Proceedings of ICFTBA 2026 Symposium: Strategic Management & Business Analytics: Data-Driven Performance Models
DOI: 10.54254/2754-1169/2026.LD38096

The Development of Machine Learning in Quantitative Finance

Xingyao Li

Mathematics, University of Bristol, Bristol, UK
dn24970@bristol.ac.uk

Abstract. The rapid growth of financial data and artificial intelligence has increased the use of machine learning in quantitative finance. Machine learning provides new ways to identify complex patterns in financial data and support financial prediction and decision-making. This paper provides an introductory review of machine learning methods and their applications in quantitative finance. It first introduces the basic machine learning process and then focuses on three main applications: stock price prediction, portfolio management and credit risk assessment. Support Vector Machine, Neural Network and Random Forest are discussed for stock prediction; Genetic Algorithm, Reinforcement Learning and Deep Reinforcement Learning are reviewed for portfolio management; and Logistic Regression, Random Forest and XGBoost are examined for credit risk assessment. The paper also discusses three major challenges, including limited model interpretability, insufficient incorporation of external information and limited adaptability to changing market conditions. Corresponding future directions include explainable artificial intelligence, the integration of alternative and multimodal data, and adaptive learning methods. Overall, this review provides a structured overview of current machine learning applications in quantitative finance and may help readers better understand existing developments, major challenges and possible future directions in this field.

Keywords: Machine Learning, Quantitative Finance, Stock Price Prediction, Portfolio Management, Credit Risk Assessment

1. Introduction

Quantitative finance applies mathematical, statistical, and computational methods to analyse financial markets and support financial decision-making. Compared with traditional financial approaches, quantitative finance relies more heavily on numerical data and mathematical models. This allows financial risks and market patterns to be analysed in a more systematic way. Modern financial markets generate large amounts of complex and high-dimensional data. This has increased the importance of data analysis and data mining in quantitative finance [1]. Artificial Intelligence (AI) algorithms can further identify complex and nonlinear patterns that may be difficult to capture using conventional models [2].

With the rapid development of artificial intelligence, machine learning has been used in more and more fields, such as healthcare, transportation and manufacturing [3]. Quantitative finance is one of

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2754-1169/2026.LD38096

these important areas because it strongly depends on data, mathematical models and computational methods. In recent years, different machine learning techniques have been studied for many financial problems, including asset pricing, market prediction, volatility forecasting, portfolio management and risk analysis [2]. For example, trees and neural networks were found to capture complex relationships between financial variables and improve prediction performance in asset pricing [4]. Machine learning has also been studied for volatility forecasting. Several machine learning models provided more accurate forecasts than traditional linear models [5]. In portfolio management, deep neural networks and LSTM models were used to predict stock returns. The results showed better portfolio performance compared with traditional linear models [6].

The applications of AI are not limited to these areas. K-nearest neighbours, support vector machines and random forests have been applied to credit risk assessment and performed better than the traditional logistic model [7]. For option pricing, an LSTM model achieved better pricing accuracy than the Black–Scholes and Heston models [8]. AI has also been used to analyse financial text. FinBERT, a language model designed for financial sentiment analysis, performed better than previous methods on financial sentiment datasets [9]. Another important development is the use of reinforcement learning. Deep reinforcement learning has been applied to derivative hedging and can consider transaction costs, liquidity constraints and risk limits [10].

Based on these developments, this paper aims to provide an introductory overview of the applications of machine learning in quantitative finance. The rest of this paper is organised into three main sections. The Methods section introduces different AI and machine learning techniques and explains how they can be applied to different problems in quantitative finance. The Discussion section considers the current limitations and challenges in this field, as well as possible future developments. Finally, the Conclusion summarises the main points of the paper.

2. Methods

2.1. Preliminaries of machine learning algorithms

Machine learning is a part of AI that learns from examples in data. The model is not given a separate rule for every possible case. Instead, it uses the available data to work out how the inputs are related to the output [11]. In quantitative finance, the inputs may be prices, trading volume, financial indicators or other market information [12].

In a basic machine learning task, the data first need to be collected and prepared. They are then usually split into a training set and a testing set. The training set is used to fit the model, while the testing set gives an indication of how the model works on data it has not seen before [11]. If several models are considered, their results can be compared and the model settings can be changed when needed.

The way performance is measured depends on the problem. For a regression task, common measures are mean absolute error (MAE), mean squared error (MSE) and root mean squared error (RMSE). For classification, accuracy, precision, recall and F1-score are often used [13]. Figure 1 gives a simple summary of this process, from preparing the data to obtaining a prediction or decision.

DOI: 10.54254/2754-1169/2026.LD38096

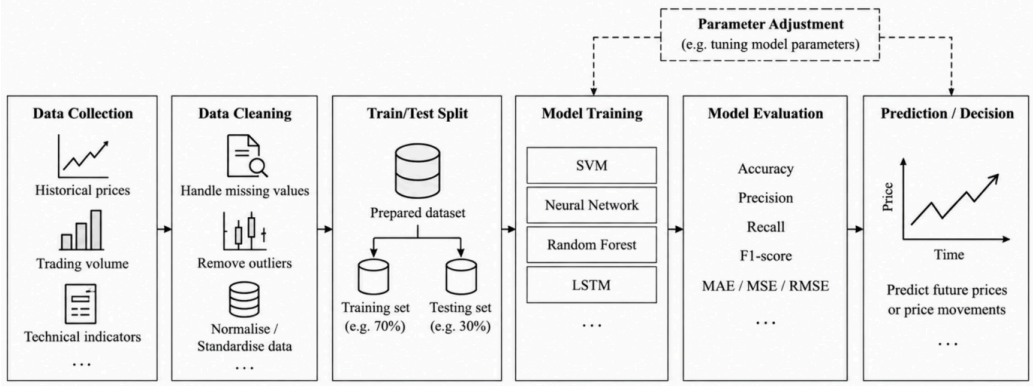


Figure 1. The main steps of the machine learning process (picture credit: original)

2.2. Stock price prediction

Stock prediction usually starts from information that is already available in the market. Historical prices, trading volume and technical indicators can be given to a machine learning model so that it can look for patterns related to later price changes [14]. The output may be a numerical price or simply the direction of the next movement.

2.2.1. Support Vector Machine

Support Vector Machine (SVM) shown in Figure 2 is a supervised learning method that can handle both classification and regression. For classification, it tries to place a boundary between groups of observations. This boundary is usually described as a hyperplane [15].

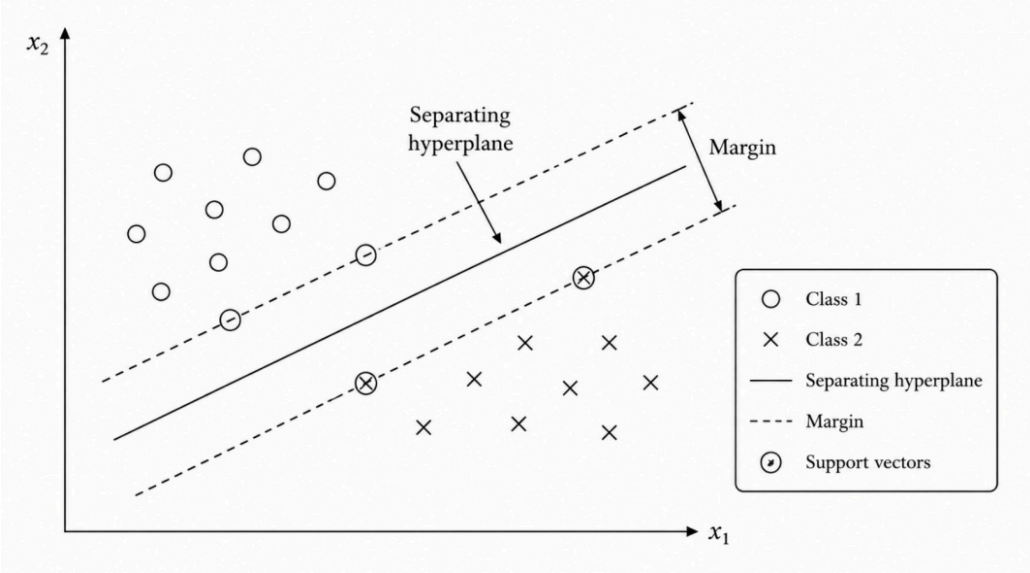


Figure 2. Basic structure of a Support Vector Machine (picture credit: original)

When the data cannot be separated well by a straight boundary, SVM can use a kernel function to deal with a more complicated relationship. In a stock market problem, the inputs may include earlier prices, trading activity and technical indicators. The model can then classify a movement as up or

DOI: 10.54254/2754-1169/2026.LD38096

down. Support Vector Regression uses the same general idea for a numerical output, such as a future price. SVM methods appear regularly in the stock forecasting literature [14].

Kim tested SVM on the daily direction of the Korea Composite Stock Price Index (KOSPI). Twelve technical indicators were used to describe the market before the next movement was classified as upward or downward [16]. The same problem was also tested with a back-propagation neural network and case-based reasoning. Kim's work mainly examined whether an existing SVM method was useful for this forecasting task rather than creating a new SVM algorithm. In that comparison, SVM gave promising results [16].

2.2.2. Neural network

A neural network shown in Figure 3 is built from connected artificial neurons. A simple network normally has an input layer, one or more hidden layers and an output layer [17].

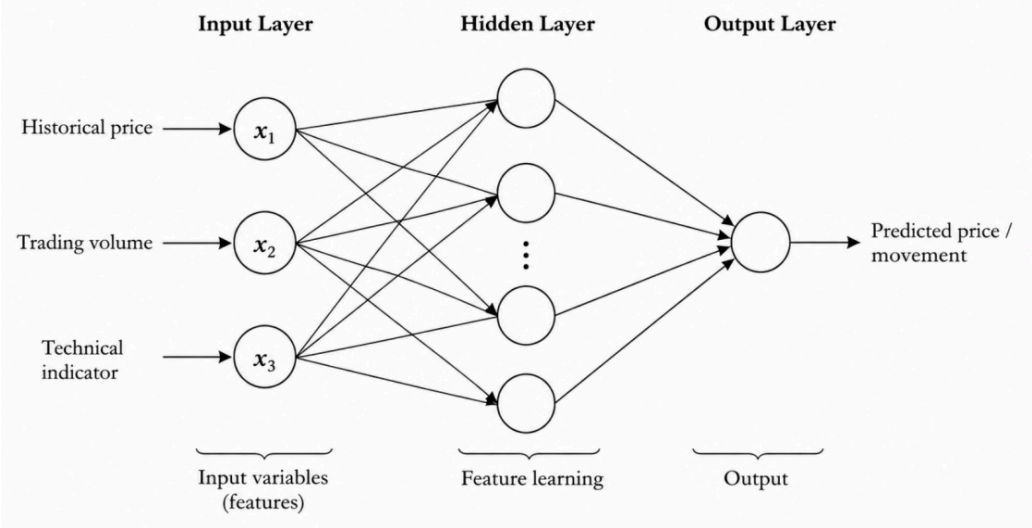


Figure 3. Basic structure of a neural network (picture credit: original)

The connections in the network are changed during training as the model learns from the data. For stock prediction, market variables can enter through the input layer and the output can represent a price, return or movement direction. Neural networks are common in this area. In the review by Kumbure et al., they were one of the methods that appeared most often in the studies considered [14].

Qiu et al. studied the Nikkei 225 and used an artificial neural network to predict market returns. Their dataset ran from November 1993 to July 2013 and included selected financial and macroeconomic variables [18]. These variables were fed into the network so that it could learn their relationship with later returns. The paper also changed the training procedure: Genetic Algorithm and Simulated Annealing were added instead of relying only on standard back-propagation. In the reported experiment, the modified training methods produced better prediction results [18].

2.2.3. Random Forest

Random Forest shown in Figure 4 combines many decision trees rather than making the final prediction from only one tree [19]. Different trees are built from different parts of the data, and their outputs are combined at the end. The method can be used for either classification or regression. In

DOI: 10.54254/2754-1169/2026.LD38096

stock forecasting, variables such as earlier prices, trading volume and technical indicators can be used to predict a later price or the direction of movement.

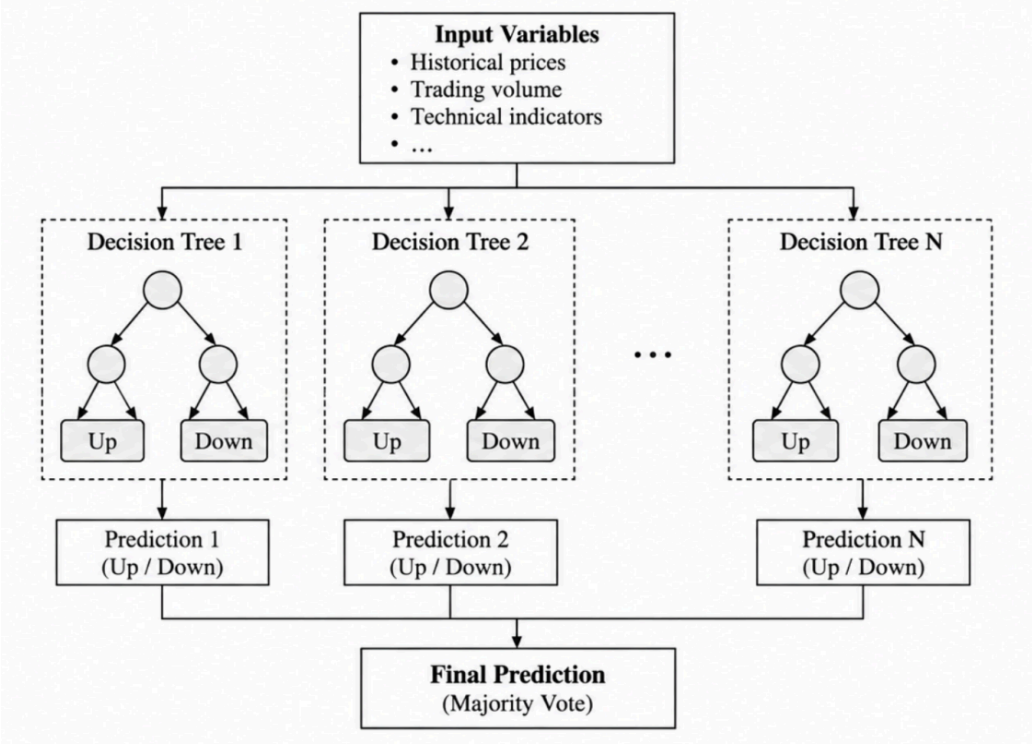


Figure 4. Basic structure of a Random Forest model (picture credit: original)

Patel et al. considered two Indian stocks and two market indices: Reliance Industries, Infosys, CNX Nifty and BSE Sensex. Their data covered 2003 to 2012 [20]. Ten technical indicators were calculated from the historical prices and used to predict whether the next movement would be up or down. An important part of the paper was the way these indicators were prepared. The authors converted them into trend deterministic data before running Random Forest and the other models. This preparation step improved the reported prediction accuracy, including the result for Random Forest [20].

2.3. Portfolio management

Portfolio management is concerned with which assets to hold and how much money to place in each one. Return is important, but the level of risk also has to be considered. AI methods can help with this problem by searching through possible allocations or by changing investment decisions when new market information is received [21].

2.3.1. Genetic Algorithm

Genetic Algorithm (GA) is an optimisation method based on the idea of natural selection. It begins with a number of candidate solutions and changes them through selection, crossover and mutation [22]. In a portfolio problem, one candidate can represent one way of dividing money across several assets. The algorithm compares these candidates and continues searching for allocations with a better return-risk trade-off. GA is useful here because portfolio problems can contain practical constraints as well as many possible combinations [22].

DOI: 10.54254/2754-1169/2026.LD38096

Lin and Liu used Genetic Algorithms to select portfolios from Taiwanese mutual funds between 1997 and 2000. Monthly returns were used to estimate expected fund returns and the relationships between them [23]. The algorithms searched across different choices of funds and investment amounts. The paper also included minimum transaction lots, which meant that the portfolios did not assume that any small amount could be traded. This made the setting closer to actual investment practice. The portfolios obtained in the study were close to the efficient frontier [23].

2.3.2. Reinforcement Learning

Reinforcement Learning (RL) shown in Figure 5 learns through repeated interaction. An agent chooses an action, sees what happens and receives a reward or penalty. By repeating this process, it can gradually learn which actions are better for a longer-term objective [24]. In portfolio management, the market can be treated as the environment and the action can be a change in portfolio allocation. The reward can be linked to portfolio return, risk or both.

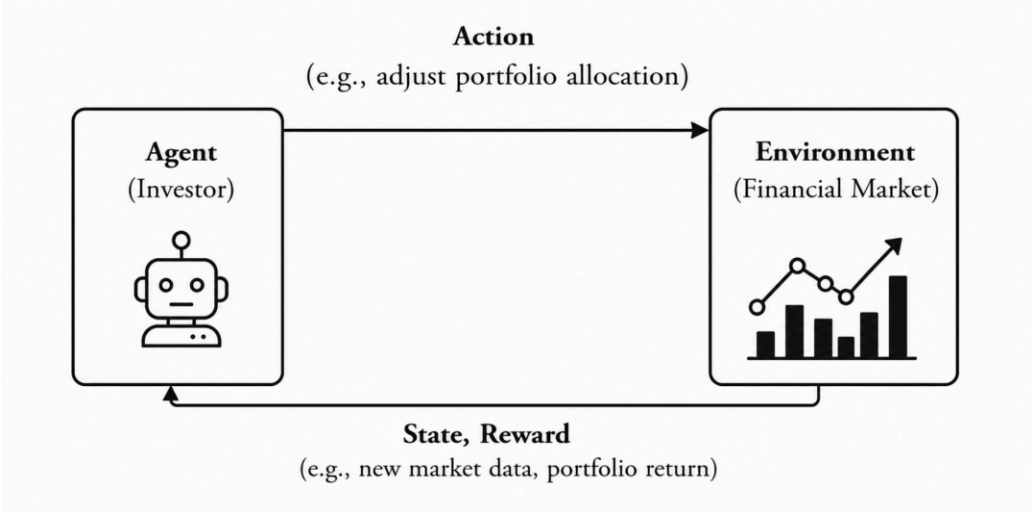


Figure 5. Basic framework of Reinforcement Learning in portfolio management (picture credit: original)

Hambly et al. explain how portfolio optimisation can be written as an RL problem. A state can contain prices, earlier returns and the current holdings, while an action changes the amount invested in the available assets [24]. A measure such as portfolio return or the Sharpe ratio can be used as the reward. The important point is that the agent is making a sequence of allocation decisions. It is therefore learning an investment policy rather than only trying to forecast the next market price [24].

2.3.3. Deep Reinforcement Learning

Deep Reinforcement Learning (DRL) adds deep neural networks to the reinforcement learning framework. The neural network allows the agent to work with more complicated market information and to connect that information with later decisions [21]. For a portfolio, this means that the model can change portfolio weights as market conditions change rather than keeping one allocation for the whole period.

Rezaei and Nezamabadi-Pour tested five DRL algorithms on daily data for 30 DJIA stocks: A2C, DDPG, PPO, SAC and TD3 [21]. The models received historical price and trading-volume information, and portfolio-value changes were used in the reward. The paper was mainly a

DOI: 10.54254/2754-1169/2026.LD38096

comparison of these existing algorithms, not a proposal for a new DRL method. Their results were compared with the DJIA and with traditional portfolio strategies. In the experiments, the DRL methods achieved higher returns and Sharpe ratios [21].

2.4. Credit risk assessment

Credit risk is the risk that a borrower will not repay a loan as agreed. In practice, lenders therefore need to estimate which applicants are more likely to default. Machine learning can use borrower information, including income, credit history and loan characteristics, to look for patterns associated with default [7].

2.4.1. Logistic regression

Logistic Regression (LR) is commonly used when the outcome has two classes. Rather than predicting any numerical value, it estimates the probability that an observation belongs to one of those classes [25]. For credit risk, the classes can be default and non-default. Borrower information becomes the model input and the estimated default probability can then be used for risk classification.

Lessmann et al. included Logistic Regression in a comparison of credit-scoring methods based on eight real-world datasets [26]. The model used borrower characteristics to estimate credit outcomes. Logistic Regression was included as a traditional benchmark, so the aim was not to design a new version of the model. Several of the newer classification methods performed better in the experiments, which allowed the authors to compare them against this established baseline [26].

2.4.2. Random Forest

Random Forest can be used for the same credit-risk task. As described in Section 2.2.3, it combines a group of decision trees instead of depending on one tree. In this setting, each tree can use borrower and credit variables to make a classification, and the separate results are combined to give the final prediction. This also allows the model to represent relationships that are not simply linear.

In the same study, Lessmann et al. also evaluated Random Forest on the eight credit-scoring datasets [26]. The trees were trained with borrower characteristics and then combined to classify the observed credit outcomes. The method was not presented as a new Random Forest algorithm; it was one of the models in the benchmark comparison. Across the datasets, Random Forest was among the methods with strong predictive performance [26].

2.4.3. XGBoost

XGBoost is based on gradient-boosted decision trees. Its trees are added one after another, with a new tree trying to correct part of the error left by the earlier trees [27]. This is different from Random Forest, where many trees are built more independently. For credit scoring, XGBoost can take borrower and loan variables as input and use the boosted trees to estimate default risk.

Xia et al. tested an XGBoost credit-scoring approach on five real datasets: German, Australian, Taiwan and two P2P lending datasets [28]. The borrower and credit variables were processed before the model was fitted. The paper also used Bayesian optimisation to select model parameters and feature importance to remove variables that contributed less. These steps meant that the study went beyond using XGBoost with default settings. In their experiments, the proposed approach performed better overall than several benchmark methods [28].

DOI: 10.54254/2754-1169/2026.LD38096

3. Discussion

3.1. Challenges

3.1.1. Limited interpretability

Interpretability is one problem that becomes more noticeable as models become more complex. A simple model can often be explained by looking directly at its variables, but this is much harder for deep neural networks and other large models. In finance, knowing the final prediction may not be enough because the reason behind a risky decision can also matter. Černevičienė and Kabašinskas discuss this difficulty in their review of explainable AI in finance and note the tension between model complexity and understandable results [29].

3.1.2. Limited incorporation of external information

A second problem is the type of information used by the model. Price histories, trading volume and financial ratios are useful, but they do not describe everything that is happening in the market. News, policy announcements, public attention and sentiment can all affect prices. Sun et al. discuss these alternative data sources and show that they can add information that is not contained in traditional financial datasets [30]. A model based only on numerical market history may therefore miss part of the current market situation

3.1.3. Limited model adaptability

A further difficulty is that financial relationships are not fixed. A pattern found in one period may weaken or disappear later when the economy, market structure or investor behaviour changes. This type of change is often described as concept drift. Wang et al. study concept drift in financial reinforcement learning and consider both gradual changes and more sudden shifts [31]. Because of this, good backtesting results do not necessarily mean that a model will remain equally reliable under a new market regime.

3.2. Future prospects

3.2.1. Explainable Artificial Intelligence

Explainable Artificial Intelligence (XAI) is one way to deal with the first problem. The aim is to add information about how a complex model reached its output rather than giving only the final prediction. For example, an explanation may show which variables had the greatest influence on a decision. Černevičienė and Kabašinskas review several uses of XAI in finance and link explainability with risk assessment, trust and practical adoption [29]. Future work therefore needs to consider both prediction quality and whether the model's reasoning can be understood.

3.2.2. Integration of alternative and multimodal data

The information given to financial models can also be expanded. Besides prices and other numerical variables, researchers can use news, social media activity and other alternative data. Sun et al. review this area and describe how these sources can provide additional and sometimes more timely information about firms and markets [30]. Multimodal methods go one step further by combining

DOI: 10.54254/2754-1169/2026.LD38096

different kinds of data in the same model. This may be useful when an event is visible in text or other sources before it is fully reflected in market prices.

3.2.3. Adaptive and dynamic learning

Another direction is to make models update more effectively when the market changes. A model trained once may become less useful if the relationships in new data are different from those in the training period. Wang et al. address this issue in a financial reinforcement-learning setting and propose a method for dealing with concept drift [31]. More adaptive approaches could therefore update what the model has learned when new evidence appears instead of assuming that old patterns will remain stable.

4. Conclusion

This review looked at how machine learning is used in three parts of quantitative finance: stock prediction, portfolio management and credit risk assessment. The methods are different, but they all use data to support a prediction or a financial decision. SVM, Neural Network and Random Forest were considered for stock forecasting. Genetic Algorithm, RL and DRL were discussed for portfolio decisions, while Logistic Regression, Random Forest and XGBoost were examined for credit risk.

The studies also show that the same algorithm can be used in quite different ways. Some papers mainly test an existing method, while others change the data preparation, training procedure or parameter selection. At the same time, model performance is only one part of the problem. Complex models may be difficult to explain, important external information may be missing from the data, and patterns learned from past markets may not continue in the future. XAI, alternative and multimodal data, and more adaptive learning methods are possible responses to these issues. They do not remove the difficulties, but they provide useful directions for further work in quantitative finance.

References

[1] Goldstein, I., Spatt, C. S., & Ye, M. (2021). Big data in finance. *The Review of Financial Studies, 34*, 3213-3225. https://doi.org/10.1093/rfs/hhab038

[2] Rundo, F., Trenta, F., Di Stallo, A. L., et al. (2019). Machine learning for quantitative finance applications: A survey. *Applied Sciences, 9*, 5574. https://doi.org/10.3390/app9245574

[3] Rashid, A. B., & Kausik, M. A. K. (2024). AI revolutionizing industries worldwide: A comprehensive overview of its diverse applications. *Hybrid Advances, 7*, 100277. https://doi.org/10.1016/j.hybadv.2024.100277

[4] Gu, S., Kelly, B., & Xiu, D. (2020). Empirical asset pricing via machine learning. *The Review of Financial Studies, 33*, 2223-2273. https://doi.org/10.1093/rfs/hhaa009

[5] Zhu, H., Bai, L., He, L., et al. (2023). Forecasting realized volatility with machine learning: Panel data perspective. *Journal of Empirical Finance, 73*, 251-271. https://doi.org/10.1016/j.jempfin.2023.07.003

[6] Babiak, M., & Baruník, J. (2026). Deep learning, predictability, and optimal portfolio returns. *Journal of Empirical Finance, 87*, 101705. https://doi.org/10.1016/j.jempfin.2026.101705

[7] Liu, Y., Yang, M., Wang, Y., et al. (2022). Applying machine learning algorithms to predict default probability in the online credit market: Evidence from China. *International Review of Financial Analysis, 79*, 101971. https://doi.org/10.1016/j.irfa.2021.101971

[8] Pimentel, R., Risstad, M., Rogde, S., et al. (2026). Option pricing with deep learning: A long short-term memory approach. *Decisions in Economics and Finance, 49*, 155-186. https://doi.org/10.1007/s10203-025-00518-9

[9] Araci, D. (2019). FinBERT: Financial sentiment analysis with pre-trained language models. arXiv:1908.10063.

[10] Buehler, H., Gonon, L., Teichmann, J., et al. (2019). Deep hedging. *Quantitative Finance, 19*, 1271-1291. https://doi.org/10.1080/14697688.2019.1571683

DOI: 10.54254/2754-1169/2026.LD38096

[11] James, G., Witten, D., Hastie, T., et al. (2021). *An introduction to statistical learning: With applications in R*. 2nd Edition, Springer, New York. https://doi.org/10.1007/978-1-0716-1418-1
[12] Nazareth, N., & Reddy, Y. V. R. (2023). Financial applications of machine learning: A literature review. *Expert Systems with Applications, 219*, 119640. https://doi.org/10.1016/j.eswa.2023.119640
[13] Rainio, O., Teuho, J., & Klén, R. (2024). Evaluation metrics and statistical tests for machine learning. *Scientific Reports, 14*, 6086. https://doi.org/10.1038/s41598-024-56706-x
[14] Kumbure, M. M., Lohrmann, C., Luukka, P., et al. (2022). Machine learning techniques and data for stock market forecasting: A literature review. *Expert Systems with Applications, 197*, 116659. https://doi.org/10.1016/j.eswa.2022.116659
[15] Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning, 20*, 273-297. https://doi.org/10.1007/BF00994018
[16] Kim, K. J. (2003). Financial time series forecasting using support vector machines. *Neurocomputing, 55*, 307-319. https://doi.org/10.1016/S0925-2312(03)00372-2
[17] Haykin, S. (2009). *Neural networks and learning machines* (3rd ed.). Pearson.
[18] Qiu, M., Song, Y., & Akagi, F. (2016). Application of artificial neural network for the prediction of stock market returns: The case of the Japanese stock market. *Chaos, Solitons & Fractals, 85*, 1-7. https://doi.org/10.1016/j.chaos.2016.01.004
[19] Breiman, L. (2001). Random forests. *Machine Learning, 45*, 5-32. https://doi.org/10.1023/A:1010933404324
[20] Patel, J., Shah, S., Thakkar, P., et al. (2015). Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques. *Expert Systems with Applications, 42*, 259-268. https://doi.org/10.1016/j.eswa.2014.07.040
[21] Rezaei, M., & Nezamabadi-Pour, H. (2025). A taxonomy of literature reviews and experimental study of deep reinforcement learning in portfolio management. *Artificial Intelligence Review, 58*, 94. https://doi.org/10.1007/s10462-024-11066-w
[22] Kalayci, C. B., Ertenlice, O., Akyer, H., et al. (2017). A review on the current applications of genetic algorithms in mean-variance portfolio optimization. *Pamukkale University Journal of Engineering Sciences, 23*, 470-476. https://doi.org/10.5505/pajes.2017.37132
[23] Lin, C. C., & Liu, Y. T. (2008). Genetic algorithms for portfolio selection problems with minimum transaction lots. *European Journal of Operational Research, 185*, 393-404. https://doi.org/10.1016/j.ejor.2006.12.024
[24] Hambly, B., Xu, R., & Yang, H. (2023). Recent advances in reinforcement learning in finance. *Mathematical Finance, 33*, 437-503. https://doi.org/10.1111/mafi.12382
[25] Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression* (3rd ed.). Wiley. https://doi.org/10.1002/9781118548387
[26] Lessmann, S., Baesens, B., Seow, H. V., et al. (2015). Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research. *European Journal of Operational Research, 247*, 124–136. https://doi.org/10.1016/j.ejor.2015.05.030
[27] Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. https://doi.org/10.1145/2939672.2939785
[28] Xia, Y., Liu, C., Li, Y., et al. (2017). A boosted decision tree approach using Bayesian hyper-parameter optimization for credit scoring. *Expert Systems with Applications, 78*, 225–241. https://doi.org/10.1016/j.eswa.2017.02.017
[29] Černevičienė, J., & Kabašinskas, A. (2024). Explainable artificial intelligence (XAI) in finance: A systematic literature review. *Artificial Intelligence Review, 57*, 216. https://doi.org/10.1007/s10462-024-10854-8
[30] Sun, Y., Liu, L., Xu, Y., et al. (2024). Alternative data in finance and business: Emerging applications and theory analysis (review). *Financial Innovation, 10*, 127. https://doi.org/10.1186/s40854-024-00652-0
[31] Wang, C. A., Huang, S. H., Chen, C. T., et al. (2026). Financial reinforcement learning under concept drift based on knowledge distillation and curriculum learning. *Decision Support Systems, 203*, 114624. https://doi.org/10.1016/j.dss.2026.114624