International Journal of Computer Applications Technology and Research
Volume 14–Issue 12, 16 – 20, 2025, ISSN:-2319–8656
DOI:10.7753/IJCATR1412.1002

Evaluating Anomaly Detection Methods for Financial Portfolio Risk Management

Narendra Lakshmana Gowda
Independent Researcher
Ashburn, VA, USA

Abstract **-** Financial institutions have a tough job managing and constantly watching the risk in many portfolios and they deal with a lot of data. To make this easier an automatic system that can spot unusual patterns in portfolio risk measures would be very useful. Therefore, this study looked at four methods for detecting these unusual patterns in time series data the Autoregressive Integrated Moving Average (ARIMA)-Generalized Autoregressive Conditional Heteroscedasticity (GARCH), and Exponentially Weighted Moving Average (EWMA), Long Short-Term Memory (LSTM) and Hierarchical Temporal Memory (HTM) Machine Learning (ML) algorithms. We tested these methods using three sets of synthetic data and one set of real-world data with manually added labels. The findings exhibit that LSTM networks are effective at this task. On the other hand, EWMA models are faster and easier to understand. The ARMA-GARCH model didn't fit the time series data well and performed poorly. The HTM method was outperformed by the other methods because it struggled to learn the time series patterns needed for effective detection. In short, LSTM models are the best at finding anomalies in portfolio risk measures. EWMA models are good for speed and clarity, but ARMA-GARCH and HTM models did not perform as well.

Index Terms Financial, Machine Learning, Portfolio, Risk

I. INTRODUCTION

The financial sector handles a lot of important data which includes the investment portfolios performance. Managing and keeping an eye on hundreds of portfolios is a big task for financial institutions. They need to track key risk measures every day. One foremost risk measure is Value-at-Risk (VaR). This becomes significant if it behaves uncommonly. Abnormal VaR values could indicate big market changes. These changes include changes in the portfolio or mistakes in the data. Spotting these anomalies is crucial because they can provide valuable information that influences investment decisions and risk management. Monitoring portfolio risk is complex and time-consuming. Automating the process of detecting unusual behaviour in risk measures could greatly reduce the amount of manual work needed. The ideal system would use an algorithm to automatically find unusual patterns in risk data based on past information. Quickly identifying potential issues would help to manage risk more effectively. Financial institutions typically use econometric models to analyze data. Common models

include Autoregressive Integrated Moving Average (ARIMA), Generalized Autoregressive Conditional Heteroscedasticity (GARCH), and Exponentially Weighted Moving Average (EWMA). These models are well-known for analyzing financial data. However, new Machine Learning (ML) techniques have shown assurance in this area. For instance, Recurrent Neural Networks (RNNs) with Long Short-Term Memory (LSTM) [1]–[8] units and Hierarchical Temporal Memory (HTM) have achieved strong results in handling time-series data and detecting anomalies. Therefore, this study investigates how these modern ML techniques can be useful to detect anomalies in financial portfolio risk measures. Comparing the performance with traditional econometric models to distinguish which method works better. Managing portfolio risk is challenging and requires a lot of time because different portfolios and risk measures vary widely. No single method is perfect without human input. Therefore, algorithms that can learn what normal risk data looks like and apply this knowledge across different portfolios with minimal manual work are needed. Although there is a lot of research on modeling risk measures there is less work on detecting anomalies in new data. The main question is which of these models is the best at detecting anomalies in the risk measures of financial portfolios. The study looks at a new problem finding anomalies in streaming time-series data without needing supervised learning. By exploring this new area the study aims to offer valuable insights that can aid both professionals and researchers. The results could improve risk management practices.

The study is as follows; similar papers are shown in the following section. The materials and methods are covered in Section III. The experimental analysis is carried out in Section IV, and in Section V, we provide some conclusions and plans for future research.

II. RELATED WORKS

Several studies have been published such as [9] looked neural and neuro-fuzzy techniques for predicting stock market trends. They looked at five main points the stock markets studied, the input data used, the techniques and settings for the prediction models, comparisons of different models, and how accuracy was measured. Similarly, [10, 11] looked at five key areas how soft computing is used in financial markets, which markets were studied, the input variables used, the trading systems proposed, and how these studies helped develop trading systems. They addressed major financial issues like forecasting stock prices, commodity prices, exchange rates, electricity prices, and predicting financial distress. [12] focused

at text mining methods to predict stock prices based on company news. They compared different text mining approaches focusing on things like feature selection. [13] looked at how Artificial Neural Networks (ANNs) are used in financial markets to predict things like exchange rates, stock prices, and financial crises. [14] looked into evolutionary computation methods, such as Genetic Programming (GP), Learning Classifier Systems (LCSs), Genetic Algorithms (GA), Multi-Objective Evolutionary Algorithms (MOEAs), and Co-evolutionary Optimization. They focused on Darwinian approaches but only covered a few methods. [15, 16] reviewed techniques for predicting the Indian stock market by discussing the parameters like benefits, and drawbacks of various methods. [17] examined research from 2000 to 2016 on text mining in finance, pointing out key issues, research gaps, challenges, and future directions.

III. MATERIALS AND METHODS

To manage large amounts of financial data effectively and ensure safe and profitable portfolio management, [18] developed the risk framework. This framework provides a clear method for managing risk in finance based on three main lines of defense. Risk owners are the people or teams responsible for handling risks in specific areas of the organization. Corporate risk functions team oversees and sets up the daily tasks related to risk management. They make sure that risk data is collected regularly and turned into useful and timely information. Internal audit is an independent team checks how well the risk management processes are working and ensures they are effective. The corporate risk functions are key to managing risk because they ensure that data about risks is constantly collected and turned into useful insights. For portfolio management, timely information means spotting and dealing with unusual portfolio behaviors as soon as they happen or as quickly as possible after they occur. If unusual patterns develop slowly like local trend outliers they should be identified as soon as the trends start to deviate from what was expected. How quickly anomalies are identified can depend on the type of portfolio and market conditions. For instance, a sudden change in trends might be seen as an anomaly while a gradual change might be a normal market trend. Accuracy is also very important. It means finding all the real anomalies without mistaking normal changes for problems. Getting this balance right is crucial because false alarms can cause unnecessary worry and disruption while missing real anomalies can lead to ignoring important risks. Moreover, modern financial institutions deal with a huge amount of data so how quickly they process this data has become increasingly important. Fast data processing is essential for keeping the quality of risk management systems high, as noted by [19]. Quick processing helps ensure that any anomalies are detected and addressed promptly, which is crucial for effective risk management [20]–[24].

A. Data Analysis

We created three different synthetic and one real-world dataset with manually added labels. The first and second synthetic datasets these datasets include 500 time series, each covering 5000 days. They are created using two different econometric models. At the start, these time series behave normally but anomalies are added randomly. Anomalies are introduced at a rate of about 0.1% per day which means one anomaly approximately every 1000 days. The anomalies are chosen from a set of predefined types and have a magnitude five times greater than the standard deviation of the past 100 days. The direction of the anomalies is also random. Third synthetic dataset this dataset features VaR time series another common financial metric. It also has 500 time series of 5000 days each with anomalies added in a similar way to the previous datasets. These synthetic datasets are used to test how well anomaly detection algorithms perform in different financial scenarios. The goal is to show the strengths and weaknesses of the algorithms and see how well they generalize to different types of financial data. However, real-world dataset includes VaR time series from investment portfolios that are actively traded on the market by a confidential source. It has manually crafted labels to test the algorithms in real-world conditions. The first synthetic dataset uses a GARCH (1, 1) model with Gaussian noise. The time series begins with random values drawn from a normal distribution. The first few hundred days are left out to make sure the data is genuinely from the GARCH process. Not influenced by initial random values. GARCH models are commonly used for financial time series, especially for returns. However, they often miss extreme events called black swan events which are rare but significant. For the purpose of anomaly detection this limitation is manageable. The synthetic data can start with a normal distribution and then have outliers added to increase the kurtosis, making it suitable for detecting anomalies.

B. Model Analysis

The ARMA-GARCH model follows the Box-Jenkins[1] method. It involves three main steps model identification this step begins with an augmented dickey-fuller test to ensure the data is stationary with 99% certainty. Then, a grid search is done over possible values for the ARMA model parameters p and q (ranging from 0 to 5). The parameters that give the lowest Akaike Information Criterion (AIC) are chosen. This process is updated every 100 timesteps because the model's parameters don't change much over short periods and finding the best parameters takes time. Model estimation this is done using maximum likelihood estimation techniques which are standard methods for estimating model parameters. Model diagnostic the model's fit is checked using a goodness-of-fit test. Once the model is set up it forecasts the mean and variance for the next day. Anomalies are found by checking if the data point deviates significantly from the forecasted values. The ARMA-GARCH model is implemented using the rugarch[2] package in R a

[1] With an emphasis on iterative model development, the Box-Jenkins technique identifies, estimates, and diagnoses ARIMA models for time series forecasting.

[2] https://www.rdocumentation.org/packages/rugarch/versions/1.5-1

International Journal of Computer Applications Technology and Research
Volume 14–Issue 12, 01 – 15, 2025, ISSN:-2319–8656
DOI:10.7753/IJCATR1412.1001

language designed for statistical computing as shown in Fig. 1. The EWMA algorithm was implemented using the equations from [25, 26]. Fig. 2 shows how EWMA detects anomalies in a time series. The LSTM network was set up using Keras a high-level API for building neural networks which runs on TensorFlow. To speed up calculations, TensorFlow's GPU support was used. However, the speed improvement was not very noticeable. This was because the neural network was relatively small and didn't gain much from GPU acceleration. The process of transferring data to the GPU caused more delay than the benefit of parallel processing. The LSTM model is trained continuously with all available data at each timestep as shown in Fig. 3. To find anomalies, the prediction error is calculated by comparing the model's predictions for the next day with the actual values. The HTM model was created using the Numenta Platform for Intelligent Computing (NuPIC), which is developed by Numenta the company that created HTM technology as shown in Fig. 4.

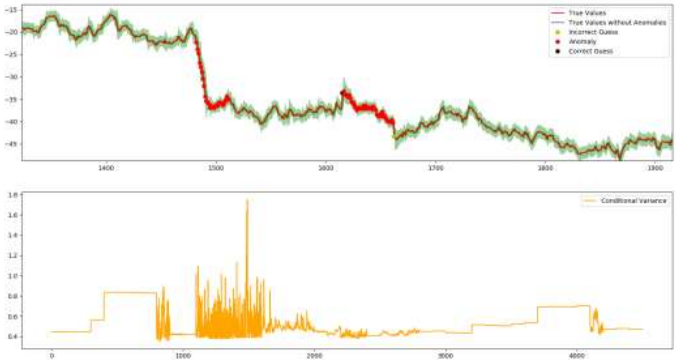

Fig. 1. ARMA-GARCH time series anomaly detection

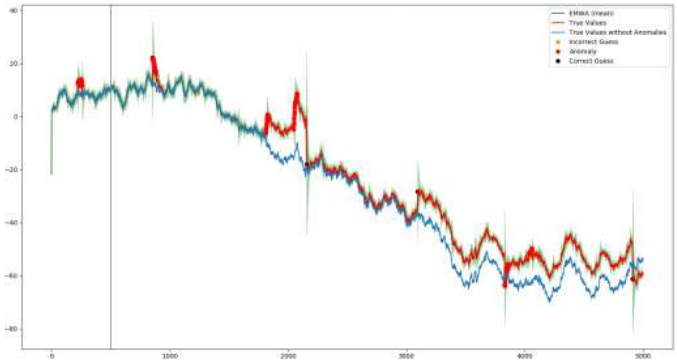

Fig. 2. EWMA time series anomaly detection

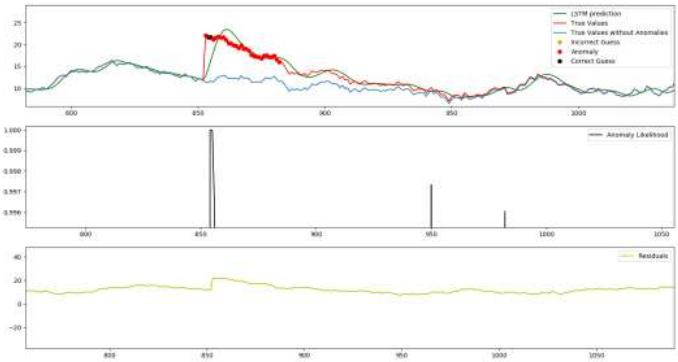

Fig. 3. LSTM time series anomaly detection

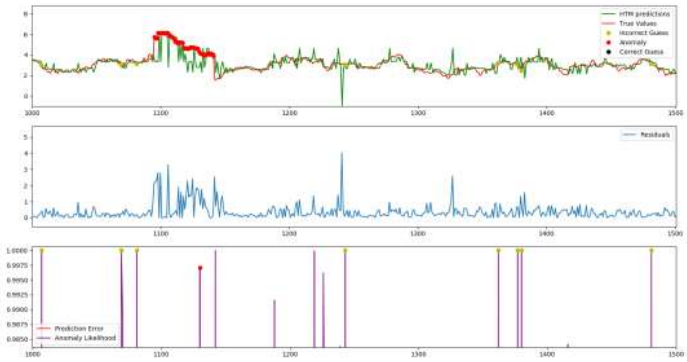

Fig. 4. HTM time series anomaly detection

IV. Experimental Analysis

A. Synthetic Data

The performance evaluation of various algorithms on synthetic datasets shows important details about their effectiveness and efficiency. The results are shown in Table I which lists the main performance metrics for each algorithm on three different datasets. Table II provides a more detailed view of how well each algorithm handles different types of anomalies. The ARMA-GARCH model achieves the highest F1-score on the return datasets. This means it is very effective at balancing precision and recall when dealing with these types of datasets. Its weaknesses is its performance drops sharply on the VaR dataset. Although it still identifies anomalies with the same recall rate its precision significantly decreases. This suggests that while it continues to find anomalies it also has many false positives on the VaR dataset. The EWMA is the fastest algorithm among those tested with the shortest elapsed time. It also has a fairly high recall meaning it successfully detects many anomalies. It's Weaknesses are despite its speed and good recall EWMA suffers from low precision especially noticeable in dataset 3. It struggles particularly with Long-Term Outliers (LTOs) as shown by the differences in anomaly detection performance across datasets. Additionally, EWMA has a low reaction delay, which helps in quick performance. The LSTM algorithm performs well overall, with a solid F1-score across both return datasets and the VaR dataset. It also effectively identifies a high percentage of LTOs, indicating good anomaly detection capabilities. Its weaknesses are the main drawbacks of LSTM are its longer processing time and reaction delay compared to ARMA-GARCH and EWMA. This means it is slower and less responsive even though it performs well in terms of detecting anomalies. The HTM algorithm performs poorly in almost all aspects when compared to the other algorithms. It consistently shows lower effectiveness in detecting anomalies, with worse results in both F1-score and precision across the different datasets. This makes HTM the least effective option among those tested.

TABLE I
PRIMARY PERFORMANCE METRICS FOR THE ALGORITHMS ON SYNTHETIC DATA

Algorithm	F1	Recall	Precision	Reaction	Elapsed Time (s)
ARMA-GARCH	0.749	0.581	0.244	0.773	5290
	0.785	0.784	0.726	0.461	9450
	0.144	0.32	1.32	1.15	5415

EWMA	0.418	0.271	0.119	0.736	11
	0.661	0.698	0.291	0.171	11
	0.065	0.147	0.324	0.233	11
LSTM	0.539	0.531	0.317	0.786	42680
	0.771	0.790	0.411	0.404	74525
	0.198	1.394	2.020	1.600	45411
HTM	0.038	0.060	0.032	0.068	131619
	0.314	0.205	0.026	0.033	110732
	0.018	6.460	5.643	6.09	127502

TABLE II
THE PROPORTION OF ANOMALIES FOUND OF THE DIFFERENT ANOMALY TYPES

Algorithm	ALO (%)	LSO (%)	TCO (%)	LTO (%)
ARMA-GARCH	92	89	90	88
	89	89	89	88
	91	39	49	46
EWMA	91	86	88	87
	85	87	88	84
	85	27	13	18
LSTM	81	81	84	82
	82	86	80	80
	85	76	66	61
HTM	8	33	19	4
	20	11	7	31
	21	8	42	32

B. Real-World Data

Table III shows the ARMA-GARCH model still has trouble with the VaR series. It has low precision, meaning it often gives false positives. Although it detects many anomalies it also identifies too many that aren't anomalies. EWMA is the fastest algorithm and works fairly well. Its precision on the VaR series is better than it was on dataset 3. It has the lowest reaction delay which means it can find anomalies almost immediately. Dataset 4 didn't have any LTOs so EWMA was able to spot all the anomalies it found quickly. The LSTM has the highest F1-score showing it balances precision and recall very well. However, it now takes even more time to process data than the HTM algorithm which makes it less efficient. This slower processing is a big downside, even though its F1-score is the best. The HTM performs the worst with a very low F1-score. It has poor recall and precision, meaning it struggles a lot to detect anomalies and is the least effective among the algorithms tested. In summary, the real-world data results support what was observed with synthetic data. The ARMA-GARCH model still struggles with precision for the VaR series. The EWMA algorithm remains the fastest with a low reaction delay and better precision on the VaR series. The LSTM, despite having the highest F1-score, is slower and less efficient. The HTM continues to perform poorly across the board.

TABLE III
PRIMARY PERFORMANCE METRICS FOR THE ALGORITHMS ON THE REAL-WORLD DATA

Algorithm	F1	Recall	Precision	Reaction	Elapsed Time (s)
ARMA-GARCH	0.04	0.889	0.02	0	27.3
EWMA	0.304	0.777	0.189	0	0.02
LSTM	0.453	0.667	0.343	1.08	169
HTM	0.05	0.10	0.03	0.5	75.7

V. CONCLUSION AND FUTURE WORKS

This study looked at four different algorithms ARMA-GARCH, EWMA, LSTM, and HTM to find out which one is best for spotting unusual patterns in financial risks. The goal was to see which algorithm works best for managing portfolio risks by checking their performance against established criteria. The study used three made-up datasets and one real-world dataset. The made-up datasets were created using models called GARCH (1, 1) and ARMA (2, 2)-GARCH (1, 1) with some fake anomalies like ALO, LSO, TCO, and LTO added in. The real-world dataset had anomalies that were labeled by hand. The results showed that EWMA, ARMA-GARCH, and LSTM worked well in different situations. However, HTM did not perform well with the conditions tested which supports previous findings that HTM has limitations with certain types of data. This means that choosing the best algorithm depends on the specifics of the data and the types of anomalies you are dealing with. The study is useful for people working with financial systems who want to add anomaly detection and for researchers interested in how Ml can be applied in finance. While traditional models like ARMA-GARCH and EWMA still have advantages, LSTM models show a lot of promise. Future research could look into other risk measures, such as expected shortfall, or explore how combining EWMA and LSTM might work together.

REFERENCES

[1] G. S. Kashyap, A. Siddiqui, R. Siddiqui, K. Malik, S. Wazir, and A. E. I. Brownlee, “Prediction of Suicidal Risk Using Machine Learning Models.” Dec. 25, 2021. Accessed: Feb. 04, 2024. [Online]. Available: https://papers.ssrn.com/abstract=4709789

[2] F. Alharbi and G. S. Kashyap, “Empowering Network Security through Advanced Analysis of Malware Samples: Leveraging System Metrics and Network Log Data for Informed Decision-Making,” *International Journal of Networked and Distributed Computing*, pp. 1–15, Jun. 2024, doi: 10.1007/s44227-024-00032-1.

[3] G. S. Kashyap *et al.*, “Revolutionizing Agriculture: A Comprehensive Review of Artificial Intelligence Techniques in Farming,” Feb. 2024, doi: 10.21203/RS.3.RS-3984385/V1.

[4] G. S. Kashyap, K. Malik, S. Wazir, and R. Khan, “Using Machine Learning to Quantify the Multimedia Risk Due to Fuzzing,” *Multimedia Tools and Applications*, vol. 81, no. 25, pp. 36685–36698, Oct. 2022, doi: 10.1007/s11042-021-11558-9.

[5] Alla, S., Soltanisehat, L., Tatar, U., & Keskin, O. (2018). Blockchain technology in electronic healthcare systems. In *IIE Annual Conference. Proceedings* (pp. 901-906). Institute of Industrial and Systems Engineers (IISE).

[6] Soltanisehat, L., & Alla, S. (2018). Centraized or distributed IT system?(blockchain concept). In *Proceedings of the International Annual Conference of the American Society for Engineering Management.* (pp. 1-7). American Society for Engineering Management (ASEM).

[7] N. L. Gowda, B. S. Banjardar, V. Manchala, and A. R. Mohammed, “Bridging Classical Conditioning and Deep Reinforcement Learning: Advancements, Challenges, and Strategies for Autonomous Systems,” May 04, 2024. Available: https://journal.esrgroups.org/jes/article/view/4121.

[8] S. Benartzi and R. H. Thaler, “Naive diversification strategies in defined contribution saving plans,” *American Economic Review*, vol. 91, no. 1, pp. 79–98, 2001, doi: 10.1257/aer.91.1.79.

[9] P. Akioyamen, Y. Z. Tang, and H. Hussien, “A hybrid learning approach to detecting regime switches in financial markets,” in *ICAIF 2020 - 1st ACM International Conference on AI in Finance*, Oct.

2020. doi: 10.1145/3383455.3422521.

[10] P. A. Bebbington, "Studies in informational price formation, prediction markets, and trading," UCL (University College London), Nov. 2017. Accessed: Feb. 04, 2024. [Online]. Available: http://discovery.ucl.ac.uk/1563501/1/thesis.pdf

[11] Alla, S., Pazos, P., & DelAguila, R. (2017). The impact of requirements management documentation on software project outcomes in health care.

[12] T. Sun, J. Wang, P. Zhang, Y. Cao, B. Liu, and D. Wang, "Predicting Stock Price Returns Using Microblog Sentiment for Chinese Stock Market," in *Proceedings - 2017 3rd International Conference on Big Data Computing and Communications, BigCom 2017*, Nov. 2017, pp. 87–96. doi: 10.1109/BIGCOM.2017.59.

[13] K. Pawar, R. S. Jalem, and V. Tiwari, "Stock Market Price Prediction Using LSTM RNN," in *Advances in Intelligent Systems and Computing*, 2019, vol. 841, pp. 493–503. doi: 10.1007/978-981-13-2285-3_58.

[14] D. Snow, "Machine Learning in Asset Management," *SSRN Electronic Journal*, Jul. 2019, doi: 10.2139/ssrn.3420952.

[15] J. P. Broussard and M. Vaihekoski, "Profitability of pairs trading strategy in an illiquid market with multiple share classes," *Journal of International Financial Markets, Institutions and Money*, vol. 22, no. 5, pp. 1188–1201, Dec. 2012, doi: 10.1016/j.intfin.2012.06.002.

[16] Lakshminarayanan, V., Ravikumar, A., Sriraman, H., Alla, S., & Chattu, V. K. (2023). Health care equity through intelligent edge computing and augmented reality/virtual reality: a systematic review. *Journal of Multidisciplinary Healthcare*, 2839-2859.

[17] A. W. Ayele, E. Gabreyohannes, and Y. Y. Tesfay, "Macroeconomic Determinants of Volatility for the Gold Price in Ethiopia: The Application of GARCH and EWMA Volatility Models," *Global Business Review*, vol. 18, no. 2, pp. 308–326, Mar. 2017, doi: 10.1177/0972150916668601.

[18] S. Wazir, G. S. Kashyap, and P. Saxena, "MLOps: A Review," Aug. 2023, Accessed: Sep. 16, 2023. [Online]. Available: https://arxiv.org/abs/2308.10908v1

[19] S. Wazir, G. S. Kashyap, K. Malik, and A. E. I. Brownlee, "Predicting the Infection Level of COVID-19 Virus Using Normal Distribution-Based Approximation Model and PSO," Springer, Cham, 2023, pp. 75–91. doi: 10.1007/978-3-031-33183-1_5.

[20] N. Marwah, V. K. Singh, G. S. Kashyap, and S. Wazir, "An analysis of the robustness of UAV agriculture field coverage using multi-agent reinforcement learning," *International Journal of Information Technology (Singapore)*, vol. 15, no. 4, pp. 2317–2327, May 2023, doi: 10.1007/s41870-023-01264-0.

[21] Alla, S., & Pazos, P. (2019). Healthcare robotics: Key factors that impact robot adoption in healthcare.

[22] Alla, S., Bheesetty, N., Prakash, Y., Sunkara, M., Chidipudi, P., Chattu, V. K., & Velur, V. R. (2023). MACHINE-LEARNING ANALYSIS OF MORTALITY DUE TO COMORBIDITIES AND RESULTING MICROVASCULAR COMPLICATIONS IN COVID PATIENTS WITH TYPE-2 DIABETES MELLITUS. In *Proceedings of the International Annual Conference of the American Society for Engineering Management.* (pp. 1-9). American Society for Engineering Management (ASEM).

[23] Alla, S. (2019). A Statistical Analysis of Surgeons' Preference on Robot-Assisted Surgeries. In *IIE Annual Conference. Proceedings* (pp. 1385-1390). Institute of Industrial and Systems Engineers (IISE).

[24] Soni, A., Alla, S., Dodda, S., & Volikatla, H. (2024). Advancing Household Robotics: Deep Interactive Reinforcement Learning for Efficient Training and Enhanced Performance. *arXiv preprint arXiv:2405.18687*.

[25] Gowda, N. L. (2025, March 3). Federated Learning a Collaborative Machine Learning Across Countries with Data Privacy. https://ijritcc.org/index.php/ijritcc/article/view/11481

[26] J. R. Chang and M. W. Hung, "Intertemporal Risk and Currency Risk," in *Encyclopedia of Finance, Third Edition*, Springer International Publishing, 2022, pp. 555–572. doi: 10.1007/978-3-030-91231-4_5.