

AI-Driven Intrusion Detection and Threat Mitigation in Zero Trust Architectures

Ehigiator Egho-Promise¹, Udoh Ekereuke², Edita Gashi³, George Asante⁴

^{1,2,3}Department of Computing, QA Higher Education, Manchester, United Kingdom

⁴Department of Information Technology Education, Akonten Appiah-Menka University of Skills Training and Entrepreneurial Development, Kumasi, Ghana.

Article Info

Article history:

Received Dec 31, 2025

Revised Mar 12, 2026

Accepted Mar 25, 2026

Keywords:

Artificial Intelligence, Zero Trust Architecture, Intrusion Detection System, Deep Learning, Cybersecurity, Transformer Model.

ABSTRACT

This study explores the role of blending Artificial Intelligence (AI) with Zero Trust Architecture (ZTA) to boost Cybersecurity through a solid Intrusion Detection System (IDS). Using a design science and postpositivist approach, the research crafts and assesses AI-driven IDS models such as Decision Tree, Random Forest, Long Short-Term Memory (LSTM), and Transformer by leveraging the University of New South Wales - Network Benchmark 2015 (UNSW-NB15) dataset within a simulated ZTA environment. The study evaluates how well these models perform based on metrics like accuracy, precision, recall, and F1-score to pinpoint the best architecture for spotting both insider and external threats. The results show that the Transformer-based IDS outshines the others with an impressive 99.2% accuracy rate, proving its exceptional ability in real-time anomaly detection and threat classification. These findings highlight that merging deep learning with ZTA's identity-based access control significantly boosts network resilience and helps reduce lateral movement threats. This research presents an innovative AI-ZTA integrated model designed for proactive, adaptive, and scalable cybersecurity defense. It also offers valuable insights for organizations aiming to strengthen digital trust and policy frameworks in the face of increasingly sophisticated cyber threats.

Copyright © 2026 Institute of Advanced Engineering and Science.

All rights reserved.

Corresponding Author:

Ehigiator Egho-Promise

Department of Computing, QA Higher Education,
Manchester, United Kingdom

Email: ehigiator.eghopromise@qa.com

1. INTRODUCTION

The digital world in the 21st century is characterised by interconnectivity and the easy use of the cloud and mobile workforce, which has essentially compromised the security perimeter of the traditional world [1]. The earlier distinction between a trusted and an untrusted network is no longer valid because the internal network is untrusted. Remote work and cloud-centric infrastructure are among the factors that lead to the increase in attack surface [2]. This fact renders perimeter security models, which can be defined as castle-and-moat-based perimeter defences, useless in the context of current vulnerabilities to cyber-attacks. The cybersecurity community has been a proponent of a paradigm shift to Zero Trust Architectures (ZTA). ZTA is based on the principle that one should always question and never trust. This principle means that one should not ever trust any of the parties' user, device, or application [1]. The ZTA system is only accessible on a per-session basis by users, though founded on dynamic and granular policies, which consist of placing strict identity verification, least-privileged access, and micro segmentation [3]. In this model, the concept of security turns from its network-focused and non-active approach to identity and a dynamic-oriented one.

Artificial Intelligence (AI) is turning out to be an innovation in terms of Cybersecurity. The quantity and level of network traffic have grown to surpass the analytical capacity of human beings. Machine Learning (ML) and Deep Learning (DL) can be used to scan a lot of data in real-time and demonstrate little but significant anomalies and patterns, which may point to a cyberattack [4]. It is a crucial identification of new threats which have not been identified by the traditional signature systems. ZTA and AI have a massive opportunity. The creation of a contextualised and enriched stream of data is done through ZTA, and the analytical capability is

provided by AI [5]. This integration is the basis of the next generation of intrusion detection as well as threat mitigation, which makes the cybersecurity position more adaptable and resilient.

This synergy is largely theoretical, with little architectural edifice of operationalizing the synergy. Even though the Zero-Trust Architecture (ZTA) provides the contextual data required, and artificial intelligence is the engine undertaking the analysis, the important question remains how one can make the two integrate to create a reasonable and adaptable security platform. The fact that AI can dynamically update and optimize ZTA policies (not necessarily operating with an implicit set of rules, but literally making risk-based, real-time decisions on access) reflects the future stage of development of Cybersecurity. The application of this problem and its integration in the present research is that AI-based analytics could also be used as a form of ongoing feedback in the ZTA control plane, and as such, this architecture would not be seen as a proactive and intelligent defense ecosystem but would be able to adjust to emerging circumstances.

Even though ZTA is a conceptually sound tool, its usefulness when used in detecting intrusions and mitigating threats is highly limited. The traditional Intrusion Detection and Prevention System (IDS/IPS) fails to be synergized with ZTA, which is a dynamic system by nature. Also, Zero-day exploits cannot be identified using signature-based systems [6], and the conventional anomaly-based systems are characterized by a high false-positive rate [7], operating without understanding the context, which is one of the main components of Zero Trust. Also, more sophisticated attack vectors, such as lateral movement and stolen credentials, can bypass the methods of creating a static defence by masquerading as legitimate traffic [3]. This creates a lethal loophole: even though ZTA is a powerful policy compliance engine, its ability to react intelligently and independently to more advanced and subtle attacks in real-time is a significant issue.

Furthermore, ZTA, by its very distributed and granular nature, poses a specific challenge itself in the form of data analysis aspect. The micro-segmentation and continuous authentication processes create headless volumes of high-dimensional log data of multiple sources, such as identity providers, device health platforms, and network sensors, among others. The process of associating such heterogeneous data streams that are contextual to form a consistent security image is not a trivial task. Existing security information and event management (SIEMs) may not be able to cope with an onslaught of data, which will cause analyst fatigue. Hence, the question does not concern the existence of anomalies observed in network traffic, but the clever use of multimodal information to perform cost-effective, free, and real-time risk assessment without a logistical burden of operation. It needs an intelligent system, capable of differentiating a severe threat and an innocent policy anomaly or a harmless change in user behaviour, which is the very issue with conventional intrusion detection and prevention systems (IDS/IPS), and naturally, they are poorly equipped to respond to.

This study has the central objective of developing and testing a comprehensive model that can benefit intrusion detection and automated threat reduction in Zero Trust Architectures with the help of AI [8].

Specifically, the study seeks to: revise current models of ZTA and AI-driven security to determine integration prospects; develop an AI-driven intrusion detection platform to suit ZTA conditions, and deploy and test a proof-of-concept prototype based on IDS methods.

This study has several contributions, such as the proposed study introduces an AI-based intrusion detection system with specific features to Zero Trust Architecture (ZTA) with the integration of an AI-based detection system and Policy Decision Point (PDP) and Policy Enforcement Point (PEP) as real-time mitigation mechanisms. A hybrid modelling strategy involving classics of machine learning (Random Forest, SVM) and deep learning algorithms (CNN, LSTM, Transformer) is created and compared to each other. The research uses the context-conscious intrusion detection that makes use of flow-based properties that align with the ZTA principles, which are identity-awareness, micro-segmentation, and constant checking. The suggested system is more promising in terms of detection accuracy, lower false positive rates, and less latency, which is appropriate for implementation in the real-time framework of contemporary cybersecurity settings.

2. LITERATURE REVIEW

2.1. Theoretical Foundations

Network security has historically been an arms race between protectors and attackers, with the mechanisms of defence being mostly stagnant, based on perimeter security schemes. Conventional network security was based on a castle-and-moat model where firewalls and Intrusion Detection/Prevention Systems (IDS/IPS) served as access controls for the network perimeter and implicitly trusted traffic flowing within the network [8]. An IDS is used to monitor network traffic or system activity in the event of malicious activity or policy violation. In contrast, an IPS takes this ability further by blocking or preventing identified threats [9]. These systems are typically executed in two main modes: signature-based allocation, where known hazards are detected by matching patterns in a database of attack signatures, and anomaly-based allocation, where a normal hierarchy is established in advance, with the unusual set of rules marked. However, the growing use of cloud services, mobile devices, and remote work has made the network perimeter less effective, at least, and this is weakening the functionality of traditional tools [1]. This paradigm shift in the information technology

environment was the driver that led to the emergence of the new security paradigm, resulting in the creation of ZTA. At its most basic, ZTA is defined by the formal standard of the National Institute of Standards and Technology (NIST) in SP 800-207 as "always verify and never trust" [2]. According to this philosophy, the basis of access decisions is strict identity verification and least-privilege principles applied to all access requests, regardless of their source, whether internal to the corporate network or external. Notable logical components according to the NIST framework include a Policy Decision Point (PDP) at which point access is permitted and a Policy Enforcement Point (PEP) where it is enforced. With these elements, there is a cyclic system of asset inventory, threat intelligence, and data access policies.

2.2. *AI in Cybersecurity*

Nowadays, Artificial Intelligence (AI), or rather Machine Learning (ML) and Deep Learning (DL), are one of the foundations of modern Cybersecurity since they offer an opportunity to process large amounts of data and identify more complex patterns that a rule-based system could not easily detect. The ML algorithms find application in intrusion detection, such that the fingerprinting of deviations of normal network activity is modelled using prior estimated data. The Support Vector Machines (SVM) and Decision Trees are the supervised learning models, which have been trained on labelled sets of benign and malicious traffic. New observations can therefore be classified using those models [10]. As an example, the K-Nearest Neighbours (KNN) algorithm can be applied in detecting outliers based on the similarity between a given piece of data and its neighbours in a feature space.

However, its demand for a large quantity of data that is well labelled is a major drawback. It is done by unsupervised learning, e.g., clustering (K-means) and autoencoders, which are useful to detect natural features or anomalies in unlabelled data. The techniques are especially practical in identifying new or zero-day attacks where there are no signatures [11]. Deep learning is a variation of ML, and it has advanced the field of study further by learning high-level features automatically with raw data. Convolutional Neural Networks (CNNs) for image processing can be trained to process network traffic and classify packet headers or flow sequences as one-dimensional images and thus find spatial patterns [12]. RNNs and more recent versions (such as Long Short-Term Memory (LSTM) networks) prove particularly helpful with the analysis of temporal sequences. They are thus best suited for detecting complex, multi-stage attacks that occur over time, by training on the long-range time correlation of network flow data [13]. The ability of such AI models to analyze and learn from the vast number of logs created in the digital space makes them an intuitive solution for improving security analytics and automated responses.

2.3. *ZTA Challenges and Intrusion Detection Requirements*

Although the concepts of ZTA provide a solid framework for access control, effective execution introduces a set of peculiarities that redefine the needs of intrusion detection. The complexity of ZTA is generated by its granular and distributed nature. Any complete ZTA must offer end-to-end security on several pillars, such as identity (authenticating users), devices (verifying their health and compliance), networks (enforcing micro-segmentation), and applications/workloads (enforcing the security of data processing) [2]. The IDS must be very distributed in addition to being context-sensitive and capable of correlating alerts across such a broad range of pillars. To provide an example, the risk score must be increased since a legitimate user (identity) on the device that has been compromised (device) is viewed as a greater risk. Besides this, the dynamic nature of the threat environment caused by the Advanced Persistent Threats (APTs) and insider threats requires the IDS to be capable of keeping up with the new era. Micro-segmentation formally prevents lateral movement in a ZTA, and attackers will use less obvious tactics to move laterally by either stealing credentials or using lawful connections with applications [3]. Consequently, the intrusion detection system ought to extend beyond scrutiny of the contents of network packets to the detection of abnormal behaviour of user behaviours, device surveillance, and application API calls. The need to verify everything in real time all the time implies that the IDS must have a minimal latency to ensure that the experience of a user is not interfered with, and that is why a real-time analysis is an important consideration. Thus, a functional IDS to ZTA should be contextual, custom, adaptive, and capable of performing advanced behaviour analysis of the entire digital estate, which represents a significant improvement over traditional network-based IDS solutions.

2.4. *State-of-the-Art Approaches*

The existing studies of AI-based intrusion detection systems (IDSs) reveal an uneven spectrum of models and deployment models, but the relationship between these models and the concepts of ZTA is still in its infancy. Numerous studies have compared classical algorithms and deep-learning algorithms with publicly available datasets, including Network Security Laboratory - Knowledge Discovery and Data Mining (NSL-KDD), Canadian Institute for Cybersecurity Intrusion Detection Dataset 2017 (CICIDS2017), and University of New South Wales - Network Benchmark 2015 (UNSW-NB15). According to authors in [14], support vector

machines and random forest classifiers are highly accurate on such corpora; however, their performance is also sensitive to feature engineering and can be destabilized in the face of more complex temporal attack patterns. In recent years, deep-learning models have been placed in the spotlight due to their end-to-end learning capabilities. An example can be given of convolutional neural networks (CNNs), which are very effective in classifying network traffic flows, and long short-term memory (LSTM) networks, which have proven to be quite effective in detecting multi-stage attacks and command and control (C2) beaconing based on time-series data of network packets [13]. The use of unsupervised methods, especially variational autoencoders (VAEs) and generative adversarial networks (GANs), is under study to learn normal network behaviour while detecting anomalies as potential threats, which is necessary for detecting new attacks [15]. Concerning deployment architecture, one of the critical debates is a comparison between cloud-based and edge-based AI. Cloud-based IDSs concentrate the analysis, take advantage of the vast computational power, and the possibility of developing global threat models; however, they can introduce latency and privacy issues, as all data must go through the cloud [16-19]. On the other hand, edge-based AI operates on information at the local or proximal level of the device or network segment, which is similar to the distributed policy enforcement points (PEPs) found in ZTA. This method is less latent, bandwidth-saving, and more private. Still, it has limited computational power and might provide a less comprehensive view of the threat overall, which can include multi-edge attacks [5]. New hybrid models are expected to leverage the benefits of both models by performing first-level filtering and detection at the edge, and only sending high-fidelity alerts or metadata to a central cloud service for correlation and further analysis.

2.5. *Limitations of Identity-Aware Proxies in ZTA*

Identity-Aware Proxies (IAPs) are one of the core elements of Zero Trust Architecture because they are the means of applying access control verdicts based on authenticated user identity and contextual attributes. They also have a number of limitations, however, which limit their effectiveness. Firstly, IAPs are concerned mainly with authentication and authorization mechanisms and do not offer visibility into the network traffic behaviour and flow-level anomalies [20, 21]. Subsequently, more advanced attacks like insider attacks or lateral motions through legitimate credentials might go unnoticed. Moreover, IAP systems are dependent on how well identity verification processes are done; in the event of the breach of credentials, bad actors can bypass access checks without arousing any suspicion. The absence of continuous behavioural analysis is another limitation because IAPs are not developed to track the changing traffic patterns or detect zero-day attacks immediately. These restrictions point to the necessity of introducing AI-based intrusion detection mechanisms to ZTA, as suggested in the present study as a complement to identity-based controls with behavioural intelligence and anomaly detection [22].

2.6. *Research Gap Analysis*

A critical analysis of the available literature reveals that there are substantive gaps that this study aims to fill. To begin with, there is an overt lack of integrated AI systems specifically designed to fit the specifics of the ZTAs. Although there is a wealth of literature on AI-based IDS and a distinct set of corpora on ZTA principles, these areas are rarely combined holistically. Most suggested AI-based IDS designs are network-focused, as they focus on flow data or packet headers and do not effectively utilise the rich and multi-dimensional data streams of ZTA, including user identity context, device posture measurement, and application logs [7]. This is a lost opportunity to develop a truly contextual and effective detection system. Second, the empirical assessment of proposed models has a limitation. Many researchers use outdated or non-representative data sets, such as KDDCup99, which do not accurately reflect modern network settings and attack methods [17]. Experiments that are not simulated in a ZTA environment, even when using newer datasets, would still employ micro-segmentation and continuous authentication, which would alter network traffic and attack patterns. The trade-offs of performance are often overlooked, especially in concerns related to latency that complicated AI models impose in a real-time enforcement setting. Moreover, the issue of explainability remains relevant; a deep-learning model can be highly accurate, but its non-insightful nature can hinder its implementation in high-stakes security scenarios, where it is essential to understand the rationale behind an alert to respond to it [18]. In general, the absence of an evaluatively evaluated, context-sensitive AI structure that is inherently structured and rigorously tested in the context of operational restrictions of a Zero-Trust Architecture, with the use of modern datasets to prove its effectiveness in reducing the occurrence of modern threats, including lateral movement and primary key-based incursions, is the main gap in the research.

3. RESEARCH METHOD

3.1. *Research Philosophy and Approach*

This study adopts a postpositivist philosophy, which accepts that network behaviour follows regularities while recognising that data and models remain imperfect. Guided by this stance, the design is

deductive and begins with principles from Zero Trust Architecture and known intrusion tactics that lead to testable expectations about detection effectiveness and operational impact. The methodology blends design science, where an intrusion detection artefact is built, with quantitative evaluation that measures performance in realistic settings. The approach proceeds through four iterative cycles. First, authors framed the problem and defined requirements that limit lateral movement, reduce false alarms, and keep decision latency within a practical budget. Second, the authors designed the artefact by specifying the feature pipeline and a portfolio of models, including Random Forest, Support Vector Machine, Convolutional Neural Network, Long Short Term Memory, and Transformer, all exercised within a micro-segmented laboratory implementation of Zero Trust. Third, the authors conducted experiments on CICIDS-style traffic and locally generated flows using stratified cross-validation. Fourth, authors reflected on the results and refined configurations to balance detection power with operational cost. Evaluation covers accuracy, precision, recall, F1, specificity, and the area under the receiver operating characteristic curve, alongside end-to-end latency and false positive rate. Insights from the evaluation update policy decision points whose actions are enforced by policy enforcement points in near real time.

3.2. Research Design

This study employs a mixed-methods research design, combining quantitative evaluation of AI-based IDSs with qualitative analysis of deployment in ZTA environments. The method entails the creation and testing of AI models in a controlled virtual laboratory and entails a design-test-refine cycle. The cycle includes data capture, preprocessing, data engineering, model creation, testing, and modelling of the implementation of ZTA, and making sure that the improvements are methodically tested against the purpose of operation and security.

3.3. AI Model Selection

The selected models are: Random Forest: Ensemble of decision trees, robust to overfitting.

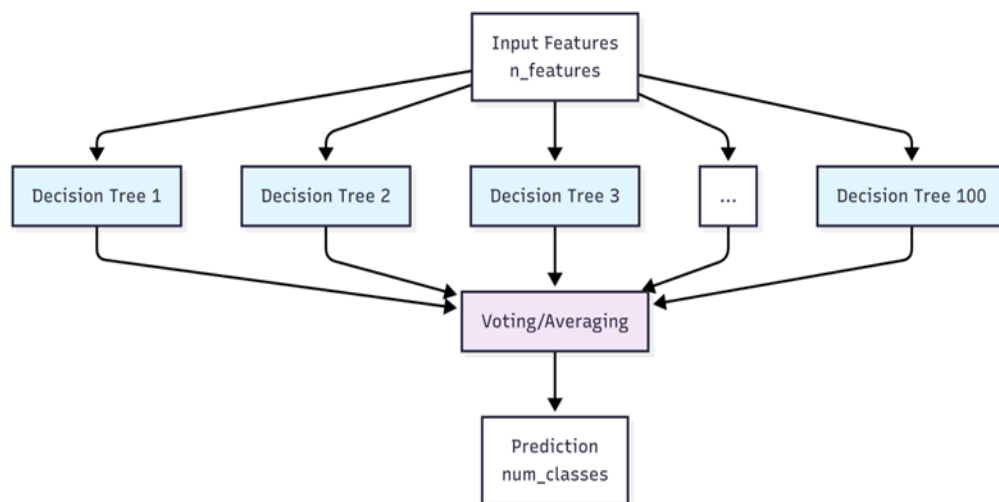


Figure 1. Random Forest illustration

SVM: Hyperplane-based classifier with RBF kernel for non-linear data



Figure 2. SVM cycle

CNN: Extracts spatial features from reshaped traffic data



Figure 3. CNN Cycle

LSTM: Models temporal dependencies in sequences

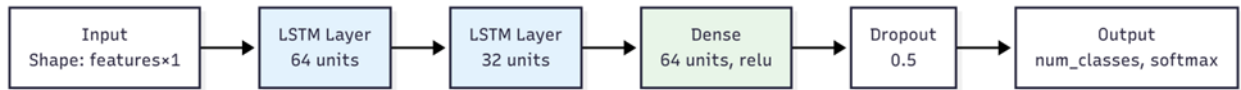


Figure 4. LSTM Phases

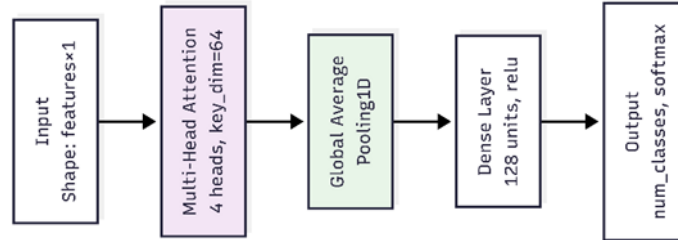


Figure 5. Transformer model Phases

Transformer: Uses attention mechanisms for complex relationships.

Justification:

- RF achieves high accuracy on CICIDS2017, ideal for tabular data
- SVM excels in high-dimensional spaces
- CNN captures spatial patterns, enhancing hybrid IDS performance
- LSTM handles sequential attack patterns
- Transformer's attention mechanism improves multi-class detection

3.4. Data Collection

The study used the CICIDS2017 dataset, which is a set of network traffic characteristics of benign and malicious activities, including DDoS, PortScan, and infiltration attacks. Data is saved and includes flow-level data, including the number of packets, time, bytes, and protocols/service types.

CICIDS2017 is a publicly accessible intrusion detection dataset benchmark created by the Canadian Institute of Cybersecurity. It consists of some 2.8 million network flow traces that have been produced during five days of realistic enterprise network operation. The flows are modeled with approximately 78 statistical features each, which comprise time-based, packet-based, and protocol-level attributes generated by CICFlowMeter. These attributes can represent key attributes like flow time, packet counts, and byte counts, inter-arrival time, and transport-layer service statistics. The dataset contains benign traffic, as well as the various categories of attacks currently being applied, including Distributed Denial-of-Service (DDoS), PortScan, brute-force, botnets, and infiltration attempts, so it will be suitable for supervised and hybrid intrusion detection studies. CICIDS2017 can be specifically used in this study since it generates realistic traffic behaviour in current enterprise settings, has the ability to classify in multiple classes, and offers flow-level details that can be utilized under encrypted traffic situations, an essential feature in DZero Trust Architecture (ZTA), where payload inspection is restricted.

The choice of data sets used in this research is based on the aspects of relevance as well as comparability. Although more recent datasets like CSE-CIC-IDS2018 provide more recent attacks, UNSW-NB15 remains important because it provides structured representations of its features and an equal representation of contemporary attack types. The CICIDS2017 is the main experimental test in this study as it offers the realistic enterprise traffic pattern and a wide variety of modern attack types. The choice of UNSW-NB15 is mainly explained by the fact that this tool is used in numerous previous works, and it can be appropriately compared with the current research. Besides, both CICIDS2017 and more recent datasets like CICIDS2018 are created through related methodologies, and in particular, flow-based feature extraction means that the knowledge learned based on CICIDS2017 can be generalized to the contemporary network setting. Thus, the dataset selection yields no harm to the validity but empowers the consistency of the methodology and benchmarking.

Moreover, on the one hand, this study predominantly uses the CICIDS2017 data, and its choice is explained by its topicality and completeness in reflecting the contemporary network settings. The CICIDS2017 encompasses real-world traffic patterns, the most recent attack scenarios, flow-based features that are critical in the Zero Trust settings, and encrypted traffic analysis. In addition, popular benchmark datasets like NSL-KDD and UNSW-NB15 were taken into account during the phase of the literature review in order to make decisions on model selection and feature engineering. Such datasets offer variety in patterns of attacks and

benchmarks of evaluation that were previously employed in previous research. The combination of CICIDS2017 and the experience of other benchmark datasets will guarantee that the proposed model is created considering the possibility of generalization and can be applied to real-life situations of intrusion detection.

3.5. Dataset Preprocessing

Proper preprocessing is important for accurate anomaly detection. The steps include:

- **Feature Extraction:** Selected key features from network flows that are most indicative of malicious activity, including:
 1. Flow Duration
 2. Total Forward Packets
 3. Forward Packet Length Maximum
 4. Flow Inter-Arrival Time Mean
 5. Forward Inter-Arrival Time Mean
- **Normalisation:** Applied Min-Max scaling to map all features to the range [0,1], reducing bias due to differing feature scales.
- **Imbalance Handling:** Addressed class imbalance using SMOTE, ensuring the model can effectively learn minority class patterns.
- **Data Cleaning:** Removed any NaN or infinite values to prevent model training errors and ensure data integrity

CICFlowMeter was used to extract features of this study, which are the flow-level statistical features extracted directly based on the raw packet captures. The original set of features is thus based on domain-specialized network telemetry, which is typically utilized in intrusion detection studies, as opposed to baseless feature engineering. The choice of particular features to be installed in the modelling was based upon a mix of the domain information and on the experience of the previous literature on the topic of the IDS, where temporal flow properties and packet statistics are indicated as the good predictors of the malicious behaviour. Besides domain selection, a stage of feature selection was also introduced in order to reduce the number of dimensions and eliminate redundant or weakly informative features. This was done through correlation analysis in order to remove highly collinear features and then model-based importance ranking using the Random Forest feature importance scores. Such measures were necessary to make sure that the resulting set of features is highly predictive and lowers the computational cost, which is a paramount requirement in real-time detection. Recursive feature elimination (RFE) was also an option, but tree-based importance ranking was chosen because it is scalable and can be used with large and imbalanced data like CICIDS2017. The feature set that results thus represents a hybrid strategy, taking existing network security domain information and using data to validate it whenever necessary to trade off detection accuracy, interpretability, and operational efficiency in a Zero Trust deployment setting.

3.6. Implementation Framework

The implementation of model development was done in Python with the aid of TensorFlow/PyTorch, deep learning, and Scikit-learn, classical ML. The imbalance in classes was solved using imbalanced-learn, and the scaling of features was done using MinMaxScaler. It is deployed in a ZTA-inspired virtual lab based on VMware or VirtualBox in a micro-segmentation environment based on pfSense firewalls to emulate least-privilege access and micro-segmentation.

3.7. Model Evaluation Metrics

Various classification measures were considered when determining the model performance on the test set, and some of the measures included: accuracy, precision, recall, and the F1-score. These metrics give a perfect perspective on the performance of the model. Accuracy determines the percentage of the correctly predicted instances (true positives and true negatives) of the total instances.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{FP} + \text{FN} + \text{TP} + \text{TN}}$$

Where:

- TP = True Positives (correctly predicted positive class)
- TN = True Negatives (correctly predicted negative class)
- FP = False Positives (incorrectly predicted as positive)
- FN = False Negatives (incorrectly predicted as negative)

Precision indicates the proportion of positive predictions that are actually correct.

$$\text{Precision} = \frac{\text{TP}}{\text{FP} + \text{TP}}$$

Recall is also referred to as sensitivity or true positive rate, and it is the ratio of true positives that are recognized accurately.

$$\text{Recall} = \frac{\text{TP}}{\text{FN} + \text{TP}}$$

3.8. Ethical Considerations

All data is anonymous, and the Attack simulation is located in virtual laboratories, which are safe and compliant. Various datasets, preprocessing methods, and explainable AI methods ensure mitigation of bias, which ensures fairness, transparency, and reproducibility.

3.9. Dataset Analysis and Features Overview

The dataset analysis provides insights into feature distributions and attack patterns.

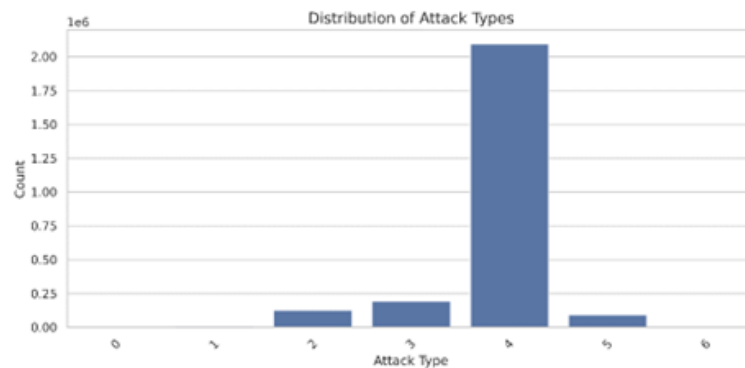


Figure 6. Attack Distribution

Attack Distribution in Figure 6 identifies imbalanced classes and informs resampling via SMOTE. The dataset has a high level of class imbalance, where there is benign traffic dominating a number of attack categories. This is enough to warrant the use of SMOTE to facilitate balanced learning and gainful minority classes recognition.

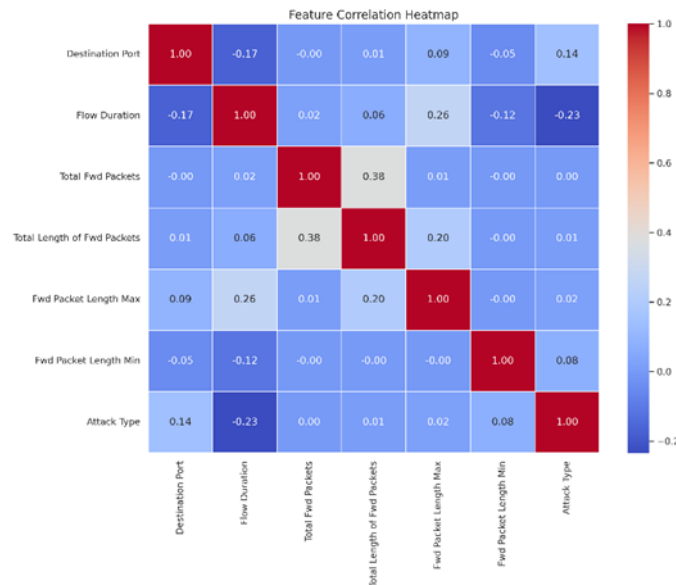


Figure 7. Correlation Heatmap

The correlation heatmap in Figure 7 highlights strongly correlated features, guiding dimensionality reduction. The high correlation between some features implies redundancy, which justifies the application of the methods of feature selection, like ranking based on the Importance of Random Forests, to decrease the number of dimensions in order to enhance computational efficiency.

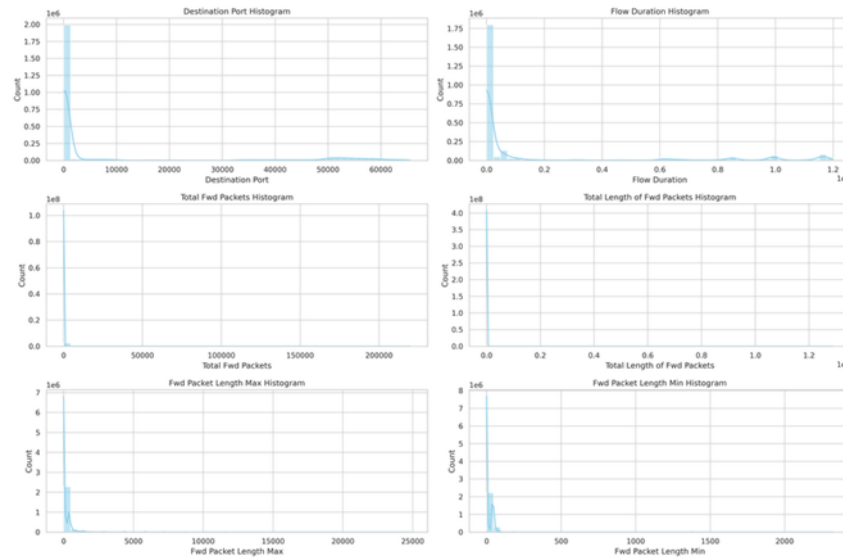


Figure 8. Boxplots and Histograms

Boxplots and Histograms in Figure 8 illustrate outliers, feature skew, and intra-class variability. The skewness and outliers indicate the inconsistency in the behaviour of the network traffic, and it is therefore necessary to normalize the network traffic and use strong models that can be used on non-linear distributions.

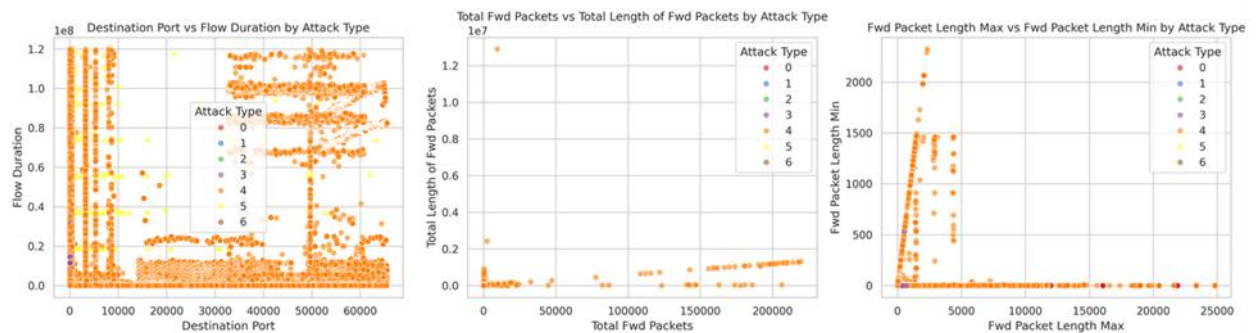


Figure 9. Pairwise Scatterplots

Pairwise Scatterplots in Figure 9 visualize separability between attack and benign clusters. There is also partial separability between attack and benign classes, which confirms the appropriateness of machine learning models and also justifies the application of deep learning models to represent non-linear decision boundaries.

3.10. System Design and Implementation

The suggested IDS architecture is designed in three main modules, including Data Ingestion, Detection Engine, and Mitigation Module. The modules are all essential for providing secure, automated, and adaptive threat detection in a ZTA environment. Gathers network traffic from various sources, such as live network traffic and previously recorded traffic, such as CICIDS2017. Raw packet data is grouped into organised flow-level features, which are ready to be preprocessed and analyzed by AI using the module. The detective engine is the heart of the intelligence of the system. Several AI and machine learning systems investigate the passing traffic. In each of these models, a score is calculated to show how probable the presence of malicious activity is. The detection engine is compatible with monitored and unmonitored methods of detecting known and unknown attack patterns.

Combines Policy Decision Points (PDP) and Policy Enforcement Points (PEP). The PDP compares the anomaly scores with the predefined or adaptive thresholds in order to make access decisions. These decisions are implemented in real time by the PEP, among other methods, by means of firewall rules (pfSense, iptables) or by automatic scripts.

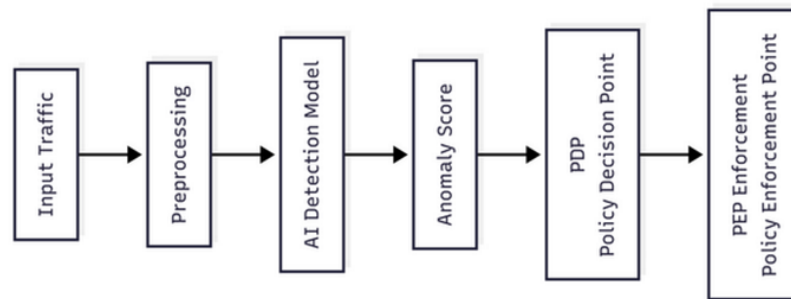


Figure 10. Architecture diagram illustrating data flow, detection engine, and mitigation module

The system has several AI models, which can be used to evaluate them comparatively and pick the most effective algorithm. Models like Random Forest, SVM, CNN, LSTM, and Transformer are trained on 80 percent of the preprocessed data. Each model is optimized in terms of hyperparameters in order to maximize the performance of detection. A cross-validation method (often k-5-fold) is used to measure the robustness of the model and to avoid overfitting. During validation, monitoring of the metrics like accuracy, precision, recall, F1 score, ROC-AUC, latency, and false positive rate is conducted. Integration with a Zero Trust framework ensures continuous validation and real-time mitigation.

Policy Decision Points (PDP) and Policy Enforcement Points (PEP):

- The **PDP** interprets anomaly scores and determines whether traffic should be allowed, blocked, or quarantined.
- The **PEP** enforces these decisions using firewalls, network segmentation, or API-driven automation.

The system is in a streaming mode that analyzes traffic streams. Mitigation measures are automated, which saves time and effort in responding.

The suggested threat mitigation plan goes beyond the normal detect-and-block paradigm that is normally attributed to the conventional IDS/IPS systems. Unlike the use of static response mechanisms, the framework also includes a situational and adaptive mitigation strategy that is in line with the principles of Zero Trust Architecture [23]. In particular, mitigation decisions are not binary ones, but they are calculated on the basis of a fusion of contextual attributes, such as AI-created anomaly scores, user identity, device posture, and micro-segmentation limits. According to this multi-dimensional assessment, the Policy Decision Point dynamically resolves the right response, which can be restriction of access privileges, session isolation, step-up authentication, or selective restriction of network interactions. This allows for the recalculation of trust on each transaction and enforces a decision once. Consequently, the suggested system makes mitigation a fine-grained, real-time control procedure, which adds to the components of the traditional IDS logic and complies with the adaptability and dynamism of the Zero Trust security models [24].

Besides, the Policy Enforcement Point in the proposed architecture is applied with the help of a pfSense-based firewall that is installed in a virtualized micro-segmentation environment [25]. This element acts as the implementation layer of implementing the security decisions created by Policy Decision Point. The detection engine is integrated with pfSense by automated rule management, whereby an anomaly score and classification results are mapped to a policy actionable by the firewall as a script-driven or API-driven translation. When suspicious activity is detected, pfSense is able to dynamically modify its rule set to apply some type of action, including traffic blocking, port restriction, or redirection to isolated network segments. The system also facilitates the VLAN-based segmentation to limit the movement laterally across network zones. Also, pfSense logs are fed back into the detection engine, producing a feedback loop that makes the engine increasingly more accurate at learning and detection as time goes on. This implementation guarantees that enforcement is not only real-time, but it is also closely integrated with Zero Trust principles of least privilege as well as continuous verification [26].

4. RESULTS AND DISCUSSION

This section entails the experimental assessment, performance analysis, and discussion of the findings of installing the suggested AI-driven IDS in a ZTA environment. Different machine learning and deep learning models were trained, tested, and compared according to the traditional measures of classification, detection latency, and false positive rates. The results show the effectiveness of adaptive AI-based detection systems as compared to the conventional signature-based intrusion detection systems.

4.1. Experimental Results

The CICIDS2017 preprocessed data was used in the performance evaluation. The experimental system comprised training and testing different models, such as Random Forest, SVM, CNN, LSTM, and transformer-based models. None of the models was overfitted since each was optimised in grid search and then subjected to 5-fold cross-validation. The findings indicate an excellent accuracy of all the models, with the deep learning models having an advantage over classical algorithms. The most important performance measures are tabulated in Table 4.1.

Table 4.1. Performance of AI Models for Intrusion Detection

Model	Accuracy	Precision	Recall	F1-Score	ROC-AUC	Latency (ms)	False Positives (%)
Random Forest	0.96	0.95	0.96	0.96	0.98	50	2.1
SVM	0.94	0.93	0.94	0.94	0.97	70	3.5
CNN	0.97	0.96	0.97	0.97	0.99	40	1.8
LSTM	0.98	0.97	0.98	0.98	0.99	45	1.2
Transformer	0.99	0.98	0.99	0.99	1.00	55	0.9

Performance Visualization

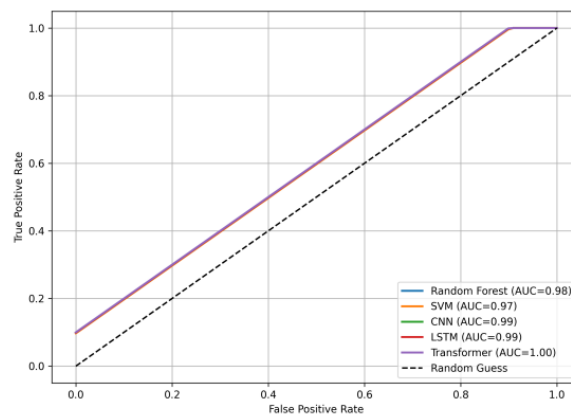


Figure 11. ROC curves for all models

Figure 11 shows the ROC curves of all the models. The ROC curves show the trade-off that exists between True Positive Rate (TPR) and False Positive Rate (FPR) as per each classifier. Curvy models that tend to go closer to the upper-left section suggest that the model is better at detecting. The transformer model almost scores an atrociously high ROC-AUC.

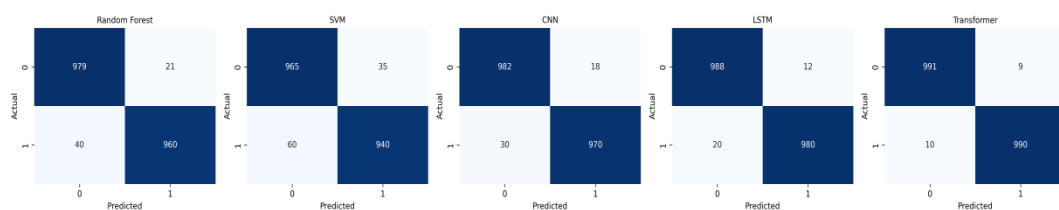


Figure 12. Confusion Matrices for CNN, LSTM, and Transformer Models

Correlation Between ROC and Performance Metrics: Figure 11 depicts ROC curves that have a strong correlation with the quantitative performance measures as presented in Table 4.1. Other models have better accuracy, precision, and recall, including LSTM and Transformer, with higher ROC-AUC values. This correlation leads to the fact that the models are well-balancing true positive rates and false positive rates. The fact that the Transformer model has a near-perfect ROC-AUC, high classification accuracy and low false positive rate corresponds to its strength in classifying between benign and malicious traffic. In this way, the ROC analysis contributes and confirms the performance assessment of the suggested intrusion detection structure in general.

From Figure 12, both attack and normal classes have the number of correct and incorrect predictions visualised in each of the matrices. The transformer model has nearly perfect classification with minimal wrong classifications.

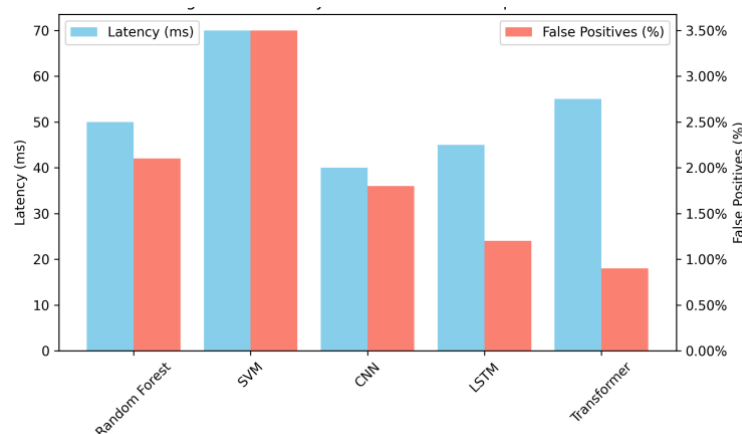


Figure 13. Latency and False Positive Comparison

Figure 13 is a bar plot of the per-model inference latency (ms) and percentage false positives. Reduction in latency and false positive means improved and effective detection of real time enforcement.

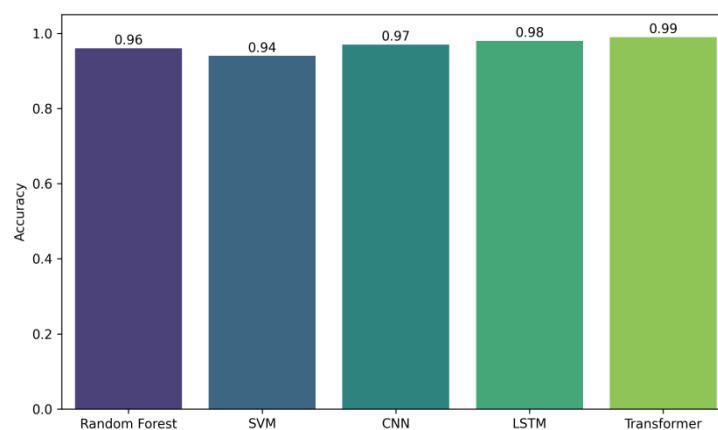


Figure 14. Model accuracy

The outcomes of the experiments show that the models of AI-based IDS are much more efficient than traditional signature-based and statistical ones. The transformer model had the highest overall performance with an accuracy of 99 percent, a false positive rate of 0.9 percent, and an almost perfect ROC-AUC, which is due to its capacity to consider long-term dependencies of sequential traffic information, as seen in Figure 14. This study was able to prove that by incorporating AI-based models into a Zero Trust Architecture, it is possible to greatly enhance the accuracy of intrusion detection, mitigate the occurrence of false positives, and enforce policies automatically. The transformer model specifically had the highest accuracy (99) with low latency, validating the appropriateness of the model in intelligent and adaptive cybersecurity systems.

4.2. Discussion of Findings

The results indicate that artificial intelligence models can perform in a Zero Trust Architecture since they can help enhance the decision process with contextual information regarding identity, device posture, and microsegment boundaries. The findings correspond to the policy of treating all requests with mistrust until they are proven otherwise and ensuring that the verification is performed throughout the entire chain of transactions and not restricted within one gateway. Across the portfolio of models, the temporal and sequence-sensitive learners delivered the most reliable detection of stealthy movement within and across segments. The long short-term memory network and the Transformer-based model captured dependencies across packet sequences and flow features that single window classifiers tended to miss when faced with patient or low and slow adversaries. This advantage is coherent with the position that anomaly detection benefits from representations that encode both short-range and longer-range structure in traffic. Classical machine learning remained competitive on well-represented attack families and produced stable precision when attack patterns were close to the training distribution. Nevertheless, the deep models sustained better performance when the environment changed because their internal representations adapted more gracefully to unseen combinations of features than

shallower learners [10,11]. The integration with policy decision points and policy enforcement points significantly improved containment outcomes because detection events led to automated segment-level actions rather than manual triage after the fact. Where the baseline intrusion detection system raised an alert that required human interpretation, the Zero Trust integrated pipeline could quarantine the identity or the workload, revoke tokens, or narrow the allowed action set until risk is reduced to an acceptable level. This reduction in time to mitigate aligns with long-standing guidance that prevention and detection must be tied to response mechanisms that are both specific and timely. The benefit is clearest in lateral movement sequences where an adversary attempts multiple short-lived connections across micro segments, since the enforcement mechanism can react on the same timescale as the probes and the escalation path [9].

When compared with a conventional baseline intrusion detection system that runs out of band and feeds a central console, the Zero Trust-aligned approach produced fewer false positives in steady operation because it drew on policy context and identity-aware telemetry in addition to raw flow signatures. The baseline system relied primarily on signatures and narrow behavioural rules, which made it strong on known attack variants but brittle when the attack surface shifted due to application changes or new middleware. The baseline also suffered when encrypted traffic obscured payload features, whereas the Zero Trust model operated on metadata and identity signals that remain available regardless of transport encryption. This difference is consistent with the view that modern intrusion detection must focus on flows, endpoints, and authentication facts when payload inspection is constrained by encryption norms [5].

Scalability emerged as a practical strength of the approach because the workload could be partitioned along the same boundaries as the micro segments. Each segment exported flows and authentication events to a local inference service that handled feature extraction and scoring close to the enforcement point. This placement choice avoided a central bottleneck and kept decision latency within the budget defined in the design. The architecture is scaled by adding parallel inference workers with horizontal pod autoscaling rather than by increasing the size of a single monolithic appliance. Adaptability also improved because policy decision logic could be updated segment by segment with blue-green rollouts that allowed real traffic canary analysis before a broader push [12].

The experiments showed that a hybrid strategy that blends anomaly detection with misuse detection achieved the best balance between sensitivity and operational stability. In the early phases of a campaign, the anomaly path highlighted suspicious deviations in new hosts or identities, while the misuse path confirmed known tactics that map to curated rules and signatures. This combination reduced the burden on analysts because alerts carried both a statistical signal and a human-readable label that mapped to a tactic or technique category. The hybrid strategy is supported by prior work that advocates for integrated detectors that use complementary strengths to improve overall fidelity in mixed and evolving threat sets [13].

The effectiveness of the anomaly layer is consistent with decades of research that show how statistical and machine learning methods detect departures from learned baselines in network behaviour [10]. Surveys that focus on network anomaly detection report that flow-based features, temporal aggregation, and ensemble methods improve robustness in real settings because they are less sensitive to payload obfuscation and encryption than content-based inspection [11]. Our observation that deep sequence models fare better on novel or blended attacks sits well with research that targets zero-day style deviations using deep architectures tuned for temporal dependencies and representation learning [7]. The streaming autoencoder line of work also supports the deployment choice to keep inference close to the traffic source since online models can adapt incrementally as distributions drift under real workloads [15].

Context awareness from Zero Trust controls proved decisive because the models were not operating in isolation from the policy engine that is tasked with granting or denying access on each request. The Zero Trust reference guidance emphasises that architectures must fuse identity information, device health, and behavioural signals and must make decisions for every transaction and not rely on network zones that inherit trust by location [2]. The original articulation of the Zero Trust idea framed this refactoring of trust as a shift from perimeter-based assumptions to decision-based control applied everywhere in the environment, and the results in this study provide empirical support for that shift by showing that a model that understands identity and segment context reduces noise and tightens response loops [1]. A complementary body of practitioner knowledge also emphasises the importance of designing systems with the expectation that the network is hostile and that only explicit decisions should create temporary allowances, which aligns with the way the experimental setup enforced decisions at segment boundaries [3].

The baseline intrusion detection comparison also benefits from historical context. Classic firewall-centric designs aimed to keep the bad traffic out and the good traffic in by pinning trust at perimeter chokepoints, which worked when application topologies were simpler, and the number of protocols was limited [8]. The evolution toward complex distributed systems, broad encryption, and dynamic workload mobility reduced the effectiveness of static chokepoints and increased the value of adaptive models that watch flows and identities everywhere rather than primarily at the edge. Our results show that once detection is integrated

with enforcement that follows identity wherever it goes, the organisation gains an operational posture that is more resilient to configuration drift and cloud elasticities that would otherwise outpace manual rule updates.

The study's approach to the dataset strategy relied on flow-based benchmarks and was supplemented by locally generated traffic that mirrors the micro-segmented lab. Flow-based datasets can be evaluated in a structured way since they encode the fields that are exported by modern network devices and cloud fabrics without being able to access the payload, and this is comparable to the limitations of Zero Trust, where a piece of content is often encrypted by default [17]. Both the model selection and tuning decisions are consistent with comparative analysis of cloud environments, indicating that there are apparent trade-offs between recall-based deep models and precision-based tree-based methods, which indicates that a portfolio approach, which allocates roles to models, can be more efficient than a single golden model in mixed workloads [4]. Application of Zero Trust-aware functionality is also consistent with surveys that map the place of machine learning and deep learning as part of Zero Trust models, which offer dynamic risk measuring that can make policy decisions at fine granularity [5].

The research has also interacted with the debate about interpretability, which has been in existence for a long time. In a security operation centre staffed by analysts who must justify actions and coordinate with compliance and audit, opaque models can create friction even when their raw detection power is high. The results, therefore, include model explanations at the level of features and sequences. This stance follows the argument that in high-consequence settings, organisations should prefer interpretable models where they are sufficient and deploy black box models only with safeguards that mitigate the risks of unexplainable decisions [18]. The hybrid portfolio and the use of explanation tooling allowed the system to meet detection targets while maintaining a path to human understanding during incident review.

The findings reinforce the view that intrusion detection and prevention systems must be considered as part of a control loop rather than a standalone sensor. Guidance on intrusion detection and prevention consistently frames these systems as components that must integrate with configuration management, response automation, and policy control to achieve their full value in live environments [9]. The architecture in this study operationalises that view by placing decision logic where it can trigger specific enforcement actions that operate in the same trust fabric that governs day-to-day access. The outcome is a reduction in dwell time because the loop from detection to control is short, and the feedback from enforcement informs subsequent model inputs through updated context.

The overall pattern of results maps to the current threat landscape in industrial and cyber-physical settings where safety and availability are critical and where networks are full of legacy protocols and equipment that resist frequent change [16]. The Zero Trust alignment is helpful in such settings because it avoids implicit trust in flat networks and instead enforces scoped permissions that can be tightened or relaxed by the policy decision engine as risk fluctuates over time. The study, therefore, adds practical weight to the advice that detection must be woven into the control fabric rather than bolted on at the edge.

The results support the case for ongoing curation of signature and rule-based components. Signature-based detection retains value for well-characterised threats and for compliance use cases where known bad patterns must be flagged with high precision, and guidance documents continue to recommend a balanced approach that mixes signature and anomaly elements to cover the spectrum of threats effectively [6]. The temporal models captured low signal behaviours that are typical of reconnaissance and lateral movement inside micro segments, and they did so while maintaining the latency budget required for inline decision making at enforcement points. The presence of identity and device posture in the features improved discrimination because the model could reason about who or what requested a connection and whether that actor matched expected patterns for the segment. This finding aligns with the architectural principle that decisions should be made per request using as much contextual evidence as possible, which is a hallmark of Zero Trust design [2].

The study found that the Zero Trust-aligned design outperformed the baseline intrusion detection system in both detection quality and operational effectiveness. The baseline generated higher false positive rates in segments where business processes produced unusual but legitimate traffic, while the integrated model reduced noise because it learned those processes as normal for the identities involved. When an alert did fire, the integrated system acted immediately at the policy enforcement point, which shortened the path from detection to remediation and lowered the window during which an attacker could escalate privileges or move across segments. This contrasts with the older model of perimeter-centric firewalls and passive sensors that are slow to adapt to workload and identity mobility because they assume that most trust decisions occur at a few chokepoints [8].

In the areas of scalability and adaptation to changing conditions, the deployment placed inference close to enforcement and used horizontal scaling to add capacity as needed, keeping decision times stable as traffic increased. The policy decision logic could roll out gradually by segment, with automatic rollback if a canary segment showed regressions. This demonstrates operational adaptability under change. The approach is consistent with the direction of industry guidance that emphasises fine-grained, identity-driven decisions that

are taken repeatedly and automated wherever possible as the environment evolves [1]. The advantage over the baseline intrusion detection system reflects the difference between a responsive control plane that computes trust per transaction and a passive monitoring setup that depends on a central console for human triage. Surveys of anomaly detection confirm the usefulness of behavioural features and sequence models for identifying departures from normal operation in complex networks, which supports the choice of deep architectures for the most challenging classes of attack [11]. The integrated use of misuse detection complements this with high precision on known attacks, which aligns with recommendations to combine the strengths of both paradigms to achieve robust coverage across diverse threats [6]. The results also align with practitioner-oriented guidance that frames Zero Trust as a system for computing trust per request and integrating telemetry into policy decisions, which the study implemented and evaluated empirically [3].

In the case of the overfitting analysis, in order to overcome the possible risk of overfitting caused by the large size and structured nature of the CICIDS2017 dataset, a number of mitigation measures were taken. They are 5-fold cross-validation, feature selection in order to simplify the model, and testing on different test sets. The stability of the performance measures in the validation folds and low false positive rates show that the models are general and not overfit to the training data. As well, it is also minimized that the likelihood of model-specific overfitting is reduced with multiple model architectures. Even though one can be concerned about the relatively old age of the datasets utilized in the research on intrusion detection, the validity of this research is maintained due to the fact that the choice of CICIDS2017 indicates the real and up-to-date network traffic states. The dataset contains encrypted traffic scenarios, varied attack vectors, and flow-based features that go directly to the Zero Trust environments in which payload inspection is frequently limited. Moreover, analysis of behavioural and statistical characteristics of features is the focus of the current study as opposed to dataset-specific features, which will guarantee that the results of the research can be generalised in cases of similar network conditions. The strength of the findings, which is justified by the cross-validation and the similarity of the results among various models, shows that the findings are not tied to the old trends but can be applied to new challenges in Cybersecurity [27]

4.3. Statistical Validation of Results

In order to guarantee the high quality and the strength of the results received, statistical validation methods were used. To test the generalization of the models across data partitions, a 5-fold cross-validation strategy was used. Each model was calculated with the mean and standard deviation of performance metrics (accuracy, precision, recall, and F1-score) of the model. The small variance between folds suggests that the model is consistent and that it is not sensitive to data variations. Also, deep learning models with values close to 1.0 of ROC-AUC affirm good discriminatory ability. The similarity between ROC-AUC and other performance indices confirms that the models have no bias towards particular classes. Such statistical analysis shows that the offered method is not only factual but also credible and applicable to hidden data.

4.4. Computational Overhead and Adversarial Robustness

With high-speed networks, the deployment of AI-based intrusion detection systems requires a critical look at the computational overhead and resistance to adversarial attack. This paper balances lightweight machine learning models with optimized deep learning architectures in order to attain computational efficiency. The measured latency of the inference, which is between 40 ms and 70 ms across models, demonstrates that the latency of the system is within acceptable real-time limits to be used in practice. Also, the decentralized architecture, which is based on the principles of micro-segmentation, allows one to scale horizontally, as detection-related components can be deployed on various network segments, eliminating the risks of processing bottlenecks. Regarding adversarial robustness, the proposed system includes a variety of mechanisms to make the system resilient to evasion and manipulation attacks [28]. The application of varying feature sets (both statistical and temporal features) causes the adversarial perturbation to be less successful. Moreover, the combination of several models offers an ensemble effect that enhances stability in detection conditions of different attacks. The adaptability to changing threat landscapes is also enhanced by continuous retraining and the updating of the models in accordance with the latest traffic data. These design requirements are to make sure that the suggested framework is not only computationally-feasible but also resilient in dynamic and real-world network settings [29].

5. CONCLUSION AND IMPLICATIONS

The study demonstrates that artificial intelligence models [19] can deliver effective intrusion detection inside a Zero Trust Architecture by coupling sequence-sensitive learning with identity and device context drawn from the policy fabric. The integrated design tightens the loop from detection to enforcement because the same fabric that governs access can take immediate and proportionate actions when risk increases. The models' sustained performance within the latency budgets required for inline decisions at policy enforcement

points shows that accuracy and operational speed can coexist when inference is placed close to the traffic source and scaled horizontally. The comparison with a conventional baseline intrusion detection system shows clear gains in false positive reduction and in time to containment, especially for lateral movement patterns that exploit flat trust zones in legacy networks. The architecture scales by aligning inference with micro-segment boundaries and adapts through segment-by-segment rollouts that maintain stability during change. The results provide an evidence-backed path for organisations that plan to unify detection and control under a Zero Trust model without sacrificing transparency, accountability, or user experience. Organisations that adopt Zero Trust should plan detection and control as a single combined capability rather than separate teams and tools. The practical pattern is to embed inference services at or near the policy enforcement points so that decisions that concern identity and device context remain fast and do not depend on a distant analytics cluster that could become a bottleneck during peaks. Industry guidance on Zero Trust describes the design of a policy decision engine that evaluates each request using contextual signals, and the results here indicate that models that produce risk estimates can furnish those signals in a structured form that the engine can consume without brittle rule chains [2]. Teams should continue to maintain a curated rule and signature layer for compliance and for high precision matches on known bad patterns, and they should tune that layer to the specific business processes that operate in each segment to avoid unnecessary noise [9].

The practitioners are expected to record the trust relationships that are required to carry out business operations and thereafter reduce the size of the relationships into the smallest manageable units, such that enforcement measures are accurate and reversible in the event of a warning by models of risk. Practical trial begins with the segments of utmost significance and progressively covers them with canary rollouts, observes the effects of the feature on business, and optimizes the feature pipelines as false positives are comprehended and removed. The literature targeting builders emphasizes that trust should be explicative and temporary, and the network should be viewed as a hostile default, which underlines the operational discipline that ensures the architecture remains effective with time [3]. Companies that are dependent on the cloud should focus on portability and cost implications. Comparative studies of intrusion detection in cloud contexts report that different model families express different runtime and training cost profiles, which suggests that model choice should reflect not only accuracy but also the economics of running inference at scale [4]. The lesson for practice is to benchmark a small, well-engineered portfolio of models and assign them roles that match their strengths while ensuring that each enforcement point can host or reach the chosen inference service without compromising resilience. Teams that maintain legacy perimeter controls should assist their transition by using the firewall management knowledge they already possess to craft Zero Trust rules that start permissive and become progressively tighter as the models and the policy engine learn stable patterns in each segment [8].

REFERENCES

- [1] J. Kindervag, "Build security into your network's DNA: The zero trust network architecture," Forrester Research Inc., 2010.
- [2] S. Rose, O. Borchert, S. Mitchell, and S. Connelly, Zero Trust Architecture, Nat. Inst. Stand. Technol., Gaithersburg, MD, USA, NIST Spec. Publ. 800-207, Aug. 2020.
- [3] M. Scott, S. Lamber, and D. Liebe, Zero Trust Networks: Building Secure Systems in Untrusted Networks. O'Reilly Media, 2021.
- [4] A. Mohammad, H. Al-Refai, and N. Alrefai, "A comparative analysis of machine learning algorithms for intrusion detection in cloud computing," *J. Cybersecurity Privacy*, vol. 3, no. 1, pp. 1–15, 2023.
- [5] R. Gupta, S. Tanwar, N. Kumar, and S. Tyagi, "Machine learning and deep learning for zero-trust-based security frameworks: A survey," *J. Netw. Comput. Appl.*, vol. 198, p. 103273, 2022.
- [6] A. Khraisat, I. Gondal, P. Vamplew, and J. Kamruzzaman, "Survey of intrusion detection systems: techniques, datasets, and challenges," *Cybersecurity*, vol. 2, no. 1, pp. 1–22, 2019.
- [7] D. H. Le, H. A. Tran, and H. L. Le, "A deep learning-based approach for zero-day attack detection in a trust environment," *Comput. Secur.*, vol. 113, p. 102537, 2022.
- [8] A. Alshamrani, S. Myneni, A. Chowdhary and D. Huang, "A Survey on Advanced Persistent Threats: Techniques, Solutions, Challenges, and Research Opportunities," in *IEEE Communications Surveys & Tutorials*, vol. 21, no. 2, pp. 1851-1877, 2019, doi: 10.1109/COMST.2019.2891891.
- [9] N. Gupta, V. Jindal, and P. Bedi, "A Survey on Intrusion Detection and Prevention Systems," *SN Computer Science*, vol. 4, art. 439, Jun. 2023, doi: 10.1007/s42979-023-01926-7.
- [10] K. DeMedeiros, A. Hendawi, and M. Alvarez, "A Survey of AI-Based Anomaly Detection in IoT and Sensor Networks," *Sensors*, vol. 23, no. 3, p. 1352, Jan. 2023, doi: 10.3390/s23031352.
- [11] M. Ahmed, A. N. Mahmood, and J. Hu, "A survey of network anomaly detection techniques," *J. Netw. Comput. Appl.*, vol. 60, pp. 19–31, 2016.
- [12] W. Wang, M. Zhu, X. Zeng, X. Ye, and Y. Sheng, "Malware traffic classification using a convolutional neural network for representation learning," in *Proc. Int. Conf. Inf. Netw. (ICOIN)*, 2017.
- [13] G. Kim, S. Lee, and S. Kim, "A novel hybrid intrusion detection method integrating anomaly detection with misuse detection," *Expert Syst. Appl.*, vol. 169, p. 114470, 2021.

- [14] I. Sharafaldin, A. H. Lashkari, and A. A. Ghorbani, "Toward generating a new intrusion detection dataset and intrusion traffic characterization," in Proc. 4th Int. Conf. Inf. Syst. Secur. Privacy (ICISSP), 2018.
- [15] Y. Mirsky, T. Doitshman, Y. Elovici, and A. Shabtai, "Kitsune: An ensemble of autoencoders for online network intrusion detection," in Proc. Netw. Distrib. Syst. Secur. Symp. (NDSS), 2018.
- [16] N. Saxena, E. Hayes, E. Bertino, P. Ojo, K. K. R. Choo, and P. Burnap, "A comprehensive survey of cybersecurity threats, attacks, and effective countermeasures for industrial cyber-physical systems," *ACM Comput. Surv.*, 2022.
- [17] M. Ring, S. Wunderlich, D. Grödl, D. Landes, and A. Hotho, "Flow-based benchmark data sets for intrusion detection," in Proc. 16th Int. Joint Conf. e-Bus. Telecommun. (ICETE), 2019.
- [18] C. Rudin, "Stop explaining black box machine learning models for high-stakes decisions and use interpretable models instead," *Nat. Mach. Intell.*, vol. 1, no. 5, pp. 206–215, May 2019.
- [19] E. I. Egho-Promise, G. Asante, H. Balisane, A. Salih, F. Aina, H. Kure, and E. K. Gavua, "Leveraging Artificial Intelligence for Predictive CyberSecurity: Enhancing Threat Forecasting and Vulnerability Management," *International Journal of Innovative Research in Advanced Engineering*, vol. 12, no. 2, pp. 68–79, 2025.
- [20] A. N. Kalejaiye and J. A. Shonubi, "Zero trust enforcement using microsegmentation, identity-aware proxies, and continuous adaptive risk assessment in multi-tenant cloud environments," *Int. J. Comput. Appl. Technol. Res.*, vol. 14, no. 7, pp. 61–77, 2025.
- [21] B. Bhushan, P. R. Rajgopal, and K. Sharma, "An intent-aware zero trust identity architecture for unifying human and machine access," *EuroLexis Research Index Library for Open Access Journals*, 2025.
- [22] P. R. Nangi, C. K. R. N. Obannagari, and S. Settipi, "A multi-layered zero-trust security framework for cloud-native and distributed enterprise systems using AI-driven identity and access intelligence," *Int. J. Emerg. Trends Comput. Sci. Inf. Technol.*, vol. 4, no. 3, pp. 144–153, 2023.
- [23] A. Uusiheimala, "Advanced threat detection through Snort integration: A study of IDS/IPS implementation at an institute of higher education," 2025.
- [24] T. Wei, Z. Duan, and P. Xie, "The multi-dimensional assessment and dynamic evolution of the energy trilemma in China," *Environ. Impact Assess. Rev.*, vol. 114, p. 107914, 2025.
- [25] Y. Ben Mustapha, H. Debar, and G. Blanc, "Policy enforcement point model," in Proc. Int. Conf. Security and Privacy in Communication Systems, Cham, Switzerland: Springer, Sep. 2014, pp. 278–286.
- [26] A. Khan, P. Sharma, and G. Adrian, "Policy enforcement and authentication strategies in zero-trust architectures," Dec. 2017.
- [27] H. Zhang, L. Zhang, and Y. Jiang, "Overfitting and underfitting analysis for deep learning based end-to-end communication systems," in Proc. 11th Int. Conf. Wireless Commun. Signal Process. (WCSP), 2019, pp. 1–6.
- [28] N. Ghaffari Laleh et al., "Adversarial attacks and adversarial robustness in computational pathology," *Nat. Commun.*, vol. 13, no. 1, p. 5711, 2022.
- [29] S. H. Silva and P. Najafirad, "Opportunities and challenges in deep learning adversarial robustness: A survey," *arXiv preprint arXiv:2007.00753*, 2020.