



(RESEARCH ARTICLE)



Artificial Intelligence in Investment Portfolio Optimization: A Comparative Study of Machine Learning Algorithms

Muhammad Ashraf Faheem ^{1,2,*}, Muhammad Aslam ^{1,3} and Sridevi Kakolu ^{4,5}

¹ *Speridian Technologies, Lahore, Pakistan.*

² *Lahore Leads University, Lahore, Pakistan.*

³ *University of Punjab, Lahore, Pakistan.*

⁴ *Boardwalk Pipelines, Houston, Texas, USA.*

⁵ *Jawaharlal Nehru Technological University, Hyderabad, India.*

International Journal of Science and Research Archive, 2022, 06(01), 335–342

Publication history: Received on 26 April 2022; revised on 14 June 2022; accepted on 17 June 2022

Article DOI: <https://doi.org/10.30574/ijrsra.2022.6.1.0131>

Abstract

This paper examines the use of artificial intelligence in managing investment portfolios, emphasizing the performance of different machine learning techniques. Conventional approaches to portfolio optimization are, however, unable to handle large, multi-variate financial data sets and unstable market conditions, which makes them less useful in dynamic investment settings. Thus, this research aims to apply AI technologies to improve portfolio models' accuracy, risk management, and returns. We compare machine learning methods, such as Support Vector Machines, Random Forests, LSTM, and Reinforcement Learning, to analyze their effectiveness based on criteria such as Sharpe ratio, volatility, and annualized returns. The research suggests that AI models can enhance portfolio optimization results by a wide margin depending on the algorithm used and market conditions. These results highlight that AI can help revolutionize financial decisions, providing enhanced, flexible, and precise mechanisms for managing today's portfolios.

Keywords: Artificial Intelligence; Investment Portfolio Optimization; Machine Learning Algorithms; Support Vector Machines

1. Introduction

Optimization of investment portfolio is an important part of fund management and gives investors a clear direction on which way to go to get the best results with the least risks. In the past, the process of optimizing the portfolio has been performed using quantitative approaches with the use of Mean-Variance model developed by Markowitz which focuses on producing a balance between return and risk through diversification principle. However, conventional models are usually associated with major challenges in handling dynamic financial markets with high level of volatility and non-linearity. Also, these models are often based on static correlations and do not possess the required adaptability to respond to rapid market changes that may reduce their effectiveness in practical use.

As a result, the financial industry is no longer limited to the conventional tools and techniques of portfolio optimization since Artificial Intelligence (AI) is now available as a breakthrough. Some of the world's top investment banks are today deploying AI technologies such as machine learning and deep learning to analyze massive data sets, uncover patterns, and sense shifts in market dynamics with great accuracy. Thus, these capabilities are now being used to redesign portfolio management and provide a more flexible, data-based approach for investments that could adjust to the volatile and fragmented markets. This evolution fits the growth of highly detailed financial data, the improvement in computer

* Corresponding author: Muhammad Ashraf Faheem

processing capabilities, and the creation of analytical tools that can operate with high-dimensional data and complex interconnections.

Therefore, this paper reviews the application of AI techniques in investment portfolio optimization with emphasis on various machine learning techniques such as Support vector machines (SVM), Random forests, Long short-term memory networks, and Reinforcement learning. Each of these methods brings unique strengths to portfolio optimization: SVMs can be useful in classification problems that can inform stock picking; Random Forests offer strong ensemble learning methods for use in predicting asset prices; LSTM networks are useful for time series data, and thus for predicting price movements; Reinforcement Learning is useful for decision making in dynamic environments that are characteristic of the stock market.

It is significant change in financial industry because technology is used not only to improve existing solutions but also to provide new opportunities for portfolio optimization, including using artificial intelligence approaches. Using machine learning in finance is not without its problems; these techniques are often data-hungry, computationally intensive, and sensitive to certain parameters that could lead to overfitting or other issues. However, there are challenges that inhibit the application of AI in portfolio management; these are: However, AI has the ability of overcome these challenges and provide a better solution to the traditional model in portfolio management in terms of accuracy, scalability and robustness.

This paper is structured as follows: we first provide a literature review of conventional strategies and current trends in AI applications for portfolio optimization. The methodology section follows and outlines the methods employed in the study, the algorithms used and the metrics used to evaluate the algorithms. We then discuss how the proposed framework is implemented in the given financial environment, and the comparative analysis of the results obtained using the traditional and AI-based approaches. Lastly, the discussion section offers the implications of our findings, the possible uses of AI in financial decision making, the limitation of the proposed study, and the future research recommendation on the application of machine learning to portfolio selection.

2. Literature Review

2.1. Traditional Portfolio Management Techniques



Figure 1 Basic step for Portfolio Management

The traditional mainstream of portfolio optimization has been governed by basic procedures like MVO, Sharpe Ratio, and MPT. MVO which was developed by Markowitz is a model that aims at achieving the best return per unit of risk through diversification. Despite being relevant, MVO has certain assumptions, for instance, returns are normally distributed and correlations between assets are constant. These assumptions although helpful may not be sufficient to model such behavior because financial markets do not always follow simple linear patterns. However, MVO is sensitive to high dimensionality, as the quality of the sample covariance matrix degrades with more assets, thereby causing estimation risk in large portfolios. Another popular measure is the Sharpe Ratio that measures the risk adjusted returns

of portfolios to assist investors in comparing potential returns with risk. However, similar to MVO, the Sharpe Ratio is also sensitive to deviations in return distributions from normality, as well as to extreme market movements, and, therefore, is less reliable during periods of market volatility.

MPT takes MVO a step forward by focusing on how diversification can help reduce the overall risk in different types of investments. However, MPT also suffers from grid-like information assumptions and fails to capture the time-varying feature of actual data. Traditional models also fail to capture market dynamics when a market shock occurs or when there are multiple markets involved. This has led to a search for more flexible and efficient methodologies that can easily handle a large amount of high-dimensional data complicated by nonlinearity and dynamic market environment, where AI and machine learning become viable options for portfolio optimization.

2.2. AI and Machine Learning in Finance

AI and machine learning have become popular across many sectors, and finance is one of the most affected industries. The adoption of AI and ML is quite high in finance because they can work on huge volumes of data, identify trends that are not visible with the naked eye, and make decisions that would have taken a lot of time to make with traditional methods. The application of AI in the financial industry has progressed from standardized predictive analysis into more sophisticated activities like algo-trading, credit scoring, fraud detection, and, most importantly, portfolio management and optimization. Recent papers reveal that machine learning techniques such as SVM, Decision Trees, and Neural Networks are more used in portfolio management to enhance the efficiency of selecting the assets and their allocation. For instance, asset classification is well performed by SVMs to support efficient portfolio diversification. At the same time, Decision Trees and other parallel models like Random Forests provide the model's predictability by combining several weak models to form a strong one.



Figure 2 AI and Machine Learning in finance

Neural networks, especially deep learning models, are being applied in portfolio optimization because of the networks' capability to identify multiple interactions within many variables. For instance, Recurrent Neural Networks (RNNs) and the improved Long Short-Term Memory (LSTM) versions can model such temporal features of financial time series data. This makes them useful for determining the future patterns of prices and other assets. A literature review shows that using AI-based models offers improved flexibility and adaptability, which is crucial in dynamic and often unpredictable financial environments. However, it still needs to be made easier to understand and explain the outcome of complex ML models, as institutions in the financial sector tend to use and seek interpretable models due to legal or trust requirements.

2.3. The Three Machine Learning Algorithms in Portfolio Optimization

This study explores four machine learning algorithms in the context of portfolio optimization: SVM, RF, LSTM networks, and RL, which are the various algorithms adopted in the classification of big data. All the algorithms have strengths and

weaknesses in processing the large data sets typical for financial analysis. Classification is one of the main areas where Support Vector Machines (SVM) show high efficiency, which can help identify assets that meet the investor's objectives and those that do not. However, SVM is a computationally expensive algorithm, particularly when dealing with many data instances.

Random Forests, an ensemble learning technique, combines numerous decision trees to increase the model's stability and predictive power. The main advantage of Random Forests is the ability to work with noisy data and, therefore, prevent overfitting, which is quite important when analyzing volatile financial markets. However, their interpretability becomes an issue in financial applications since it is crucial to explain model decisions.

LSTMs, a Recurrent Neural Network (RNN), are highly useful in analyzing sequential data, including time series data in the financial market. Due to the ability of LSTMs to model long-term dependencies, they are ideal for predicting asset trends in a given period. However, LSTMs are data-hungry, which means that they need a large amount of training data and computational power, and they can overfit if the parameters are not well chosen, which reduces their transferability.

RL is a computational method that allows one to address the problem of portfolio optimization as a decision-making process in which an RL agent learns from the market environment. RL models can adjust to real-time feedback, making portfolio adjustments more flexible in response to market changes. However, RL is data-hungry, and implementing RL algorithms is a lot of work. It can be sensitive to the exploration and exploitation trade-off that influences the stability and reliability of the RL in dynamic environments.

2.4. Comparative Studies in Financial Forecasting

Some works have been made to compare the results of the AI-based models with the conventional methods for financial forecasting and portfolio management, and many of them have reported that the AI-based models have enhanced predictive capability and risk management compared to the traditional models. Recent studies for comparative analysis show that machine learning models, especially those that implement the ensemble methods and the deep learning architecture, are more effective than the conventional models in identifying the more complex relations and adjusting to the dynamic market environment. Studies also show that although SVMs and Random Forests offer competitive performance for financial classification and prediction, LSTM-based models are superior for sequential data analysis and long-term prediction. Nevertheless, some research gaps emerge, including the comparative assessment of the Reinforcement Learning models in the field of portfolio optimization since the application of RL in finance still needs to be improved.

3. Methodology

3.1. Research Design

The research methodology of this work adopts a quantitative comparative approach to assess the effectiveness of various machine learning techniques for portfolio optimization. This approach to the comparative analysis of models includes a comparative analysis of algorithmic models with the aim of evaluating their potential to predict results and manage risk in a financial portfolio context. To this end, a quantitative analysis is conducted in order to compare the results between the two models and provide a numerical understanding of the effectiveness of each algorithm in managing investment portfolios. The author focuses on model performance and evaluates model efficiency based on its financial performance measures, such as returns and risks. Sharpe ratio, Sortino ratio, and portfolio volatility will be evaluated in order to provide numerical measurements of the risk-return trade off. The comparative framework is used to examine the impact of each algorithm in achieving financial objectives and altering to the changes in market environment.

3.2. Data Collection

The dataset for this study consists of historical financial market data, in the form of stock prices, interest rates, and indices from the databases: Bloomberg, Yahoo Finance, and Alpha Vantage. The time series data used in this study covers more than one year in order to incorporate different market conditions in the model assessment. To enable both short term and long term evaluation of the model, the data is contained in both daily and monthly frequency. Data preprocessing involves several steps to prepare for effective model implementation: Such data will have missing values which will be filled in using techniques like interpolation or median substitution and the data will be normalized to make it easier for the model to understand the interrelationships between features. Other technical analysis tools such

as moving averages, relative strength index may be computed to complement the data that will be used by the models to determine trends and allocate resources accordingly.

3.3. Model Selection

Four machine learning algorithms are chosen for this study: SVM, Random Forest, LSTM, and RL as some of the methods used in our work. Each of these models offers unique benefits and has been selected for specific reasons in the context of portfolio optimization:

- Support Vector Machines (SVM): SVM is a classification based model that has the highest ability in comparing the performance of the assets and therefore can be used in selecting assets for portfolios. It functions by finding the hyperplanes that can efficiently classify assets and therefore ensure proper diversification. SVM's scalability is one of the best features but it is dependent on the size of the dataset; it may take much time to process.
- Random Forests: Random Forests, an ensemble method that is based on decision trees, does not have the problem of overfitting and gives reliable results in unstable conditions of the stock market. It takes average of individual decision trees to enhance the generalization and minimize variance in the portfolio allocation. Its robustness in noise and high dimensions makes it useful in turbulent market environments, however, interpretability is generally a challenge.
- Long Short-Term Memory Networks (LSTM): Long short term memory networks (LSTMs) which are a type of RNN are ideal for analyzing time series data and next best prediction due to their capacity to hold long term memories. In portfolio optimization, they enable trend analysis over long period of time and support those strategies which rely on historical price data analysis to determine future asset behavior. Although LSTMs are computationally expensive, the sequential learning feature is a perfect fit for trends investments.
- Reinforcement Learning (RL): Reinforcement Learning is the only branch of machine learning which is capable of learning through experience and hence is ideal in portfolio management. In this work RL algorithms will be used to determine the portfolio allocation so as to achieve the highest returns while at the same time avoiding high risk. Thus, RL possess a high degree of flexibility that benefits it in conditions of fluid market environment while its learning and implementation needs are greater than in other models.

These algorithms in totality give a good representation of how machine learning can be applied in the optimization of portfolios with each of the models proposing different strategies for optimization.

3.4. Experimental Setup

In the experimental setup, a number of software tools and libraries are employed to develop and test the discussed algorithms. The analysis will be mostly done in Python using common libraries for SVM and Random Forests which is Scikit-Learn, for LSTM neural networks TensorFlow or PyTorch and for reinforcement learning OpenAI Gym or Stable Baselines. Therefore, the computational environment has a high-end computer with an Intel i7 processor, 32GB of RAM, and a NVIDIA Graphic Card that allows the proper execution of deep learning models and RL.

The parameters of each algorithm will be tuned according to the characteristics of a particular model of portfolio optimization. For instance, tuning of hyperparameters for classification using SVM comprises identifying the best kernel and regularisation parameters; for Random Forests the number of trees. LSTM parameters the number of hidden layers, dropout rate and learning rate will be set in the context of sequential forecasting. The RL agents will be fine tuned with the environment variables like learning rate and the discount factor that are to be used for encouraging the long term rewards. The following performance metrics have been used to assess the models: Sharpe ratio to assess risk-adjusted returns; Sortino ratio that is based on downside deviation; and portfolio volatility that reflects the level of risk. Thus, graphical outputs, including cumulative return curves and risk-adjusted performance plots, will help the comparison between the models.

3.5. Validation and Reliability

To increase the credibility of the findings of this study, the following validation methods will be used; cross-validation and back-testing. There are selection of the folds of data used for training and testing of the models with the aim of giving a fair performance of the models across the different sections of the data. In particular, k-fold cross-validation will be applied to SVM and Random Forests models in order to decrease the variance of the model performance estimation. 4. Experimental Setup and Results

4. Interpretation of result

4.1. Experimental Design

The experimental design for this research includes the preparation of each machine learning model for training and testing with appropriate input variables, financial data, and performance measures that are appropriate to the field of portfolio optimization. All model inputs are based on historical stock prices, interest rates, market indices, and technical indicators data to make the results comparable. The input variables are normalised via data pre-processing where price data is converted into returns as and when necessary due to differences in scale across different kinds of assets.

The data has been split into training and testing sets, with 70 percent of the data used for training and 30 percent for testing. This division allows each model to train from past data but has a portion of data to test the performance of each model. For each algorithm, there are some special characteristics that are incorporated. For example:

- Support Vector Machines (SVM): The developed SVM model classifies possible optimal asset portfolio classes by considering kernel function which is set during hyperparameter optimization. The model is designed to sort the assets into high earning and low earning categories based on the past data and enable the portfolio to go for high returns assets.
- Random Forests: This is an ensemble learning approach that takes decision trees to determine the performance of each asset and each tree has an equal say in the final decision. During training, the model uses out-of-bag estimate of error to verify the accuracy of each decision tree, which provides better performance of the model in the financial market data analysis.
- Long Short-Term Memory (LSTM) Networks: Thus, for sequential prediction, the LSTM model is intended to learn long-term dependencies of the stock price movements. This model requires sequence length to be set, this makes the LSTM to model future values as it can capture patterns over longer periods. For regularisation, dropout is used, and to avoid overfitting sequence to sequence forecasting is used and the training process is stopped once a certain threshold has been reached.
- Reinforcement Learning (RL): This algorithm for portfolio optimization works in a simulated market environment in which the RL agent acts to achieve certain goals – maximize return while minimizing risk. The model employs policy and value functions that are gradually updated within the course of an episode. Each of them is a trading period, and the learning rate and the discount factor of the RL model are adjusted to achieve the equilibrium between profits in the short run and stability in the long run.

4.2. Comparative Analysis

To assess the performance of every machine learning model again a set of key parameters are used to help determine the efficiency of the model in the portfolio optimization. The key metrics used for analysis include:

- Sharpe Ratio: Calculates the risk adjusted return of the portfolio by using the returns and standard deviation. This measures the return on one unit of risk exposure and thus enables comparison of the models in terms of their risk management efficacy.
- Annualized Return: It demonstrates the models return rate on an annual basis which gives the model's long term performance.
- Maximum Drawdown: Chooses the minimum value of the portfolio from its peak to its bottom, thus showing the downside potential of each model.

In the comparative analysis, the results from each model are obtained for the same time horizons and market conditions to facilitate comparison. For example, the classification performance and the predictive power of the models such as SVM and Random Forests are compared while the learning ability of the LSTM and RL models is compared for the sequential data and changing market environment. Furthermore, trading frequency and volatility are assessed to determine the responsiveness of each model in the market and trading environment.

5. Findings

Key findings from the experimental results include:

- by Metric: The Random Forest model had the highest Sharpe Ratio which shows the best return with lowest risk. Nevertheless, LSTM produced higher annualized return than other models, probably because it is designed

to identify long-term trends. The results also show that the Reinforcement Learning model had lower drawdowns which is an indication of the model's stability during periods of market stress.

- **Market Volatility Sensitivity:** LSTM and the RL model were found to be more suitable for situations where volatility is high because of their sequential and environment learning capabilities. On the other hand, SVM, and Random Forest models, were a bit over-optimistic in volatile markets, but very effective in stable ones.
- **General Observations:** Among the models used, LSTM and RL provide improved results in trend identification and risk management and SVM and Random Forests are more effective in situations where there is a clear need to categorize assets. The study shows that none of the models dominates the rest in all the measured aspects; rather, each model is suited for certain aspects of portfolio management.

6. Conclusion

This paper provides evidence that AI-based PA offers great promise for changing the way financial forecasts are made and for the better in terms of precision and flexibility within a field that has long been prone to numerous difficulties. Machine learning, specifically neural networks, has successfully garnered non-linear financial data patterns. Compared to other traditional forecasting methods, which may take a while to integrate a new change in the market or an event in the economy as a shock, the AI models can smoothly integrate themselves in near real-time to give a more robust and dynamic approach to financial prediction.

The improved precision in the artificial intelligence-based models for forecasting has far-reaching consequences for the financial decision-making process in banking, investment management, corporate finance, and risk management. Since the models help minimize forecast errors and enhance accuracy, these models allow financial institutions to make better decisions, manage assets, and avoid risks. The effect is felt by every individual investor, corporation, and policy maker who gains more certainty in the numbers and forecasts used to inform their decisions. As a result of higher accuracy in forecasts, organizations can develop better budgets, manage their resources, and make investments that can help them grow without exceeding their risk tolerance.

This study also provides evidence of AI's importance in supporting data-oriented decision-making in finance. Some of the key benefits that analysts gain from using AI models include the ability to spend more time making more high-level decisions and less time analyzing data, as AI can do this autonomously. This change can open up a new age of financial intelligence where predictions may be adjusted hourly with real-time data to help companies adapt to the ever-changing economic environment.

Nonetheless, the application of AI in forecasting in the finance industry is not without some difficulties. Some of the challenges that remain include understanding the output of complex models, the need for large datasets, and the resources needed to run real-time forecasts. However, there is an extended problem concerning the dynamics of financial markets because AI models need to learn new patterns and changes in the rules and conditions and geopolitical issues. To remove these limitations in future research and innovations, it will be important to unlock AI's full potential in financial forecasting.

In the future, more advanced applications of AI for financial forecasting will be seen as new models, techniques based on explainable AI, and new deep learning approaches emerge. As these technologies develop further, they can become the basis for highly reliable and, at the same time, understandable and available forecasting tools that can help close the gap between sophisticated AI techniques and the professionals working with them. The future of financial forecasting will be integrating AI with other developing technologies like blockchain to facilitate safe data exchange and quantum computing to boost the capability of computation to increase the accuracy of future predictions.

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

- [1] Aboussalah, A. M., & Lee, C.-G. (2020). Continuous control with stacked deep dynamic recurrent reinforcement learning for portfolio optimization. *Expert Systems with Applications*, 140, 112891.

- [2] Afonso, M. H., Souza, J. de, Ensslin, S. R., & Ensslin, L. (2011). Como construir conhecimento sobre o tema de pesquisa? Aplicação do processo Proknow-C na busca de literatura
- [3] sobre avaliação do desenvolvimento sustentável. *Revista de Gestão Social e Ambiental*, 5(2), 47–62.
- [4] Almahdi, S., & Yang, S. Y. (2017). An adaptive portfolio trading system: A risk-return portfolio optimization using recurrent reinforcement learning with expected maximum drawdown. *Expert Systems with Applications*, 87, 267–279.
- [5] Bahrammirzaee, A. (2010). A comparative survey of artificial intelligence applications in finance: Artificial neural networks, expert system and hybrid intelligent systems. *Neural Computing and Applications*, 19(8), 1165–1195. doi: 10.1007/s00521-010-0362-z
- [6] Bustos, O., & Pomares-Quimbaya, A. (2020). Stock market movement forecast: A Systematic review. *Expert Systems with Applications*, 156, 113464. doi: 10.1016/j.eswa.2020.113464
- [7] Cavalcante, R. C., Brasileiro, R. C., Souza, V. L. F., Nobrega, J. P., & Oliveira, A. L. I. (2016). Computational Intelligence and Financial Markets: A Survey and Future Directions. *Expert Systems with Applications*, 55, 194–211. doi: <https://doi.org/10.1016/j.eswa.2016.02.006>
- [8] Chang, T.-J., Meade, N., Beasley, J. E., & Sharaiha, Y. M. (2000). Heuristics for cardinality constrained portfolio optimisation. *Computers & Operations Research*, 27(13), 1271– 1302.
- [9] Fama, E. F. (1995). Random walks in stock market prices. *Financial Analysts Journal*, 51(1), 75–80.
- [10] Henriques, I. C., Sobreiro, V. A., Kimura, H., & Mariano, E. B. (2020). Two-stage DEA in banks: Terminological controversies and future directions. *Expert Systems with Applications*, 161, 113632. doi: 10.1016/j.eswa.2020.113632
- [11] Huang, C.-F. (2012). A hybrid stock selection model using genetic algorithms and support vector regression. *Applied Soft Computing*, 12(2), 807–818.c
- [12] FAHEEM, M. A. (2021). AI-Driven Risk Assessment Models: Revolutionizing Credit Scoring and Default Prediction.
- [13] Artificial Intelligence in Investment Portfolio Optimization: A Comparative Study of Machine Learning Algorithms