

A Comprehensive Review of Machine Learning Approaches in Stock Market Portfolio Management: Challenges, Innovations, And Future Directions

Shivkumar R. Kashid¹, Satish P. Pise²

^{1,2}*Department of Computer Science & Engineering, Ashokrao Mane Group of Institutions, Vathar, Maharashtra, India*

Abstract—Portfolio management for the stock market portfolio is an intricate problem of constrained optimization that aims at efficiently allocating limited resources among numerous investments in order to yield maximum gains while minimizing risks. The present paper provides a detailed overview of the development from conventional optimization methodologies, such as the mean variance model of Harry Markowitz, to state-of-the-art machine learning algorithms that overcome the limitations inherent to traditional methods. Traditional approaches suffer from several drawbacks, which include ignoring trading expenses, cardinality constraints, market imperfections, and overly simplistic assumption regarding returns distribution. Conversely, machine learning algorithms benefit from employing a wide variety of heterogeneous datasets, which include past prices, fundamentals, sentiment analysis, macroeconomic variables, and other alternative data sources.

The article reviews in a systematic way supervised, unsupervised and deep learning approaches used for predicting returns and risks, selecting assets and diversifying portfolio investments. Issues that stand out and pose barriers to the application of such machine learning methods in portfolio optimization include robustness and overfitting, interpretability issues, quality of data, and computational complexity. Finally, emerging trends are highlighted in the areas of multi-agent systems, explainable artificial intelligence, federated learning, graph neural networks, and evolutionary algorithms among others, all geared towards enhancing intelligent portfolio management systems.

Index Terms—Portfolio optimization, machine learning, deep learning, risk management, explainable AI, reinforcement learning, financial data analytics.

I. INTRODUCTION

Definition of stock market portfolio management problem:

Stock market portfolio management can be described as a constrained optimization problem that seeks to find the best capital allocation strategy for a certain amount of money amongst different financial instruments to maximize returns and minimize risks, which may be defined as variance, value at risk (VaR), conditional value at risk (CVaR), or liquidation-adjusted risks; from a computational perspective, this would involve the processing of mixed, inconsistent, and time-sensitive financial information such as past asset prices, sentiments, and future asset prices using artificial intelligence methods, including machine learning, reinforcement learning, or multiple criteria decision making (MCDM), etc. (Al Janabi, 2020; Ye et al., 2020).

Portfolio optimization models like the classical Markowitz mean-variance model operate under assumptions where the investor makes a choice between expected return and risk, which is measured as the variance, under the premise of having complete and stationary information regarding asset returns in a one-period investment strategy. However, such models have several shortcomings, including the exclusion of transaction costs, cardinality constraints, and market frictions which are very relevant in real life scenarios, especially in the domain of high-frequency trading; furthermore, the models make use of a normally distributed return and assume quadratic measures of risk, which are not appropriate for representing the investor's preferences and incorporating non-normal risk distributions; in

addition, classical models fail to consider liquidity constraints and dynamics in markets, which limit their usability and require extension through the integration of multiple criteria decision making, stochastic dominance, expected utility approaches, and the incorporation of other measures of risk like CVaR, and possibly linear programming or heuristics and neural networks for computational efficiency (Cao & Li, 2024).

Constraints related to reality, such as risk limitations, budgetary limitations, and transaction costs, make a huge difference in portfolio construction since they add a layer of complexity to the optimization problem by imposing practical limitations on the portfolio, with risk limitations frequently involving bounding the portfolio variance or downside risk metrics, which restricts the choice of assets for a given portfolio and the weighting thereof, whereas budgetary constraints involve setting the total weight of the asset weights equal to the total amount of capital available, with extra constraints such as prohibition on short selling imposed in some cases; transaction costs, whether linear or non-linear and fixed or proportional, reduce the gross return of an investor and encourage less frequent trading activity by creating areas of no transactions in the state space because of their costliness, which makes them a difficult addition to optimization problems because of their non-linearity and possibly concavity; these constraints call for advanced optimization methods such as convex optimization and dynamic programming that incorporate market impacts (Cao & Li, 2024).

The process of portfolio management is transformed into a learning or optimization problem through the formulation of the issue of asset allocation and investments as an optimization problem that involves maximization of expected utilities or gains under certain risk conditions, which are commonly formulated in dynamic terms; the transformation helps facilitate the use of sophisticated algorithms, including reinforcement learning and artificial intelligence techniques such as deep neural networks and hierarchy-based logic learning systems that learn optimal strategies or weights for portfolio selections through iterative learning based on historical and current financial data without necessarily adhering to conventional assumptions; the optimization problem is represented by a sequence of decisions made by the decision-maker in choosing actions for the portfolio to

maximize its gains using reward functions that can incorporate risk-related considerations in order to overcome the limitations of conventional static formulations (Gaggero et al., 2024; Rezaei & Nezamabadi-Pour, 2025).

II. DATA SOURCES AND FEATURE ENGINEERING FOR ML-BASED PORTFOLIO SYSTEMS

Portfolio management using machine learning typically incorporates various financial data sources that are essential for decision-making and optimization tasks, such as asset prices and returns, which reflect time-series properties and trends that are useful for predicting future outcomes; fundamental indicators such as profitability, dividend yields, and valuation ratios provide information about asset intrinsic values; market sentiment data extracted from news reports, social media platforms, and financial statements reflect investor behavior and market mood; macroeconomic variables such as interest rates, inflation rates, and growth indicators provide contextual information about market conditions; and alternative data such as volume data, order flow, and liquidity indicators help understand market microstructure characteristics; the diversity of data sources helps machine learning algorithms to model complex nonlinear relationships and adapt to changing market environments (Gaggero et al., 2024; Jangra et al., 2024; Kour, 2024).

Portfolio models combine technical, fundamental, and alternative datasets in integrating multiple sources of data for the purpose of improving asset selection and asset allocation strategies. Technical data, like price patterns and volume data, is often analyzed using machine learning techniques or statistical methods that can analyze momentum in the markets. Fundamental data, obtained from firm financial statements and relative financial strength measurements, helps in obtaining insight into the inherent value and competitive advantage of firms in the market and may be employed using methods like Data Envelopment Analysis (DEA). Alternative datasets include things like sentiment analysis of news articles and social media posts or entropy calculations from large language models, providing additional context to market data and capturing market psychology and complex market interactions beyond what is found in

traditional data sets. These alternative techniques, sometimes in combination with semantic augmentation using natural language processing techniques and robust optimization, produce better-performing portfolios through better prediction and portfolio models that mitigate issues such as estimation error and non-stationarity (Aiche et al., 2024; Cohen et al., 2025; Mohseny-Tonekabony et al., 2024).

The key preprocessing steps that are important in dealing with noise, missing values, and non-stationarity in financial data are data cleaning procedures that identify and eliminate irregularities, outliers, and errors in order to minimize noise and increase data accuracy; missing value imputation strategies that involve the replacement of missing values using simple methods like mean/mode replacement, interpolation, or complex modeling techniques, which aid in recovering missing data without compromising the integrity of the dataset; scaling and normalization procedures that standardize data scales to facilitate better comparison and learning by models; in order to manage non-stationarity, transformation strategies including differencing, detrending, or other specialized time series methodologies like decomposition and seasonal adjustment are applied to make data stationary and maintain consistent statistical behavior; moreover, feature extraction and engineering methods transform raw data into stationary and valuable representations; finally, automated preprocessing systems integrate various techniques based on data attributes to enhance forecasting and machine learning outcomes (Ortiz et al., 2024; Usmani et al., 2024).

III. MACHINE LEARNING MODELS FOR PORTFOLIO CONSTRUCTION

Some of the best supervised learning models suitable for predicting returns and risks in the financial market include deep learning models, such as GANs, which are good at predicting risks in terms of measures such as VaR, and VAEs, which are appropriate for predicting return rates; hybrid models that use traditional econometrics techniques, such as GARCH, and extreme value theory have been successful in predicting risks; ensemble methods, such as LightGBM augmented with attention-based mechanisms and boosting approaches, enable

comprehensive credit risk prediction by recognizing complex nonlinear patterns and handling class imbalances; recurrent neural networks, specifically LSTMs, efficiently integrate temporal financial information and improve price prediction capabilities when combined with unstructured data, such as financial news; white-box models, including logistic regression, continue to be effective for explaining credit risk prediction but are generally supplemented with other advanced black-box models to capture complex feature correlations; ultimately, the selection of modeling techniques is based on balancing predictive accuracy and explainability, and deep learning and hybrid techniques have become prevalent owing to their high accuracy in modeling financial market dynamics and risks (Yang et al., 2024; Ying et al., 2025;).

Unsupervised learning approaches can contribute to asset selection and diversification by revealing the hidden patterns and relationships present in financial markets without the need for labels; clustering and dimensionality reduction approaches can help discover groups of assets sharing common risk-reward characteristics or latent factors, allowing investors to create diversified portfolios that are exposed to different sources of risks; unsupervised algorithms can be used to select assets on the basis of topological or statistical characteristics of the financial market, thereby increasing the risk-reward ratio through diversification between separate asset groups or factors; moreover, unsupervised algorithms, specifically entropy maximization algorithms, can optimize the allocation of capital in a portfolio to ensure maximum diversification in order to mitigate the effects of parameter uncertainties on returns; finally, sophisticated machine learning algorithms, such as hierarchical risk parity algorithms, use interpretable algorithms to optimally allocate capital among multiple asset classes and maximize risk-adjusted returns while providing an explanation of individual asset contributions to the portfolio (Bhasin et al., 2024).

In finance, deep learning algorithms offer several benefits compared to conventional machine learning techniques because of their capacity to automatically extract hierarchical representations from vast amounts of high-dimensional data without needing substantial feature selection or engineering; various types of

artificial intelligence frameworks, including CNN, RNN, LSTM, transformer, GANs, can detect nonlinear relationships and dynamics in time series and learn latent patterns that improve the accuracy of machine learning tasks in areas such as algorithmic trading, risk management, fraud detection, and portfolio management; deep reinforcement learning can also optimize decision-making strategies through adaptive learning for automated trading systems; moreover, deep learning can process unstructured data such as financial reports, market sentiment, and alternative data sets, enhancing the predictive power of traditional quantitative data input by adding new features; yet, there are still limitations with interpretability and high computational costs, but developments in Explainable AI and model optimizations are overcoming these problems (Gao et al., 2024; E. Mienie et al., 2024).

IV. CHALLENGES, LIMITATIONS, AND FUTURE RESEARCH DIRECTIONS

Some of the significant obstacles in adopting ML-based approaches to stock portfolios involve problems related to model robustness and overfitting because of the noisy, volatile, and non-stationary financial market environment which may result in poor predictions; interpretability challenges, which arise due to the complexity of the ML models that tend to act like black boxes, and therefore lack the level of transparency needed in finance; data quality, which refers to incomplete, inaccurate, and even biased data that may influence training processes and reduce the model's predictive capacity; the dynamic nature of the stock market that requires models to be updated frequently due to changing correlations and relationships in the market; computational complexity associated with processing vast amounts of financial data; regulation, which necessitates model explainability and fairness; finally, the multi-dimensional nature of stock price determinants, such as psychological factors, macroeconomic trends, geopolitical events, among others, adds to the level of uncertainty that cannot be effectively accounted for by ML models.

Overfitting affects portfolio performance over a long period of time since it makes the model pick up noise and randomness inherent in the past training dataset instead of real signals coming from the markets, thus

failing to generalize well when applied to future unknown market scenarios; the outcome is that an overly complex and unpredictable portfolio will emerge, which may not be able to handle any possible shifts in market dynamics or structural changes, thereby being exposed to higher risks and lower returns per unit of risk over time; as such, portfolios formed using overfitted models will underperform benchmarks and will not withstand the changing economic environment due to their inability to perform consistently through market regimes (Eliza, 2024; Mats, 2024).

Explanation is vital for machine learning-powered financial systems due to the increased level of transparency, which allows all involved parties, including regulators, investors, and financial organizations, to comprehend, evaluate, and control the reasoning process behind their models' operations; since financial systems function within highly regulated settings, the need for transparent models is crucial not only from an ethical standpoint but also to comply with relevant legislation; besides, explanation allows for the discovery and rectification of potential biases, contributes to the effective management of risks by elucidating the model's behavior under various circumstances, and aids in debugging and validating models; therefore, explainable AI ensures trustworthiness and fair deployment of sophisticated ML algorithms within financial settings (Khan et al., 2025; Kour, 2024).

Potential research areas in developing intelligent portfolio management systems include applying multi-agent systems and distributed artificial intelligence, which will foster collaborative decision-making and risk assessment capabilities among varied assets. Another potential area is the utilization of advanced natural language processing algorithms to conduct a holistic analysis of financial news content to provide supplementary qualitative information to numerical data and facilitate sentiment-based investment allocation. Additionally, the integration of explainable AI technology and human-AI partnerships will aid in fostering.

V. OVERALL DISCUSSION

However, in the case of the EHBO algorithm, the convergence characteristics are quite favorable, and its performance does not depend much on the number of

independent runs. Despite the fact that the EHBO algorithm cannot guarantee the attainment of globally optimal solutions in all cases, the adaptability of the parameters contributes to its insensitivity to the initial conditions and performance instability, particularly in the case of multimodal and deceptive functions. This feature indicates that the EHBO algorithm places more emphasis on convergence stability and robustness rather than optimization. The test results indicate that the adaptability of the search parameters contributes to achieving a balance between exploration and exploitation in the optimization process. Moreover, the performance analysis demonstrates that the EHBO algorithm can act as a good starting point for further improvement, and hybridization techniques or problem-specific methods can be applied to increase its efficiency in solving multimodal optimization problems. transparency, regulatory compliance, and investor confidence levels in the process of asset management. The other potential research direction in portfolio management is utilizing federated learning to share data securely and privately between organizations to bolster modeling robustness while protecting sensitive data. Graph neural network approaches can be used to study the intricate interconnectivity between assets and transactions in financial markets. Finally, evolutionary algorithms, including genetic algorithms, can help in optimizing investment portfolios through dynamic portfolio construction by searching vast solution spaces efficiently.

VI. CONCLUSION AND FUTURE SCOPE

An Extended Honey Badger Optimization algorithm with a dynamic adaptive control mechanism to balance exploration and exploitation phases is suggested. The EHBO algorithm differs from the rest of the variants of HBA in terms of using dynamic adaptive control mechanism which is capable of balancing the exploration and exploitation phases of the search process. The proposed algorithm is investigated on several standard benchmark functions in comparison with other meta-heuristics such as PSO, DE, ABC, BA, and HBA. Results of experiments indicated that the new algorithm does not yield the best fitness values, but its advantages lie in providing a smoother convergence process, reduced performance variations, and better robustness. Statistical analysis revealed the

significant differences between the performances of the algorithms. The usage of adaptive control parameters enhances the robustness of algorithms and decreases their dependence on initializations. Future studies may include feedback control mechanisms for adaptive tuning, large-scale optimization, multi-objective optimization, and real-world applications like image segmentation, feature selection, and scheduling.

REFERENCES

- [1] W. Addy, A. Ajayi-Nifise, B. Bello, S. Tula, O. Odeyemi, and T. Falaiye, "Machine learning in financial markets: A critical review of algorithmic trading and risk management," *International Journal of Science and Research Archive*, vol. 11, no. 1, pp. 1853–1862, 2024, doi: 10.30574/ijrsra.2024.11.1.0292.
- [2] A. Aiche, Z. Winer, and G. Cohen, "Constructing cybersecurity stocks portfolio using AI," *Forecasting*, vol. 6, no. 4, pp. 1065–1077, 2024, doi: 10.3390/forecast6040053.
- [3] N. K. Bhasin, S. Kadyan, K. Santosh, R. Hp, R. Changala, and B. K. Bala, "Enhancing quantum machine learning algorithms for optimized financial portfolio management," in *Proc. Int. Conf. Intelligent and Innovative Computing Applications (INCOS)*, 2024, pp. 1–7, doi: 10.1109/INCOS59338.2024.10527612.
- [4] X. Cao and S. Li, "Neural networks for portfolio analysis with cardinality constraints," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 35, no. 12, pp. 17674–17687, 2024, doi: 10.1109/TNNLS.2023.3307192.
- [5] G. Cohen, A. Aiche, and R. Eichel, "Artificial intelligence models for predicting stock returns using fundamental, technical, and entropy-based strategies: A semantic-augmented hybrid approach," *Entropy*, vol. 27, no. 6, Art. no. 550, 2025, doi: 10.3390/e27060550.
- [6] E. Eliza, "Sustainable investment practices: Assessing the influence of ESG factors on financial performance," *Global International Journal of Innovative Research*, vol. 2, no. 7, pp. 1445–1454, 2024, doi: 10.59613/global.v2i7.228.
- [7] G. Gaggero, P. G. Giribone, M. Muselli, E. Ünal, and D. Verda, "Portfolio optimization and risk

- management through hierarchical risk parity and logic learning machine: A case study applied to the Turkish stock market,” *Risk Management Magazine*, vol. 19, no. 1, pp. 26–49, 2024, doi: 10.47473/2020rmm0137.
- [8] H. Gao, G. Kou, H. Liang, H. Zhang, X. Chao, C.-C. Li, and Y. Dong, “Machine learning in business and finance: A literature review and research opportunities,” *Financial Innovation*, vol. 10, no. 1, 2024, doi: 10.1186/s40854-024-00629-z.
- [9] G. Jangra, M. Irfan, M. Jangra, and C. Verma, “Artificial intelligence approach to portfolio management,” in *Artificial Intelligence Applications in Finance*, pp. 59–73. Hershey, PA, USA: IGI Global, 2024, doi: 10.4018/979-8-3693-3282-5.ch004.
- [10] F. S. Khan, S. S. Mazhar, K. Mazhar, D. A. Alsaleh, and A. Mazhar, “Model-agnostic explainable artificial intelligence methods in finance: A systematic review, recent developments, limitations, challenges and future directions,” *Artificial Intelligence Review*, vol. 58, no. 8, 2025, doi: 10.1007/s10462-025-11215-9.
- [11] M. Kour, “Challenges and opportunities of machine learning in the financial sector,” in *Artificial Intelligence Applications in Finance*, pp. 61–73. Hershey, PA, USA: IGI Global, 2024, doi: 10.4018/979-8-3693-1746-4.ch004.
- [12] V. Mats, “Hedge performance of different asset classes in varying economic conditions,” *Radioelectronic and Computer Systems*, vol. 2024, no. 1, pp. 217–234, 2024, doi: 10.32620/reks.2024.1.17.
- [13] L. J. Mena, V. García, V. G. Félix, R. Ostos, R. Martínez-Peláez, A. Ochoa-Brust, and P. Velarde-Alvarado, “Enhancing financial risk prediction with symbolic classifiers: Addressing class imbalance and the accuracy–interpretability trade-off,” *Humanities and Social Sciences Communications*, vol. 11, no. 1, 2024, doi: 10.1057/s41599-024-04047-5.
- [14] E. Mienye, N. Jere, G. Obaido, I. D. Mienye, and K. Aruleba, “Deep learning in finance: A survey of applications and techniques,” *AI*, vol. 5, no. 4, pp. 2066–2091, 2024, doi: 10.3390/ai5040101.
- [15] I. D. Mienye and N. Jere, “Deep learning for credit card fraud detection: A review of algorithms, challenges, and solutions,” *IEEE Access*, vol. 12, pp. 96893–96910, 2024, doi: 10.1109/ACCESS.2024.3426955.
- [16] N. Mohseny-Tonekabony, S. J. Sadjadi, E. Mohammadi, M. Tamiz, and D. F. Jones, “Robust, extended goal programming with uncertainty sets: An application to a multi-objective portfolio selection problem leveraging DEA,” *Annals of Operations Research*, vol. 346, no. 2, pp. 1497–1552, 2024, doi: 10.1007/s10479-023-05811-7.
- [17] M. K. Nallakaruppan, H. Chaturvedi, V. Grover, B. Balusamy, P. Jaraut, J. Bahadur, V. P. Meena, and I. A. Hameed, “Credit risk assessment and financial decision support using explainable artificial intelligence,” *Risks*, vol. 12, no. 10, Art. no. 164, 2024, doi: 10.3390/risks12100164.
- [18] B. L. Ortiz, V. Gupta, R. Kumar, A. Jalin, X. Cao, C. Ziegenbein, A. Singhal, M. Tewari, and S. W. Choi, “Data preprocessing techniques for AI and machine learning readiness: Scoping review of wearable sensor data in cancer care,” *JMIR mHealth and uHealth*, vol. 12, no. 5, Art. no. e59587, 2024, doi: 10.2196/59587.
- [19] M. Rezaei and H. Nezamabadi-Pour, “A taxonomy of literature reviews and experimental study of deep reinforcement learning in portfolio management,” *Artificial Intelligence Review*, vol. 58, no. 3, 2025, doi: 10.1007/s10462-024-11066-w.
- [20] Z. Rojan, “Financial fraud detection based on machine and deep learning: A review,” *Indonesian Journal of Computer Science*, vol. 13, no. 3, 2024, doi: 10.33022/ijcs.v13i3.4059.
- [21] S. K. Sahu, A. Mokhade, and N. D. Bokde, “An overview of machine learning, deep learning, and reinforcement learning-based techniques in quantitative finance: Recent progress and challenges,” *Applied Sciences*, vol. 13, no. 3, Art. no. 1956, 2023, doi: 10.3390/app13031956.
- [22] A. Singh, P. Gupta, and N. Thakur, “An empirical research and comprehensive analysis of stock market prediction using machine learning and deep learning techniques,” *IOP Conference Series: Materials Science and Engineering*, vol. 1022, no. 1, Art. no. 012098, 2021, doi: 10.1088/1757-899X/1022/1/012098.
- [23] M. Usmani, Z. A. Memon, A. Zulfiqar, and R. Qureshi, “Preptimize: Automation of time series

data preprocessing and forecasting,” Algorithms, vol. 17, no. 8, Art. no. 332, 2024, doi: 10.3390/a17080332.

- [24] H. Vachhani, M. S. Obiadat, A. Thakkar, V. Shah, R. Sojitra, J. Bhatia, and S. Tanwar, “Machine learning based stock market analysis: A short survey,” in Proc. Int. Conf. Data Management, Analytics and Innovation, pp. 12–26. Singapore: Springer, 2020, doi: 10.1007/978-3-030-38040-3_2.