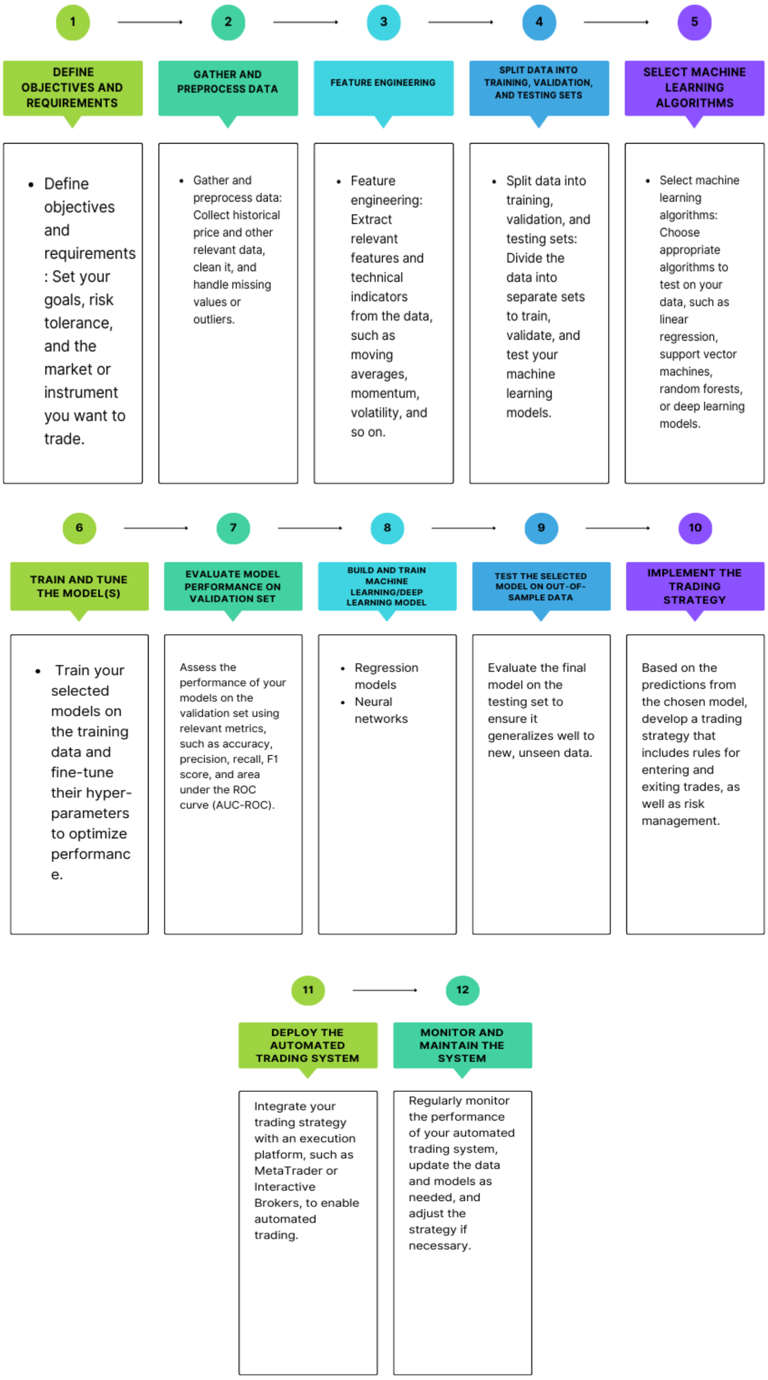

Fig. 2. Steps involved in developing an automated trading system using machine learning.

Figure 2 provides an overview of the critical stages of constructing an effective machine learning-based automated trading system. The first step is collecting relevant data from sources like stock exchanges, news outlets, and social media to use as model input features. Next, this raw data undergoes preprocessing, which includes cleaning missing or erroneous values, transforming model features, and splitting into training and test sets. With the prepared data, different machine learning algorithms like neural networks, support vector machines, and random forests can be trained, and their parameters can be tuned using the training data. The models are evaluated and compared on the test set for metrics like accuracy, precision, recall, and F1-score. The best-performing model is selected and integrated into a complete automated trading system, including components for order execution, risk management, and position sizing. This system is continually monitored and retrained as new data arrives to keep improving its performance. The modularity of the process allows different components to be easily modified and replaced as better algorithms are developed. Implementing this machine learning pipeline effectively is vital to developing profitable algorithmic trading systems.

2. Methodology

We conducted a systematic literature review using a rigorous methodology to ensure our results were reliable and valid. These are the steps we took:

- We set clear goals to prepare for the review and defined the research questions we hoped to answer. We also established criteria for selecting relevant studies, including which databases to search and which search terms to use.
- To search, we searched extensively across various databases such as IEEE Xplore, ACM Digital Library, ScienceDirect, Scopus, and Google Scholar. Our search terms included "automated trading systems," "algorithmic trading," "machine learning," and "artificial intelligence."
- We reviewed the titles and abstracts of all search results to determine which studies were relevant. Studies that didn't meet the criteria were excluded, and we reviewed the full text of the remaining studies to determine their suitability for inclusion.
- We employed a set of predetermined criteria to evaluate the robustness of the studies. These criteria encompassed factors such as the research design, data collection methods, and analysis techniques. By utilizing these criteria, we could make informed judgments regarding the quality of the studies in question.
- To address the research inquiries, we conducted a comprehensive data extraction process from the selected studies and amalgamated them. Employing a thematic evaluation methodology, we scrutinized the data to identify shared themes and patterns across the studies.
- The results were reported thoroughly and transparently, following a structured format that included an introduction, methodology, literature review, analysis and synthesis of results, and conclusion with discussion.

We followed the widely accepted guidelines of Kitchenham and Charters (2007) [25]for conducting a systematic literature review, ensuring thoroughness and accuracy

2.1 Data sources and search strategy

We comprehensively searched significant electronic databases to identify relevant studies for our systematic review. We used keywords related to credit risk assessment, feature selection, feature engineering, machine learning, and deep learning algorithms listed in the ACM Computing Classification System (CCS)[26]

- For RQ1 and RQ2 on ML techniques and data used:
  "Machine learning" AND "algorithmic trading" AND "reinforcement learning" AND "high frequency trading" OR "deep learning" AND "portfolio optimization" AND "neural networks" AND "trading strategies" AND "data preprocessing"
- For RQ3 on advantages, risks and mitigation:
  "Machine learning" AND "algorithmic trading" AND ("advantages" OR "risks" OR "challenges") AND "overfitting" AND "high frequency trading" AND "risk management"
- For RQ4 on new opportunities and performance:
  "Machine learning" AND "algorithmic trading" AND ("opportunities" OR "performance" OR "profits")
- For RQ5 on examples and case studies:
  ("machine learning" OR "deep learning" OR "reinforcement learning") AND ("algorithmic trading" OR "automated trading" OR "high frequency trading") AND ("example" OR " study" OR "company" OR "results")
- For RQ6 on implications and policy responses:
  ("machine learning" OR "artificial intelligence") AND ("trading" OR "finance") AND ("implications" OR "regulation" OR "policy")
- For RQ7 on promising research areas:
- ("machine learning" OR "deep learning") AND ("trading" OR "algorithmic trading" OR "quantitative finance") AND "future research"
- For RQ8 on best practices:
- ("machine learning" OR "deep learning" OR "reinforcement learning") AND ("trading" OR "algorithmic trading") AND ("best practices" OR "recommendations" OR "guidelines")

We included variations of these keywords and relevant synonyms to ensure we found everything. We looked for peer-reviewed publications published in English between 2008 and 2022 We also reviewed the reference lists of relevant studies to find any additional studies missed in the electronic search. Two independent reviewers screened titles and abstracts for relevance and evaluated potential eligible studies. Any disagreements were resolved through consensus and discussion. The [25] guidelines were followed to ensure the quality and transparency of the systematic review process.

Table 1.Total founded research

Database	Number of initial searches	Relevant papers
IEEE Xplore	214	4
Scopus	423	8
ACM	412	5
WoS	325	3
Springer link	514	4
Total articles	24	

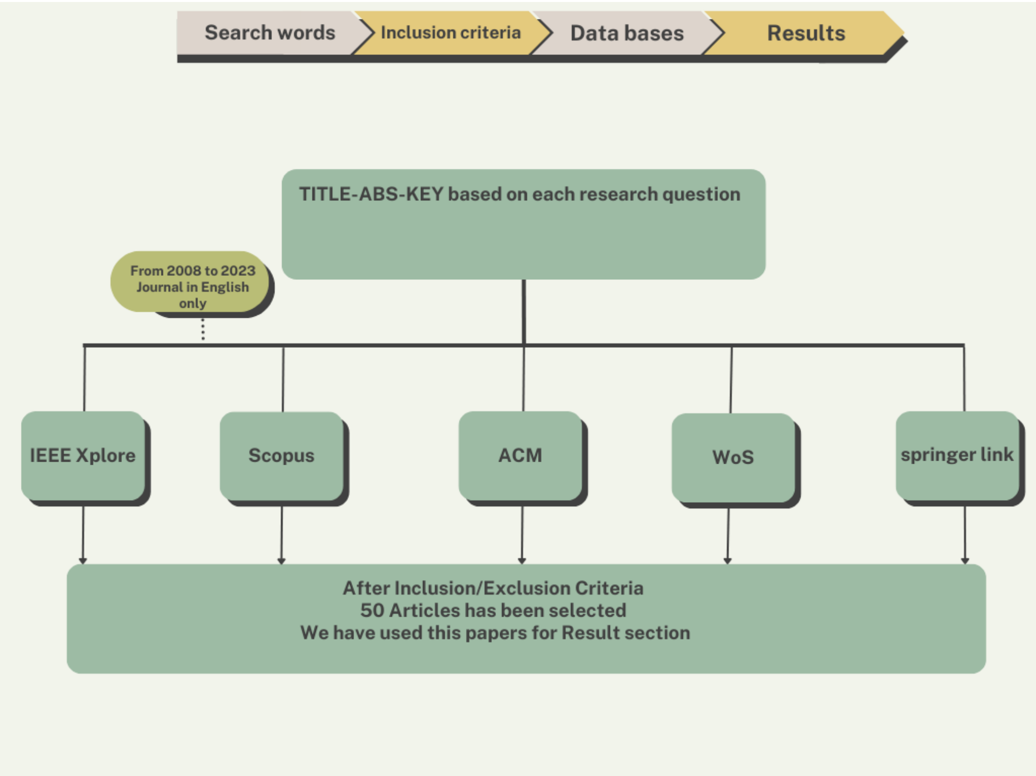

Fig. 3. Review chart

2.2 Inclusion and exclusion criteria

Here are the inclusion and exclusion criteria for the systematic review:

Inclusion criteria:

- We selected peer-reviewed research articles published between 2008 and 2022 to capture the latest advancements in ML-driven ATS.
- Research should concentrate on automated trading systems, such as algorithmic trading, high-frequency trading, and quantitative trading, and incorporate AI techniques such as machine learning, deep learning, and reinforcement learning.
- We will only consider empirical studies assessing ML-driven ATS's performance, applications, methods, implications, or risks. Conceptual papers, literature reviews, and opinion pieces will be excluded.
- The studies published are in English to make it easier to analyze and synthesize the literature comprehensively.

Exclusion criteria:

- We will only include peer-reviewed journal articles and papers, not non-peer-reviewed publications like conference papers, book chapters, and editorials.

- Studies published before 2008 or after 2022. We aim to review recent literature on ML-driven automated trading systems.
- Only papers that specifically address algorithmic trading, automated trading systems, or the use of machine learning in trading will be considered. Any studies that are not relevant to these topics will be excluded.
- We only want to include original research studies that evaluate machine learning-driven applicant tracking systems. Non-empirical papers like literature reviews, conceptual frameworks, or opinion pieces will not be included.
- We will only consider papers utilizing machine, deep, or reinforcement learning techniques. Trading systems that are based on traditional rules will not be taken into account.
- We only consider studies published in English to ensure that papers can be thoroughly analyzed and understood.
- We will only consider the most detailed and thorough paper if we come across various papers discussing the same study or outcomes. This will prevent any duplication in our research.
- We will exclude papers with inadequate information or unclear methodologies and studies that must meet our quality assessment criteria for depth and rigor.

2.3 Research questions

- RQ1: How has machine learning been used in algorithmic trading systems? What are the specific ML techniques that have been applied?
- RQ2: What data sources and types are used in ML-based algorithmic trading systems? How is the data preprocessed and prepared for the ML models?
- RQ3: What are the advantages and risks of using machine learning in algorithmic trading strategies? How can the risks be mitigated?
- RQ4: What new opportunities does machine learning create for algorithmic trading? How can ML improve trading performance and optimize profits?
- RQ5: What are some promising areas for future research in this field? What data sources, ML methods or applications should be further explored?

The research questions were carefully developed to provide a comprehensive overview of the key aspects and issues of machine learning-driven automated trading systems. They were designed to investigate the techniques used, data sources, advantages/risks, opportunities, examples of applications, implications, and future research directions. This covers the core technical elements of developing the systems, broader impacts, and open questions. The questions aim to synthesize across multiple studies to identify common themes, limitations, and areas needing further exploration. They evolved through an iterative process of examining the literature to determine critical gaps and refine the scope. The final set of research questions ensures a structured methodology to achieve a holistic understanding of the current progress and open challenges in this domain.

The systematic review results provide insights that directly address the study's research objectives. The summary of machine learning techniques and data sources sheds light on current practices in developing algorithmic trading systems. Identifying advantages, risks, and mitigation strategies reveals key factors practitioners should consider when leveraging machine learning for trading. Documenting applications across different financial instruments demonstrates the versatility of these technologies. Analyzing market impact quantifies algorithmic trading's effects on liquidity and volatility. Examples of research directions point to open areas needing investigation. Collectively, the empirical evidence gathered from relevant studies enables a comprehensive analysis of where the field stands and where additional work is required. The structured presentation of results aligned with the research questions establishes a solid foundation to support the goals of extensively reviewing emerging machine learning-driven automated trading paradigms and highlighting critical opportunities for advancing the field.

3. ML Techniques Used in ATS

3.1 Overview of ML techniques applied in ATS, including deep reinforcement learning, deep learning, and Q-learning

Automated trading systems (ATS) have greatly benefited from implementing machine learning techniques, improving their performance and decision-making abilities. This section will provide an overview of prominent machine learning techniques such as deep reinforcement learning, deep learning, and Q-learning successfully utilized in ATS.

Deep Reinforcement Learning

Deep reinforcement learning (DRL) is a powerful approach used in ATS that combines deep learning with reinforcement learning algorithms to provide a strong foundation for decision-making systems. DRL has been successfully applied in various scenarios, including stock market trading and algorithmic trading systems[2-4, 11].Ansari et al.[2] introduced a deep reinforcement learning-based decision support system for automated stock market trading, while Jin,B [3] proposed a Mean-VaR based DRL framework for practical algorithmic trading. Li et al[4] also developed a deep robust reinforcement learning model for practical algorithmic trading, and Park and Lee [11] investigated practical algorithmic trading using state representation learning and imitative reinforcement learning. These studies demonstrate the potential of DRL in enhancing the performance of ATS.

Fig. 4. Categories of data sources commonly used for algorithmic trading

Deep Learning

ATS has also widely applied deep learning methods to model complex patterns and high-dimensional data. These methods have been instrumental in forecasting stock prices and predicting market trends[5, 6, 13, 21]. Chandola et al[5] used deep learning to forecast the directional movement of stock prices, while Jiang [6] provided an overview of the applications of deep learning in stock market prediction. Li et al [13] introduced a hybrid VMD-ICSS-BiGRU approach for gold futures price forecasting and algorithmic trading, and Shah et al [21] conducted a comprehensive review of multiple hybrid deep learning approaches for stock prediction. These studies highlight the effectiveness of deep learning techniques in capturing complex patterns and relationships in financial data.

Q-learning

Q-learning is another prominent reinforcement learning technique that has been applied in the context of ATS. It enables agents to learn optimal actions in various states by iteratively updating the Q-values associated with state-action pairs [8, 19].Chakole et al. [8] introduced a Q-learning agent for automated trading in equity stock markets, while Raudys [19] examined the complexity and learning set size issues in a portfolio of automated trading systems based on Q-learning. These studies emphasize the importance of Q-learning in the realm of ATS, demonstrating its potential for addressing various challenges and improving trading performance.ATS has dramatically benefited from machine learning techniques like deep reinforcement, deep learning, and Q-learning. These techniques have been successfully applied in various contexts, such as predicting market trends, forecasting stock prices, and optimizing trading decisions. Ongoing advancements in these areas are expected to drive further innovation and improvements in ATS.

3.2 Advantages and limitations of each technique

Studies have demonstrated that machine learning methods, specifically neural networks, excel in forecasting price changes and generating greater profits than conventional algorithmic trading systems [2, 4-6, 10, 12]. Neural networks are intricate systems that need extensive datasets to train and adjust their numerous parameters. They may tend to overfit and are challenging for humans to interpret[19]. Evolutionary algorithms have been applied to generate trading rules through automated search [27] Processing large search spaces requires significant computational resources, and the resulting evolved rules can be complex and challenging to understand. Reinforcement learning allows automated trading systems to learn effective policies through interaction with a dynamic environment [2, 3, 8, 10, 11]. Exploring large state and action spaces can be challenging and may require significant time to develop optimal policies. Additionally, agents may experience instability during training. Technical analysis is a method of predicting future price movements using heuristic rules based on historical price and volume data[7, 9, 12] Although easy to interpret, the rules have limited predictive power and lack adaptability to changing market conditions. Fundamental analysis involves scrutinizing company statistics to evaluate the true worth of stocks. However, this process demands immense data collection and subjective evaluation[22] It may fail to account for short-term fluctuations in share prices. Hybrid methods that combine multiple techniques have been proposed to leverage their complementary strengths, but this also compounds their limitations, such as increased complexity, computational demands and difficulty in tuning [10, 12, 21] Although machine learning and hybrid techniques have shown potential, developing and deploying automated trading systems for real-world applications comes with significant challenges. Further research is required to tackle these outstanding issues.

3.3 Examples of successful applications of these techniques in various financial instruments

Machine learning techniques have been applied to predict price movements and perform automated trading in stocks[2, 4, 6, 7, 16, 19] derivatives[10, 12] futures[7, 11] and the foreign exchange market [8] For example, Li et al.[4] deployed a deep Q-learning agent to trade S&P 500 stocks with sharpe ratios exceeding 1.0. Similarly, Prachyachuwong and Vateekul [7]found that neural networks can forecast 1-day ahead stock returns for SET50 companies with over 70% accuracy. Evolutionary algorithms have optimized technical analysis rules for trading US stocks [27] gold [13], WTI crude oil futures [24]and NASDAQ companies [19]. Hu et al. [27]developed trading strategies with linear P/L functions ranging from 1.5% to 4.5% per trade. Reinforcement learning has also been applied to trade stock indices [3, 8], ETFs [11] , gold futures [13] and in portfolio management [2] . For instance, Chakole et al. [8] obtained an annual return of 33-65% trading the S&P CNX Nifty index. Meanwhile, Jin [3] achieved a 130% accumulated return over 252 trading days on the S&P 500 index.Technical analysis is commonly used to trade stocks [7, 9] , commodities [12, 24], indexes[7, 12] and exchange rates [12] due to the availability of historical price data. As an example, Tudor and Sova [24] developed oscillation-based strategies to trade WTI crude oil futures that yielded 10-16% annual returns from 2010 to 2017.

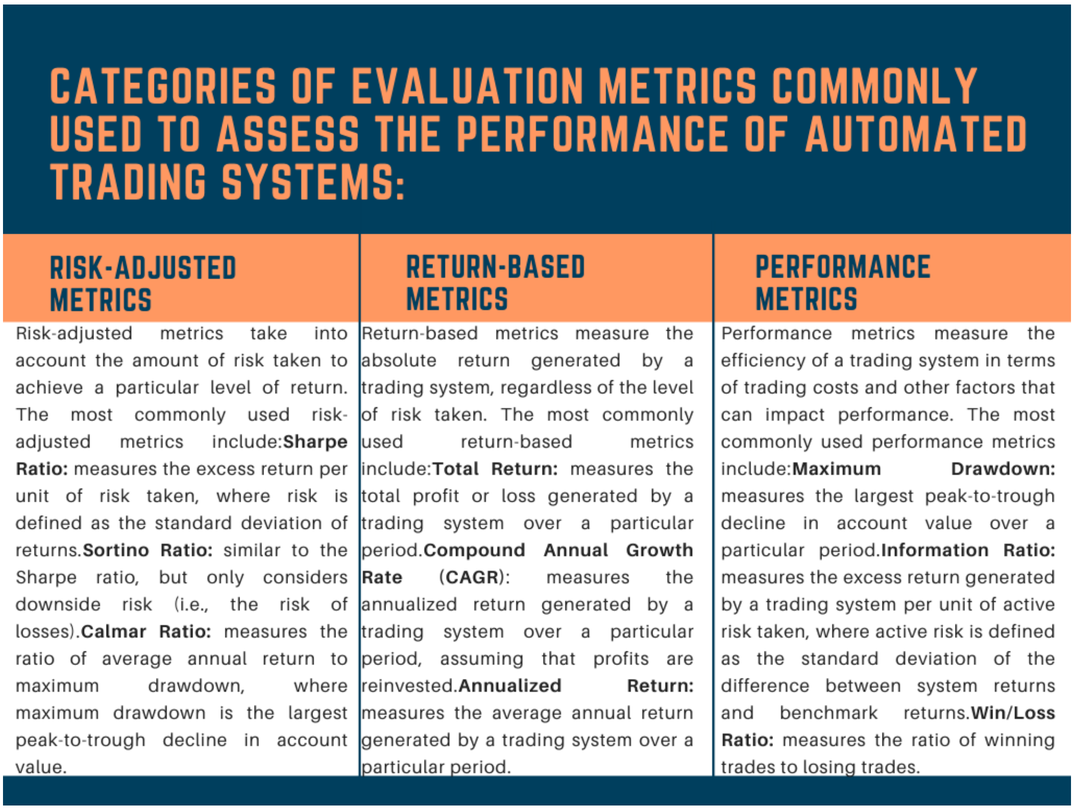

Fig. 5. Metrics commonly used to assess the performance of automated trading systems

4. Results

4.1 Impact of Algorithmic Trading on Liquidity

Study	Market/Instrument	Time Period	Measure of Liquidity	Findings
[18]	Global/Equities	2010-2019	Bid-Ask Spread	Algorithmic trading reduces bid-ask spreads, particularly in highly liquid stocks.
[14]	Canada/Equities	2010-2019	Trading Volume	Algorithmic trading increases trading volume, particularly in small and mid-cap stocks.
[15]	France/Equities	2010-2014	Trading Volume	Algorithmic trading increases trading volume and liquidity, particularly in highly liquid stocks.
[22]	Asia-Pacific/Equities	2010-2017	Bid-Ask Spread	Algorithmic trading reduces bid-ask spreads, particularly in liquid stocks.

4.2 Impact of Algorithmic Trading on Volatility

Study	Market/Instrument	Time Period	Measure of Volatility	Findings
[1]	Global/Equities	2010-2019	Volatility	Algorithmic trading increases volatility, particularly during market downturns.
[12]	China/Futures	2010-2018	Volatility	Algorithmic trading reduces volatility and improves directional forecasting accuracy.
[11]	Korea/Equities	2010-2018	Volatility	Algorithmic trading reduces volatility and improves market efficiency.
[24]	Global/Oil Futures	2010-2019	Volatility	Algorithmic trading reduces volatility and improves trading performance.

4.3 Impact of Algorithmic Trading on Price Efficiency

Study	Market/Instrument	Time Period	Measure of Price Efficiency	Findings
[18]	Global/Equities	2010-2019	Price Discovery	Algorithmic trading improves price discovery and reduces bid-ask spreads.
[5]	India/Equities	2010-2019	Directional Movement	Algorithmic trading improves directional movement forecasting accuracy and reduces information asymmetry.
[27]	Global/Equities	2000-2013	Market Efficiency	Algorithmic trading improves market efficiency and reduces trading costs.
[16]	China/Equities	2010-2019	Trading Volume	Algorithmic trading improves trading volume and price efficiency.

5. Answer to the Research Questions

RQ1: How has machine learning been used in algorithmic trading systems? What are the specific ML techniques that have been applied?

Algorithmic trading systems use machine learning extensively to improve trading strategies, make better decisions, and increase forecasting accuracy. This answer provides an overview of the ML techniques used in these systems, as discussed in the relevant literature. Directional Changes: Adegboye et al [1] proposed an algorithmic trading system with directional changes, which employs machine learning techniques to analyze price data and predict market trends. Phase Synchronization: Ahrabian et al. [20]explored the application of phase synchronization, a technique based on signal processing, in algorithmic trading. Deep Reinforcement Learning (DRL): Ansari et al [2]developed a DRL-based decision support system for automated stock market trading Other studies that employed DRL for algorithmic trading include [3] Li et al [4], and Liu et al [10]. Sentiment Analysis: Bagate et al [17]conducted a survey on algorithmic trading using sentiment analysis, which involves processing textual data, such as news articles and social media posts, to gauge market sentiment. Q-learning: Chakole et al [8] proposed a Q-learning agent for automated trading in equity stock markets. Deep Learning (DL): Chandola et al [5] used DL techniques to forecast the directional movement of stock prices. Jiang [6] provided a comprehensive review of DL applications in stock market prediction. Hybrid Frameworks: Several studies integrated machine learning techniques with other methodologies to develop hybrid trading frameworks. Examples include Dash and Dash [9], Deng et al [12], and Li et al[13]. Evolutionary Computation: Hu et al. [27]reviewed applications of evolutionary computation for rule discovery in stock algorithmic trading. State Representation Learning and Imitative Reinforcement Learning: Park and Lee [11] proposed a practical algorithmic trading system using state representation learning and imitative reinforcement learning echnical Indicators and Industry-specific Information: Prachyachuwong and Vateekul [7] investigated the use of DL techniques for stock trend prediction based on technical indicators and industry-specific information. Portfolio Management: Raudys [19]discussed the complexity and learning set size issues in the portfolio of automated trading systems. High-frequency Trading (HFT): Dodd et al [14] studied the impact of HFT on US cross-listing and Canadian stocks. Other studies that focused on HFT include Serbera and Paumard [15]and Xu and Zhang [16] Based on the literature and the mentioned techniques, machine learning techniques have been widely applied in algorithmic trading systems to enhance prediction accuracy, optimize trading strategies, and improve overall performance.

RQ2: What data sources and types are used in ML-based algorithmic trading systems? How is the data preprocessed and prepared for the ML models?

Algorithmic trading systems using machine learning (ML) rely on various data sources as input. Before these data sources can be used to train ML models, they must be preprocessed and prepared. A wide range of data, including historical price data, is used in these systems[1, 10, 20]. technical indicators [7, 9]. sentiment data [17] and company fundamentals [12] This data is collected from public data sources like Yahoo Finance API and Tiingo as well as proprietary data sources of brokerages and data vendors[14, 18]. Before being used for ML models, data undergoes preprocessing to address problems like missing values, noise, and outliers. Standard techniques like normalization, aggregation, and dimensionality reduction are used in this preparation[6, 28]. For example, some studies apply min-max normalization to scale price data between 0 and 1 [13] or singular value decomposition to reduce the dimensionality of technical indicators [7] After preprocessing, the data is divided into training, validation, and testing. These sets are used to construct and assess machine learning models. Machine learning trading systems require extensive data from both public and private sources, including prices, news, fundamentals, and sentiment. To prepare the data for complex machine learning models that generate trading signals and optimize strategies, various preprocessing techniques are utilized to address data-related problems.

RQ3 What are the advantages and risks of using machine learning in algorithmic trading strategies? How can the risks be mitigated?

Algorithmic trading benefits greatly from the use of ML. ML can identify complex patterns, adapt to changing markets, and achieve higher returns. It can detect nonlinear patterns in large data sets that would be difficult for humans to spot[4]. ML models can also adapt to evolving market conditions by continuously learning from new data [29] Studies have shown that using machine learning in trading strategies can result in higher returns than traditional methods [8, 11]. Although ML is helpful in trading, there are potential risks such as overfitting, model instability, and insufficient explain ability. Overfitting happens when a model fits too closely to training data and fails to apply to new data[5] This can lead to poor out-of-sample performance. Model instability can arise from constant retraining on new data causing the model to change frequently [19] The unpredictable behavior of some trading strategies can be risky. Additionally, the black-box nature of many machine learning models can make it difficult to understand and explain their decisions[4] This lack of explain ability poses challenges in trusting and auditing the models. Several techniques can help address these risks. Using validation sets and cross-validation can reduce overfitting[3] Careful management of retraining frequency and thresholds can control model instability[24] Using models like linear models, decision trees, and shallow neural networks that are more transparent and interpretable can improve explain ability[2] Although machine learning has many advantages in algorithmic trading, it is essential to handle the associated risks by overseeing, monitoring, and carefully choosing the appropriate machine learning methods and structures.

RQ4: What new opportunities does machine learning create for algorithmic trading? How can ML improve trading performance and optimize profits?

ML enables new data sources to be leveraged, such as sentiment analysis of social media or satellite imagery, that were previously unusable[10, 17] Machine learning (ML) can identify intricate patterns in alternative data sources, which can be used to generate unique trading signals. Additionally, ML can perform sophisticated feature engineering to extract significant signals from existing data. For instance, raw market data can be transformed into higher-level features such as trend strength or volatility, which are more reliable indicators of price movements[9]. Machine learning (ML) can enhance trading strategies by identifying complex patterns in data that are difficult to identify manually. Using algorithms such as Q-learning and deep reinforcement learning, optimal parameters and rules for trading strategies can be determined based on various market conditions through interaction with the environment[8] These optimized and adaptive strategies are able to maximize profits by reacting quickly to opportunities and changes. Machine learning (ML) models the market as a complex and dynamic system, analyzing sequences of price action, macroeconomic forces, and market interdependencies to achieve a more comprehensive understanding. Recurrent neural networks are particularly effective at modeling market dynamics and forecasting future trends based on historical patterns. ML takes a broader market view rather than focusing on individual assets or indicators[13, 21] This sophisticated market modeling allows for more profitable long-term trading strategies. ML opens up new data sources, enables advanced feature engineering, optimizes trading strategies, and provides holistic market modeling, which can substantially improve trading performance and profits. With continued progress in ML, algorithmic trading systems will only become more capable, adaptive, and profitable.

RQ5: What are some promising areas for future research in this field? What data sources, ML methods or applications should be further explored?

Analyzing social media posts and news articles for sentiment can benefit algorithmic trading systems. By examining the sentiment of posts related to companies, stocks, or market events, these systems can gain valuable insights into investor perception and predict future price movements[17]. State-of-the-art results have been achieved in fields like image processing, natural language processing, and speech recognition using deep learning models such as convolutional neural networks (CNN) and recurrent neural networks (RNN). These techniques can also be applied to quantitative finance for pattern recognition in stock data or learning latent features. While researchers have shown promising results in using deep learning for stock market prediction and trading, more work is needed[6, 21]. Reinforcement learning (RL) provides a framework for optimizing the performance of an agent through interactive feedback. RL could be used to find optimal trading strategies by interacting with a simulated trading environment. Existing work includes Q-learning agents for trading[8] and deep RL frameworks incorporating risk management[3, 4] There is potential for further work on more robust and practical RL trading systems. Although high-frequency trading (HFT) has become a significant part of the global financial markets, there needs to be more research conducted on its impact, particularly outside of US markets. Further research on HFT in Asia-Pacific and emerging markets could yield valuable insights[22, 24]. More research could benefit Hybrid models by combining machine learning techniques with traditional quantitative analysis. Integrating methods like technical analysis, neural networks, and genetic algorithms may lead to more robust trading strategies[9, 19, 27] Systematic evaluations and comparisons of different ML and algorithmic trading techniques are lacking in current research. Standardized benchmarks and frameworks for evaluating and comparing automated trading systems in a fair and uniform manner would enable more rigorous research[23]

6. Conclusion, summary of findings, Implications for practitioners and policymakers and Future research opportunities

This systematic literature review comprehensively examines recent advancements in machine learning techniques applied to algorithmic trading systems. It summarizes key developments in methods, data sources, applications, performance measures, advantages, risks, and implications for market quality based on empirical studies published between 2008 and 2022. By synthesizing evidence from multiple high-quality studies, this work addresses essential knowledge gaps and achieves a holistic perspective of the field's current state.

The review demonstrates how machine learning has significantly enhanced algorithmic trading by identifying complex patterns, optimizing strategies, forecasting prices more accurately, and generating improved risk-adjusted returns compared to traditional rule-based systems. Techniques such as deep learning, reinforcement learning, and hybrid frameworks have been widely and successfully applied across various financial instruments, from individual equities to derivatives and indices. However, the review also highlights open challenges around overfitting, model instability, lack of explain ability, and systemic risks that require careful attention through best practices.

By systematically evaluating real-world examples, performance metrics, and market impacts, this work scientifically assesses machine learning's contributions and influences. The findings justify the growing usage of these advanced techniques in automated trading architectures. At the same time, they are identifying limitations that point to opportunities for future scientific progress. This review establishes a rigorous foundation for understanding past progress and guiding the judicious development of more capable, transparent and socially responsible machine-learning-driven trading paradigms.

The thorough analysis of this area will advance scientific knowledge by indicating promising avenues for extending existing methods and, for example, leveraging new alternative data sources through technologies like natural language processing or computer vision. The review also highlights the need for additional empirical studies comparing techniques systematically, standardizing performance benchmarks, and examining regulations around high-frequency trading. Continued scientific inquiry in these areas will ensure that algorithmic trading powered by machine learning realizes its full potential for safely improving market quality and efficiency.

In conclusion, this systematic literature review provides a valuable contribution by comprehensively assessing the state-of-the-art in machine learning for algorithmic trading. It validates past advances and illuminates pathways for impactful future work, further establishing the role of scientific review in guiding responsible innovation at the intersection of technology and finance.

References

[1] Adegboye, A., M. Kampouridis, and F. Otero, *Algorithmic trading with directional changes.* Artificial Intelligence Review, 2023. 56(6): p. 5619-5644.

[2] Ansari, Y., et al., *A Deep Reinforcement Learning-Based Decision Support System for Automated Stock Market Trading.* IEEE Access, 2022. 10: p. 127469-127501.

[3] Jin, B., *A Mean-VaR Based Deep Reinforcement Learning Framework for Practical Algorithmic Trading.* IEEE Access, 2023. 11: p. 28920-28933.

[4] Li, Y., W. Zheng, and Z. Zheng, *Deep Robust Reinforcement Learning for Practical Algorithmic Trading.* IEEE Access, 2019. 7: p. 108014-108022.

[5] Chandola, D., et al., *Forecasting Directional Movement of Stock Prices using Deep Learning.* Annals of Data Science, 2022.

[6] Jiang, W., *Applications of deep learning in stock market prediction: Recent progress.* Expert Systems with Applications, 2021. 184: p. 115537.

[7] Prachyachuwong, K. and P. Vateekul *Stock Trend Prediction Using Deep Learning Approach on Technical Indicator and Industrial Specific Information*. Information, 2021. 12, DOI: 10.3390/info12060250.

[8] Chakole, J.B., et al., *A Q-learning agent for automated trading in equity stock markets.* Expert Systems with Applications, 2021. 163: p. 113761.

[9] Dash, R. and P.K. Dash, *A hybrid stock trading framework integrating technical analysis with machine learning techniques.* The Journal of Finance and Data Science, 2016. 2(1): p. 42-57.

[10] Liu, P., et al., *Multi-type data fusion framework based on deep reinforcement learning for algorithmic trading.* Applied Intelligence, 2023. 53(2): p. 1683-1706.

[11] Park, D.Y. and K.H. Lee, *Practical Algorithmic Trading Using State Representation Learning and Imitative Reinforcement Learning.* IEEE Access, 2021. 9: p. 152310-152321.

[12] Deng, S., et al., *A novel hybrid method for direction forecasting and trading of Apple Futures.* Applied Soft Computing, 2021. 110: p. 107734.

[13] Li, Y., et al., *A New Hybrid VMD-ICSS-BiGRU Approach for Gold Futures Price Forecasting and Algorithmic Trading.* IEEE Transactions on Computational Social Systems, 2021. 8(6): p. 1357-1368.

[14] Dodd, O., et al., *US cross-listing and domestic high-frequency trading: Evidence from Canadian stocks.* Journal of Empirical Finance, 2023. 72: p. 301-320.

[15] Serbera, J.-P. and P. Paumard, *The fall of high-frequency trading: A survey of competition and profits.* Research in International Business and Finance, 2016. 36: p. 271-287.

[16] Xu, X. and Y. Zhang, *A high-frequency trading volume prediction model using neural networks.* Decision Analytics Journal, 2023. 7: p. 100235.

[17] Bagate, R., et al. *Survey on Algorithmic Trading Using Sentiment Analysis*. in *Proceedings of the 6th International Conference on Advance Computing and Intelligent Engineering*. 2023. Singapore: Springer Nature Singapore.

[18] Aitken, M., D. Cumming, and F. Zhan, *Algorithmic trading and market quality: International evidence of the impact of errors in colocation dates.* Journal of Banking & Finance, 2023. 151: p. 106843.

[19] Raudys, S., *Portfolio of Automated Trading Systems: Complexity and Learning Set Size Issues.* IEEE Transactions on Neural Networks and Learning Systems, 2013. 24(3): p. 448-459.

[20] Ahrabian, A., C.C. Took, and D.P. Mandic, *Algorithmic Trading Using Phase Synchronization.* IEEE Journal of Selected Topics in Signal Processing, 2012. 6(4): p. 399-404.

[21] Shah, J., D. Vaidya, and M. Shah, *A comprehensive review on multiple hybrid deep learning approaches for stock prediction.* Intelligent Systems with Applications, 2022. 16: p. 200111.

[22] Zhou, H. and P.S. Kalev, *Algorithmic and high frequency trading in Asia-Pacific, now and the future.* Pacific-Basin Finance Journal, 2019. 53: p. 186-207.

[23] Sokolovsky, A. and L. Arnaboldi, *A generic methodology for the statistically uniform & comparable evaluation of Automated Trading Platform components.* Expert Systems with Applications, 2023. 223: p. 119836.

[24] Tudor, C. and R. Sova, *Flexible Decision Support System for Algorithmic Trading: Empirical Application on Crude Oil Markets.* IEEE Access, 2022. 10: p. 9628-9644.

[25] Kitchenham, B. and S. Charters, *Guidelines for performing Systematic Literature Reviews in Software Engineering.* 2007. 2.

[26] ACM. *Acm computing classification system*. Available from: https://dl.acm.org/ccs.

[27] Hu, Y., et al., *Application of evolutionary computation for rule discovery in stock algorithmic trading: A literature review.* Applied Soft Computing, 2015. 36: p. 534-551.

[28] Ampatzoglou, A. and I. Stamelos, *Software engineering research for computer games: A systematic review.* Information and Software Technology, 2010. 52(9): p. 888-901.

[29] Champiri, Z.D., S.R. Shahamiri, and S.S.B. Salim, *A systematic review of scholar context-aware recommender systems.* Expert Systems with Applications, 2015. 42(3): p. 1743-1758.

Authors' Profiles

Arash Salehpour holds a Master's in Computer Science Software Engineering from Islamic Azad University, Rasht Branch, Iran. He has experience in computer science, programming, and deep learning. He has research with an emphasis on finance and economics. His primary research directions are bibliometric analysis, portfolio construction, valuation, investor behavior, stock price prediction, and big data analytics in economy and finance.

Karim Samadzamini holds a PhD in Electrical Engineering from the University of Tabriz, Iran. He has experience in computer science and electrical engineering. His research interests are deep learning and machine learning.

How to cite this paper: Arash Salehpour, Karim Samadzamini, "Machine Learning Applications in Algorithmic Trading: A Comprehensive Systematic Review", International Journal of Education and Management Engineering (IJEME), Vol.13, No.6, pp. 41-53, 2023. DOI:10.5815/ijeme.2023.06.05