ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25mar1287

Robustness and Adversarial Resilience of Actuarial AI/ML Models in the Face of Evolving Threats

Niha Malali[1]; Sita Rama Praveen Madugula[2]

[1,2]Independent Researcher

Publication Date: 2025/03/25

Abstract: The application of artificial intelligence (AI) and machine learning (ML) in actuarial science yields data-driven financial decision-making processes, as well as transformed predictive modeling and risk assessment. Security threats that occur due to increasing AI/ML model adoption create significant risks for actuarial applications through data poisoning and both evasion techniques and model inversion attacks. Breach points in systems create substantial risks for misjudged risks, price distortions, and regulatory issues, which damage the dependability of actuarial modeling outcomes. Adversarial resilience and robustness of AI/ML models in actuarial science receive detailed exploration in this paper through assessments of existing defense mechanisms which primarily include adversarial training, anomaly detection and robust feature engineering methods as well as identification of main threat vectors. This paper covers the essential regulatory structures and ethical matters because such frameworks protect the integrity of trustable AI-driven actuarial systems. The effectiveness of various adversarial threat defenses against actuarial AI models is evaluated through experimental results. The research confirms that security measures in the actuarial domain of AI need ongoing development to protect its systems from current and future threats which require sustainable reliability and threat resistance.

Keywords: *Artificial Intelligence, Machine Learning, Actuarial Science, Risk Assessment, Predictive Modeling, Robustness, Adversarial Attacks.*

How to Cite: Niha Malali; Sita Rama Praveen Madugula (2025). Robustness and Adversarial Resilience of Actuarial AI/ML Models in the Face of Evolving Threats. *International Journal of Innovative Science and Research Technology*, 10(3), 910-916. https://doi.org/10.38124/ijisrt/25mar1287

I. INTRODUCTION

The Artificial Intelligence systems, in partnership with Machine Learning applications for actuarial science, have transformed all risk analysis activities, decision processes and predictive modeling operations. AI details support a large range of data sources to generate exact predictions and improve the detection of fraudulent schemes and the management of asset groups. AI-ML advancements provide better efficiency levels together with enhanced financial decision-making accuracy [1]. The use of AI/ML systems by companies generates intensified security threats because the protection of these models remains challenging. Maintaining the reliability and security of actuarial AI/ML models in altered financial environments has become critical to sustaining operational effectiveness.

An AI/ML model exhibits robustness through stable performance that remains consistent despite any changes to input data uncertainties affecting the model or interruptions from outside sources [2]. The robustness feature of actuarial applications guarantees models retain reliable risk prediction abilities no matter what economic conditions exist alongside demographic shifts or unexpected market disturbances [3]. Threats aimed at late-stage actuarial AI/ML models include data poisoning attacks as well as evasion approaches and model inversion techniques which pose severe dangerous risks. These attacks generate financial mispricing, biased risk assessment, and regulatory violations that result in the loss of credibility along with the reliability of AI-based actuarial systems.

The defense mechanisms found within adversarial resilience systems of actuarial AI/ML models work to safeguard models from threats by strengthening their security. Studies recommended that combining adversarial training with anomalous detection techniques and enhancing feature resistance constitute measures to protect against adversarial attacks [4]. Every AI-driven actuarial system operates under rules to meet industry demands that support transparency as well as maintain fairness standards while following security protocols. Security protocols need to adapt to technological progress because increasing threats demand continuous development of protection measures for actuarial models against potential risks.

This paper provides evaluates AI/ML model approaches in actuarial science through complete assessments of their

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25mar1287

resilience against adversaries. The research considers the main adversarial threats while examining existing defense techniques and proposes methods to boost model resistance. Security procedures and ethical standards earn attention during the implementation process of protected AI-driven actuarial models. By addressing these challenges, the actuarial industry can ensure the continued reliability and trustworthiness of AI/ML applications in financial risk assessment and decision-making.

The paper is organized as follows: The issue and goals of the study are presented in Section I. Section II examines relevant research and current approaches. Section III details the proposed approach, including data collection and model selection. Experiments, analyses, and findings are presented in Section IV. Findings and difficulties are covered in Section V. Limitations and potential directions are described in Section VI. Finally, Section VII concludes with key contributions.

II. FOUNDATIONS OF AI/ML IN ACTUARIAL SCIENCE

➢ *This section Provides Fundamental Technical Knowledge of AI/ML*

To start, although the definition of AI varies from person to person, it commonly refers to the study of how computers might mimic human intelligence see Figure 1. Among the elements that may be used to do this are natural language processing, text-to-speech, visual capabilities, robotics, and decision-making [5]. Many of the aforementioned aspects of AI rely on machine learning. The term ML refers to a broad category that includes several models and the methods used to tailor them to specific data or scenarios. Supervised, unsupervised, and reinforcement learning are all forms of learning. One branch of machine learning, deep learning, primarily deals with different kinds of neural networks [6]. There has been some success in the past with reducing runtime using categories of conventional predictive analytics methods.

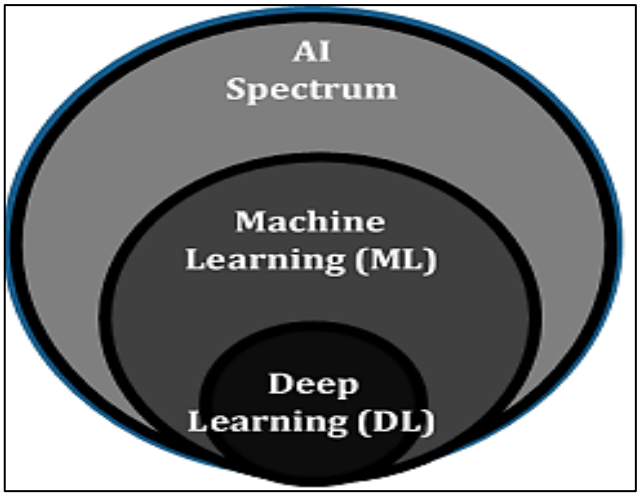

Fig 1 AI/ML Diagram

➢ *Role of Artificial Intelligence and Machine Learning in Actuarial Models*

In ML, supervised learning, unsupervised learning, and reinforcement learning are the three primary categories; actuaries often have the greatest experience with supervised learning. In actuarial work, supervised learning is prevalent and is used for tasks such as "predicting the frequency and severity of claims by fitting [generalized linear models] to claims datasets."[7], in order to forecast the frequency of policyholder lapses, or to databases of policyholders." As shown in Figure 2, supervised learning models can range from simple linear regression models to more complicated ones. The subfield of machine learning known as "unsupervised learning" focuses on pattern and sequence recognition[8].

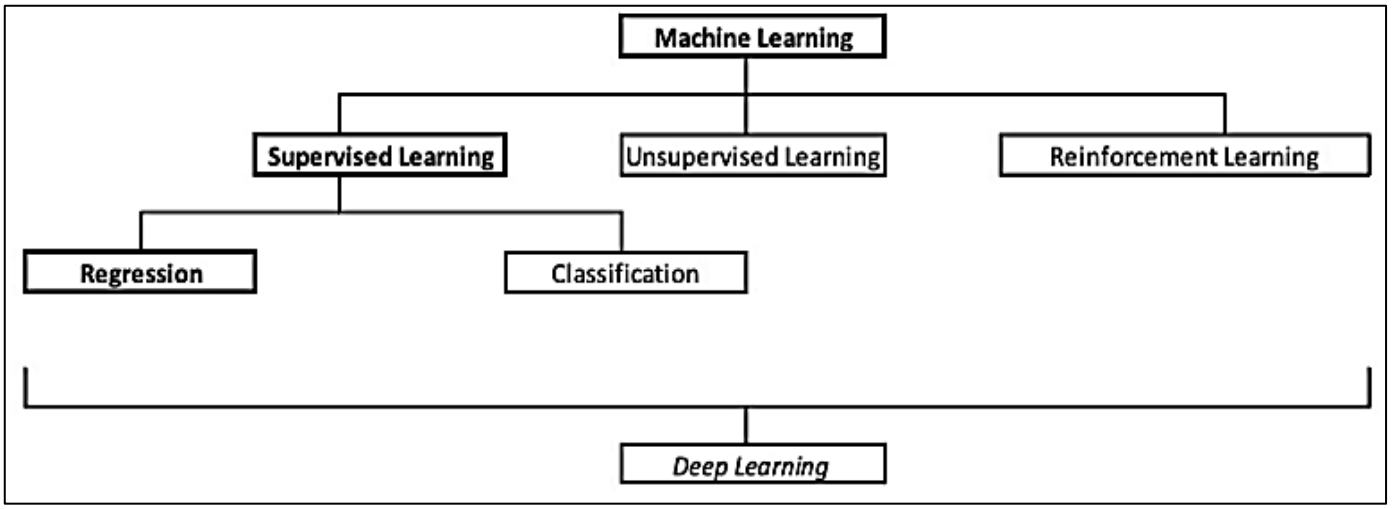

Fig 2 Artificial Intelligence and Machine Learning Usage in Actuarial Science

➢ *Common Techniques and Algorithms Used*

- *ML Algorithms*

In order to deal with data issues, ML uses a variety of methods. How many parameters there are, the kind of problem you're trying to answer, the optimal design, and other factors determine the technique used. Sample data is subject to both quality and quantity criteria. The "sample quality" metric measures how well the sample reflects the target population. In a mass-free scenario, the sample size is ideal [9]. The

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25mar1287

predictions produced by different algorithms vary according to the data sets, application approaches, and prediction goals. This necessitates repeated testing for better application results. Different learning methods make use of different algorithms. Different algorithms yield different prediction results for different uses [10]. The study focused on the standard design and the collected data; the results affect other linkages and additional machine learning can make the model change these events so they are risk-free and ignore big risk occurrences. The employment of state-of-the-art statistical approaches can greatly enhance the accuracy of forecasts. Many algorithms perform better on training sets than on testing ones, albeit the outcome isn't necessarily the same.

- *Deep Learning Techniques*

Deep learning encompasses a broad class of models and techniques for learning hierarchical representations from data. A few examples of the most prevalent deep learning architectures include [11]:

- Fully-connected neural networks (FCNs): The FCN architecture is similar to that of MLPs, but instead of using a linear activation function, each node in the network calculates the weighted sum of its inputs.
- Convolutional neural networks (CNNs): CNNs are made to handle grid-like data, such as time series and pictures[12]. They employ pooling layers to combine data across geographical or temporal dimensions and convolutional layers to identify local patterns.
- Recurrent neural networks (RNNs): RNNs are designed to handle sequential data, such as time series or text. Based on the prior hidden state and the current input, they update their concealed state at each time step.

➢ *Challenges and opportunities in Actuarial Science*

Although, actuarial science has improved over the years through so many developments, these have posed challenges as well. The nature of technological advancement trends that are very dynamic means that actuaries have to sharpen their skills all the time and embrace working tools. Due to its sensitivity to issues like data privacy and bias [13], consideration of the ethical use of machine learning and artificial intelligence has become an important concern. However, the future of actuarial science is promising despite the challenges mentioned above. New opportunities include climate risk modeling, cyber risk assessment, and personalized insurance where actuary can bring significant value. Thus, actuaries are able to remain engaged in the management of risk and the delivery of value in today's more complex environment by embracing innovation while staying true to their methods[14].

III. ROBUSTNESS OF ACTUARIAL AI/ML MODELS

In order to generate a shared definition of robustness, the notion of robustness and its relationships to other pertinent words are examined in this study. To maintain optimal performance, a strong production system must be able to handle disruptions. This may be accomplished by either responding appropriately to changing circumstances (flexibility, changeability) or by being resilient and agile in the face of disruptions) [15]. A definition of robustness will be created from these findings, which will also highlight the parallels and contrasts between the term's resilience, agility, flexibility, changeability, and performance.

➢ *Causal Perspective on Defining Robustness*

A causal view of robustness would offer a shared conceptual framework for comprehending different definitions or viewpoints of robustness in the literature, as their review did not identify any appropriate substitutes. Robustness is essentially a relative metric for model performance rather than an absolute one. The three main components of robustness that want to take into account are the actual image change or corruption, the model's design and optimization to counteract this corruption and the type of assessment and performance metrics. The shape and characteristics of and, in particular, frequently influence the design decisions and must be utilized to restrict the scope of their review. A causal approach to resilience in the context of deep learning for computer vision [16]. Assume that an SCM with a matching DAG and SCM may represent the DGP [17]. The definition of such a model is naturally made possible by knowledge of physics, scene creation, and other components of the picture-generating process. The causal method enables researchers to intuitively articulate assumptions and construct suitable priors using causal models connected to their application, as various imaging domains apply distinct generation processes.

➢ *Evaluating Robustness in AI/ML Models*

The aforementioned three datasets and the attack techniques found in the three APIs were used to construct a robustness validation mechanism. The method took adversarial attack bias into account as well. As to the IEEE Standard Glossary of Software Engineering Terminology, "robustness" refers to "the extent to which a system or component can operate accurately even when faced with faulty inputs or stressful environmental circumstances." The assessment should offer a straightforward and effective technique of calculating robustness without taking too long to collect data, such as perturbations or specific parameters if it employs an open testing approach. As a result, adversarial assaults were carried out within a predetermined range, and the robustness assessment score was produced using the attack accuracy of each model. These ratings show the relative robustness and enable comparing the defense capabilities of different models. Each trained model's resilience was assessed using perturbations of different sizes[18].

➢ *Factors Contribute to Model Robustness*

- The term "DevOps" was coined in 2009 by Patrick
- Debois, and since then, it has evolved into a comprehensive set of practices that emphasize
- automation, collaboration, and continuous improvement. The foundational principles of DevOps
- draws from Agile methodologies, Lean practices, and the Theory of Constraints. Humble and Farley
- (2010) in their seminal work "Continuous Delivery: Reliable Software Releases through Build,

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25mar1287

- Test and Deployment Automation" laid the groundwork for understanding CI/CD as integral
- components of the DevOps pipeline. They highlighted the importance of automating the build, test,
- and deployment processes to achieve faster and more reliable software releases
- The term "DevOps" was coined in 2009 by Patrick
- Debois, and since then, it has evolved into a comprehensive set of practices that emphasize
- automation, collaboration, and continuous improvement. The foundational principles of DevOps
- draws from Agile methodologies, Lean practices, and the Theory of Constraints. Humble and Farley
- (2010) in their seminal work "Continuous Delivery: Reliable Software Releases through Build,
- Test, and Deployment Automation" laid the ground work or understanding CI/CD as integral
- components of the DevOps pipeline. They highlighted the importance of automating the build, test,
- and deployment processes to achieve faster and more reliable software releases.

➢ *Generalization*

The generalization of a robust model to new data points should remain strong after training has completed through avoidance of overfitting to its training data. The model requires detecting meaningful data patterns then filtering out noise segments and unimportant data variations [19]. Multiple elements must be analyzed to achieve proper generalization:

- Sufficient and representative training data
- Avoiding overfitting
- Feature representation
- Hyperparameter tuning

➢ *Noise Tolerance*

A robust model remains unaffected by random data errors and unimportant input features. The model's design allows it to differentiate signal from noise and clearly focus on important data points in order to generate precise predictions [20]. These methods and techniques function to boost noise tolerance in machine-learning models:

- Feature engineering
- Regularization
- Data augmentation
- Ensemble method

➢ *Adversarial Robustness*

An adversarial example serves as an input whose design objective is to deviate from model prediction results. Models with robust structures can defend against such attacks because they stay accurate while facing perturbation attacks. The ability of a model to stay unharmed through such attacks is referred to as adversarial robustness [21]. A reliable model needs to perform accurately under circumstances where adversaries introduce perturbation to the input data. The research identifies multiple types of adversarial attacks that exist today:

- Defensive distillation
- Robust optimization
- Adversarial detection and rejection

IV. ADVERSARIAL THREATS TO ACTUARIAL AI/ML MODELS

Deep learning models encounter crucial security challenges from adversarial attacks during their deployment for cybersecurity purposes. The vulnerabilities within DL algorithms allow attackers to introduce subtle perturbations to input data that both humans cannot detect and lead to significant prediction changes from the model. The results from adversarial example testing showed that tiny modifications made to phishing emails, like adding noise, would make DL models mistake threats [22]. A CNN that detects phish sees a 25% decrease in its accuracy level when adversarial noise is applied to headers and email text content during an attack demonstration.

➢ *Types of Adversarial Attacks*

Attackers perform adversarial attacks through two methods: white-box and black-box. White-box attacks from attackers present the greatest threat because they acquire complete details regarding model architectures and parameters which enables them to take advantage of specific model vulnerabilities. Black-box attacks become possible through input testing of the model to uncover its security weaknesses. Both attacks represent major threats to DL-based cybersecurity systems, but white-box attacks prove more damaging since their effectiveness rises while black-box attacks provide practical usage in actual deployments.

➢ *Data Poisoning Attacks*

Adversarial attacks can only alter the test instance and cannot alter the model's training process; in contrast, data poisoning attacks can alter the training process [23]. To impact the learning model, attackers specifically try to change the training data (e.g., by flipping labels, poisoning features, changing the model weights, and adjusting the model configuration parameters). Attackers are presumed to possess the capacity to either contribute to or influence the training data for the exercise. The primary goal of injecting poison data is to affect the learning outcome of the model.

➢ *Model Inversion Attacks*

In order to replicate sensitive training data, model inversion attacks aim to leverage the model outputs depicted in Figure 3. This poses a significant privacy risk, particularly in applications involving sensitive data. Demonstrated how ML models trained on healthcare data could be queried to reveal private information about individual patients. Similarly, model extraction attacks involve an adversary attempting to replicate a model's functionality by querying it. These attacks highlight the necessity of understanding the trade-offs between model transparency and security[24].

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25mar1287

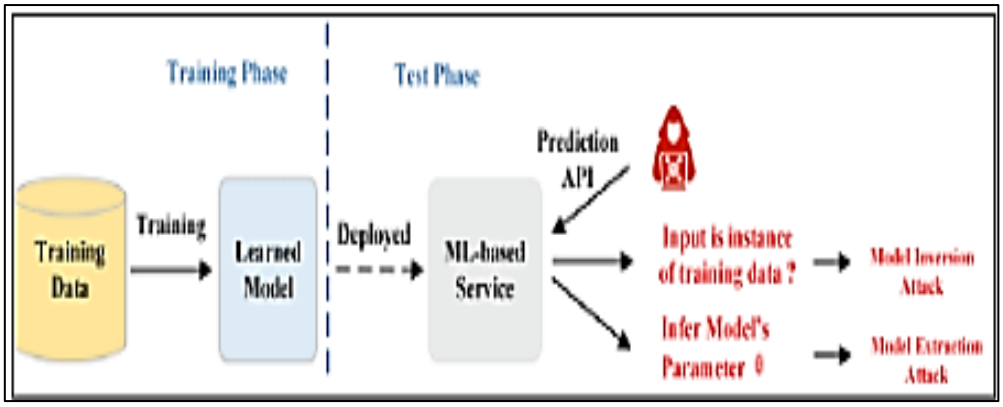

Fig 3 Model Inversion Attack

- *Evasion Attacks*

The primary topic of this work is Evasion Attacks, whereby the attacker modifies the input data using well-planned perturbations to deceive the ML model throughout the decision-making process. When a feature perturbation is introduced, it causes the IoT ML intrusion detection model to provide an incorrect classification [25]. In a poisoning attack, the attackers contaminate the training datasets with malicious samples. Where the attackers were able to introduce a backdoor into the machine learning detection model and incorrectly categorize harmful traffic as benign using a data poisoning assault against an IoT intrusion detection system that relied on federated learning. An attacker can use model stealing attacks to generate a "clone" model that closely resembles the target model by probing a black box model and extracting all the necessary information. It is important to note that creating ML models for a large system like NIDS is costly and resource-intensive. Therefore, the associations that established them would suffer a great loss if they were stolen. This paper's primary focus is on evasion attacks.

- *Impact of Adversarial Attacks on Actuarial Decision-Making*

A sample of input data that has been subtly modified to deceive an ML system is known as an adversarial example. As a result, the AI program predicts things incorrectly. Even while AI applications for text, speech, photos, and videos are getting more complex, Adversarial assaults focused on certain perturbations of their incoming data might nevertheless affect them. These disruptions can occasionally be so slight that they are invisible to the naked eye. In this situation, not only are ML systems tricked into detecting them, but a greater degree of these disturbances can also make the assault more successful by decreasing the system's accuracy [26]. The picture of a panda is a well-known adversarial example that demonstrates how subtle, imperceptible changes to the image's input pixels cause it to be incorrectly classified as a gibbon. The Appendix explains how various algorithms are used to produce adversarial assaults.

V. LITERATURE OF REVIEW

This study highlights the literature on robustness and adversarial resilience in Actuarial AI/ML and explores various methodologies, challenges, and advancements in enhancing model reliability and security against adversarial threats.

Gujar (2024) introduces an innovative approach to adversarial defense that diverges from traditional methods by proposing a defense strategy based on stable diffusion. Their method avoids training with adversarial examples and instead leverages continuous learning and comprehensive threat modeling to build inherently resilient AI systems. By addressing the limitations of existing defenses and emphasizing a dynamic, adaptive strategy, their approach aims to provide a more generalized and robust solution to adversarial threats [27].

Divya et al. (2024) focus on the application of adversarial autoencoders to improve the robustness of image generation. Therefore, in this study, adversarial training is proposed to be incorporated into the autoencoder structure in order to enhance the quality and robustness of the synthesized images. The approach includes using adversarial autoencoders with different datasets, CIFAR-10, CelebA and ImageNet, and estimating models' quality with the use of IS, FID, and MSE indicators. It is ascertained from the outcomes that adversarial autoencoders attain an Inception Score of approximately 8. 14, while the FID as assessed using the Frechet Inception Distance is 15 and the value of 19. 22 and Mean Squared Error of 0. 018, outperforming traditional autoencoders [28].

Tasneem and Islam (2024) In order to improve the adversarial training system, researchers should integrate explainable AI techniques with data augmentation techniques to fortify AI model predictions in remote sensing data against adversarial attacks. The suggested approach showed great robustness transfer capabilities against untested assaults in addition to having the highest PGD attack resilience in the Euro SAT and AID datasets [29].

Hannon et al. (2024) provide additional hostile tests as well as the RQS, a metric created especially to evaluate the subtleties of AI reactions. The study also includes Freedom, an AI tool designed to maximize the alignment between AI interpretation and user intent. The study's empirical results are crucial for assessing the security and robustness of AI models in use today. They highlight the necessity of thorough testing and ongoing development to fortify AI defenses against a

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25mar1287

variety of adversarial attacks. It's interesting that this study also looks at the social and ethical ramifications of using intricate "jailbreak" methods in AI testing. The findings are essential for comprehending the shortcomings of AI and developing methods to improve its ethics and dependability, opening the door for safer and more secure AI applications [30].

Navita, S. Srinivasan and Nitin (2024) combine two essential components—hostile defense and stability—to aid in the creation of AI systems capable of managing the intricate realm of cyber threats. Maintaining the reliability and integrity of these systems becomes both a technological and a societal requirement as AI continues to play a significant role in technological advancement. This is to guard against possible threats and vulnerabilities in the intricate digital landscape. These days, AI systems are used in many important sectors. Therefore, it's imperative to make these systems more resistant to cyberattacks [31]. A summary of the research, methodology, main conclusions, difficulties, and constraints of the robustness and adversarial literature on actuarial AL/ML is shown in Table I.

Table 1 Summary of Literature Review Based on Robustness and Adversarial Resilience of Actuarial Ai/Ml Models

References	Study On	Approach	Key Findings	Challenges	Limitations
Gujar (2024)[27]	Adversarial Defense Using Stable Diffusion	A defense strategy based on stable diffusion rather than adversarial example training	Introduces a dynamic, adaptive strategy to enhance resilience against adversarial threats	Requires comprehensive threat modeling and continuous learning	Generalizability to different AI applications needs further validation
Divya et al. (2024) [28]	Robustness of Image Generation using Adversarial Autoencoders	Incorporates adversarial training into autoencoder structures	Achieves improved Inception Score (8.14), FID (15), and MSE (0.018), outperforming traditional autoencoders	Dataset-specific performance variations	Potential computational overhead due to adversarial training
Hannon et al. (2024)[29]	Adversarial Training for Remote Sensing AI Models	Uses explainable AI with data augmentation techniques for adversarial robustness	AID and EuroSAT datasets provide the strongest adversarial resistance against PGD assaults.	Validation of the transferability of resilience to invisible assaults is still required.	Requires specialized datasets for effective training
Hannon et al. (2024)[30]	Evaluating AI Model Robustness and Security	highlights the Response Quality Score (RQS) for AI answers as well as additional hostile testing.	Highlights Freedom GPT's improvements in intent alignment and response security	Ethical concerns over AI "jailbreaking"	Ongoing need for meticulous testing to enhance AI security
Navita, S. Srinivasan and Nitin (2024)[31]	AI Resilience Against Cyber Threats	Combines stability and adversarial defense to enhance AI system security	Emphasizes the social necessity of securing AI in critical sectors	Complexity in integrating stability with adversarial robustness	Balancing security with AI interpretability remains a challenge

VI. CONCLUSION AND FUTURE WORK

The effectiveness of machine learning approaches in improving fault detection accuracy in semiconductor manufacturing. By leveraging advanced algorithms and data-driven methodologies, the proposed techniques enhance defect identification and classification, ultimately contributing to higher manufacturing yield and reduced production costs. The actual measurements displayed better performance than previous fault detection systems. Some restrictions affect its deployment because the system depends on high-quality inputs, requires extensive computing capacity, and struggles when applied to changing manufacturing platforms. In order to improve the ethical performance and dependability of AI-based actuarial approaches, more research is required to resolve data source compatibility difficulties. In hybrid AI systems, the integration of cutting-edge machine learning models with conventional actuarial techniques would improve forecast accuracy and dependability. XAI technologies represent a fundamental requirement that will support open decision-making transparency for actuarial processes.

REFERENCES

[1]. S. Mohamed, Artificial Intelligence implementations in Actuarial Science: Empirical Study for Mortality Rate Forecasting, no. September. 2024. doi: 10.52789/0302-043-164-007.

[2]. S. Arora, S. R. Thota, and S. Gupta, "Artificial Intelligence-Driven Big Data Analytics for Business Intelligence in SaaS Products," in 2024 First International Conference on Pioneering Developments in Computer Science & Digital Technologies

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25mar1287

(IC2SDT), IEEE, Aug. 2024, pp. 164–169. doi: 10.1109/IC2SDT62152.2024.10696409.

[3]. R. Richman, "AI in Actuarial Science," SSRN Electron. J., no. October, pp. 24–25, 2018, doi: 10.2139/ssrn.3218082.

[4]. E. al. Nand Kumar, "Enhancing Robustness and Generalization in Deep Learning Models for Image Processing," Power Syst. Technol., vol. 47, no. 4, pp. 278–293, 2023, doi: 10.52783/pst.193.

[5]. S. Tyagi, "Analyzing Machine Learning Models for Credit Scoring with Explainable AI and Optimizing Investment Decisions," Am. Int. J. Bus. Manag., vol. 5, no. 01, pp. 5–19, 2022.

[6]. J.-P. Larochelle et al., "Predictive Analytics and Machine Learning – Practical Applications for Actuarial Modeling (Nested Stochastic)," Soc. Actuar., 2023.

[7]. S. Tyagi, T. Jindal, S. H. Krishna, S. M. Hassen, S. K. Shukla, and C. Kaur, "Comparative Analysis of Artificial Intelligence and its Powered Technologies Applications in the Finance Sector," in Proceedings of 5th International Conference on Contemporary Computing and Informatics, IC3I 2022, 2022. doi: 10.1109/IC3I56241.2022.10073077.

[8]. J. Riley, "AI and Machine Learning Usage in Actuarial Science," University of Akron, 2020.

[9]. S. S. S. Neeli, "Optimizing Data Management and Business Intelligence: Integrating Database Engineering with AI-driven Decision Making," Int. J. Commun. Networks Inf. Secur., vol. 17, no. 01, p. 24, 2025.

[10]. H. Smith, "Applications of Machine Learning in Actuarial Pricing and Underwriting-A Review," Acad. Master, 2025.

[11]. A. Abdur, R. Khan, L. K. Tanwani, and S. Kumar, "Applications of Deep Learning Models for Insurance Pricing," no. June, 2024, doi: 10.13140/RG.2.2.36760.40965.

[12]. S. Masarath, V. Waghmare, S. Kumar, R. Joshitta, and D. Rao, "Storage Matched Systems for Single-click Photo Recognitions using CNN," 2023 Int. Conf. Commun. Secur. Artif. Intell., pp. 1–7, 2024.

[13]. S. Murri, "Optimising Data Modeling Approaches for Scalable Data Warehousing Systems," Int. J. Sci. Res. Sci. Eng. Technol., vol. 10, no. 5, pp. 369–382, Oct. 2023, doi: 10.32628/IJSRSET2358716.

[14]. A. P. A. Singh and N. Gameti, "Improving Asset Information Management in the Oil & Gas Sector: A Comprehensive Review," in 2024 IEEE Pune Section International Conference (PuneCon), 2024, pp. 1–6. doi: 10.1109/PuneCon63413.2024.10895047.

[15]. N. Stricker and G. Lanza, "The concept of robustness in production systems and its correlation to disturbances," Procedia CIRP, vol. 19, no. C, pp. 87–92, 2014, doi: 10.1016/j.procir.2014.04.078.

[16]. N. Drenkow, N. Sani, I. Shpitser, and M. Unberath, "A Systematic Review of Robustness in Deep Learning for Computer Vision : Mind the gap ?," 2021.

[17]. Suhag Pandya, "A Machine and Deep Learning Framework for Robust Health Insurance Fraud Detection and Prevention," Int. J. Adv. Res. Sci. Commun. Technol., pp. 1332–1342, Jul. 2023, doi: 10.48175/IJARSCT-14000U.

[18]. 8]C. Chang, J. Hung, C. Tien, C. Tien, and S. Kuo, "Evaluating Robustness of AI Models against Adversarial Attacks," pp. 47–54, 2020.

[19]. K. Than, D. Phan, and G. Vu, "Gentle Local Robustness implies Generalization," 2024.

[20]. K. Ding, J. Shu, D. Meng, and Z. Xu, "Improve Noise Tolerance of Robust Loss via Noise-Awareness," J. Latex Cl. Files, vol. 14, no. 8, pp. 1–15, 2021.

[21]. J. Puigcerver, R. Jenatton, C. Riquelme, and S. Bhojanapalli, "On the Adversarial Robustness of Mixture of Experts," NeurIPS, 2022.

[22]. F. Ekundayo, "International Journal of Research Publication and Reviews Leveraging AI-Driven Decision Intelligence for Complex Systems Engineering," vol. 5, no. 11, pp. 5489–5499, 2024.

[23]. J. Lin, L. Dang, M. Rahouti, and K. Xiong, "ML Attack Models : Adversarial Attacks and Data Poisoning Attacks," pp. 1–30, 2021.

[24]. C. Hiruni, "Securing Machine Learning Models : A Comprehensive Review of Adversarial Attacks and Defense Mechanisms," no. November, pp. 0–6, 2024, doi: 10.13140/RG.2.2.30723.92965.

[25]. A. Matrawy, "Adversarial Evasion Attacks Practicality in Networks : Testing the Impact of Dynamic Learning," pp. 1–13, 2024.

[26]. S. Pandya, "Predictive Analytics in Smart Grids : Leveraging Machine Learning for Renewable Energy Sources," Int. J. Curr. Eng. Technol., vol. 11, no. 6, pp. 677–683, 2021.

[27]. S. S. Gujar, "Enhancing Adversarial Robustness in AI Systems: A Novel Defense Mechanism Using Stable Diffusion," in 2024 2nd DMIHER International Conference on Artificial Intelligence in Healthcare, Education and Industry (IDICAIEI), 2024, pp. 1–6. doi: 10.1109/IDICAIEI61867.2024.10842888.

[28]. S. Divya, A. Sathishkumar, S. J. Jemila, R. Vanitha, and S. MC, "Improving Image Generation Robustness with Adversarial Autoencoder for Enhanced Quality and Stability," in 2024 Second International Conference Computational and Characterization Techniques in Engineering & Sciences (IC3TES), 2024, pp. 1–5. doi: 10.1109/IC3TES62412.2024.10877526.

[29]. S. Tasneem and K. A. Islam, "Improve Adversarial Robustness of AI Models in Remote Sensing via Data-Augmentation and Explainable-AI Methods," Remote Sens., vol. 16, no. 17, p. 3210, 2024, doi: 10.3390/rs16173210.

[30]. B. Hannon, Y. Kumar, D. Gayle, J. J. Li, and P. Morreale, "Robust Testing of AI Language Model Resiliency with Novel Adversarial Prompts," Electron., vol. 13, no. 5, 2024, doi: 10.3390/electronics13050842.

[31]. Navita, D. S.Srinivasan, and D. Nitin, "Resilient AI Systems : Robustness and Adversarial Defense in the Face of Cyber Threats," Int. J. Intell. Syst. Appl. Eng., vol. 12, no. 19, pp. 355–365, 2024.