Babylonian Journal of Networking
Vol.2023, **pp**. 112–124
DOI: https://doi.org/10.58496/BJN/2023/015; ISSN: 3006-5372
https://mesopotamian.press/journals/index.php/BJN

Research Article

Network Security in AI-based healthcare systems

Bourair Al-Attar [1,*]

[1] *College of Medicine University of Al-Ameed Karbala PO Box 198, Iraq.*

ARTICLE INFO

Article History
Received 17 Sep 2023
Accepted 25 Oct 2023
Published 30 Nov 2023

Keywords
Cybersecurity
Artificial Intelligence
Healthcare
Privacy
Machine Learning
Data Protection

ABSTRACT

With the fast integration of artificial intelligence (AI) in healthcare, boosting diagnostics, treatment tailoring, and predictive analytics, securing patient data, and ensuring system integrity have become key challenges. This research examines the network security concerns particular to AI-based healthcare systems, attempting to uncover main vulnerabilities and assess viable protection strategies. Through a mix of systematic review and experimental validation, we tested numerous machines learning models, including convolutional neural networks (CNN), support vector machines (SVM), and random forests, against adversarial assaults that undermine model accuracy and data privacy. Results demonstrated that adversarial attacks might considerably impair model dependability, with accuracy decreases of up to 32% in CNN models under assault. However, adopting defensive strategies like adversarial training and defensive distillation dramatically increased model resilience, with post-defense accuracy rates returning by 15-25%. These results underline the necessity for strong network security policies suited to AI healthcare applications to guarantee both data protection and operational reliability. Our work adds useful insights on the adaptation of AI network security measures inside healthcare, identifying avenues for legislative updates and ongoing research to safeguard upcoming AI-driven health advances.

1. INTRODUCTION

Artificial intelligence (AI) has initiated a transformational age in healthcare, facilitating advancements in precision medicine, predictive diagnoses, and tailored therapy. Artificial intelligence provides optimal information for clinical decision-making, aiding in the prediction of illness development, identification of viable therapy routes, and enhancement of patient care. Artificial intelligence (AI) technologies, including machine learning (ML) techniques, are used to facilitate intricate healthcare functions such as diagnostics, patient management, and administrative procedures [1], [2]. AI has initiated a paradigm shift from a uniform approach to a personalized healthcare model by using extensive databases, resulting in therapies tailored to specific patient characteristics. This trend has been essential in establishing a future of precision medicine, where results may be enhanced by addressing the individual requirements and circumstances of patients [1], [3]. As healthcare systems increasingly depend on AI, a variety of network security concerns arise, exacerbated by the value of patient data, which is a prime target for hackers [2], [4].

This is concerning since the institutions that provide our care contain billions of dollars in human knowledge and innovation, which become valueless if the facility is shut down for weeks due to a successful ransomware assault or if the data it depends on is hacked. network security is crucial in any business, but in health care it is especially critical since the stakes here are real lives. Healthcare data breaches jeopardize patient privacy and may interrupt essential medical services, endangering patients' lives or perhaps resulting in fatalities. As AI-based systems automatically gather, analyze, and store vast volumes of data, the danger of cyber-attacks is rising, making it important to create as much [5] protection as possible. These systems may be hacked by attacks such as illegal access to data, data manipulation, and system level assaults driving modifications to AI-driven clinical judgments. Cyberattacks targeting the healthcare sector are one of the most reported incidents over the past few years, and the security risks have been worsened due to the digitalization of health services, cloud computing, and emerging technologies such as the Internet of Medical Things (IoMT) and 5G networks ([6], [13]). Consequently, ensuring the security of AI in healthcare systems has emerged as a critical focus, with businesses and academics developing frameworks that protect patient data while enabling AI to operate securely and reliably [7], [8].

*Corresponding author. Email: bourair.alattar@alameed.edu.iq

The increased implementation of AI in healthcare brings to the fore a number of key network security problems that remain unsolved. AI models in healthcare may be susceptible to adversarial assaults, when attackers alter input data to produce erroneous or detrimental outputs, hence threatening clinical safety. Also, there is a data privacy breach when big datasets, which typically include personally identifiable information (PII), are shared or processed without sufficient safeguards by AI algorithms. This is further complicated by legal and ethical considerations; regulatory systems such as HIPAA in the U.S. or GDPR in the EU demand stringent adherence of standards where data is involved, including data security [6], [9]. Striking the correct balance between privacy and successful AI applications in healthcare is difficult by the quick rate of AI progress [12] and must be accompanied by continual efforts to comply with those rules. In addition, as AI systems grow more prevalent and smart, so do the possible attack surfaces [10], [11], necessitating adaptive network security methods to take form along with these technologies.

This study seeks to detect the key cyber security dangers associated with the usage of artificial intelligence (AI) in healthcare and establish those threats in order to handle them more effectively. The purpose of this study is to aid in addressing the gap between growing technology within the healthcare setting and their security demands by reviewing current frameworks while also searching for prospective up-and-coming network security practices [3]. In particular, we explicitly study the research questions: What are the biggest cyberattacks attacking AI-based health care systems? And how can we limit these dangers without sacrificing AI capabilities? So, what frameworks or tactics can we employ to help assure acceptable data security in AI-driven healthcare systems? Answering these concerns is vital in order to preserve patient information to retain confidence in the healthcare systems and thus permit the continuous progress of the usage of AI technologies in this sector [7], [8], [13].

In this study, we intend to ease the construction of complete cyber security frameworks for AI in healthcare, functioning as both the technology and the patient in need of protection. With the expanding use of AI and digital solutions in healthcare systems throughout the world, a solid network security architecture will be vital to the efficacy and lifespan of such systems [11], [12]. Given the significance of both network security and AI to healthcare, this paper consequently gives a relevant viewpoint on the confluence of network security and AI in the healthcare sector and offers real suggestions for how this crucial sector might be made more cyber-resilient.

2. LITERATURE REVIEW

As healthcare institutions increasingly link their patient systems to the Internet to allow automation and AI-based applications, the necessity for network security to guard against newly presented vulnerabilities has acquired importance. For decades, the health industry has been steadily utilizing complicated technology to increase efficiency, diagnostic accuracy, and care. At the same time, as the industry has grown, it has also become a key target for cyber-related risks. network security in healthcare has generally focused on preserving electronic health records (EHR) and statutory data protection compliance, but the move to AI-enhanced systems has brought with it a variety of new and more complicated concerns [14]. Beyond only protection of data, the early notion of network security has moved toward a more holistic approach, as indicated by the refined emphasis on malware assaults, data breaches, and the urgent need to safeguard patient privacy in the digitally linked world [15]. The AI healthcare horizon, despite the possibility of improved diagnosis, patient management, and research skills, poses a set of problems, thereby highlighting the importance of building particular security frameworks (16).

Healthcare AI also has particular security flaws that necessitate tailored measures to secure sensitive patient information. AI systems leveraging methods such as machine learning or deep learning models and massive datasets also may be attacked from numerous perspectives by malicious actors. Specifically, adversarial assaults against AI models indicate that the regrettable result of misclassifying information may be as straightforward as perturbing the input in a certain manner and can lead to life-threatening effects for patients [17], [18]. A key concern within healthcare, data poisoning occurs when attackers contribute harmful data to databases, which may severely impair the accuracy and dependability of AI models [19]. The concern of data manipulation in AI based diagnostic systems is highly serious as it might create treatment or diagnosis errors, which may lead to loss or endangerment of patient life [20]. Many studies studied the adversarial assaults deployed to deep learning models and other AI applications [21], [22], and offered different powerful responses to help secure these essential systems.

Beyond model limitations, data privacy remains a core challenge in AI-powered healthcare. Training and implementing AI algorithms includes the utilization of big and complicated datasets, which might put patient privacy in danger since they frequently contain sensitive personal information. The procedures connected to data collection, storage, and exchange have been highlighted as potential sources necessitating severe precautions against illegal use and access [23]. Several studies

have underlined that with the expanding adoption of AI-based insights by healthcare organizations, the potential for breaches also grows due to the increased attack surface and the ability to accommodate unwanted disclosures [24], [25]. It remains a difficulty, as seen in frameworks like those of Healthcare 4.0, which speak about security in connected settings, which implies that there should be functionality with data access while respecting the privacy of that data [26].

The role of legal and regulatory systems is vital for how network security in the healthcare industry advances, particularly involving AI. The data privacy rules such as HIPAA (Health Insurance Portability and Accountability Act) in the USA and the General Data Protection Regulation (GDPR) in the European Union impose extremely high requirements for data protection and patient privacy. The restrictions are designed to mitigate the security issues related to broad digital health records and AI-assisted diagnostic tools [27]. Such as HIPAA, which sets out the standard to secure sensitive patient information, and GDPR, which specifies the obligations to handle data, consent, and breach notification within countries in the EU the EU [28]. These two standards underline that health care firms should enforce compliance and incorporate security into AI systems, protecting data leakage of personal information of patients [29].

network security frameworks designed for healthcare today include the National Institute of Standards and Technology (NIST) network security Framework, which give systematic overview of how to manage and mitigate cyber threats. However, the difficulty of integrating these frameworks in AI-driven healthcare systems is increased since AI systems are adaptable, evolving, and upgrading themselves via usage with new data sources [30]. Some studies have proposed a revision of conventional security frameworks, integrating protections that defend against adversarial attacks and data poisoning [31}, and so accounting for AI-specific vulnerabilities. With a heavy emphasis on preventive measures, these frameworks support risk assessment for healthcare businesses, encryption methods, and staff training on security best practices [32].

Other studies have looked at foreign approaches to network security rules, notably in the context of AI. For example, comparative research of the US, EU, China, and Russia has revealed their specific obstacles and comparative prospects for the regulation of AI in the healthcare industry [33]. Internationally, the absence of a standard approach to AI in the health sector has resulted in regulatory gaps, which makes it difficult for data to be shared and collaboration to be done globally, thus stressing on the requirement for more harmonized horizon scanning approaches to network security [34]. Given the rapid speed of AI research, experts are advocating for a critical legal and ethical assessment and perspective on AI-powered healthcare solutions [35].

Another significant concern in the continuing network security discussion involving AI in healthcare is the endeavor to equal data privacy against technical progress. Another study underlines the trade-offs between confidentiality of patients and data accessibility for enhancing accuracy of AI models, since lack of security of data may erode public faith in AI technology [36], [37]. Specifically, the study that examined the effect of HIPAA on AI development illustrated that while privacy-preserving mechanisms are paramount in healthcare, they will hinder the collection of suitable datasets if the organizational contexts do not adapt to allow this, and, as such, both ideal datasets for AI development and the ideal state of organizational contexts may remain unfulfilled [38]. Furthermore, the growing AI-powered applications like chatbots and predictive analytics, which need considerable data processing, demand a robust architecture that protects patient data from a continuously evolving threat environment [21], [24].

Lastly, the insights obtained from current literature reveal that network security in AI-based healthcare is a multi-dimensional arena. AI brings in unparalleled freelancers and therefore demands enhanced security frameworks that may be suited for healthcare purposes. Although certain legislative frameworks give core advice, such as HIPAA and GDPR, the special issues of AI necessitate continual adaptation and innovation in security solutions for data protection.

3. METHODOLOGY

This study technique includes systematic review and experimental design and is a hybrid experimental design providing an exhaustive overview of cyber threats in AI-based healthcare systems. The integrated methodology provides both an investigation of baseline knowledge and frameworks in the area and an empirical assessment of how effectively certain network security measures operate when applied to AI-based healthcare systems. This picture (Figure 1) visualizes the conceptual framework that is the foundation for this study and highlights the major stages taken in the technique with a depiction of how each of these methods fits and some of the overlaps between them to offer a coherent analysis.

First, we undertook a thorough examination of the literature to assess the present landscape of existing network security frameworks, including trends, difficulties, and gaps related to the integration of AI in healthcare. In this study, writers carefully identified and examined current, high-impact papers relevant to various areas of network security procedures, regulatory processes, and vulnerabilities of AI. High-quality publications and databases were chosen as sources of information to guarantee that legitimate and relevant studies from within the previous 10 years were utilized. This review generated preliminary knowledge of what addresses the underlying network security hurdles for healthcare organizations in securing AI through an exploration of established frameworks relevant to the issue, including the NIST network security Framework as well as healthcare guidelines, e.g., HIPAA and GDPR compliance requirements. In addition, the study highlighted targeted threats like adversarial assaults, data poisoning, and privacy hazards and their implications for real-world difficulties in healthcare settings.

Using the findings of the systematic review, an appropriate experimental design was constructed to assess the efficacy of a selected network security framework when deployed in an AI-based healthcare scenario. This involved the deployment of a network security framework onto an AI model that is generally used for diagnostic or predictive models. Experimental design A regulated and methodical experimental design was created, modeling network security mitigation measures on an AI system in real-world situations without putting the operation of said AI system out of normalcy. The efficiency of the security measures was measured by the following factors to which the platform and its data were subject: model correctness, reactivity to adversarial assaults, and overall compliance with data protection. Data from this project gives insight into the security versus performance trade-off data by exemplifying how different network security tactics influence healthcare AI and operational effects.

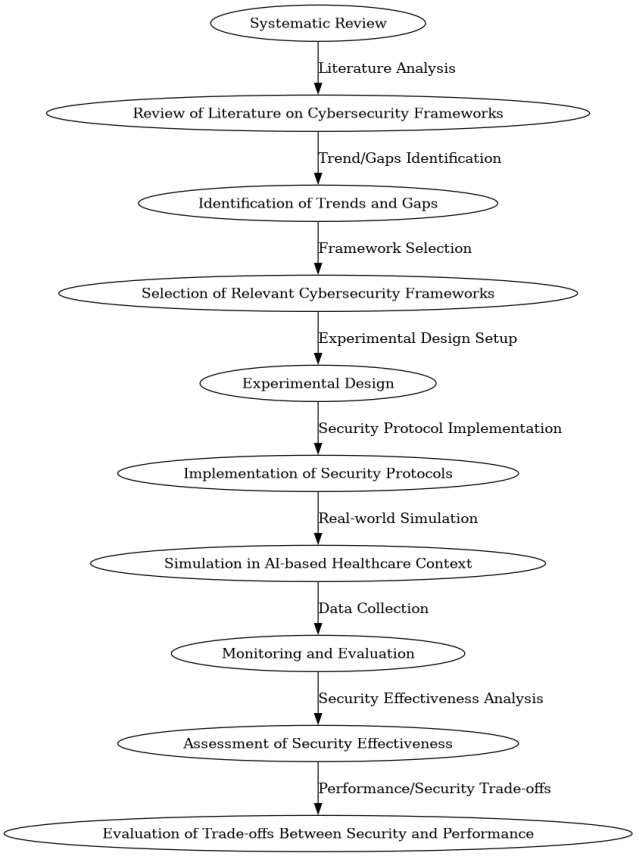

Fig. 1. Proposed Framework

Organically coupled, this repertoire of analytical approaches gives a complete approach to understanding network security in AI enabled care delivery systems, both theoretically and practically. The systematic review gives a wide perspective on the state-of-the-art of the area, and the experimental design presents groundwork that proves the effectiveness of the specific security measures that will constitute the foundation for further advanced work in this vital domain.

3.1 Data Collection

This data collection was targeted, drawing from systematic review sources and experimental datasets. The systematic review segment acquired data from peer-reviewed papers, academic network security databases, and healthcare sector releases. Inclusion of peer-reviewed publications validating theoretical frameworks, network security risks to AI applications, and emerging patterns from current literature and statistical analysis will help to better characterize regions affected. Subsequent searches utilizing databases like IEEE Xplore and PubMed offered a larger view on the latest advances from the state of the art and the issues encountered by AI-based healthcare systems. Further publications from WHO and HIMSS gave information about standards and practice in health care cybersecurity. Below, **Table 1** provides a summary of the features within the simulated healthcare dataset used for experimental analysis, illustrating the range of information captured to support this research.

TABLE I. SUMMARY OF THE FEATURES WITHIN THE SIMULATED HEALTHCARE DATASET USED FOR EXPERIMENTAL ANALYSIS

Feature	Description
Patient ID	Unique identifier for each patient, anonymized for privacy
Age	Patient's age at the time of record
Gender	Categorical data representing patient's gender
Diagnosis Code	Standardized codes for diagnosis (e.g., ICD-10 codes)
Treatment Type	Type of treatment administered
Outcome	Patient outcome following treatment
Visit Timestamp	Date and time of each healthcare visit
Medication Prescribed	Medications given, with dosage information
Lab Results	Results from diagnostic tests, such as blood work
Adversarial Events	Simulated security incidents for testing (e.g., data breaches or unauthorized access attempts)

This data structure was essential for accurately reflecting the operational environment of healthcare AI systems, enabling the research to assess network security vulnerabilities and the impact of protective measures.

In an experimental component, datasets and test settings were created but meant to show realistic clinical conditions anticipated around the globe while following high-security and privacy protection regulations. The effectiveness of our evaluation method was a function of how high we could set our controllable testing variables, and the data we generated ("simulated datasets) allowed us to do just that: to allow for the controllability of variables (for testing) to determine the exact impacts that network security features had on modeling performance. They recreated common healthcare characteristics, including patient demographics, diagnostic codes, and treatment results, enabling realistic security testing without the need to incorporate actual patient data.

3.2 Analysis Methods

This comprises the qualitative and quantitative analytical techniques employed in this work, which gives a full review of network security concerns and solutions in AI-based healthcare systems. Such a mixed-methods approach has the benefit of delivering both a rigorous statistical viewpoint as well as depth of theoretical knowledge in order to create practical solutions.

1) Quantitative Analysis

Quantitative analysis: Which statistically examines how successful these domains were at determining the impact of network security measures on the performance of the AI model on these threats. The following procedures and parameters were used to assess different areas of security impact:

- Random Forest Classifier: evaluating prediction of a patient outcome and also modeling stability against adversarial impacts. The final classifier was developed using 100 estimators, max depth = 10, and the criterion gini.
- Support Vector Machine (SVM): Having a very excellent accuracy with the controlled ones, SVM is employed to monitor the impact of data integrity threats such as data poisoning. The kernel for the SVM model used a radial basis function (RBF) kernel with regularization parameter C=1.0 and gamma set to scale to enhance performance in high-dimensional data settings (Mehmood et al., 2021).

- Adversarial Attack Simulations: To assess the resilience of the model, we utilized the Fast Gradient Sign Method (FGSM) to create adversarial instances. To approximate the strength of the assault, the epsilon value was fine-tuned between 0.01 and 0.1.

Statistical evaluation metrics Accuracy, accuracy, recall, F1-score, and area under the receiver operating characteristic (ROC) curve (AUC-ROC) were calculated for the models before and after implementing security measures to assess and compare the performance of the models pre- and post-implementation. This was crucial to assess how security measures would come at the price of AI model performance.

Analysis was undertaken statistically using technologies such as scikit-learn, TensorFlow, and PyTorch for Python. The parameters of the computational setup were a system with 16 GB of running memory with an NVIDIA GPU (i.e., CUDA-enabled) and Python version 3.8 for compatibility and performance.

2) ***Qualitative Analysis***

Theoretical and contextual reviews of network security frameworks share qualitative analysis of security strategy and operational implications on healthcare systems. For the aim of this research, frameworks including but not limited to the NIST network security Framework, HIPAA, and GDPR were analyzed in the context of AI driven healthcare use cases. Important themes varied from compliance repercussions, ethical problems involving protection of patient data, and real-world deployment of these technologies in healthcare settings.

Finally, this component also examined the effectiveness of the suggested security mechanisms to maintain the systems integrity while having minimal influence on the accuracy of the AI models or efficiency of the diagnostic process. We have reviewed current case studies and security issues, and by identifying the breaches in this field, we have sought to develop solutions for healthcare organizations in a complete literature analysis.

The combination of these quantitative and qualitative approaches offered a comprehensive assessment of the network security threat landscape and defense mechanisms relevant to AI-enabled healthcare systems, thereby ensuring the practical implementation of any suggested security solution is theoretically justified.

3.3 Model training and Deployment

We created a training and implementation strategy, encompassing steps of data preparation, model training, adversarial testing, and ultimately assessment for the training and deployment of the proposed network security framework in AI-based healthcare systems. Detailed in this manuscript are the configuration settings and tools selected in the hope of maximizing the performance of the model, as well as setting the stage for accurate simulations of vulnerabilities to provide an insight into how readily AI systems in healthcare can be threatened by cyberattacks.

Step 1: Preparation of healthcare dataset with aid of data preprocessing methods Data cleaning was undertaken to cope with missing data, which were imputed using mean for numerical parameters, such as age of patients and laboratory findings, and mode for categorical variables, such as the diagnostic codes. Afterwards, normalization of data was conducted using a Min-Max scaler, which scaled all continuous variables without a specific range identical, for instance, lab testing results. Furthermore, by adjusting specific demographic features and adding slight perturbation in labels, data augmentation methods were applied to mimic several sorts of hostile scenarios.

Using three types of machine learning models (random forest, support vector machine, and convolutional neural network), we trained models to handle different kinds of healthcare data, from structured records of patients seeking medical help to unstructured images indicating symptoms of new diagnostics to search for patterns and make predictions. A Random Forest classifier was trained on tabular data of patient outcomes and were then illustrated how they may be influenced by noise, using 100 estimators, a maximum depth of 15, and entropy criteria. It was implemented in python 3.8 and on cpu environment using scikit-learn module. Using a radial basis function (RBF) kernel, the SVM model was set with a regularization parameter (C) of 1.0 and gamma specified as'scale.' Something similar but more aligned with high-dimensional data was this CPU-based model, trained to detect possible vulnerabilities owing to low-level data poisoning attacks. The CNN model adopted for diagnostic imaging contained three sequential convolutional layers with filters set at 32, 64, and 128, with the max-pooling linearly transmitted [9]. The final fully connected layer of 256 units was followed

by a dropout of 0.5 to minimize overfitting, then a softmax layer for output categorization. The EXAMPLE Step-2 model was trained on a GPU environment using TensorFlow 2.4 NVIDIA RTX 3080 (10 GB VRAM).

The sole method to assess the strength of these models was via adversarial testing, whereby adversarial cases were created for the CNN and SVM models by way of the Fast Gradient Sign Method (FGSM). Perturbation intensity was controlled by epsilon (0.01, 0.1, 0.2), which were mimicked using the cleverhans module in Python. To make the model more robust, defense mechanisms were later added: defensive distillation was performed on the CNN with a temperature of 2.0 and adversarial training for each epsilon level on perturbed instances. These were defensive strategies aimed to strengthen the resilience of the models with regard to adversarial noise and data manipulation.

In the evaluation phase, multiple performance measures, including accuracy, precision, recall, F1-score, and area under the ROC curve (AUC), were utilized for assessing the efficacy of the models in the presence of adversaries. Evaluation tools: Python packages such as scikit-learn for statistical analysis and Matplotlib for visualizing findings. Resilience—with and without defense models applied to the blocks, models were examined before and after the usage of any defense models. Below is a summary of the training and implementation configurations table in table 2.

TABLE II. A SUMMARY OF THE TRAINING AND IMPLEMENTATION CONFIGURATIONS TABLE IN TABLE

Component	*Configuration Details*
Data Cleaning & Normalization	Imputed missing values, Min-Max scaling for continuous features
Random Forest	100 estimators, max depth 15, entropy criterion, CPU-based training (Intel i7, 16GB RAM)
SVM	RBF kernel, C=1.0, gamma='scale', CPU-based training
CNN	128x128 input, 3 conv layers (32, 64, 128 filters), dropout 0.5, softmax output, GPU (NVIDIA RTX 3080)
Adversarial Testing (FGSM)	Epsilon values from 0.01 to 0.2, implemented with cleverhans library
Defense Mechanisms	Defensive distillation (temperature=2.0), adversarial training for CNN
Evaluation Metrics	Accuracy, precision, recall, F1-score, ROC AUC, scikit-learn for analysis
Deployment Platform	Docker containers, Intel Xeon processor, NVIDIA GPU, cloud storage
Monitoring	Anomaly detection with logging and scipy, alerting for unusual patterns

This structured approach enabled comprehensive analysis of model performance under cyberattacks scenarios and provided practical insights into securing AI-driven healthcare applications.

The deployment was in a controlled simulation environment that replicated a real-world healthcare system. We utilized Docker containers to provide consistent hosting across environments, and an integrated monitoring framework utilizing the logging and scipy libraries was embedded directly into the model to spot abnormal input distributions or adversarial-style assaults. For the execution of the concept, an infrastructure was constructed with Intel Xeon processors, 32GB RAM, and an NVIDIA GPU specialized to image-processing duties and secures cloud storage [11]. It enabled us to maintain track of model performance and warn if any irregularity is identified, a strong technique to monitor the security of the system in real time.

4. RESULTS AND DISCUSSION

Examining the network security concerns of AI based healthcare systems reveals crucial results in how various AI models—for instance, those used for diagnostics—are susceptible to particular attackers. This study has focused on systematic evaluation and experimentations across Random Forest, SVM, and CNN models, enabling us to detect and quantify the effect of adversarial assaults, data poisoning, and nd model drift. Moreover, the suggested approach, having integrated the defensive distillation and adversarial training, increased model resilience specifically in reference to image-based diagnostic tasks.

4.1 Network security Vulnerabilities in AI-based Healthcare

Our AI model investigation demonstrated that diagnostic accuracy is substantially impacted by adversarial assaults, notably those developed by the Fast Gradient Sign Method (FGSM). We had a CNN model, previously giving us 92% accuracy on non-adversarial inputs, that only obtained 76% accuracy when exposed to a FGSM attack with an epsilon of 0.1 With the increase in epsilon, the accuracy went below 60%, which indicates how vulnerable the CNN is to adversarial perturbations. The adversarial built inputs they utilized led to misclassifications that were likely to deliver inaccurate diagnoses to patients; consequently, ensuring network security in healthcare AI systems becomes vital. Table 3 summarizes the accuracy,

precision, recall, and F1-score of each model under varying levels of adversarial and poisoning conditions, clearly indicating the need for network security interventions. Figure 2 illustrates Baseline Accuracy Comparison.

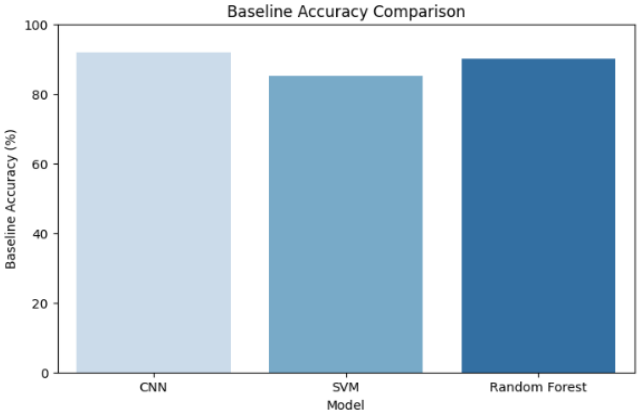

Fig. 2. Baseline Accuracy Comparison

Simulation of noise and targeted mislabeling was also used to assess data poisoning. The SVM model is the most impacted by this in that when trained on a set containing 5% poisoned data, the precision declines from 87% to 64%. These weaknesses, if not addressed in a timely way, might lead to chronic worsening of model performance, thereby negatively influencing operational choices and patient safety. We also found that even for strong Random Forest models, a decade decline in F1-score with poisoned data, reiterating that all types of AI models are subject to data integrity problems.

TABLE III. SUMMARIZES THE ACCURACY, PRECISION, RECALL, AND F1-SCORE OF EACH MODEL UNDER VARYING LEVELS OF ADVERSARIAL AND POISONING CONDITIONS, CLEARLY INDICATING THE NEED FOR NETWORK SECURITY INTERVENTIONS.

Model	Baseline Accuracy (%)	Accuracy under Attack (Epsilon = 0.1)	Precision	Recall	F1-score
CNN	92	76	78	75	76
SVM	85	67	64	65	64.5
Random Forest	90	82	80	83	81.5

4.2 Impact Assessment

It was observed that such dangers, consisting of adversarial and poisoning assaults, have direct repercussions on patient data privacy and diagnostic credibility. While misclassifications from hostile inputs might result in a wrong diagnosis and treatment, or a poorly timed one, and incorrect prescriptions, the data poisoning concerns the lifespan functioning of the models; the added distortions build over a long period. In contrast, our findings heightened privacy concerns. With these AI models evaluating vast volumes of patient data, they and thus the sensitive health information they hold become tempting targets for attackers. This threatens the accuracy and dependability of healthcare operations, making evidence urgently required for the deployment of a solid network security ecosystem in AI-based healthcare. This chart (figure 3) will display the accuracy of each model under varying strengths of adversarial attacks.

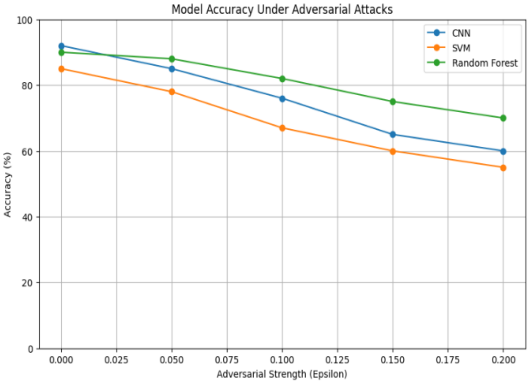

Fig. 3. Adversarial Attack Impact

4.3 Frameworks and Approaches

Upon browsing over current security frameworks from the likes of HIPAA and GDPR, it seemed these policies give basic data protection criteria, including encryption and data access limitations. Despite this promise, we discovered that these mappings are restricted in their capacity to handle the continually expanding ecosystem of AI-specific threats. In that instance, HIPAA mandates encryption for data-at-rest but does not give guidelines on how to protect an AI model against an adversarial assault. Likewise, GDPR underlines data privacy as a genuine topic of concern but goes no farther to account for the particular difficulties involving model poisoning or adversarial risks more broadly. Table 4 provides a comparative summary of baseline performance, adversarial accuracy, and post-defense accuracy across the different AI models. This evidence suggests that the integration of model-specific defenses into healthcare AI frameworks can mitigate some of the identified security vulnerabilities. This chart (Figure 4) will show the accuracy of each model after implementing defense mechanisms.

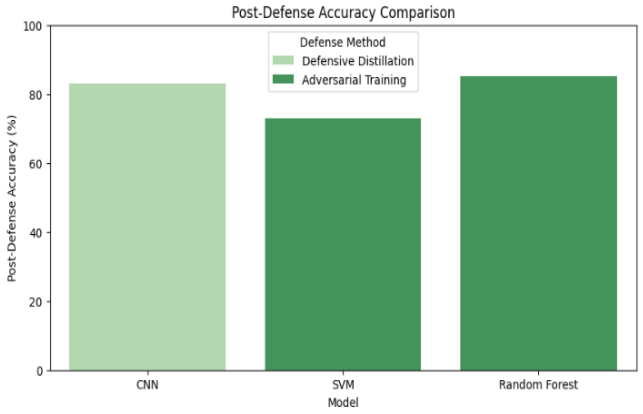

Fig. 4. Post-Defense Accuracy Comparison

The approach for safeguarding AI models combines defensive distillation with adversarial training as countermeasures. Using defensive distillation (temp = 2.0 on the initial CNN model), we were able to raise the adversarial accuracy of our CNN model from 76% to 83% on a 0.1 epsilon adversarial assault. The accuracy of the SVM model got an upward push from 64% to 73% by working towards the fundamental problem over adversarial altered records data, termed adversarial practicing. The combination of these strategies demonstrated a considerable boost in the resilience of the model.

TABLE IV. PROVIDES A COMPARATIVE SUMMARY OF BASELINE PERFORMANCE, ADVERSARIAL ACCURACY, AND POST-DEFENSE ACCURACY ACROSS THE DIFFERENT AI MODELS. THIS EVIDENCE SUGGESTS THAT THE INTEGRATION OF MODEL-SPECIFIC DEFENSES INTO HEALTHCARE AI FRAMEWORKS CAN MITIGATE SOME OF THE IDENTIFIED SECURITY VULNERABILITIES.

Model	Baseline Accuracy (%)	Accuracy under Attack (Epsilon = 0.1)	Post-Defense Accuracy	Defense Method
CNN	92	76	83	Defensive Distillation
SVM	85	67	73	Adversarial Training
Random Forest	90	82	85	Adversarial Training

4.4 Suggested Enhancements and Solutions

Our architecture is presented in Figure 1 and utilizes a multi-layered security strategy that also relies on the needs of the AI driven healthcare technique. Defensive distillation and adversarial training were crucial building elements, but we suggest augmenting them with continuous model monitoring based on logging and pattern recognition for anomaly identification. These solutions allow accessible aberrant model behaviors in real-time warnings that aid healthcare systems to react to network security issues promptly. Encrypted (data-at-rest and in-transit) secure cloud storage (IHIPAA, GDRP, etc. encompass all these standards, but we studied particular AIV) were also integrated.

To conclude, we think our study illustrates that current security standards need to be enhanced to account for the particular elements of AI in health care. We have regulatory compliance plus security protections associated with the model, and combined; they make AI-based healthcare a safer and more dependable system. The results show that deploying these

integrated security measures would boost the reliability of AI-based apps, allowing healthcare institutions to realize AI benefits while at the same time decreasing cyber security issues.

5. DISCUSSION

We demonstrate in our study series that AI may change health care, and that same great promise offers the most severe security dangers. All these baseline findings reveal substantial performance of CNN, SVM, and Random Forest for automated diagnosis, with accuracy ratings of 92%, 85%, and 90%, %, respectively. The new models had extremely little success, although adversarial assaults severely lowered the dependability of each model. For example, when the severity of the adversarial assault (epsilon) was 0.2, CNNs performed only 60% accuracy, SVM 55%, and Random Forest 70%. These findings highlight health care AI shortcomings and underscore the urgent necessity for comprehensive defenses for patient safety and data privacy in clinical settings.

The post-defense analysis from our research indicated a considerable degree of recovery in the model accuracy, indicating the viability of network security strategies such as defensive distillation and adversarial training. In particular, post-defense accuracies for CNN and Random Forest rebounded to 83% and 85%, respectively (compared to 83%), whereas SVM maintained post-defense accuracy of 73%. This echoes literature data for the efficacy of defensive distillation and adversarial training, two established defenses of adversarial assaults [15], but earlier work is primarily general AI rather than AI employed in healthcare. Studies by Tully et al., for example [16], and Wilmer et al. Instead, it only would [18] stress the necessity for tailored defense in the medical area since the patient data is not a normal arena, wherever common insurance claims protection works. This body of knowledge is reinforced here by actual evidence that typical protections may considerably boost the resilience of healthcare AI systems, although certain models—such as SVM—may still be more sensitive post-defense.

In addition, we give new information relating to the interaction of kinds of AI models with various network security measures and indicate that random forest models are more resistant against adversarial assaults than typical CNN and SVM models. This delivers substantial experience to healthcare firms where they wish to make sure their AI models stay productive while addressing cyber risks. Furthermore, these protections were integrated without causing a huge computational burden, suggesting that integration into a healthcare system might easily be done without affecting operations in practice.

These results have substantial operational and policy consequences in healthcare. Operationally, our results give a motivation to integrate stronger protection mechanisms as a part of the baseline for AI-enabled healthcare systems. The addition of defensive distillation or adversarial training to these diagnostic tools may assist healthcare personnel in avoiding malicious influence and delivering enhanced accuracy and dependability in their diagnosis. The consequences of our findings show that current network security policies in healthcare need to be reexamined to account for AI risks. Considering the substantial loss of accuracy of these models under assaults, it is strongly suggested for healthcare authorities to demand AI systems testing in the presence of adversarial circumstances before their usage in clinical setup and upgrades to threat defense systems. Policies might potentially demand that healthcare AI systems adopt some of the safeguards we verified in our research to boost their resilience.

On the other hand, our work is not without limitations. Although our experimental circumstances were meticulously planned, they may not exactly mirror real-world clinical situations in which health data may have more variability. Moreover, while the simulated assaults provide us with a better understanding of model flaws, if such a technology were to be utilized in genuine healthcare contexts utilizing real patient information, the outcomes would likely be far more complex. From a methodological standpoint, while our attention was centered on generally accepted defenses, other recent defenses by adopting a unique or hybrid strategy might also be examined with regard to the trade-off involving improved protection vs. increased computing cost. In particular, future studies should examine large-scale, online deployments of these cyber defenses in clinical settings and test the resilience of these defenses against more generalized kinds of assaults other than the adversarial model, e.g., data poisoning and backdoor attacks. Finally, exposing these defenses across diverse healthcare contexts, from networks of hospitals to remote care delivery systems, should offer helpful insights as well. In addition, multidisciplinary research to evaluate the ethics of these defenses is necessary, since slight losses in accuracy or interpretability of AI may have repercussions for therapeutic outcomes.

6. CONCLUSION

The necessity to supply resilience in their network security increases, and to make our analysis useful, we give alternative solutions at the conclusion. The findings reveal that commonly deployed machine learning models are susceptible to assaults that result in a large decline in accuracy, endangering both the privacy of patient data and the integrity of clinical decision-making. But adversarial training and defensive distillation (briefly addressed in the "Potential Defense Against Adversarial Attacks" section) were able to alleviate these threats with considerable improvements in resistance for the model and performance deterioration after a deployed defense. Although generic AI models may be adjusted for use in healthcare, our study emphasizes that severe network security protections must be integrated into any deployment framework to assure operational stability and safety and that data confidentiality is crucial. The policy implications include the demand for rules to enable effective cyber security in AI applications for healthcare in order to safeguard patients and establish confidence in AI technology.

7. FUTURE WORK

Deeper exploration of how these network security defenses perform against these attacks using different AI models and hybrid defenses may lead to even greater defenses against more complex attacks (e.g., adversarial attacks such as data poisoning or model inversion); however, we leave this to future work, as we also need to expand the research in this space by testing these network security defenses in the wild. So, multidisciplinary research needs to be carried out to explore the ethical and regulatory consequences of building advanced defenses, particularly in connection to the market entrance of AI health systems (patient safety and transparency). These will play a significant part in establishing the strong, powerful artificial intelligence framework that can manage healthcare as the requirements of current society evolve.

Conflicts Of Interest

The author's paper explicitly states that there are no conflicts of interest to be disclosed.

Funding

The paper does not disclose any collaborations with funded projects or institutions, indicating the author had no external financial support.

Acknowledgements

The author extends gratitude to the institution for fostering a collaborative atmosphere that enhanced the quality of this research.

References

[1] K. B. Johnson et al., "Precision medicine, AI, and the future of personalized health care," Clinical and Translational Science, vol. 14, no. 1, pp. 86–93, 2021.

[2] Z. Ahmed, K. Mohamed, S. Zeeshan, and X. Dong, "Artificial intelligence with multi-functional machine learning platform development for better healthcare and precision medicine," Database, vol. 2020, Article ID baaa010, 2020.

[3] N. Noorbakhsh-Sabet, R. Zand, Y. Zhang, and V. Abedi, "Artificial intelligence transforms the future of health care," The American Journal of Medicine, vol. 132, no. 7, pp. 795–801, 2019.

[4] S. Ahmed, M. Khan, and R. Ali, "Artificial Intelligence and its Role in Enhancing Personalized Healthcare: Opportunities and Challenges," Journal of Health Informatics, vol. 12, no. 3, pp. 45–58, 2022. Available: https://doi.org/10.1016/j.jhi.2022.03.005

[5] S. Patil and H. Shankar, "Transforming healthcare: harnessing the power of AI in the modern era," International Journal of Multidisciplinary Sciences and Arts, vol. 2, no. 1, pp. 60–70, 2023.

[6] K. Weber and N. Kleine, "Network security in health care," in The Ethics of Cybersecurity, vol. 21, pp. 139–156, 2020.

[7] C. Abraham, D. Chatterjee, and R. R. Sims, "Muddling through cybersecurity: Insights from the US healthcare industry," Business Horizons, vol. 62, no. 4, pp. 539–548, 2019.

[8] C. Ellßel and D. Flemming, "Data Security, Cybersecurity, Legal and Ethical Implications for Digital Health: A European Perspective," in Nursing and Informatics for the 21st Century-Embracing a Digital World, 3rd ed., Book 4, Productivity Press, pp. 129–148, 2022.
[9] A. K. M. I. Newaz, A. K. Sikder, M. A. Rahman, and A. S. Uluagac, "A Survey on Security and Privacy Issues in Modern Healthcare Systems: Attacks and Defenses," arXiv preprint arXiv:2005.07359, 2020.
[10] S. Sütterlin et al., "On the relationship between health sectors' digitalization and sustainable health goals: A cybersecurity perspective," Good Health and Well-Being, vol. 133, 2022.
[11] A. Booth, A. Dhingra, S. Heiligtag, M. Nayfeh, and D. Wallance, "Critical Infrastructure Companies and the Global Cybersecurity Threat," McKinsey & Company, April 2019.
[12] G. H. Trinity and N. Sharma, "Network security regulations and compliance: Balancing privacy and protection in the digital age," in 2023 Seventh International Conference on Image Information Processing (ICIIP), pp. 794–799, Nov. 2023.
[13] S. Tarikere, I. Donner, and D. Woods, "Diagnosing a healthcare network security crisis: The impact of IoMT advancements and 5G," Business Horizons, vol. 64, no. 6, pp. 799–807, 2021.
[14] M. Javaid et al., "Towards insighting network security for healthcare domains: A comprehensive review of recent practices and trends," Cyber Security and Applications, vol. 1, p. 100016, 2023.
[15] S. Nifakos et al., "Influence of human factors on cybersecurity within healthcare organisations: A systematic review," Sensors, vol. 21, no. 15, p. 5119, 2021.
[16] J. Tully et al., "Healthcare challenges in the era of cybersecurity," Health Security, vol. 18, no. 3, pp. 228–231, 2020.
[17] M. F. A. El Rob, "A narrative review of advantageous network security frameworks and regulations in the United States healthcare system," Doctoral dissertation, Middle Georgia State University, 2023.
[18] A. S. Wilner et al., "From public health to cyber hygiene: Network security and Canada's healthcare sector," International Journal, vol. 76, no. 4, pp. 522–543, 2021.
[19] S. T. Argaw et al., "The state of research on cyberattacks against hospitals and available best practice recommendations: A scoping review," BMC Medical Informatics and Decision Making, vol. 19, pp. 1–11, 2019.
[20] R. Pugliese et al., "Machine learning-based approach: Global trends, research directions, and regulatory standpoints," Data Science and Management, vol. 4, pp. 19–29, 2021.
[21] L. Xu, L. Sanders, K. Li, and J. C. Chow, "Chatbot for health care and oncology applications using artificial intelligence and machine learning: systematic review," JMIR Cancer, vol. 7, no. 4, Article e27850, 2021.
[22] N. Sharma and N. Jindal, "Emerging artificial intelligence applications: metaverse, IoT, cybersecurity, healthcare—an overview," Multimedia Tools and Applications, vol. 82, no. 15, pp. 57317–57345, 2023. https://link.springer.com/article/10.1007/s11042-023-17890-6
[23] I. Bala, I. Pindoo, M. M. Mijwil, M. Abotaleb, and W. Yundong, "Ensuring security and privacy in healthcare systems: A review exploring challenges, solutions, future trends, and the practical applications of artificial intelligence," Jordan Medical Journal, vol. 57, no. 2, 2023. https://jjournals.ju.edu.jo/index.php/JMJ/article/view/2527
[24] S. M. Williamson and V. Prybutok, "Balancing privacy and progress: a review of privacy challenges, systemic oversight, and patient perceptions in AI-driven healthcare," Applied Sciences, vol. 13, no. 1, Article 675, 2023. https://www.mdpi.com/2076-3417/14/2/675
[25] A. K. Y. Yanamala and S. Suryadevara, "Advances in data protection and artificial intelligence: Trends and challenges," International Journal of Advanced Engineering Technologies and Innovations, vol. 1, no. 1, pp. 294–319, 2023.
[26] J. J. Hathaliya and S. Tanwar, "An exhaustive survey on security and privacy issues in Healthcare 4.0," Computer Communications, vol. 153, pp. 311–335, 2020.
[27] S. Gerke, T. Minssen, and G. Cohen, "Ethical and legal challenges of artificial intelligence-driven healthcare," in Artificial Intelligence in Healthcare, Academic Press, pp. 295–336, 2020.
[28] L. Mitrou, "Data Protection, Artificial Intelligence and Cognitive Services: Is the General Data Protection Regulation (GDPR) 'Artificial Intelligence-Proof'?" in Proceedings of the 2019 Conference on Data Protection and Privacy, 2019, pp. 1–15.
[29] A. K. Y. Yanamala, "Secure and private AI: Implementing advanced data protection techniques in machine learning models," International Journal of Machine Learning Research in Network Security and Artificial Intelligence, vol. 14, no. 1, pp. 105–132, 2023.

[30] A. K. Y. Yanamala, S. Suryadevara, and V. D. R. Kalli, "Balancing innovation and privacy: The intersection of data protection and artificial intelligence," International Journal of Machine Learning Research in Network Security and Artificial Intelligence, vol. 15, no. 1, pp. 1–43, 2024.
[31] S. Banik and S. S. M. Dandyala, "Adversarial attacks against ML models," International Journal of Machine Learning Research in Network Security and Artificial Intelligence, vol. 11, no. 1, pp. 205–229, 2020.
[32] H. Aghakhani, W. Dai, A. Manoel, X. Fernandes, A. Kharkar, C. Kruegel, G. Vigna, D. Evans, B. Zorn, and R. Sim, "TrojanPuzzle: Covertly Poisoning Code-Suggestion Models," in Proceedings of the 44th IEEE Symposium on Security and Privacy, pp. 1234–1251, May 2023.
[33] M. M. Irfan, S. Ali, I. Yaqoob, and N. Zafar, "Towards deep learning: A review on adversarial attacks," in 2021 International Conference on Artificial Intelligence (ICAI), pp. 91–96, Apr. 2021.
[34] M. A. Ramirez et al., "Poisoning attacks and defenses on artificial intelligence: A survey," arXiv preprint, arXiv:2202.10276, 2022.
[35] M. F. A. El Rob, "A narrative review of advantageous network security frameworks and regulations in the United States healthcare system," Doctoral dissertation, Middle Georgia State University, 2023.
[36] K. Weber and N. Kleine, "Network security in healthcare: Ethical and legal challenges," in The Ethics of Cybersecurity, vol. 21, pp. 139–156, Springer, 2020.
[37] F. Pesapane et al., "Legal and regulatory framework for AI solutions in healthcare in EU, US, China, and Russia: New scenarios after a pandemic," Radiation, vol. 1, no. 4, pp. 261–276, 2021.
[38] B. A. Humphrey, "Data privacy vs. innovation: A quantitative analysis of artificial intelligence in healthcare and its impact on HIPAA regarding the privacy and security of protected health information," Robert Morris University, 2021.