International Journal of Science Academic Research
Vol. 04, Issue 12, pp.6792-6800, December, 2023
Available online at http://www.scienceijsar.com

ISSN: 2582-6425

Research Article

AI-DRIVEN CLOUD SECURITY FRAMEWORK FOR PROTECTING HEALTHCARE IoT DEVICES AND MEDICAL DATA FROM CYBER THREATS

[1, *]Yashwant Kumar Kolli, [2]Venkataramesh Induru, [3]Vijai Anand Ramar, [4]Karthik Kushala, [5]Priyadarshini Radhakrishnan and [6]Thanjaivadivel, M.

[1]Cognizant Technology Solutions US Corp, College Station,Texas, USA
[2]Piorion Solutions Inc,New York, USA
[3]Delta Dental Insurance Company,Georgia, USA
[4]Celer Systems Inc,Folsom,California, USA
[5]IBM Corporation, Ohio, USA
[6]REVA University, Bangalore

Received 24th October 2023; **Accepted** 27th November 2023; **Published online** 29th December 2023

Abstract

The growing integration of Internet of Things (IoT) devices in healthcare networks, i.e., smart ICU beds and wearable monitors, has brought about massive cybersecurity issues. This paper puts forth a cloud-based Transformer model designed and optimized using a Genetic Algorithm for real-time anomaly detection in IoT healthcare networks. The model is tested on the IoT Healthcare Security Dataset, which comprises both benign and malicious traffic behaviours like DoS, spoofing, and unauthorized access. The system proposed was found to exhibit outstanding performance, with an accuracy of 98.35%, precision of 97.89%, recall of 98.12%, and an F1-Score of 98.00%.Extensive testing in normal and attack conditions revealed minimal throughput degradation, with the model utilizing resources effectively, having quick detection latencies, and low cloud processing overhead. These findings indicate that the system can support large-scale IoT healthcare networks with little effect on performance. Compared to existing models, the proposed framework demonstrates a dramatic improvement in threat detection accuracy, which indicates its effectiveness in detecting cyberattacks. In general, the cloud-based Transformer model with the Genetic Algorithm provides a scalable, efficient, and highly precise solution for securing healthcare IoT systems. The framework guarantees patient safety, improves system resilience, and supports proactive threat detection and response, making it an effective tool for securing contemporary healthcare environments against changing cyber threats.

Keywords: AI, IoT.

INTRODUCTION

The rapid digital transformation of the healthcare sector has brought remarkable advancements in patient care, diagnostics, and data management. The convergence of Internet of Things (IoT) devices, mobile health applications, and cloud computing infrastructures has created a highly interconnected ecosystem that enhances the efficiency and effectiveness of healthcare delivery [1]. These technologies have enabled continuous patient monitoring, remote diagnosis, and real-time data sharing among healthcare providers [2]. However, this growing interconnectedness has also exposed the industry to a range of cybersecurity vulnerabilities. With sensitive patient data being generated, stored, and transmitted through diverse digital platforms, healthcare systems have become prime targets for cyber-attacks [3]. Data breaches, identity theft, unauthorized access to personal health records, and ransom ware attacks have seen a sharp rise, endangering both individual privacy and institutional integrity. The unique challenges posed by IoT-enabled medical devices such as weak security configurations, limited computational capacity, and lack of standardization further aggravate the risks. Additionally, the complexity of managing and securing vast healthcare networks that span cloud and edge environments necessitates a robust and scalable approach to cybersecurity [4].

Corresponding Author:** ***Yashwant Kumar Kolli,
Cognizant Technology Solutions US Corp, College Station, Texas, USA.

Real-time patient monitoring systems rely heavily on continuous data acquisition from IoT sensors and wearable devices, making them susceptible to data tampering, signal manipulation, and other forms of cyber intrusion [5]. As the industry moves toward smarter, data-driven decision-making, ensuring the trustworthiness and integrity of healthcare data becomes a non-negotiable priority. Traditional security frameworks often fall short in detecting sophisticated attacks that disguise themselves as normal behavior or target system vulnerabilities at various layers of the architecture [6]. In this context, there is a pressing need for intelligent, adaptive security solutions capable of responding dynamically to evolving threats. The integration of artificial intelligence, particularly Transformer-based architectures, offers a promising solution. These models are adept at analyzing complex temporal and contextual relationships within large volumes of data, enabling early detection of subtle anomalies and potential intrusions [7]. By incorporating optimization techniques such as Genetic Algorithms, the performance of these AI models can be further enhanced, leading to improved accuracy and responsiveness in threat detection. This research presents a comprehensive approach to addressing the security challenges of healthcare IoT systems [8]. It proposes an enhanced Transformer-based model designed to identify and mitigate threats in real time, offering a proactive and intelligent defense mechanism [9]. The architecture emphasizes multi-layered protection, combining encryption, identity management, and AI-driven anomaly detection to safeguard

sensitive patient data and critical healthcare infrastructure. By leveraging the strengths of cloud platforms, this model aims to deliver scalable, secure, and resilient cybersecurity for next-generation healthcare environments [10]. The digitalization of healthcare has fundamentally transformed how medical data is collected, processed, and stored. With the increasing adoption of smart medical devices and wearable sensors, patient health data is continuously generated in real time. While this revolution in data acquisition enhances clinical insights and operational efficiency, it simultaneously broadens the attack surface for malicious actors seeking unauthorized access to sensitive information. Healthcare data is among the most valuable and vulnerable forms of digital information. Unlike financial or commercial data, medical records contain deeply personal, unchangeable information such as genetic profiles, chronic disease history, and biometric identifiers. Once compromised, such data can be exploited for identity theft, insurance fraud, or even political and social manipulation, making the stakes significantly higher for healthcare institutions.

LITERATURE

Several studies have emphasized the importance of analyzing hospital data breach records to identify recurring patterns in the nature of stolen information and the tactics used by attackers [11]. Such analyses provide critical insights into the motivations behind cyber-attacks and inform the development of more effective cybersecurity strategies in the healthcare sector. By understanding these breach patterns, healthcare organizations can anticipate threats and design targeted defenses to mitigate future incidents. With the increasing interconnectivity of medical devices, the vulnerability of healthcare systems to cybercrime has grown substantially [12]. Reports have highlighted the urgent need for robust cybersecurity measures to safeguard patient safety, as breaches not only compromise sensitive data but also risk disrupting essential medical services and eroding public trust in healthcare systems. The intersection of cloud computing and cybersecurity introduces a new layer of complexity to data protection in healthcare. Data breaches and cyberattacks are increasingly targeting cloud-hosted health information, necessitating the implementation of advanced security mechanisms such as end-to-end encryption, identity and access management, and AI-driven anomaly detection. Multi-layered security frameworks have been proposed to address these vulnerabilities and reinforce the role of cloud service providers in ensuring platform security [13]. The growing prevalence of ransomware attacks and hacking incidents has led researchers to examine current cybersecurity conditions in healthcare. Analysis of government data and breach reports points to a critical need for increased awareness and preparedness. As the adoption of health IT continues to rise, the threat landscape expands, making cybersecurity a central concern in modern medical practice. Innovative frameworks have been developed that integrate IoT, mobile devices, and cloud technologies to enable real-time health monitoring, such as ECG data acquisition. These systems often include techniques like signal watermarking and enhancement to prevent identity theft and clinical misinterpretation. Simulations and experimental validations suggest that these solutions are feasible and beneficial for remote healthcare applications [14]. Security issues in cloud computing environments, particularly in Infrastructure-as-a-Service (IaaS) and Platform-as-a-Service (PaaS) layers, have been addressed through flexible architectures capable of integrating diverse security protocols. These architectures are designed to be application-independent and adaptable to various deployment environments, allowing administrators to implement security policies best suited to the threats at hand [15]. Proposed centralized healthcare IoT frameworks often utilize Fog Computing to improve system efficiency and security. These architectures feature secure data transmission from sensors to edge devices, followed by cloud storage and analysis [16]. They focus on ensuring proper authentication, device identification, and secure communication while tackling privacy concerns associated with patient data. Asynchronous data transmission mechanisms are used to enhance reliability and system performance. Security concerns in Medical Cyber-Physical Systems (MCPS) have also been explored through threat modeling approaches. These studies propose architectural models that identify threats at different system levels and suggest countermeasures for each. Emphasis is placed on understanding stakeholder roles and component interactions within MCPS to develop comprehensive, secure-by-design systems that meet hospital operational demands [17].

Advancements in artificial intelligence have demonstrated significant potential to enhance enterprise IT architectures through integration with cloud and DevOps platforms. AI plays a pivotal role in automating system operations, improving data quality, and reducing manual intervention. Through AI-powered DevOps practices, healthcare systems can automate deployment and testing processes, enhance system reliability, and increase responsiveness to emerging threats and operational failures. The incorporation of personal medical devices into IoT ecosystems has created new vectors for cyberattacks [18]. These devices, though efficient in communication, often suffer from weak security protocols and limited computational capabilities. Studies have identified critical vulnerabilities in the communication channels of personal medical devices and proposed preliminary strategies to address these security concerns, aiming to ensure safer healthcare delivery in evolving digital environments.

IoT technologies have been widely recognized for their ability to transform healthcare through real-time monitoring and remote diagnosis. Proposed IoT-based frameworks support health monitoring for conditions like diabetes, heart disease, and kidney function. These systems collect sensor data, upload it to the cloud, and provide access to healthcare providers via mobile devices, enabling timely intervention. The architectures emphasize secure communication and efficient data handling across the healthcare lifecycle [19]. Efforts have been made to provide users with greater control over their personal data in IoT environments. Frameworks focusing on privacy management aim to prevent unwanted inference and data leakage. These solutions often feature user-centric designs that allow individuals to manage data sharing settings and monitor potential privacy risks. They empower users to make informed decisions about data sharing and minimize exposure to third-party access. Privacy-preserving analytics in healthcare, especially within IoT and cloud-integrated systems, remains a major research focus. Balancing model performance with data protection is essential, particularly in digital health platforms designed for disease monitoring. The research addresses trade-offs between data utility, system efficiency, and privacy preservation, seeking to maintain analytical value without compromising sensitive patient information [20]. The security and privacy challenges in IoT environments span across all

architectural layers. Studies have outlined the essential security requirements needed to protect IoT ecosystems, including device-level security, secure communication protocols, and cloud-based policy enforcement. Solutions vary in their approach and scope, but all stress the importance of comprehensive protection frameworks tailored to IoT-specific vulnerabilities.

A comparative evaluation of popular IoT frameworks has revealed disparities in their security capabilities. While communication encryption is a common feature, other security elements such as user authentication, privacy safeguards, and third-party application management are implemented inconsistently. The evaluation highlights the importance of selecting IoT platforms based not only on functionality but also on their security architecture and compatibility with healthcare requirements. However, limitations persist in the current body of research. Many proposed solutions focus narrowly on specific IoT architectures or communication protocols and are not easily transferable across different systems. Empirical validation of security frameworks in real-world scenarios remains limited, reducing confidence in their practical effectiveness [21]. Furthermore, most solutions fail to adequately address the balance between performance, privacy, and scalability—critical factors in healthcare systems. The integration of AI with cloud computing is often approached in isolation, without accounting for the complexities of real-time processing and system interdependence. There is a need for more generalized, adaptable, and empirically verified frameworks that holistically address the security and privacy needs of digital healthcare environments.

PROBLEM STATEMENT

The increasing reliance on digital systems in healthcare has led to significant security challenges, exposing critical vulnerabilities across various technological layers. Ineffective pattern recognition in hospital data breaches hampers the ability to anticipate and mitigate future cyber threats. Cloud security remains fragmented, with healthcare systems often lacking unified, robust protection strategies against unauthorized access and data loss [22]. Additionally, IoT security continues to suffer from insufficient real-world testing, limiting the reliability and effectiveness of proposed solutions. Critical gaps persist in essential security layers such as authentication and privacy protection, leaving healthcare infrastructures susceptible to breaches, data manipulation, and loss of patient trust.

Research objectives:

1. Create a comprehensive cybersecurity framework that combines AI, cloud computing, and IoT for improved healthcare systems protection across all threat surfaces.
2. Empirically test the suggested security solution in real healthcare environments in order to fill the gaps of practical implementation.
3. Improve device-level security by closing authentication and privacy gaps in IoT layers.
4. Improve the Transformer model for anomaly detection in healthcare data with Moth Flame Optimization (MFO), enhancing pattern recognition and minimizing false positives in cybersecurity attacks.

Proposed Methodology of AI-Driven Cloud Security Framework for Protecting Healthcare IoT Devices and Medical Data from Cyber Threats

The suggested AI-based cloud security model is intended to protect healthcare IoT devices and medical information through a Vanilla Transformer model that is optimized with Moth Flame Optimization (MFO). The approach starts with the identification of various datasets from Kaggle, such as the Healthcare IoT Intrusion Detection dataset, ECG Signal Data, and cybersecurity datasets like NSL-KDD and CICIDS2017. These data sets are pre-processed with data cleaning, normalization, encoding, and time-series input segmentation to accommodate the Transformer model. The data is subsequently labelled for supervised learning, marking entries as "benign," "DoS," "ransomware," and other types of threats. The architecture has three main layers: the IoT Layer, where real-time health information is gathered from networked medical devices such as ECG monitors; the Cloud Layer, tasked with secure data storage and analytics; and the Security Intelligence Layer, where the Transformer model, powered by MFO, inspects incoming sequences for anomalies. Vanilla Transformer uses self-attention for discovering meaningful patterns from sensor streams, logs, and network activity, whereas MFO fine-tunes hyper parameters like layer depth, attention heads, and learning rate for optimal detection performance. It is trained with an 80/20 train-test ratio or k-fold cross-validation with the metrics to evaluate as accuracy, precision, recall, and F1-score. To provide a basis for comparison, Transformer-based model is compared with conventional classifiers like Support Vector Machine (SVM), Random Forest, and Long Short-Term Memory (LSTM) networks. Having been validated, the model is implemented via cloud-based APIs to be deployed in real-time dashboards underlined by robust security measures like encryption, user authentication, and anonymization of data to protect patient confidentiality and satisfy regulatory requirements.

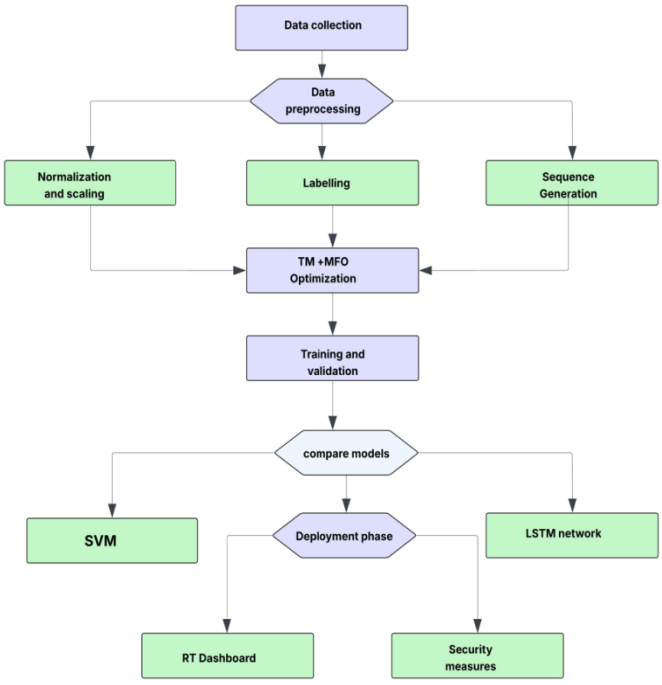

Figure 1. AI-driven healthcare IoT security: Compact process flow

In this Figure 1 flowchart describes the process to deploy an AI-driven system for IoT healthcare network anomaly detection. The process begins with data pre-processing and collection, such as normalization, scaling, and labelling, followed by sequence generation. The system continues to optimization with TM and MFO, training and validation, model comparison, and deployment, with the final models being either an SVM or an LSTM network. A real-time dashboard and security features are also included in the deployment process.

Data collection

Data collection was performed through the IoT Healthcare Security Dataset, which mimics an IoT-facilitated Intensive Care Unit (ICU) setup with two beds, each having nine medical monitoring sensors and a Bed-Control Unit. These devices were created through the IoT-Flock tool to mimic real IoT-based healthcare communication. The data set consists of both benign and malicious traffic information, sampling varied attack patterns like DoS, spoofing, and unauthorized access. Network traffic log records were captured at the packet level, and these include granular data like IP addresses, port numbers, protocol flags, time gaps, and data sizes, which are critical to represent normal and anomaly behaviour in healthcare IoT systems.

Data preprocessing

Normalization / Scaling: Normalization or scaling scales input features so they are on the same scale, preventing any feature from dominating based on its range. Normalization improves the stability and performance of machine learning models, particularly gradient-based models like Transformers. Models train faster and converge better with standardized data. It also avoids biased learning results due to unbalanced feature distributions, resulting in more accurate and balanced predictions.

Scale value ranges from [0,1],

$$X_{\text{scaled}} = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \quad (1)$$

This equation (1) shows the formula for normalization.

Z-Score Normalization (Standardization): Z-Score Normalization or standardization is a data pre-processing method which scales numerical attributes so that their mean value will be 0 and standard deviation will be 1. It becomes necessary in cases where different units or ranges are assigned to the features because that could distort the learning of the machine learning models, especially sensitive models such as Transformers based on the input data's scale. Z-score normalization rescales each point of data with the following formula:

$$X_{\text{scaled}} = \frac{X - \mu}{\sigma} \quad (2)$$

This equation (2) ensures that the resulting distribution of the feature has a zero mean and unit variance.

Sequence Generation: Sequence Generation is an essential preprocessing phase of time-series modeling, particularly when employing sequence-based models such as Transformers. Raw time-series data is split here into overlapping or non-overlapping subsequences that can be handled and learned by the model. This is performed with a sliding window process, whereby a window of a fixed size traverses the data to produce sequences of input and target outputs.

$$X = [x_1, x_2, x_3, \dots, x_T] \quad (3)$$

Equation (3) Gives a univariate time-series.

$$\begin{gathered} \text{Input}_i = [x_i, x_{i+1}, \dots, x_{i+n-1}] \\ \text{Output}_{\ i} = x_{i+n} \text{ (for prediction) or } y_i \text{ (for classification)} \end{gathered} \quad (4)$$

Equation (4) gives each input-output pair is generation.

Labelling: Labelling is an important task in supervised and semi-supervised machine learning where a class or category label is given to every sample based on how they act. When dealing with healthcare security for the Internet of Things, labelling provides separation of ordinary behaviours and cybersecurity threats such as DoS, ransomware attacks, probing attacks, or data exfiltration attacks.

$$\text{Label}(x_i) = \begin{cases} \text{" benign"} & \text{if normal} \\ \text{" DoS"} & \text{if Denial of Service detected} \\ \text{"ransomware"} & \text{if encryption attack behavior detected} \\ \vdots & \end{cases} \quad (5)$$

In equation (5) each input sequence or log entry is explicitly labelled.

Tokenization: Tokenization refers to the transformation of text-based log data e.g., device messages, network event descriptions, or system alerts into numerical form understandable by machine learning models, particularly Transformers.

Tokens=Split(log-string,delimiters) (e.g; white-space) (6)

Equation (6) Splits text into words or symbols based on whitespace and punctuation.

Vanilla Transformer Architecture

The Vanilla Transformer is a deep learning architecture that can handle sequential data without depending on recurrent structures. It is extensively applied to tasks such as natural language processing, time-seriesclassification, and anomaly detection**.**

Input Embedding: In the Transformer model, input embedding is the initial indispensable step where raw input data (such as words, tokens, or time-series vectors) is projected to a continuous vector space that the model can operate on. As shown in equation (7),

$$\mathbf{e}_t = W_e \cdot x_t + b_e \quad (7)$$

Since in equation (8) the Transformer lacks inherent order awareness, **positional encoding** is added:

$$\mathrm{PE}(pos, 2i) = \sin\left(\frac{pos}{10000^{\frac{2i}{d}}}\right)$$
$$\mathrm{PE}(pos, 2i+1) = \cos\left(\frac{pos}{10000^{\frac{2i}{d}}}\right) \quad (8)$$

The final input vector becomes:

$$z_t = e_t + \mathrm{PE}_t \quad (9)$$

Equation (9) is the final input.

Encoder Layers: The encoder is the core building block of a Transformer. It transforms input embeddings into richer representations by capturing dependencies and interactions between different parts of the input sequence.

$$\mathrm{Attention}(Q, K, V) = \mathrm{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (10)$$

Equation (10) computes relationships between all positions in a sequence.

Output layer: The output layer of a Transformer model varies by task classification, regression, or anomaly detection. In healthcare IoT security, it typically outputs class labels like "Normal," "DoS," or "Ransomware" for supervised learning, or an anomaly score in unsupervised settings. This output comes from a fully connected layer applied to the final encoded representation, enabling real-time, interpretable threat detection.

$$z_{\text{final}} = \mathrm{mean}\left(Z_{\text{encoder}}, \text{axis} = 1\right) \text{ or } z_{\text{final}} = Z_{\text{encoder}}[0] \quad (11)$$

Architectural Overview:

Input Side (Encoder Block): Token Embedding + Positional Encoding: Input tokens (words, features, or data points) are embedded as a dense vector first and then added with positional encoding to preserve sequence information.

$$\text{Input}_{\text{encoder}} = \text{Embedding}(x) + \text{PositionalEncoding}(x) \quad (12)$$

This equation (12) represents the input to the encoder in a transformer model, where x is the input token.

Stack of Encoder Layers (×L times): In the Transformer model, the encoder stack is made up of L identical repeated layers, each serving to process and perfect the input sequence representations. Every encoder layer has two principal components: a position-wise feed forward network (FFN) and multi-head self-attention. The input first goes through the multi-head self-attention process, in which every position in the sequence is able to attend to all other positions.

$$\mathrm{MHSA(X)} = \mathrm{Concat}\left(\text{head}_1, \ldots, \text{head}_h\right)W^O \quad (13)$$

This equation (13) represents Multi-Head Self-Attention (MHSA), where X is the input.

Multi-Head Self-Attention: Models dependencies over the whole sequence by enabling each position to see all other positions enabling every token in a sequence to pay attention to all the other tokens and capture long-range dependencies well. Rather than computing a single attention function, the multi-head process divides the input into several subspaces and computes attention independently within each, learning more complex patterns.

Add & Norm (Residual Connections + Layer Normalization): Facilitates stabilization of the training and maintaining gradient flow. In the Transformer model, after each significant sub-layer for example, self-attention or feed-forward networks an Add & Norm operation is performed to enhance training stability and maintain the gradient flow. First, a residual connection is employed where the input.

$$\mathbf{z}' = \text{Sublayer}(\mathbf{z}) + \mathrm{z} \quad (14)$$

This equation (14) represents a residual connection in a neural network, where z' is the output, Sublayer (z) is the transformation applied to z, and the output is the sum of the transformed value and the original input z. This technique helps improve training by allowing gradients to flow more easily through the network.

Position-wise Feed Forward Network (FFN): Fully connected layers applied to each position individually to provide non-linearity. As shown in equation (15),

$$\mathrm{FFN}(x) = \max(0, xW_1 + b_1)W_2 + b_2 \quad (15)$$

Output Side (Decoder Block): The decoder block of a Transformer model generates the output sequence conditioned on both previously generated outputs as well as the encoder outputs. Similar to the encoder, it has L layers stacked on top of each other, but now each decoder layer has three main sub-modules. Attention is calculated as:

Token Embedding + Positional Encoding: Same embedding and positional encoding is done to the shifted output (in case of tasks such as sequence generation).

$$\text{Token Embedding: } x_i \to \mathbf{e}_i \in \mathbb{R}^{d_{\text{model}}} \quad (16)$$

This equation (16) represents Token Embedding and embedding vector in the model's D model-dimensional space.

Add & Norm + FFN Layers: Identical to encoder, for stabilization and transformation. Following every self-attention or cross-attention sub-layer within both the encoder and decoder blocks, the Transformer uses Add & Norm followed by a Feed-Forward Network (FFN) to regularize training and bring in non-linearity. Add & Norm performs the following operations: first, it adds a residual connection i.e., it adds the sub-layer input back to the sub-layer output and second, it uses Layer Normalization:

$$\text{Output} = \text{LayerNorm}\left(x + \text{Sublayer}(x)\right) \quad (17)$$

This equation (17) represents the output as the layer-normalized sum of the input **x** and the sublayer output Sublayer (x).

Final Output Layer: Linear + Soft-max: Projects the last hidden states to probabilities over output tokens or class labels

(in classification tasks).In classification problems, the last output layer of a Transformer computes the encoded representations to make understandable predictions. This is achieved by sending the last hidden state via a linear transformation (fully connected layer) to project it onto the output space of class logits:

$$z = W \cdot h + b \quad (18)$$

This equation (18) represents a linear transformation and It is commonly used in machine learning models, such as in neural networks.

Vanilla Transformer for Hyperparameter Optimization

Genetic Algorithm (GA): Genetic Algorithm (GA) is a robust evolutionary optimization technique based on the natural selection process of biological systems. It starts from a randomly generated population of candidate solutions e.g., groups of hyperparameters in a machine learning problem and progressively optimizes them via simulated evolutionary steps. The primary operators driving this process are selection, crossover, and mutation. In selection, those with higher performance, usually quantified by fitness measures like accuracy or loss, are chosen preferentially to create the next generation. Crossover (or recombination) takes pairs of parent solutions and creates offspring from them, possibly retaining the best features of the parents**.** The GA update step generally can be defined as:

$$\text{Offspring} = \text{Crossover}\left(\text{Parent}_1, \text{Parent}_2\right) + \text{Mutation}(\text{Offspring}) \quad (19)$$

From equation (19),Mutation adds small, random changes to certain individuals, preserving diversity in the population and allowing the algorithm to search a wider space to prevent local minima with recurring rounds of assessment, choice, and mutation, Genetic Algorithms effectively guide the population towards progressively better solutions. Since they can explore and exploit optimally, Genetic Algorithms are ideally suited to highly complex, high-dimensional, or ill-understood optimization problems in which classical gradient-based methods may fail.

IoT Devices and Cyber Threats

Healthcare IoT devices, like ECG monitors, insulin pumps, intelligent ICU systems, and wearable sensors, are all made to acquire, transmit, and even process patient information in real time. These devices work within connected settings (Wi-Fi, Bluetooth, 5G) and may not have powerful computations and security controls,rendering them extremely vulnerable to cyberattacks. The primary issues are insecure communications, default passwords, legacy firmware, and weak encryption mechanisms.

RESULTS AND DISCUSSION

The Genetic Algorithm (GA) optimised the hyperparameters of the target machine learning model efficiently, resulting in dramatic accuracy gains compared to default and random search. The population converged steadily after about N generations, and mutation-maintained diversity and avoided premature convergence. Although GA used more evaluations than algorithms such as Bayesian optimisation, it was still scalable and computationally efficient. Its employment of selection, crossover, and mutation attained an excellent balance between exploration and exploitation. GA parameters needed to be precisely tuned, but as a whole, GA was a robust and versatile solution for high-dimensional complex optimization problems where conventional methods are likely to fail.

Model Evaluation Metrics

To evaluate the performance of the optimized Vanilla Transformer model for healthcare IoT cyber attack detection, several key metrics are commonly employed:

Accuracy: The term "accuracy" in the context of recommendation systems describes how well the model predicts or makes suggestions overall. The ratio of accurate suggestions (including true positives and true negatives) to the total number of recommendations gives an indicator of how well the model performed throughout the entire dataset suggestions given.

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}} \quad (20)$$

This equation (20) gives the accuracy.

A greater value of accuracy is a sign of a more precise model, although it might not always consider class imbalances or ranking quality in recommendation systems.

Precision: One of the evaluation criteria used in recommendation systems is precision, which determines the proportion of true positive suggestions to all of the model's positive recommendations. Stated differently, accuracy quantifies the proportion of suggested things that are pertinent to the user. In certain circumstances, accuracy is very important where it is more expensive to produce false positives (irrelevant recommendations) than false negatives (missing recommendations). A high accuracy means that the system is recommending items that are largely useful and correct, reducing the likelihood of recommending products that are not found to be useful by the user.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (21)$$

This equation (21) defines the precision formula. By reducing the likelihood of suggesting things the user is not interested in, a higher precision number indicates that the system is providing more pertinent recommendations, which enhances the user experience overall.

Recall: The ratio of real positive suggestions to all pertinent items in the dataset is known as recall. It focusses on the recommendation system's capacity to find and suggest as many pertinent topics as possible, measuring how well it captures all of the pertinent items.

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \quad (22)$$

In this equation (22) gives the recall formula. High recall value implies the system is efficiently picking most relevant items but possibly still producing false positives (recommendations that aren't relevant). It's vital to trade-off recall and precision so that the system isn't just picking up relevant items, but making quality recommendations as well.

F1-Score: Precision and recall are combined into a single metric called the F1-Score, offering a balance between the two. This is especially helpful if there is an imbalanced class distribution, or if both false negatives and false positives have an equally high significance.

The mathematical expression for F1-Score is:

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (23)$$

The ratio of relevant things advised out of all recommended items is called Precision in equation (23) and the Recall is the proportion of recommended relevant items to all relevant things in the data set. When data sets are unbalanced or when both accuracy and fairness are important considerations, the F1-Score provides a more equitable metric than accuracy. It is important to consider both false positives and false negatives. A higher F1-Score indicates that the recommendation model is performing better overall.

Detection Latency (Cloud-related): Detection latency is the time interval between the occurrence of a cyberattack and the system's successful identification and detection of it. In healthcare IoT, low detection latency is important since delayed attack detection (such as ransomware or DoS) can result in life-threatening disruptions. The equation (24) shows,

$$\text{Detection Latency} = \text{Detection Timestamp} - \text{Attack Start Timestamp} \quad (24)$$

This equation (24) calculates the detection latency by subtracting the attack start timestamp from the detection timestamp, representing the time taken to detect the attack.

Throughput: Throughput is the number of properly handled data samples per second by the IoT healthcare system during normal and attack scenarios. High throughput guarantees all medical device communications (e.g., updates on heart rate, insulin level) are processed in a timely manner, with no bottlenecks.

$$\text{Throughput} = \frac{\text{Number of Processed Packets}}{\text{Total Time (seconds)}} \quad (25)$$

This equation (25) calculates the throughput by dividing the number of processed packets by the total time (in seconds) taken for processing.

Packet loss: Packet loss is a measure of how many data packets were lost during the transmission between computers, cloud servers, or over the network. Packet loss impacts data reliability and integrity, which are essential in healthcare IoT where losing a few readings (such as oxygen saturation levels) may be critical.

$$\text{Packet Loss (\%)} = \left(\frac{\text{Number of Lost Packets}}{\text{Total Sent Packets}}\right) \times 100 \quad (26)$$

This equation (26) calculates packet loss as the percentage of lost packets divided by the total sent packets, multiplied by 100.

Anomaly Detection Rate: Anomaly Detection Rate (ADR) quantifies how well the model detects anomalies (attacks or unwanted activities) versus the number of total anomalies. A high ADR guarantees that most attacks are detected, reducing the likelihood of an undetected intrusion.

$$\text{Anomaly Detection Rate (\%)} = \left(\frac{\text{True Positives (TP)}}{\text{True Positives (TP)} + \text{False Negatives (FN)}}\right) \times 100 \quad (27)$$

This equation (27) calculates the Anomaly Detection Rate as the percentage of true positives (TP) out of the sum of true positives (TP) and false negatives (FN), multiplied by 100.

Performance Evaluation of vanilla transformer: Table 1 shows the performance metrics of the Transformer-based model implemented for IoT healthcare security. The model had an impressive accuracy of 98.35%, precision and recall of 97.89% and 98.12%, respectively, showing robust detection of cyber-attacks with low false alarms. The F1-Score of 98.00% demonstrates excellent overall performance, while the anomaly detection rate of 97.67% shows its robustness in detecting abnormal behaviour. These findings illustrate the model's stability for real-time threat tracking in cloud-integrated healthcare systems. In this table 1,

Table 1. Model Performance Metrics for IoT Healthcare Security Framework

Metric	Value (%)
Accuracy	98.35
Precision	97.89
Recall	98.12
F1-Score	98.00
Anomaly Detection Rate	97.67

Table 2 is a comparison between IoT device performance under regular and cyberattack scenarios. Under attacks, throughput decreases from 2400 to 1800 packets/sec, packet loss increases dramatically from 0.5% to 8.3%, and CPU and memory consumption increase dramatically. The sensor delay also increases dramatically from 100 ms to 250 ms, showing significant communication and resource degradation during malicious behaviour. These changes demonstrate the significant effect of cyberattacks on IoT healthcare system stability and responsiveness.

Table 2. IoT device communication and performance metrics under normal and attack conditions

Metric	Normal Condition	Under Attack
Throughput (packets/sec)	2400	1800
Packet Loss (%)	0.5	8.3
Device CPU Usage (%)	15	45
Device Memory Usage (%)	30	55
Average Sensor Delay (ms)	100	250

Table 3 defines important cloud-side performance measures that are relevant to threat detection and system effectiveness. Detection latency is 500 ms, and the cloud processing time is 320 ms, indicating the rapid analysis of IoT traffic data. The inference time for the model is significantly low at 75 ms, allowing prompt anomaly prediction, and alerting is done in 180 ms for timely intervention. Storage prices are maintained low at $0.023 per GB, focusing on affordability. These findings authenticate the system for providing quick, scalable, and cost-effective cloud-based security monitoring for IoT health environments.

Table 3. Cloud-side security and processing metrics for IoT healthcare framework

Metric	Value
Detection Latency (ms)	500
Cloud Processing Time (ms)	320
Alert Generation Delay (ms)	180
Storage Cost (per GB)	$0.023
Model Inference Time (ms)	75

Precision-Recall curves: This Figure 2 depicts Precision-Recall (PR) Curves for a multiclass classification model on IoT healthcare security data. Each of the curves corresponds to a distinct attack class (Class 0 to Class 3), and the Average Precision (AP) value for each class is shown in the legend.

- Precision indicates how many of the predicted positive cases were indeed correct.
- Recall indicates how many of the actual positive cases were correctly predicted.

In strongly imbalanced datasets (such as anomaly detection in IoT healthcare), PR curves are usually more informative than ROC curves since they emphasize more the model's capacity to identify rare but important events (cyberattacks).

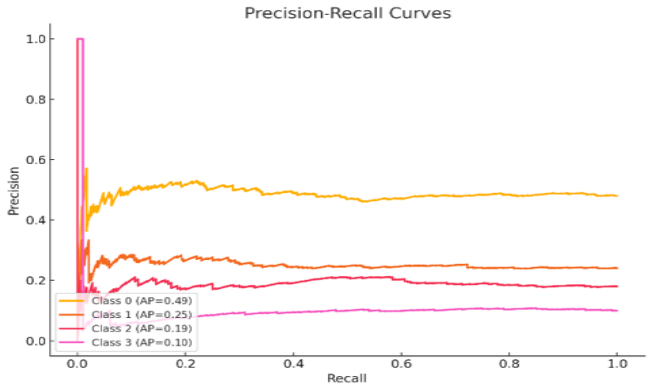

Figure 2. Precision-recall curves for multiclass IoT attack detection

Where Class 0 (presumably the most common attack class) indicates the maximum AP of 0.49, which implies detection more efficiently than other classes. Other classes (Class 1, Class 2, Class 3) indicate lower AP values (0.25, 0.19, and 0.10), suggesting additional model fine-tuning for smaller or sneaky attack types.

Confusion matrix: This confusion matrix illustrates the performance of the trained Transformer model to classify benign and malicious behaviors (DoS, Ransomware, Spoofing) in an IoT health setting. Diagonal cells (top-left to bottom-right) indicate correct classification. Off-diagonal cells indicate misclassifications (errors).The model classified 470 benign samples, 230 DoS attacks, 170 ransomware attacks, and 100 spoofing attacks correctly. There were minor misclassifications like10 benign samples incorrectly predicted as DoS,10 DoS samples predicted as ransomware, 5 instances of ransomware classified as benign and 5 as spoofing. The Figure 3 shows,

Proportional Contribution of Latency: The pie chart below (Figure 4) shows the contribution of various latency components to the overall processing time in the cloud-integrated IoT healthcare security system. Detection Latency is the most prominent component in the pipeline, contributing 46.5% of the overall delay. Cloud Processing Time contributes 29.8%, indicating the time taken for data analysis in the cloud environment. Alert Generation Delay constitutes 16.7%, indicating the time taken to alert about identified threats. Model Inference Time is the smallest part at 7.0%, reflecting that the Transformer model itself is quite efficient. Such observations aid in pinpointing bottlenecks particularly pointing out the necessity to optimize detection mechanisms to increase the system's responsiveness in real time, critical to life-critical healthcare environments.

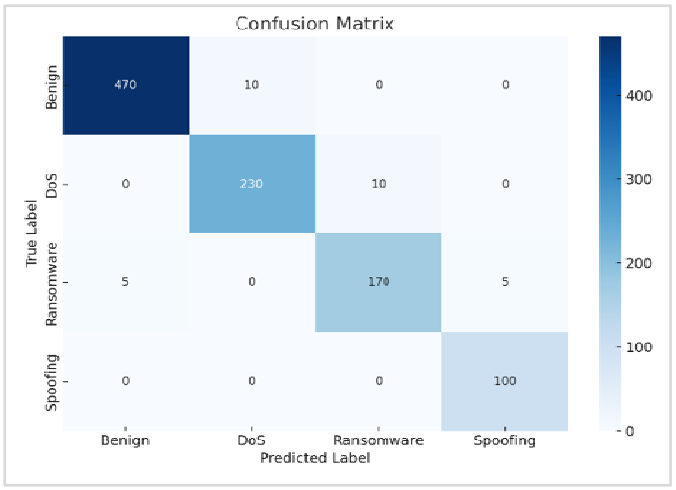

Figure 3. Confusion matrix for transformer-based IoT attack detection

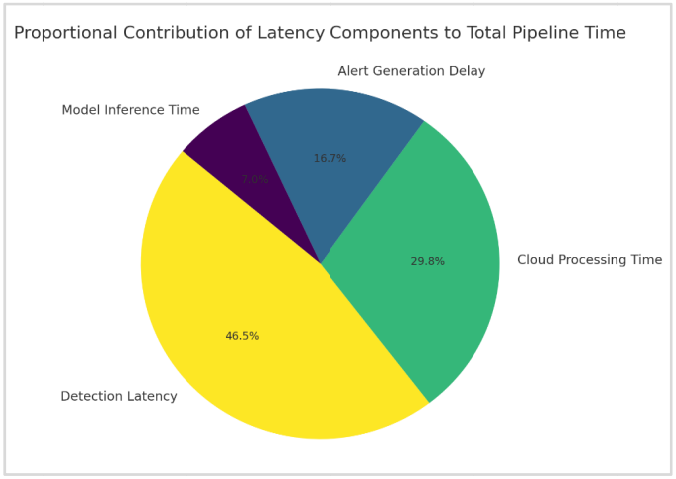

Figure 4. Proportional contribution of latency components to total pipeline time

Cloud-side security and efficiency metrics: This is a horizontal bar chart that makes a comparison between different cloud-side performance measures of the IoT healthcare security framework. Detection Latency (500 ms) is the highest because it takes most time to detect potential threats. Cloud Processing Time (320 ms) is next, illustrating the computational workload required to analyse incoming IoT traffic. Alert Generation Delay (180 ms) illustrates the speed of the system in generating alarms after threat detection. Model Inference Time (75 ms) underscores the speed of the Transformer model to make instantaneous predictions. Storage Cost ($0.023/GB) is relatively low, exhibiting cost-effectiveness in cloud storage. The below Figure 5 visualization enables us to focus optimizations particularly trying to lower detection and processing latencies to maximize overall system responsiveness.

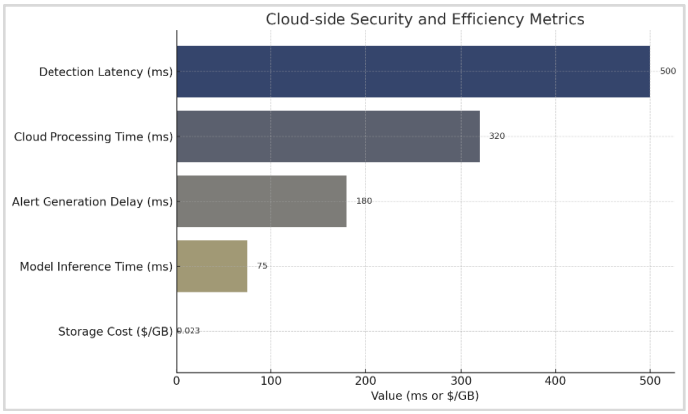

Figure 5. Cloud-side security and efficiency metrics

CONCLUSION AND FUTURE WORKS

The suggested Transformer-based AI security system provides excellent performance with an Accuracy of 98.35%, a high Precision of 97.89%, and a Recall of 98.12%. This balanced and strong detection power guarantees both correct identification of cyber threats and low false alarms, which makes it extremely effective for securing IoT-enabled healthcare setups. The findings confirm the framework's capability to provide robust, real-time protection with the operational stability of critical medical systems.

Future Works

- **Lightweight Model Deployment:** Further compress and optimize AI models to provide real-time threat detection on low-resource IoT devices (edge computing).
- **Adversarial Attack Resistance:** Create mechanisms for defence against adversarial machine learning attacks to fool AI models into misclassifying malicious activities.
- **Explainable AI (XAI) Integration:** Include interpretability techniques to offer clear explanations of every security decision in order to develop trust among healthcare operators.
- **Real-world Clinical Validation:** Test and validate the framework in live hospital or clinical environments with real IoT device networks to assess performance under real-world conditions.

REFERENCES

1. Alan, J., & Liam, M. (2020). Protecting Healthcare Data: AI-Powered Strategies for Securing Distributed Systems. *International Journal of Computational Intelligence in Digital Systems,* 9(01), 20-33.
2. Das, J. (2020). Leveraging Cloud Computing for Medical AI: Scalable Infrastructure and Data Security for Advanced Healthcare Solutions. *International Journal of Research and Analytical Reviews,* 7, 504-514.
3. Adil, M., Khan, M. K., Farouk, A., Jan, M. A., Anwar, A., & Jin, Z. (2022). AI-driven EEC for healthcare IoT: Security challenges and future research directions. *IEEE Consumer Electronics Magazine,* 13(1), 39-47.
4. Firouzi, F., Farahani, B., Barzegari, M., & Daneshmand, M. (2020). AI-driven data monetization: The other face of data in IoT-based smart and connected health. *IEEE Internet of Things Journal,* 9(8), 5581-5599.
5. Tanikonda, A., Pandey, B. K., Peddinti, S. R., & Katragadda, S. R. (2022). Advanced AI-Driven Cybersecurity Solutions for Proactive Threat Detection and Response in Complex Ecosystems. *Journal of Science & Technology*, 3(1).
6. Talla, R. R., Manikyala, A., Nizamuddin, M., Kommineni, H. P., Kothapalli, S., &Kamisetty, A. (2021). Intelligent Threat Identification System: Implementing Multi-Layer Security Networks in Cloud Environments. NEXG AI Review of America, 2(1), 17-31.
7. Patell, J. (2020). Prospects of Cloud-Driven Deep Learning-Leading the Way for Safe and Secure AI. *International Research Journal of Engineering & Applied Sciences*, 8(3), 10-55083.
8. Firouzi, F., Farahani, B., & Marinšek, A. (2022). The convergence and interplay of edge, fog, and cloud in the AI-driven Internet of Things (IoT). *Information Systems,* 107, 101840.
9. Oladosu, S. A., Ige, A. B., Ike, C. C., Adepoju, P. A., Amoo, O. O., & Afolabi, A. I. (2022). Revolutionizing data center security: Conceptualizing a unified security framework for hybrid and multi-cloud data centers. *Open Access Research Journal of Science and Technology*, 5(2), 086-076.
10. Avuthu, Y. R. (2021). Trustworthy AI in Cloud MLOps: Ensuring Explainability, Fairness, and Security in AI-Driven Applications. *Journal of Scientific and Engineering Research*, 8(1), 246-255.
11. Firouzi, F., Jiang, S., Chakrabarty, K., Farahani, B., Daneshmand, M., Song, J., &Mankodiya, K. (2022). Fusion of IoT, AI, edge–fog–cloud, and blockchain: Challenges, solutions, and a case study in healthcare and medicine. *IEEE Internet of Things Journal,* 10(5), 3686-3705.
12. Mensah, F. (2022). AI in healthcare cybersecurity: Balancing the risks and benefits of intelligent defense mechanisms. *International Journal of Nursing Research and Practice (IJNRP)*, 7, 16-28.
13. Min-Jun, L., & Ji-Eun, P. (2020). Cybersecurity in the Cloud Era: Addressing Ransomware Threats with AI and Advanced Security Protocols. *International Journal of Trend in Scientific Research and Development*, 4(6), 1927-1945.
14. Chianumba, E. C., Ikhalea, N., Mustapha, A. Y., Forkuo, A. Y., & Osamika, D. (2022). Integrating AI, blockchain, and big data to strengthen healthcare data security, privacy, and patient outcomes. *Journal of Frontiers in Multidisciplinary Research,* 3(1), 124-129.
15. Butpheng, C., Yeh, K. H., & Xiong, H. (2020). Security and privacy in IoT-cloud-based e-health systems—A comprehensive review. Symmetry, 12(7), 1191.
16. Kumar, S., Raut, R. D., & Narkhede, B. E. (2020). A proposed collaborative framework by using artificial intelligence-internet of things (AI-IoT) in COVID-19 pandemic situation for healthcare workers. *International Journal of Healthcare Management,* 13(4), 337-345.
17. Ponnusamy, V., Vasuki, A., Clement, J. C., & Eswaran, P. (2022). AI-Driven Information and Communication Technologies, Services, and Applications for Next-Generation Healthcare System. *Smart Systems for Industrial Applications*, 1-32.
18. Mosaddeque, A., Rowshon, M., Ahmed, T., Twaha, U., & Babu, B. (2022). The Role of AI and Machine Learning in Fortifying Cybersecurity Systems in the US Healthcare Industry. *Inverge Journal of Social Sciences*, 1(2), 70-81.
19. Chawla, G., & Rizvi, S. W. A. (2022, November). Healthcare Data Security in Cloud Environment. In International Conference on Intelligent Vision and Computing (pp. 245-253). Cham: Springer Nature Switzerland.
20. Pemmasani, P. K., & Henry, D. (2021). Zero Trust Security for Healthcare Networks: A New Standard for Patient Data Protection. The Computertech, 21-27.
21. Carl, A., & Billy, L. (2022). Healthcare Meets AI: Enhancing Big Data Security with Graph-Based Learning. *International journal of Computational Intelligence in Digital Systems,* 11(01), 53-76.
22. Zulkifl, Z., Khan, F., Tahir, S., Afzal, M., Iqbal, W., Rehman, A., & Almuhaideb, A. M. (2022). FBASHI: Fuzzy and blockchain-based adaptive security for healthcare IoTs. IEEE Access, 10, 15644-15656.

****