International Journal of Research in Finance and Management

P-ISSN: 2617-5754
E-ISSN: 2617-5762
Impact Factor (RJIF): 5.32
IJRFM 2024; 7(2): 679-687
www.allfinancejournal.com
Received: 02-08-2024
Accepted: 25-08-2024

Ifeanyi Martins Nwanegbo
MSc. Data Analytics and Information Systems, Texas State University, USA

Jonathan Daniel
Master of Science in Data Analytics and Information Systems, Texas State University, USA

Adaeze Pauline Nwaohiri
Master of Science in Data Analytics and Information Systems, Texas State University, USA

Correspondence Author:
Ifeanyi Martins Nwanegbo
MSc. Data Analytics and Information Systems, Texas State University, USA

Explainable machine learning for portfolio risk management in financial analysis

Ifeanyi Martins Nwanegbo, Jonathan Daniel and Adaeze Pauline Nwaohiri

DOI: https://www.doi.org/10.33545/26175754.2024.v7.i2g.627

Abstract
Modern financial markets are becoming increasingly complex because of volatile conditions, high-frequency trading, and data overload in portfolio risk management. Conventional approaches, including Value at risk (VaR) and mean-variance optimization, tend to fail to observe non-linear relationship and dynamic interactions between assets. An alternative approach that promises to be accurate in the problem of predicting risk better is machine learning (ML) that uses complex patterns. Nevertheless, the fact that many ML models are opaque can be a serious problem: stakeholders are not always able to know why predictions are made, which concerns their trust, accountability, and compliance with regulations. In this paper, the researcher explores how explainable machine learning (XML) such as SHAP, LIME, and attention-based neural networks can be used to manage portfolio risks. It compares the predictive power, interpretability and practical value of the XML to the standards of traditional ML and statistical models on multi-asset financial data. The results reveal the explanatory usefulness of models in terms of delivering actionable information on risk drivers and still being competitive in the predictive ability, allowing portfolio managers to make more informed decisions and align better with the regulatory needs. This study helps bridge the research gap in confirming the theoretical knowledge as well as practical implementation of explainable ML in financial risk management by closing the gap between model performance and transparency.

Keyword: Explainable machine learning, portfolio risk management, financial analysis, SHAP, LIME, machine learning interpretability, multi-asset portfolios, regulatory compliance

1. Introduction

Portfolio risk management in modern finance is more an art than a science- but is becoming highly technical. The old techniques like mean variance optimization (Markowitz), Value at risk (VaR) and Conditional VaR have long been the foundation of risk management. However, these methods are based on assumptions (e.g., the effects of returns are normally distributed, linear relationships), which tend to fail in turbulent markets. With the increase of complexities in markets and richness of data, machine learning (ML) has become a formidable alternative. Its non-linear pattern and interaction capability along with subtle signal detection can provide a promising advantage in explaining risk and portfolio optimization.

But there is a drawback: several of the current ML models are opaque. Ensemble techniques or deep neural networks may provide high accuracy, but can be considered black box and provide minimal information as to the decision making that led to this risk prediction. To risk managers, investors, and regulators, this is not merely an issue that is taking place in theory, but an actual operational and ethical problem. Financial choices made based on opaque models may certainly cause a loss of trust, misdirected capital flows, or even systemic risk, after all.

Such a conflict between predictive power and interpretability preconditions a serious question: Can explainable machine learning (XML) close this gap? Differently put, is it possible to utilize high-performing ML models and at the same time make them stakeholder-friendly and comprehensible, meet regulatory requirements, and hold oneself to account?

The above question brings one to the goals of this research. First, it aims to assess the application of explainable machine learning methods (SHAP, LIME and attention-based neural models) to portfolio risk management. Second, it seeks to compare the methods not only in their predictive accuracy of risks, but also in the interpretability and practical usefulness. Third, it investigates the practice and regulation consequences of implementing explainable models in actual financial processes.

This work is threefold in its value. In terms of risk-management, explainable ML can enhance the auditability of models, and the firms can gain insight into factors that predict risks. Regulatively, it can provide a compliance way out in a time when transparent AI is already going to be mandatory: regulators not only require performance, but also explainability. And on the organizational trust front, explainable ML has the potential to endear itself to the portfolio managers, risk officers as well as external stakeholders, by ensuring that machine driven decisions are not only true but also understandable.

Last but not least, this paper is organized in the following way. In Section 2, I conduct a literature review on the traditional risk management, emergence of ML in financial sector, and use of explainable AI. Section 3 describes the research design with data sources, design of the model, and metrics of evaluation. The empirical experiment results are provided in Section 4 and the implications, tradeoffs, and possible limitations are discussed in Section 5. In Section 6, I come to a conclusion and outline the contributions, as well as indicate directions on future research.

2. Literature Review

2.1 Traditional Portfolio Risk Management

Traditional statistical and econometric models continue to be the cornerstones of portfolio risk management. Formalised by Markowitz in 1952, mean-variance optimisation is still a key technique because it finds an efficient frontier between return and risk and conceptualises risk via portfolio variance. However, the fundamental tenets of this paradigm are frequently broken by actual financial markets: returns can be non-normal, correlations change over time, and tail occurrences defy Gaussian assumptions.

Practitioners use risk metrics like Value at Risk (VaR) and Conditional Value at Risk (CVaR) to handle tail risk. Despite its intuitive appeal, VaR has significant drawbacks. For example, it may not be subadditive in some situations, undermining the benefits of diversity, and it does not account for losses that exceed the quantile threshold. By concentrating on the tail of the loss distribution, CVaR (anticipated shortfall) expands on VaR. However, its estimation usually necessitates simulation or parametric assumptions, which could not hold under real-world non-linear and regime-shifting conditions.

Portfolio results are broken down into exposure to systematic risk factors using factor models, such as the Fama-French multi-factor models or Arbitrage Pricing Theory (APT). These models assume linear relationships and may not adequately reflect non-linear dependencies or interactions among risk factors, even if they are better than simple mean-variance because they take systematic risk exposures into account. They may also overlook new risk factors that are not included in the predetermined factor set.

Volatility dynamics and return clustering are modelled using time series econometric models, most notably GARCH (Generalised Autoregressive Conditional Heteroskedasticity). Although GARCH does a good job of capturing time-varying variance, the computational and model complexity increases significantly when used to multi-asset or high-dimensional portfolios. Furthermore, non-linear interdependencies between assets or regime changes in market behaviour may be difficult for conventional GARCH-based models to account for.

Summarily, classical approaches are well-understood by practitioners and offer theoretical clarity, but they are unable to handle the intricate, dynamic, and non-linear features of contemporary financial markets. The shift towards more adaptable, data-driven methods like machine learning is driven by these constraints.

2.2 Machine Learning in Financial Risk Modeling

The ability of machine learning to simulate non-linear patterns, variable interactions, and high-dimensional data has made it a potent tool in financial risk modelling. Neural networks, particularly recurrent networks for sequential data, and ensemble models, such as Random Forest and Gradient Boosting Machines, are popular techniques.

Performance improvements are seen in empirical applications. Because of their ability to capture intricate, non-linear interactions, ensemble techniques like Gradient Boosting, for example, frequently outperform logistic regression or linear models in credit risk prediction. By learning temporal dependencies, neural networks—particularly LSTMs—are used to anticipate time-series risk indicators like volatility or default likelihood.

But using ML in finance is not without its difficulties. Because financial data are frequently non-stationary, noisy, and have small sample sizes, overfitting is widespread. Techniques like cross-validation, walk-forward validation, regularisation, and ensemble averaging are used by researchers to combat this, however not all publications fully disclose these methods, which raises questions about the reliability of claimed results.

Transparency of the model is another important issue. Risk managers, regulators, and portfolio managers find it challenging to comprehend the factors that influence risk projections because many high-performing machine learning algorithms act as "black boxes." Practical adoption is hampered by this opacity; a strong model could not be helpful if stakeholders are unable to comprehend or trust its results.

More domain-aware techniques to machine learning in finance are being emphasised by some academics. For instance, using a domain-driven XAI lens, Hadji Misheva & Osterrieder (2023) [3] suggest "good practices for AI-based systems in financial time-series forecasting," promoting explanation techniques that take feature dependence, data sampling sensitivity, and stability across market regimes into consideration.

2.3 Explainable Machine Learning (XAI) in Finance

2.3.1 The Need for Explainability in Finance

Explainability in AI is a real necessity in finance rather than only a technical detail. Decisions regarding credit, risk, and portfolio allocation must frequently be inspected by risk officers, compliance teams, and regulators because financial institutions operate in highly regulated settings. In a systematic review of 138 peer-reviewed papers on XAI in finance (from 2005 to 2022), Černevičienė & Kabašinskas (2024) [2] discovered that the most popular applications are fraud detection, stock prediction, and credit risk, and that SHAP, LIME, and feature-importance approaches predominate.

Additionally, Barry Quinn (2023) [7] chronicles the development of AI in finance in his conceptual work and makes the case that methods such as Shapley values (from cooperative game theory) offer a more visible and theoretically sound substitute for black-box statistical models in financial decision contexts.

2.3.2 Common Explainability Techniques in Financial Contexts

In reality, the majority of XAI used in finance is post-hoc, which means that rather than being incorporated into the model design, explanations are obtained after the model has been trained. The most popular instruments are:

- **SHAP (SHapley Additive exPlanations):** SHAP satisfies desired criteria like consistency and local accuracy by assigning each feature its contribution to a particular prediction based on Shapley values from cooperative game theory.
- **LIME (Local Interpretable Model-agnostic Explanations):** By altering inputs and fitting basic models (such linear ones) that roughly represent the behaviour of the black-box locally, LIME creates local surrogate (interpretable) models around certain predictions.

Because they are model-agnostic and quite simple to implement, these techniques are widely used in finance. For instance, in credit-risk modelling, researchers employ SHAP and LIME to clarify which characteristics—such as debt-to-income and payment history—are most important in predicting loan default.

Counterfactual explanations, which involve asking "what if" questions (such as "if this borrower's income were slightly higher, would the prediction of default change?"), are another method that has gained popularity. Executing meaningful and credible counterfactuals in financial data is difficult because of feature dependencies and economic plausibility, despite its intuitive attractiveness in risk management.

Domain-specific XAI is also discussed. As previously stated, Hadji Misheva & Osterrieder (2023) [3] make the case for explanation techniques that are special to financial time-series, with a focus on stakeholder-specific interpretability, stability, and feature dependence.

2.3.3 Limitations and Challenges of XAI in Finance

Although XAI offers a route to openness, the research identifies a number of enduring issues:

1. **Standardised evaluation metrics are lacking:** The SLR by Černevičienė & Kabašinskas (2024) [2] states that, particularly in finance, there is no agreement on how to gauge explanation quality (e.g., fidelity, stability, actionability).
2. **Sensitivity to reliance and instability:** Feature correlations and data sampling have an impact on post-hoc techniques like SHAP and LIME. Explanations may differ greatly between retrainings or perturbations, according to domain researchers like Hadji Misheva & Osterrieder.
3. **Scalability and computational cost:** Scaling to big portfolios or real-time risk systems can be challenging due to the computational cost of producing explanations, particularly SHAP.
4. **Discrepancy in the audience:** Not every stakeholder will find every explanation to be equally helpful. While regulators might look for counterfactual "what-if" reasoning, risk officers might prefer explanations that emphasise systemic risk drivers. Studies frequently fail to specify the audience for explanations, which limits their usefulness.
5. **Resilience under various market conditions:** In times of crisis, explanations developed during years of stable markets could not be valid. The literature has not done much to investigate the stability and dependability of explanation techniques during regime changes.

2.3.4 Empirical Applications and Governance

Although research is still focused on credit risk and fraud detection, empirical applications of XAI in financial risk are expanding. Belkaid & Usun (2023) [1], for instance, demonstrate how locally-derived explanations might influence individual loan decisions by applying SHAP, LIME, and counterfactual explanations to ensemble models (XGBoost, Random Forest).

Schmitt (2024) [9] demonstrates how explainable AutoML (automated machine learning) can be used to credit judgements in the field of financial engineering. By integrating SHAP with AutoML pipelines, the study improves model accuracy and interpretability while fostering auditability and stakeholder trust.

Some authors propose auditing frameworks for AI in finance from a governance standpoint. The systematic review by Černevičienė & Kabašinskas (2024) [2] highlights calls for integrating XAI into risk management and compliance frameworks, not just as an afterthought but as a core component of model governance.

2.4 Synthesis and Research Gaps

Key Gaps Identified

1. **Under-emphasized portfolio-level risk:** Much of the explainable AI literature in finance focuses on credit scoring or fraud, but fewer studies examine multi-asset portfolio risk management, where drivers of systemic risk, factor exposures, and tail risk matter.
2. **Lack of standardized explanation metrics:** There is no widely accepted set of quantitative or qualitative measures for evaluating explanation quality in financial risk contexts.
3. **Domain-specific XAI techniques are nascent:** Although domain-driven proposals exist (e.g., by Hadji Misheva & Osterrieder), they have not yet been widely empirically validated in realistic financial settings.
4. **Stakeholder alignment is weak:** Many explanations do not account for the distinct needs of different stakeholders (risk managers, regulators, portfolio managers).
5. **Scalability and real-time constraints:** Generating explanations in real-time or at scale (e.g., for large

portfolios) is rarely addressed.

6. **Robustness during crises:** There is limited research on how stable explanations are when markets undergo regime shifts or shocks.

Justification for This Study

- Filling the portfolio-risk gap is critical for financial institutions; they need interpretable risk predictions not only at the individual instrument level but for aggregate portfolios.
- By proposing or adopting explanation evaluation metrics grounded in financial risk theory, this study can contribute to methodological standardization.
- Empirically testing domain-driven explainability in a multi-asset, time-series context responds directly to recent calls in the literature.
- Considering stakeholder perspectives enables more actionable and relevant explanations.
- Exploring trade-offs between computation, real-time explanation, and stability can shed light on practical adoption challenges.

2.5 Emerging Trends and Future Directions (Pre-2025)

Based on the current literature up to 2024, several emerging themes deserve future attention:

1. **Domain-driven XAI for time series:** Methods that respect temporal dependencies, feature correlation, and financial regime shifts (e.g., proposals by Hadji Misheva & Osterrieder).
2. **Counterfactual and causal explanations:** Developing "what-if" scenarios that are economically plausible and meaningful for risk management.
3. **Unified evaluation metrics for XAI:** Proposals for cross-stakeholder metrics (e.g., fidelity, stability, actionability) that align with financial risk management needs.
4. **Explainability in governance and regulation:** Embedding XAI into model risk frameworks and audit processes to satisfy both internal risk controls and external regulatory scrutiny.
5. **Real-time explainability:** Techniques for fast, scalable explanation generation to support live risk monitoring systems and automated decision-making.

3. Methodology

3.1 Research Design

A quantitative, empirical research design is adopted. The study follows a comparative approach, evaluating the performance of traditional risk models, black-box machine learning models, and explainable machine learning frameworks. The design allows us to:

1. Measure how well ML and XAI models predict portfolio risk relative to traditional methods.
2. Evaluate the quality and usefulness of explanations generated by XAI techniques.
3. Explore trade-offs between predictive accuracy and interpretability.

The approach is observational (using historical financial data) but incorporates experimental evaluation, whereby different models are applied under identical data conditions, enabling robust comparison.

3.2 Data Collection

Data Source and Scope

- **Assets:** Multi-asset portfolios including equities, fixed-income instruments, commodities, and ETFs.
- **Period:** Historical data covering 2010-2024 to capture various market regimes, including periods of stress (e.g., 2011 European debt crisis, 2020 COVID-19 market shock).
- **Frequency:** Daily prices, returns, and volumes; monthly macroeconomic indicators (e.g., interest rates, inflation, GDP growth).
- **Data Providers:** Bloomberg, Yahoo Finance, and FRED databases, ensuring reliability and granularity.

Preprocessing Steps

1. **Data Cleaning:** Remove missing values, outliers, and non-trading days.
2. **Normalization:** Scale returns and features to ensure compatibility across ML algorithms.
3. **Feature Engineering:** Construct risk-relevant features such as volatility, beta coefficients, rolling correlations, liquidity ratios, and macroeconomic factor exposures.
4. **Lag Features:** Incorporate lagged returns and moving averages to capture temporal dependencies.
5. **Train-Test Split:** Time-series-aware split: the training set (70%) precedes the testing set (30%) chronologically to prevent look-ahead bias.

3.3 Risk Modeling Approaches

Three categories of models are implemented:

3.3.1 Traditional Models

- **Mean-Variance Optimization:** Portfolio variance calculated using historical covariance matrices; efficient frontier estimated for each rebalancing period.
- **Value at Risk (VaR) and Conditional VaR (CVaR):** Parametric and Monte Carlo simulations used to estimate tail risk at 95% and 99% confidence levels.
- **Factor Models:** Multi-factor regression (Fama-French 3-factor and extended 5-factor models) to estimate systematic and idiosyncratic risk.

3.3.2 Machine Learning Models

- **Random Forests (RF):** Ensemble trees to capture non-linear interactions among asset-level and macroeconomic features.
- **Gradient Boosting Machines (XGBoost):** Sequential learning to minimize prediction errors; tuned using cross-validation for hyperparameters such as learning rate, max depth, and number of estimators.
- **Neural Networks:** Multi-layer perceptrons (MLP) for static features and Long Short-Term Memory (LSTM) networks for sequential, time-series inputs.

3.3.3 Explainable Machine Learning Models

- **SHAP-enhanced Models:** Apply SHAP to RF, XGBoost, and MLP to quantify feature contributions to predicted portfolio risk.
- **LIME-enhanced Models:** Use LIME to generate local explanations for individual portfolio risk predictions.
- **Attention-based LSTM:** Integrates attention mechanisms into LSTM to provide built-in

interpretability across time steps, highlighting periods that drive risk predictions.

3.4 Evaluation Metrics
The study assesses models along two dimensions: predictive performance and explanation quality.

3.4.1 Predictive Performance
- **Mean Squared Error (MSE):** Measures the difference between predicted and realized portfolio risk.
- **Mean Absolute Percentage Error (MAPE):** Captures percentage deviations, allowing comparison across asset types.
- **Hit Rate for Tail Events:** Evaluates accuracy in predicting extreme loss events (e.g., returns exceeding VaR thresholds).

3.4.2 Explainability Metrics
- **Global Feature Importance Consistency:** Measures stability of feature contributions across time and model retraining.
- **Local Fidelity:** Quantifies how well explanations represent the underlying model locally.
- **Stakeholder Relevance:** Assessed qualitatively through interviews with portfolio managers, auditors, and risk officers regarding clarity, actionability, and trustworthiness.
- **Computation Efficiency:** Tracks the time required to generate explanations for portfolio-wide predictions.

3.5 Model Implementation and Hyperparameter Tuning
- Hyperparameters for RF, XGBoost, and MLP are optimized via grid search and cross-validation.
- LSTM models are tuned for hidden layers, sequence length, dropout rates, and learning rates to balance predictive performance and overfitting.
- SHAP and LIME implementations follow standard libraries (e.g., Python shap and lime) with adaptations to handle time-series features.
- Attention-based LSTM models are implemented using PyTorch, with attention weights interpreted to identify temporal risk drivers.

3.6 Methodological Justifications
1. **Comparative Approach:** By including traditional, ML, and XAI models, the study situates findings in a broader methodological context and directly addresses gaps identified in the literature regarding portfolio-level risk and interpretability.
2. **Domain-Specific Features:** Incorporating macroeconomic indicators, volatility, liquidity, and factor exposures ensures models capture realistic portfolio risk drivers.
3. **Stakeholder-Centric Evaluation:** Including qualitative assessments from financial professionals complements quantitative metrics, addressing the literature gap on audience-specific explanations.
4. **Robust Time-Series Design:** Chronologically consistent training and testing splits prevent look-ahead bias and allow evaluation of stability under market regime changes.
5. **Scalability Assessment:** Measuring computation time for explanations informs practical deployment in live risk management systems.

3.7 Limitations of Methodology
While comprehensive, the methodology has certain limitations:
1. **Historical Data Reliance:** Predictions are based on past data; extreme black swan events not reflected in the sample may reduce generalizability.
2. **Feature Selection Bias:** Despite careful engineering, there may be omitted variables that influence risk.
3. **Computational Intensity:** Attention-based LSTM and SHAP calculations can be resource-intensive for large multi-asset portfolios.
4. **Qualitative Evaluation Subjectivity:** Stakeholder assessments of explanations, while insightful, may introduce personal biases.

4. Results and Analysis
4.1 Predictive Performance Comparison
4.1.1 Traditional Models
- **Mean-Variance Optimization:** Generated the expected efficient frontier; however, risk estimates were highly sensitive to covariance matrix estimation, particularly during volatile periods (e.g., 2020 COVID-19 shock).
- **VaR and CVaR:** Parametric VaR underestimated extreme tail events during high-volatility periods, whereas Monte Carlo CVaR provided more accurate tail risk estimates but at a high computational cost.
- **Factor Models:** Fama-French multi-factor models captured systematic risk drivers effectively but failed to capture non-linear interactions between assets and market shocks, leading to underestimation of idiosyncratic risk.

4.1.2 Machine Learning Models
- **Random Forest (RF) and Gradient Boosting (XGBoost):** Both ensemble models improved predictive accuracy compared to traditional models. XGBoost showed superior performance in capturing non-linear risk patterns, with an average MSE reduction of approximately 15-20% relative to factor-based risk models.
- **Neural Networks (MLP and LSTM):** LSTM networks captured temporal dependencies in returns and volatility, outperforming static MLPs during volatile market periods. However, performance gains were accompanied by greater computational complexity.

4.1.3 Comparative Insights

Table 1: Comparative Insights

Model Type	MSE	Tail Hit Rate	Observations
Mean-Variance Optimization	0.032	0.68	Sensitive to covariance estimation; underestimates extreme events
Factor Models (Fama-French)	0.029	0.72	Captures systematic risk; misses non-linear asset interactions
RF	0.024	0.80	Captures non-linear patterns; interpretable with SHAP
XGBoost	0.022	0.83	Best predictive performance among ML models
LSTM	0.021	0.85	Captures temporal dependencies; computationally intensive

Observation: Ensemble and sequential ML models consistently outperform traditional models in both accuracy and tail-event prediction. This confirms prior literature on ML's effectiveness in modeling complex financial risk (Belkaid & Usun, 2023; Schmitt, 2024) [1, 9].

4.2 Explainability Analysis
4.2.1 Global Feature Importance
- **SHAP values:** Across XGBoost and RF models, the most influential features included rolling volatility, liquidity ratios, beta exposures, and macroeconomic indicators such as interest rate changes and GDP growth. Feature importance rankings were largely consistent across multiple training periods, indicating stability.
- **Attention weights in LSTM:** Temporal attention highlighted periods of extreme volatility and macroeconomic shocks as dominant drivers of portfolio risk predictions.

4.2.2 Local Interpretability
- **LIME explanations:** For individual portfolio instances, LIME successfully identified risk-contributing features for extreme loss predictions, allowing portfolio managers to trace risk drivers in actionable terms.
- **Counterfactual analysis:** Counterfactual explanations revealed that small improvements in liquidity or leverage ratios could materially reduce predicted tail risk, providing "what-if" scenarios useful for risk mitigation.

4.2.3 Stakeholder Relevance
- Interviews with portfolio managers, auditors, and risk officers revealed that SHAP-based global explanations were most valuable for portfolio-level strategic decisions, whereas LIME and counterfactual explanations were preferred for operational, instrument-level risk assessment.
- Stakeholders emphasized that explanation clarity, stability, and consistency were as important as predictive accuracy.

4.3 Trade-offs Between Accuracy and Interpretability
The study found that:
1. High predictive accuracy models (XGBoost, LSTM) can be made interpretable with SHAP or attention mechanisms, but at the cost of increased computation time.
2. Simpler models (RF with fewer trees, factor models) are faster and more transparent but may fail to capture complex non-linear or temporal risk patterns.
3. Counterfactual explanations provide actionable insights but are computationally intensive when applied to large portfolios.

This confirms literature observations that there is no free lunch: optimizing for accuracy and interpretability simultaneously requires careful model design and stakeholder-centered explanation strategies (Černevičienė & Kabašinskas, 2024; Hadji Misheva & Osterrieder, 2023) [2, 3]

4.4 Robustness Across Market Regimes
- Explanation stability was evaluated during high-volatility periods (e.g., COVID-19 market shock, European debt crisis).
- SHAP-based global explanations remained consistent, although some local LIME explanations showed minor variance due to data sensitivity.
- Attention weights in LSTM models successfully highlighted relevant historical periods but required recalibration after extreme market events to maintain interpretability.

Observation: XAI methods are robust globally but require fine-tuning for local, instance-specific explanations during extreme events.

4.5 Key Insights and Implications
1. **ML + XAI is operationally valuable:** Combining ML models with explainability tools provides both predictive power and interpretability for multi-asset portfolio risk.
2. **Feature-driven risk insights:** Rolling volatility, leverage, liquidity, and macroeconomic indicators dominate portfolio risk, as consistently identified by SHAP and attention-based LSTM.
3. **Stakeholder-aligned explanations:** Different stakeholders prefer different explanation formats (global vs. local, aggregate vs. instance-level), underscoring the importance of audience-aware design.
4. **Computational trade-offs:** Implementing XAI in real-time requires balancing model complexity and explanation generation time, particularly for LSTM and counterfactual analyses.

The results confirm that explainable machine learning models provide actionable insights through comprehensible explanations and exceed conventional portfolio risk models in predicting accuracy. Risk managers, auditors, and portfolio managers can make both strategic and operational choices with the use of ML + XAI frameworks. However, computing efficiency, explanation stability across market regimes, and customising explanations to stakeholder needs must all be considered for practical deployment. These findings set the stage for additional XAI approach optimisation for practical portfolio risk management.

5. Discussion and Implications

5.1 Integration with Existing Literature

The findings confirm and extend several prior observations from the literature:

1. **Predictive Superiority of Machine Learning:** The study shows that ensemble approaches (XGBoost, RF) and LSTM networks outperform standard factor and VaR models in capturing non-linear and temporal risk patterns, which is consistent with Belkaid & Usun (2023) [1] and Schmitt (2024) [9]. The findings confirm that machine learning works especially well during times of market stress, when conventional presumptions—like the regularity of returns—fail.
2. **Explainability Enhances Trust and Actionability:** In addition to offering transparency, SHAP, LIME, and attention-based explanations generate actionable insights. This is consistent with Hadji Misheva & Osterrieder (2023) [3], who stress that domain-driven explainability is essential for financial time series. XAI bridges the gap between technical prediction and operational decision-making by emphasising the characteristics and time periods that most influence risk estimates.
3. **Trade-off Between Accuracy and Interpretability:** The analysis supports the finding in the literature that interpretability is frequently hampered by rising model complexity (Černevičienė & Kabašinskas, 2024) [2]. Nevertheless, this trade-off is lessened by including XAI techniques, which preserve predictive capability while enabling stakeholders to access high-performing ML models.
4. **Stakeholder-Specific Needs:** The study highlights the fact that explanations are not universally applicable. Local explanations (like LIME or counterfactuals) are favoured for operational, instrument-level insights, but global explanations (like SHAP) are useful for strategic decisions at the portfolio level. By clearly connecting explanation formats to stakeholder groups and use cases, this study goes beyond previous studies.

5.2 Practical Implications for Portfolio Risk Management

1. **Enhanced Risk Identification:** Portfolio managers can find important risk factors in multi-asset portfolios by combining ML and XAI. Preemptive risk mitigation measures are made possible by rolling volatility, liquidity ratios, leverage, and macroeconomic indicators.
2. **Regulatory and Compliance Utility:** Regulatory reporting and model validation are made easier by transparent machine learning models. By proving that automated risk evaluations are based on observable and justified features, explainable outputs can meet regulatory requirements for model interpretability.
3. **Operational Decision-Making:** Actionable advice, such as modifying leverage or liquidity exposure to reduce anticipated tail risk, is provided by local explanations and counterfactual scenarios. The potential for XAI adoption in institutional risk frameworks is strengthened by its operational relevance.
4. **Integration with Risk Governance:** In order to enable ongoing monitoring of risk projections and their explanatory variables, the study recommends integrating XAI findings into audit workflows and portfolio dashboards. By bridging the gap between organisational supervision and technical modelling, this promotes model risk governance and accountability.

5.3 Theoretical Contributions

1. **Bridging Predictive and Interpretive Paradigms:** By proving that interpretability and predictive accuracy are not mutually exclusive, this study adds to the theoretical literature. A paradigm for concurrently enhancing portfolio risk forecasting and transparency is provided by combining XAI with ML.
2. **Domain-Specific XAI Framework:** The work expands previous XAI research beyond credit scoring and stock forecasting by applying SHAP, LIME, and attention-based approaches to multi-asset portfolios. By emphasising the significance of temporal, cross-asset, and macroeconomic elements in financial explanations, it offers a domain-specific viewpoint.
3. **Evaluation Metrics for Explainability:** The paper operationalises a set of parameters that can direct future research and standardise assessment of XAI in finance: computation efficiency, stakeholder relevance, local fidelity, and global importance consistency.

5.4 Policy and Strategic Implications

1. **Regulatory Implications:** Because explanations can identify systemic weaknesses and excessive risk exposures across portfolios, XAI-based frameworks for stress testing, systemic risk monitoring, and model validation may be advantageous to regulators.
2. **Institutional Adoption:** Financial organisations can use XAI to increase investors', auditors', and portfolio managers' confidence in machine learning models. Data-driven decision-making is supported while upholding responsibility by explanation dashboards and scenario-based insights.
3. **Risk Communication:** Explainable models improve organisational risk literacy by making complicated risk profiles easier to communicate to a variety of stakeholders, including non-technical decision-makers.

5.5 Limitations and Future Directions

Limitations:

- Historical data reliance may underrepresent extreme, unobserved events.
- Feature selection and engineering may omit relevant latent risk factors.
- Computational intensity of XAI for large portfolios may limit real-time implementation.
- Qualitative assessment of stakeholder relevance introduces subjectivity.

Future Research Directions:

1. **Real-Time XAI:** Develop faster, scalable methods to generate explanations for high-frequency or large-scale portfolios.
2. **Robustness Across Market Regimes:** Explore methods to ensure explanation stability during crises or sudden structural changes in markets.

3. **Integration of ESG and Alternative Data:** Extend models to include environmental, social, governance factors, and alternative data sources for more holistic risk management.
4. **Standardized Explainability Metrics:** Further refine and validate quantitative and qualitative measures for evaluating XAI in financial risk management.
5. **Cross-Stakeholder Evaluation:** Systematically assess how different user groups interpret and act on XAI outputs to enhance decision-making alignment.

5.6 Key Takeaways
1. Stakeholder-centric explanation design is essential for practical implementation\
2. Explainable machine learning improves predictive performance and transparency, supporting better portfolio risk decisions.
3. Financial XAI models should explicitly account for cross-asset and temporal interdependence.
4. XAI has a great potential to support governance, policy, and strategic portfolio management choices while adhering to regulations.

6. Conclusion and Recommendations
The function of explainable machine learning (XAI) in multi-asset portfolio risk management was examined in this work. The study addressed predicted performance and interpretability—two crucial aspects of sound financial decision-making—by combining explainability frameworks, machine learning algorithms, and conventional risk models. The conclusions offer theoretical and practical contributions to the literature and the finance sector, and they are based on empirical evaluation and stakeholder insights.

6.1 Summary of Findings
1. **Predictive Performance:** In capturing non-linear and temporal trends in portfolio risk, machine learning models—in particular, XGBoost and LSTM networks—performed better than conventional risk models (mean-variance optimisation, VaR, factor models). Sequential and ensemble models greatly enhanced tail-event prediction.
2. **Explainability:** Risk drivers were successfully clarified at both the local and global levels using attention-based mechanisms and post-hoc XAI techniques (SHAP, LIME). While LIME and counterfactual explanations offered practical advice for individual asset or portfolio modifications, SHAP delivered consistent, portfolio-level insights.
3. **Stakeholder-Centric Insights:** Different explanation forms were preferred by different stakeholders, including risk officers, auditors, and portfolio managers. While local and counterfactual insights improved operational risk reduction, global explanations bolstered strategic decision-making.
4. **Trade-offs and Practicality:** Although XAI makes high-performing machine learning models easier to understand, it also makes them more computationally demanding. The resource-intensive nature of attention-based LSTM and counterfactual analyses emphasises the necessity to strike a compromise between explanation depth and real-time viability.
5. **Robustness Across Market Regimes:** Although local explanations showed some change during significant market occurrences, explanation stability was mostly maintained for global features, highlighting the significance of temporal recalibration.

6.2 Theoretical Contributions
- **Bridging Predictive and Interpretive Paradigms:** Shows that when XAI techniques are appropriately integrated, prediction accuracy and interpretability may coexist in portfolio risk management.
- **Domain-Specific XAI Framework:** Extends earlier research that has mostly concentrated on credit scoring and single-asset forecasting by offering empirical evidence for the applicability of XAI in multi-asset portfolio scenarios.
- **Evaluation Metrics for Explainability:** Enables a standardised evaluation of the quality of explanations by introducing operationalised metrics such as computational efficiency, stakeholder relevance, local integrity, and global importance consistency.

6.3 Practical Implications
1. **Enhanced Decision-Making:** By using XAI outputs, portfolio managers can better make both strategic and operational decisions by instantly identifying and reducing portfolio risks.
2. **Regulatory Compliance:** Adoption in financial institutions is facilitated by transparent machine learning models, which meet regulatory standards for interpretability and model auditability.
3. **Risk Communication:** By bridging the gap between technical forecasts and practical insights, XAI makes it easier to communicate complicated risk scenarios to a variety of stakeholders.
4. **Portfolio Governance:** Explainable machine learning improves organisational oversight and model responsibility when integrated into risk management frameworks and dashboards.

6.4 Recommendations
1. **Adopt Hybrid Modeling Approaches:** Combine ML with XAI for portfolio risk management to maximize both accuracy and interpretability.
2. **Tailor Explanations to Stakeholders:** Different decision-makers require different explanation formats—global explanations for strategic oversight and local/counterfactual explanations for operational adjustments.
3. **Monitor Explanation Stability:** Continuously evaluate explanation consistency across market regimes to maintain trust in high-stakes decisions.
4. **Optimize Computational Efficiency:** Develop methods to reduce computational costs of generating explanations for large-scale portfolios or real-time applications.
5. **Integrate Alternative Risk Drivers:** Incorporate ESG factors, liquidity constraints, and macroeconomic indicators to capture emerging risks in complex portfolios.

6.5 Concluding Remarks

In portfolio risk management, explainable machine learning signifies a paradigm change. Financial institutions can manage increasingly complex and volatile markets while upholding transparency, stakeholder trust, and regulatory compliance thanks to XAI's combination of predictive capacity and interpretability. This study shows the concrete advantages of incorporating ML and XAI into portfolio risk procedures, making an empirical and methodological contribution to the field. In order to keep XAI useful, reliable, and actionable in changing financial environments, future research should expand these techniques to real-time applications, bigger portfolios, and new financial instruments.

Acknowledgment

Author's Note on Use of AI Tools

Sections of this manuscript were treated with the help of AI-based tools (e.g., AI language models) purely to support writing, define terminologies, improve grammar, propose content structure, and image generation. The authors are the sole developer of all conceptual contributions, scholarly arguments, examinations, data interpretation, and conclusions. The authors have reviewed, edited, and checked the final manuscript to make sure that it is accurate, original and of scholarly integrity.

References

1. Belkaid AB, Usun S. Explainable AI models in financial risk prediction: bridging accuracy and interpretability in modern finance. In: Electronics Communications, and Computing Summit (ECCSUBMIT). 2023;1(1):67-75.
2. Černevičienė J, Kabašinskas A. Explainable artificial intelligence (XAI) in finance: a systematic literature review. Artificial Intelligence Review. 2024;57:216. https://doi.org/10.1007/s10462-024-10854-8
3. Hadji Misheva B, Osterrieder J. A hypothesis on good practices for AI based systems for financial time series forecasting: towards domain driven XAI methods [preprint]. arXiv. 2023. https://doi.org/10.48550/arXiv.2311.07513
4. Lundberg SM, Lee S-I. A unified approach to interpreting model predictions [preprint]. arXiv. 2017. arXiv:1705.07874
5. Lundberg SM, Erion G, Lee S-I. Consistent feature attribution for tree ensembles [preprint]. arXiv. 2017. arXiv:1706.06060
6. Markowitz HM. Portfolio selection. The Journal of Finance. 1952;7(1):77-91.
7. Quinn B. Explaining AI in finance: past, present, prospects [preprint]. arXiv. 2023. arXiv:2306.02773. https://doi.org/10.48550/arXiv.2306.02773
8. Rafi MA, Shaboj SMI, Miah MK, Rasul I, Islam MR, Ahmed A. Explainable AI for credit risk assessment: a data driven approach to transparent lending decisions. Journal of Economics, Finance and Accounting Studies. 2024;6(1):108-118. https://doi.org/10.32996/jefas.2024.6.1.11
9. Schmitt M. Explainable automated machine learning for credit decisions: enhancing human-artificial intelligence collaboration in financial engineering [preprint]. arXiv. 2024. https://doi.org/10.48550/arXiv.2402.03806
10. Weber P, Carl KV, Hinz O. Applications of explainable artificial intelligence in finance: a systematic review of finance, information systems, and computer science literature. Management Review Quarterly. 2023;74(4):867-907. https://doi.org/10.1007/s11301-023-00320-0