

Analysis of the Capabilities of Predictive Artificial Intelligence Models in Corporate Risk Management

Kacper Ziemiński*

Department of Computer Science, Lublin University of Technology, Nadbystrzycka 36B, 20-618 Lublin, Poland

Abstract

This study examines the use of predictive artificial intelligence models in corporate risk management. Recurrent Neural Networks (RNN), Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) architectures are evaluated in forecasting five widely used market risk indicators based on daily financial time series for Apple Inc. Model performance is assessed using a unified experimental framework and out-of-sample data, with accuracy measured by Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE). Results show that predictive effectiveness varies across risk measures and model types, and no single model consistently outperforms the others. These findings highlight the need to align model choice with the structural properties of the specific market risk indicators being predicted.

Keywords: market risk; financial risk management; RNN; LSTM; GRU

*Corresponding author

Email address: s95623@pollub.edu.pl (K. Ziemiński)

Published under Creative Common License (CC BY 4.0 Int.)

1. Introduction

Financial risk management has become a fundamental aspect of decision-making process in modern enterprises operating in global financial markets. Among various categories of financial risk, market risk plays a particularly significant role, as it directly affects asset prices, firm valuation and investment performance. Comprehensive frameworks for identifying, measuring and managing market risk have been discussed in both academic literature and professional practice [1, 2]. Therefore, effective identification and forecasting of market risk is essential for maintaining financial stability and supporting strategic corporate decisions.

In recent decades, financial markets have been characterized by increasing volatility, accelerated dissemination of information and stronger interdependencies between assets and market. Such conditions give rise to complex and non-linear patterns in financial time series, making risk assessment more challenging. Classical analytical approaches often struggle to capture such dynamics, particularly during periods of market stress or structural breaks [3, 4]. As a result, enterprises and financial institutions increasingly depend on analytical tools that not only measure historical risk indicators but also provide reliable forecasts of future risk exposure.

Traditional approaches to market risk assessment are largely grounded in statistical and econometric methods, including volatility measures, Value-at-Risk methodologies and linear factor models. Although these approaches are well established and offer high degree of interpretability, they often rely on restrictive assumptions, such as linear dependencies, time series stationarity and predefined distributional forms. In practice, actual financial data often deviate from these assumptions, which may reduce the predictive accuracy of traditional risk modelling techniques.

Recent advances in artificial intelligence, particularly in the field of machine learning, have created new

opportunities for the analysis of financial risk. Neural network architectures designed for sequential data have shown large potential in modelling complex, non-linear relationships and high-dimensional data structures [4, 5]. However, despite their increasing adoption, the practical applicability and relative predictive performance of various sequential machine learning models in the context of corporate market risk management remain insufficiently explored. This creates a clear motivation for further investigation in this area.

2. Literature review

The role of artificial intelligence in financial risk management has gained importance in recent years due to its ability to process large volumes of data and capture complex, non-linear patterns present in financial time series. Although traditional risk assessment methods have been widely analysed and applied for decades, the increasing complexity of financial markets has encouraged researchers to explore more advanced data-driven methodologies.

Early studies on financial risk prediction were primarily based on statistical models, particularly linear and logistic regression. Pioneering studies by Altman [6] and Ohlson [7] demonstrated that financial ratios combined with linear probabilistic frameworks could be successfully used to predict corporate bankruptcy. Despite their simplicity and ease of interpretation, these models rely on restrictive assumptions related to linearity and stationarity, which limits their effectiveness in dynamic market conditions.

The development of machine learning significantly expanded the range of tools available for financial risk analysis. Recurrent Neural Networks (RNNs) were among the earliest neural architectures designed to process sequential data, what made them one of the most suitable tools for predicting financial time series. Empirical studies confirmed the potential of RNN-based models in market prediction tasks [8, 9]. However, they also

highlighted well-known limitations of standard RNNs, particularly the vanishing gradient problem, which restricts the ability of modelling long-term dependencies.

To overcome these limitations, Long Short-Term Memory (LSTM) networks were developed, incorporating gating mechanism that enable more effective retention of long-term information. Fischer and Kraus [10] reported that LSTM architectures achieved superior predictive performance in comparison to traditional statistical approaches in financial market forecasting tasks. Similarly, Bao et al. [11] proposed hybrid approach, which combines stacked autoencoders with LSTM networks, achieving improved forecasting accuracy. Subsequent studies confirmed the effectiveness and stability of LSTM-based models in various financial risk prediction contexts [12, 13, 14].

Recent studies have increasingly focused on comparative analyses of sequential deep learning architectures. Both articles and analyses suggest that models incorporating gating mechanism, such as LSTM and GRU, generally outperform traditional statistical techniques in capturing non-linear patterns in financial data [15, 16]. GRU-based models have also been successfully applied to specific risk management tasks, including liquidity risk assessment and banking stability analysis [17]. Additionally, alternative deep learning architectures, such as transformer-based time series models, have been proposed for financial risk prediction, further expanding the methodological landscape [18, 19].

At the same time, these models are frequently criticized for their limited interpretability, high data requirements and sensitivity to hyperparameter configurations. Although hybrid approaches have been proposed to address these challenges [17, 20], their increased complexity may reduce transparency and practical applicability in corporate risk management settings.

In summary, the existing literature provides substantial evidence of growing relevance and effectiveness of sequential deep learning models in financial risk management. However, despite numerous studies on individual architectures, there is still limited evidence comparing the predictive performance of RNN, LSTM and GRU models across multiple market risk indicators within a unified experimental setting. This research gap motivates the present study.

2.1. Research hypothesis

Based on the reviewed literature, the following research hypotheses are formulated:

- H1: Predictive artificial intelligence models (RNN, LSTM, GRU) exhibit significantly different predictive performance when applied to market risk indicators derived from financial time series.
- H2: Gated recurrent architectures (LSTM and GRU) achieve lower prediction errors than classical RNN models in forecasting market risk indicators.
- H3: The predictive effectiveness of RNN, LSTM and GRU models varies across different market risk measures, indicating that no single architecture is usually optimal.

3. Materials and methods

3.1. Data description

The analysis is based on daily financial market data retrieved from the Yahoo Finance database using the `yfinance` Python library. The study focuses on market risk exposure of a representative publicly listed company, Apple Inc. (AAPL), which serves as an example of a large-cap corporation operating in the U.S. equity market. To capture overall market movements, the S&P 500 index (^GSPC) is employed as a benchmark reflecting market performance.

The dataset consists of daily closing prices spanning a ten-year period from January 1st 2016 to January 1st 2026 yielding 2515 trading-day observations. The effective training set consists of 80% of the data, while the test set 20% of it. The selected time frame captures a broad range of market conditions, including periods of increased volatility, such as the COVID-19 market disruption or the interest rate tightening cycle of 2022, as well as periods of relative stability.

3.2. Market risk indicators

The study focuses on market risk, defined as the company's exposure to fluctuations in asset prices and overall market conditions. Five widely recognized market risk indicators were analysed:

- Volatility, measuring the variability of asset returns.
- Beta (β), capturing the sensitivity of a stock market movement.
- Value-at-Risk (VaR), estimating potential losses under adverse market conditions.
- Maximum Drawdown, reflecting the largest observed decline from peak to a trough.
- Sharpe Ratio, reflecting risk-adjusted performance.

3.3. Risk Indicator Computation

All market risk indicators were calculated using rolling windows based on daily return. A window of 20 trading days was applied to Volatility, VaR and Maximum Drawdown, corresponding approximately to one calendar month of trading activity. A longer window of 60 days was applied to Beta, as estimating covariance between two return series requires a larger sample to reduce the estimation noise. The following definitions were applied:

- Volatility was calculated as the standard deviation of returns with given rolling window:

$$Vol_t = \sigma(r_t) \quad (1)$$

- Beta (β) was estimated as the product of the correlation between stock and market returns and the ratio of their respective standard deviations:

$$\beta_t = Corr[r_{stock}, r_{market}] \times \frac{\sigma_{stock}}{\sigma_{market}} \quad (2)$$

where σ_{stock} and σ_{market} denote the standard deviations of stock and market returns, respectively.

- Value-at-Risk (VaR) at the 5% confidence level was estimated as the 5th percentile of returns:

$$VaR_{0.05} = Q_{0.05}(r) \quad (3)$$

- Maximum Drawdown was defined as the maximum observed loss from a peak to a subsequent trough:

$$MDD_t = \min \left[\frac{P_t}{\max(P_{1:t})} - 1 \right] \quad (4)$$

- The Sharpe Ratio was calculated as the ratio of excess average returns over the risk-free rate to return volatility. Due to the daily frequency of the data, the risk-free rate was assumed to be equal to zero:

$$Sharpe_t = \frac{\mu_r - r_f}{\sigma_r} \quad (5)$$

where μ_r is the average return, r_f is the risk-free rate and σ_r represents return volatility.

3.4. Data preprocessing and feature construction

Prior to model training, all calculated risk indicators were scaled using min-max normalization to ensure numerical stability and comparability across variables. Normalization was applied separately to each indicator to prevent scale dominance in the learning process.

The dataset was divided into training and testing subsets using a chronological split, allocating 80% of the observations to the training set and 20% to testing set. This procedure preserves the temporal ordering of the data and mitigates the risk of information leakage.

3.5. Sequential input windows

To enable the learning of temporal dependencies, sequential input windows were constructed using a sliding window approach. Each input sequence consisted of a fixed number of past observations (look-back window), representing historical values of a given risk indicator, while the prediction target corresponded to the subsequent value of the same indicator.

This setup enables recurrent neural networks to capture both short-term and long-term dependencies present in financial time series.

3.6. Predictive artificial intelligence models

The analysis considers three recurrent neural network architectures:

- Recurrent Neural Network (RNN), used as a baseline sequential model
- Long Short-Term Memory (LSTM), incorporating gating mechanisms to effectively capture long-term dependencies
- Gated Recurrent Unit (GRU), designed as a computationally efficient alternative to LSTM

To ensure a fair comparison, all models were implemented within a consistent experimental setup, employing the same network depth, number of hidden units and training procedures. Model parameters were optimised using the Adam optimiser and training was performed via backpropagation through time.

3.7. Experimental setup and evaluation metrics

To ensure the comparability and reproducibility of the conducted experiments, all three architectures were configured using identical hyperparameter settings. Each model consists of two recurrent layers, with 64 hidden units in the first layer and 32 in the second. The larger first layer is responsible for capturing higher-dimensional temporal dependencies, while the smaller second layer reduces the total parameter count and consolidates the extracted representations. A dropout rate of 0.2 was applied between the layers to reduce the risk of overfitting. All models were trained using the Adam optimiser with a learning rate of 0.001 for 50 epochs, with a batch size of 32. The look-back window was set to 20 trading days and the Mean Squared Error (MSE) was used as the loss function during the training.

4. Results

This section presents the performance of the three predictive artificial intelligence models (RNN, LSTM and GRU) applied to five market risk indicators for Apple Inc. (AAPL). Model evaluation was performed on out-of-sample test data, with forecasting accuracy measured using Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE).

For each risk indicator, all models were trained and evaluated under identical settings and experimental conditions.

4.1. Volatility prediction

Daily volatility was computed as the rolling standard deviation of returns over a 20-day window. Table 1 shows the predictive performance of the three models for volatility. RMSE and MAE values are reported for the out-of-sample test set.

Table 1: Models predictive performance for volatility

Model	RMSE	MAE
RNN	0.00157	0.00078
LSTM	0.00177	0.00092
GRU	0.00167	0.00081

The results for daily volatility indicate relatively low error values for all three models. The RNN architecture achieved the lowest RMSE (0.00157) and the lowest MAE (0.00078) among the evaluated models. The GRU model produced results most comparable to the RNN in terms of MAE (0.00081), while the LSTM architecture exhibited the highest prediction errors for both evaluation metrics (RMSE 0.00177, MAE 0.00092).

4.2. Beta prediction

Beta measures the sensitivity of the stock to overall market movements and was calculated relative to the S&P 500 index using a 60-day rolling window. Table 2 presents RMSE and MAE for three models.

Table 2: Models predictive performance for Beta

Model	RMSE	MAE
RNN	0.04858	0.03243

LSTM	0.05741	0.03966
GRU	0.05011	0.03278

Forecasting the beta coefficient resulted in noticeably higher error magnitudes compared to volatility. This outcome shows the greater complexity of beta dynamics. Among the evaluated models, the RNN achieved the lowest error values for both RMSE and MAE, followed closely by the GRU model. In this case, the LSTM returned noticeably higher errors for both metrics.

4.3. Value-at-Risk (5%) prediction

Value-at-Risk at 5% (VaR₅) was calculated as the 5th percentile of daily returns over a rolling 20-day window. Table 3 summarizes the RMSE and MAE for all models on the set.

Table 3: Models predictive performance for VaR at 5%

Model	RMSE	MAE
RNN	0.00505	0.00318
LSTM	0.00644	0.00421
GRU	0.00472	0.00292

For the 5% Value-at-Risk, the models' error values were slightly larger than those for volatility, but significantly smaller than the results for Beta. The GRU achieved the lowest RMSE and MAE, while LSTM exhibited the highest values of prediction errors. The RNN produced results very similar to those of the GRU.

4.4. Maximum Drawdown prediction

Maximum drawdown represents the largest peak-to-trough within a rolling window of 20 days. Table 4 reports RMSE and MAE for this indicator.

Table 4: Models predictive performance for Maximum drawdown

Model	RMSE	MAE
RNN	0.00865	0.00516
LSTM	0.00936	0.00464
GRU	0.00915	0.00450

The prediction results for maximum drawdown show similar error values across all three architectures. Differences in RMSE and MAE are minor and no single model consistently outperforms the others across both metrics.

4.5. Sharpe Ratio prediction

The Sharpe ratio was calculated as the mean return over standard deviation of returns for each rolling window. It is presented in Table 5 with RMSE and MAE provided for all models.

Table 5: Models predictive performance for Sharpe ratio

Model	RMSE	MAE
RNN	0.07475	0.05629
LSTM	0.08279	0.06260
GRU	0.07909	0.06013

The Sharpe ratio returned the highest error values among all analysed indicators, regardless of model. The RNN architecture achieved the lowest RMSE and MAE,

while the LSTM and GRU produced slightly larger errors.

5. Discussion

The present study evaluated the predictive capabilities of RNN, LSTM and GRU in forecasting multiple market risk indicators derived from daily financial time series. The findings enable a direct assessment of the strengths and limitations of the recurrent neural network in the context of financial risk management and allow for explicit verification of the proposed research hypothesis.

Regarding H1, which assumes significant differences in predictive performance among RNN, LSTM and GRU models, the results indicate that performance variations are present but not even across all indicators. For risk measures such as volatility and Value-at-Risk, error differences between models were relatively small, with no model consistently outperforming the others. In contrast, the performance variations turned out to be more noticeable for complex risk indicators like Beta or Sharpe ratio. These findings suggest that H1 is partially supported, as model architecture affects forecasting performance, but the significance of its impact depends on the modelled risk measure.

The second hypothesis (H2) proposed that gated recurrent architectures (LSTM and GRU) outperform classical RNN models in terms of prediction accuracy. The actual results do not provide consistent evidence to support this assumption. While LSTM and GRU occasionally achieved lower prediction errors for selected indicators, the classical RNN demonstrated comparable or even superior performance in specific cases, particularly for volatility-based measures. This indicates that theoretical advantages of gating mechanisms do not significantly impact the predictive performance in the considered experimental setup. Consequently, the results do not provide unambiguous evidence to confirm H2, as LSTM and GRU did not systematically achieve lower prediction errors than the RNN across all evaluated indicators and metrics.

The third hypothesis (H3) referred to the variability of model effectiveness across different market risk indicators. The results strongly confirm this hypothesis: predictive accuracy depends strongly on the specific risk measure. Measures derived directly from return distributions, such as volatility and Value-at-Risk, forecasted with higher accuracy, while indicators that integrate market relationships or risk-adjusted performance demonstrated higher prediction inaccuracy. Moreover, no single architecture consistently outperformed the others, highlighting the dependence of model suitability on the characteristics of the risk measure.

Overall, all the findings highlight the importance of picking or adjusting the model to the structural properties of financial risk indicator. Furthermore, increasing the architectural complexity of the models does not guarantee improved predictive performance in market risk forecasting tasks.

6. Conclusions

This study investigated the performance of predictive artificial intelligence models in forecasting multiple market risk indicators that are commonly used in corporate financial risk management. The analysis provides a thorough assessment of the strengths and limitations of the recurrent neural networks (particularly RNN, LSTM and GRU) in financial risk modelling and allows for direct evaluation of the proposed hypothesis.

The results indicate that while differences in predictive performance among the models exist, they are not regular or consistent across all risk measures. As a result, hypothesis H1 was partially confirmed, suggesting that model choice influences forecasting accuracy, but the scale of the impact depends on the selected risk indicator. In subsequent analyses, hypothesis H2 was rejected, as gated recurrent architectures did not systematically outperform classical RNN models. In contrast, H3 was fully supported, demonstrating that predictive effectiveness varies across different market risk indicators and that no single architecture is universally optimal.

From a practical perspective, the results suggest that simpler recurrent architectures may offer a favourable trade-off between accuracy and computational efficiency in market risk management systems.

This work leaves room for the future investigations that may extend the study by analysing broader set of assets, incorporating diverse market conditions and including additional independent variables in models. Furthermore, the application of statistical significance tests, hybrid or attention-based architectures may offer deeper insights into which artificial intelligence models offer the best performance in context of market risk prediction.

References

- [1] M. Crouhy, D. Galai, R. Mark, *The Essentials of Risk Management*, McGraw-Hill Education, New York, 2014.
- [2] A. Karasan, *Machine Learning for Financial Risk Management with Python*, O'Reilly Media, Sebastopol, 2021.
- [3] T. Hastie, R. Tibshirani, J. Friedman, *The Elements of Statistical Learning*, Springer, New York, 2009.
- [4] K. P. Murphy, *Machine Learning: A Probabilistic Perspective*, MIT Press, Cambridge, 2012.
- [5] I. Goodfellow, Y. Bengio, A. Courville, *Deep Learning*, MIT Press, Cambridge, 2016.
- [6] E. I. Altman, Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy, *J. Finance* 23 (4) (1968) 589-609, <https://doi.org/10.2307/2978933>.
- [7] J. A. Ohlson, Financial Ratios and the Probabilistic Prediction of Bankruptcy, *Journal of Accounting Research* 18 (1) (1980) 109-131, <https://doi.org/10.2307/2490395>.
- [8] S. Selvin, R. Vinayakumar, E. A. Gopalakrishnan, V. K. Menon, K. P. Soman, Stock price prediction using LSTM, RNN and CNN-sliding window model, In 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI) (2017) 1643-1647, <https://doi.org/10.1109/ICACCI.2017.8126078>.
- [9] D. Nurmaltasari, D. Pantagama, D. Pramanaswari, Predicting LQ45 financial sector indices using RNN-LSTM, *Journal of Big Data* 8 (2021) 1-15, <https://doi.org/10.1186/s40537-021-00495-x>.
- [10] T. Fischer, C. Krauss, Deep Learning with long short-term memory networks for financial market predictions, *European Journal of Operational Research* 270(2) (2018) 654-669, <https://doi.org/10.1016/j.ejor.2017.11.054>.
- [11] W. Bao, J. Yue, Y. Rao, A deep learning framework for financial time series using stacked autoencoders and long short-term memory, *PLoS ONE* 12(7) (2017) e0180944, <https://doi.org/10.1371/journal.pone.0180944>.
- [12] K. Xu, Y. Cheng, S. Long, J. Guo, J. Xiao, M. Sun, Advancing Financial Risk Prediction through Optimized LSTM Model Performance and Comparative Analysis, In 2024 IEEE 2nd International Conference on Sensors, Electronics and Computer Engineering (ICSECE) (2024) 1264-1270, <https://doi.org/10.1109/ICSECE61636.2024.10729338>.
- [13] Y. Gao, Stock Prediction Based on Optimized LSTM and GRU Models, *Scientific Programming* 2021 (2021) 4055281, <https://doi.org/10.1155/2021/4055281>.
- [14] G. R. Naufal, A. Wibowo, Time Series Forecasting Based on Deep Learning CNN-LSTM-GRU Model on Stock Prices, *International Journal of Engineering Trends and Technology* 71(6) (2023) 126-133, <https://doi.org/10.14445/22315381/IJETT-V71I6P215>.
- [15] C. Zhang, N. N. A. Sjarif, R. Ibrahim., Deep learning models for price forecasting of financial time series: A review of recent advancements: 2020-2022, *WIREs Data Mining and Knowledge Discovery* 13(1) (2023) e1519, <https://doi.org/10.1002/widm.1519>.
- [16] E. Tölö, Predicting systemic financial crises with recurrent neural networks, *Journal of Financial Stability* 49(2020) 100746, <https://doi.org/10.1016/j.jfs.2020.100746>.
- [17] Z. Xu, J. Pan, S. Han, H. Ouyang, Y. Chen, M. Jiang, Predicting Liquidity Coverage Ratio with Gated Recurrent Units: A Deep Learning Model for Risk Management, In 2024 5th International Conference on Machine Learning and Computer Application (ICMLCA) (2024) 108-112, <https://doi.org/10.1109/ICMLCA63499.2024.10753702>.
- [18] P. Chen, M. Ji, Deep learning-based financial risk early warning model for listed companies: A multi-dimensional analysis approach, *Expert Systems with Applications* 283 (2025) 127746, <https://doi.org/10.1016/j.eswa.2025.127746>.
- [19] W. Sun, Z. Xu, W. Zhang, K. Ma, Y. Wu, M. Sun, Advanced Risk Prediction and Stability Assessment of Banks Using Time Series Transformer Models, *Proceedings of the 2024 5th International Conference on Big Data Economy and Information Management (BDIEM)* (2024) 657-661, <https://doi.org/10.1145/3724154.3724264>.
- [20] E. Ramos-Pérez, P. J. Alonso-González, J. J. Núñez-Velázquez, Forecasting volatility with a stacked model based on a hybridized Artificial Neural Network, *Expert Systems with Applications* 129 (2020) 1-9, <https://doi.org/10.1016/j.eswa.2019.03.046>.