National College of Ireland

Intelligent Anomaly Detection in Network Traffic Using Advanced Machine Learning Techniques

MSc Research Project
Programme Name:- Practicum 2

Name: - Shaik Junaid

Student ID: x23354631

School of Computing
National College of Ireland

Supervisor: Lavish Thomas

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student NAME:- SHAIK JUNAID

Student ID: X23354631

Programme: Master of Science in AI **Year:** 2025

Module: Practicum 2

Supervisor: Lavish Thomas

Submission Due Date: 15/09/2025

Project Title: Intelligent Anomaly Detection in Network Traffic Using Advanced Machine Learning Techniques

Word Count: ..... **Page Count**.....

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: Shaik Junaid

Date: 15/09/2025

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Intelligent Anomaly Detection in Network Traffic Using Advanced Machine Learning Techniques

Student Name: Shaik Junaid
Student ID: x23354631

Table of Contents

1.0 Introduction 5

1.1 Background 5

1.2 Problem Statement 5

1.3 Research Aim 6

1.4 Research Objectives 6

1.5 Research Questions 6

1.6 Research Significance 6

1.7 Research Motivation 6

1.8 Structure of the Dissertation 6

2.0 Literature Review 7

2.1 Introduction 7

2.2 Concept of Anomaly Detection in Network Traffic using ML 7

2.3 Managing Class Imbalance in ML-Based Anomaly Detection 7

2.4 ML-Facilitated Solutions for Scalable and Comprehensive Anomaly Detection 7

2.5 Significance of Advanced ML Techniques for Anomaly Detection in Network Traffic 8

2.6 Competence of Supervised and Unsupervised ML Techniques in Anomaly Detection 8

2.7 Evaluating the Efficiency of ML in Resource-Constrained Ambience 9

2.8 Adaptability of ML Models for Real-World Network Traffic Scenarios 9

2.9 Optimising Computational Resource Usage in ML-Based Anomaly Detection 10

2.10 Efficiency Calculation of ML Models Using Key Metrics 10

2.11 Theoretical Frameworks 11

2.11.1 Information Theory 11

2.11.2 Computational Learning Theory (CLT) 11

2.11.3 Pattern Recognition Theory 11

2.12 Literature Gap 11

2.13 Summary 11

Chapter 3: Methodology 12
3.1 Introduction 12
3.2 Research Philosophy 12
3.3 Research Approach 12
3.4 Research Design 12
3.5 Data Collection Methods 12
3.6 Research Tools 12
3.7 Data Analysis Technique 12
3.8 Validity and Reliability 13
3.9 Ethical Considerations 13
3.10 Summary 13
Chapter 4: Design Specification 13
4.1 Methodology 13
4.2 Data Collection and Preparation 13
4.3 Balancing the Dataset 13
4.4 Exploratory Data Analysis (EDA) 14
4.5 Data Preprocessing 14
4.6 Model Building and Evaluation 14
4.7 Training and Validation 14
4.8 Evaluation Summary 15
4.9 Result Analysis 15
Chapter 5: Implementation 15
5.1 Enhancing Detection Accuracy 15
5.2 Supervised vs. Unsupervised Models 16
5.3 Model Sustainability in Low-Resource Environments 16
5.4 Managing Class Imbalance 16
5.5 Scalable Solutions 16

Chapter 6: Conclusion and Future work 16

6.1 Summary 16

6.2 Linking with Objectives 17

6.3 Recommendations 17

6.4 Limitations 17

6.5 Future Scope 17

References 18

1.0 Introduction

1.1 Background

The rise of network attacks and the formation of different internet technologies significantly made the intrusion detection process a key focus and underscored the need for efficient anomaly detection to safeguard network integrity. Signature-based intrusion detection systems often struggle while combating new attack strategies as well as produce high false positives and fail to identify zero-day exploits. Machine Learning (ML)has emerged as an efficient alternative approach as it uses both supervised and unsupervised ML techniques in order to determine deviations from normal network behaviour. Recent studies in their findings have shown the efficiency of ML models such as IsolationForest, SVM, XGBoost, Naive Bayes as well as Support Vector Machine (SVM) to detect anomalies while LightGBM achieves almost the highest training accuracy such as 1.0 and comprehensive test accuracy is 0.85 (Ness *et al*., 2025).

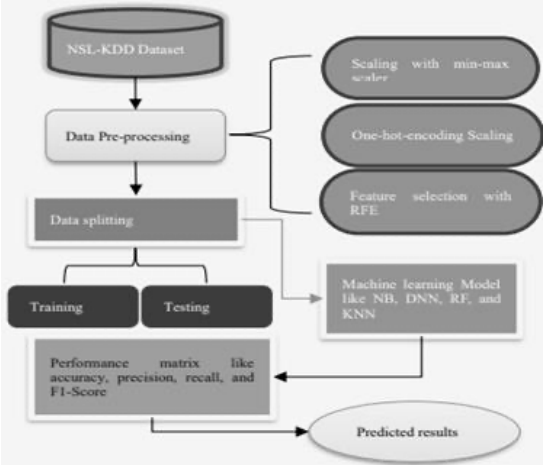


Figure: Process of Network Traffic Analysis in Cybersecurity
Source: (Gopalsamy, 2022)

Hybrid approaches like Deep Learning (DL) and Semi-Supervised Learning have the highest accuracy in identifying spatial-temporal results based on one condition such as optimum computational power. Furthermore, it has been theoretically explored that DL-based hybrid models often enhance detection accuracy but lack interpretability as well as (Abdullah *et al*., 2024) have shown trade-offs between scalability and false positives.

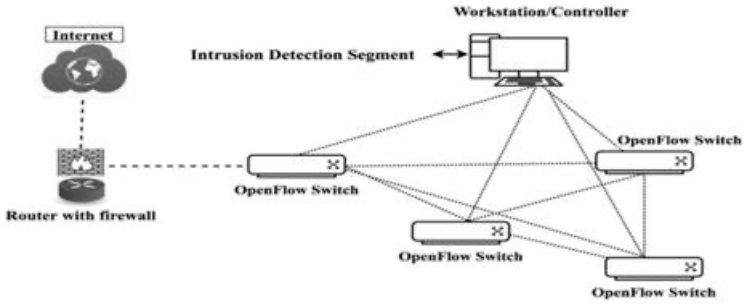


Figure: Anomaly Detection by Using ML Based Networks
Source: (Abdullah *et al.*, 2024)

This research critically emphasizes these meaningful insights for assessing ML techniques comprehensively and addresses potential gaps in scalability, real-time application and imbalance mitigation for current cybersecurity frameworks (Gopalsamy, 2022).

1.2 Problem Statement

In this digital world, the rapid increase of cyber threats and dynamic behaviour of network traffic have significantly shown some crucial limitations in traditional anomaly detection

systems that are signature-based systems. This traditional approach remains incompetent in some specific cases such as zero-day attacks as well as often generating extreme false positives. Well-recognized ML techniques such as LightGBM, XGBoost and DL hybrids hold potential solutions but they experience significant challenges in terms of class imbalance, poor interoperability and low-resource availability.

This research seeks to address these gaps by evaluating and suggesting adaptive and lightweight sustainable anomaly detection frameworks by highlighting key characteristics such as practical applicability and high accuracy of anomaly detection from traffic networks in dynamic network environments.

1.3 Research Aim

This research aims to determine how emerging ML algorithms and techniques increase the efficiency and accuracy of anomaly detection in network traffic to enhance cybersecurity measures. This research seeks to assess the efficiency, sustainability and overall performance of different supervised and unsupervised ML techniques in real-world contexts with imbalanced data and computational constraints.

1.4 Research Objectives

- To critically assess the significance of advanced ML techniques to increase the effectiveness and accuracy of anomaly detection in network traffic for cybersecurity
- To assess the competence of supervised and unsupervised ML techniques while detecting anomalies in network traffic
- To determine the sustainability of Machine Learning techniques in low-resource environments
- To identify and examine the efficiency of these ML models in managing class imbalance using precision, accuracy, recall and F1 score
- To suggest potential ML-facilitated solutions to maintain scalable and comprehensive anomaly detection in network security.

1.5 Research Questions

- How can emerging machine learning approaches increase the precision and efficiency of anomaly detection?
- How do supervised and unsupervised ML models detect anomalies in network traffic?
- How feasible are these ML techniques and models for real-world deployment specifically in environments with poor computation resources?
- How well do these models act as key evaluation metrics such as recall, precision, accuracy to address challenges like imbalance datasets?

1.6 Research Significance

This research makes a critical evaluation of both supervised and unsupervised models that include XGBoost, Isolation Forest and others by using required metrics such as recall, precision and accuracy. This research addresses key challenges like class imbalance in training data as well as computation constraints in deployment environments. The findings of this research offer innovative approaches for optimising model performance while maintaining operational feasibility specifically in resource-constrained systems.

1.7 Research Motivation

The motivation behind conducting this research is to develop comprehensive ML-based detection systems and evaluate the competencies of different ML models and techniques to implement in the real-world context in a low-resourced deployment environment.

1.8 Structure of the Dissertation

Chapter 1 introduces the research background, aim, objective, problem statement, contributions, and motivations behind conducting this research. ***Chapter 2*** has focussed on theoretical information from existing studies by critically evaluating previous works. ***Chapter 3*** represents the methodology employed in the research. ***Chapter 4*** represents data analysis

and data interpretation to extract key findings from diverse existing studies. ***Chapter 5*** is the discussion chapter that critically discusses the identified findings. ***Chapter 6*** Conclusion links findings with each objective, and gives research limitations, future research directions and practical recommendations.

2.0 Literature Review

2.1 Introduction

Anomaly detection in network traffic, or the detection of a pattern of data which is representative of normal behaviour for that system, is fundamental to cybersecurity. Network Intrusion Detection Systems (NIDS) aims to monitor the network's traffic and present anomalies in action as potentially malicious activity that is representative or indicative of an attack, such as Denial of Service (DoS), malware use, or data exfiltration (Xu *et al.*, 2021). Existing (and traditional) approaches (e.g. signature-based IDS) are no longer able to sustain evolving and variant threats, meaning that intelligent anomaly detection and ML (Machine Learning) were adopted as standard approaches. Several recent surveys comment on the promise of deep learning particularly because of its ease of capturing complex patterns/traffic for modern networks. This review has brought together existing literature to provide a thorough survey of intelligent anomaly detection methods in network traffic involving key models, the problem of class imbalance, real-world applicability and future directions.

2.2 Concept of Anomaly Detection in Network Traffic using ML

Anomaly detection can discover any abnormalities in normal network operations that can indicate potential threats such as DDoS, malware, or theft of data (Thawaini, 2022). The bandwidth, packet structures, and protocols are analysed by ML to detect the patterns in real time. Supervised methods include Random Forest and SVM models, while there are unsupervised methods for identifying unknown threats. CNNs and LSTMs are examples of deep learning procedures that represent complex relations. Practical examples involve DDoS mitigation whereby AI-based systems can detect traffic spikes up to 2-5 Gbps in less than a few seconds and respond. Such methods increase the speed, the accuracy, and flexibility of detection in comparison with the conventional systems.

2.3 Managing Class Imbalance in ML-Based Anomaly Detection

Anomaly detection based on ML faces the significant challenge of class imbalance, where the normal network traffic can predominate and the attack traffic can be a small number of minorities (Keerthana *et al.,* 2024). Because of this bias, models are likely to classify rare but important anomalies into the wrong category, so accuracy is not a good metric. Rather, precision, recall, and F1-score are more balanced measures and should offer a more reasonable conclusion (Mirsadeghi *et al.,* 2023). Precision determines the false alarm rates by indicating the percentage of predicted positives that are not anomalous, and recall measures the percentage of real anomalies that are found, indicating model sensitivity. The F1-score strikes a balance between the two. SMOTE has also been widely used as a technique to oversample the minority category, but its efficiency largely depends on the complexity of the data. The integration of cost-sensitive learning with SMOTE raises recall and F1 by punishing erroneous classifications of the important classes. Balanced Random Forests and EasyEnsemble are ensemble algorithms that train on balanced partitions to improve minority detection. Imbalance is also solved in deep learning models through weighted loss and focal loss variations. Nevertheless, the danger of overfitting and decreased generalizability still occurs since oversampling may produce synthetic data that are too close to the existing ones. Future progress in imbalance-sensitive architecture is necessary to realize effective anomaly detection in real time.

2.4 ML-Facilitated Solutions for Scalable and Comprehensive Anomaly Detection

Scalable anomaly detection needs to handle large volumes of traffic, a variety of protocols, encrypted data, and changing usage patterns and should strike the right balance between accuracy and computational costs (Liu *et al*., 2024). Hybrid ML algorithms have the advantage of using supervised accuracy when dealing with known attacks and unsupervised flexibility when it comes to emerging threats, mitigating the overall dependence on huge labelled datasets (Pillai and Sharma, 2023). One significant example was a combination of Isolation Forests and deep autoencoders, resulting in a 30 percent reduction in false positives and a 20 percent increase in zero-day detection. Recent developments, including LightGBM and the use of MobileNetV2, offer fast, low-resource detection that would suit an IoT setting (Yang *et al.,* 2025). With representation learning, it is no longer required to manually extract features, which leads to better detection of encrypted traffic and a higher level of scalability (Nassif *et al.,* 2021). Using transfer learning mitigates training expenses by re-utilizing pre-trained models, and using edge computing will decrease latency and bandwidth requirements. Federated learning provides privacy-sensitive collaborative training, which reduces training by 35% with no drop in performance on par (Taik *et al*., 2022). In order to increase trust, explainable AI (XAI) tools, such as SHAP, help to interpret the output of models, which can be beneficial to SOC operations and compliance requirements (Gummadi *et al.,* 2024; Ullah *et al.,* 2022). Nonetheless, a lack of relevance to the real world is constrained by using old data, such as KDD Cup 99. Strengths of deployment require modern datasets that capture encrypted and multi-protocol traffic. The models should also incorporate scalability through all the metrics such as KPR, KRecall, and KF1-score, and all the methods of balancing the classes such as SMOTE, cost-sensitive learning, and ensemble-based methodology. A combination of these approaches makes the detection of anomalies in dynamic networks possible in real time.

2.5 Significance of Advanced ML Techniques for Anomaly Detection in Network Traffic

Zero days are difficult to detect by traditional signature-based IDS, and a transition to adaptive ML approaches based on data is therefore necessary. Supervised models like SVM, Random Forest, and Gradient Boosting Machines do their job with enough labelled data; however, they are not generally applicable and cannot be applied to new attacks (Azam, 2023; Adhikari *et al.,* 2024). Unsupervised methods such as clustering and autoencoders do not require prior labels and are thus effective in changing environments (Singh & Jang-Jaccard, 2022; Chevrot, 2022). Complex spatial-temporal patterns are addressed by using deep learning approaches such as CNNs and RNNs, and enhance detection accuracy. Some of these innovations, like ARCADE (adversarially regularized convolutional autoencoder), improve robustness and minimize the possibility of overfitting, but require many computational resources (Sowmya and Anita, 2023). Transformer-based models and architectures deliver high detection rates and low latency, with architectures reporting a 96 percent accuracy and a 94 percent precision (Ullah *et al.,* 2024) and hybrid transformer-based CNN–BiLSTM architectures with a 99.8 percent accuracy (Alsaleh *et al.,* 2025). Also, reinforcement learning modelling, Q-learning with LSTM, demonstrates good results: 90 percent precision, F1 scores of 85 percent (Oyewola *et al.,* 2024). There is still performance hampered by class imbalance, solved with SMOTE, cost-sensitive learning, and other metrics in addition to accuracy (Mirsadeghi *et al.,* 2023). The issues with scalability and sustainability remain, in part, because of the old dataset availability, such as KDD Cup 99, which does not account for encrypted, multi-protocol traffic. Hybrid models that use both supervised learning, unsupervised learning, and reinforcement learning are intended to increase adaptability (Adhikari *et al.,* 2024). Although ML has improved anomaly detection significantly, a balance between accuracy, intelligibility, computational load, and versatility is critical so that it can combat the changing cybersecurity threats (Chevrot, 2022).

2.6 Competence of Supervised and Unsupervised ML Techniques in Anomaly Detection

Signature-based IDS has a problem with zero-day attacks, and instead, ML techniques are used in predicting anomalous behaviour in the network. Supervised models like SVM, Random Forest, and Gradient Boosting Machines can be effective when enough labelled data is accessible, as they can identify pre-defined known forms of attacks effectively (Zhang *et al.,* 2023). Nevertheless, they are not very effective in real-time dynamic environments where the threats are invisible (Miguel-Diez *et al.,* 2025). Unplanned procedures, such as clustering and autoencoders, recognize anomalies in unlabelled data. Multi-scale CNN-LSTM autoencoders exceed the performance of conventional algorithms in detecting subtle anomalies (Singh and Jang-Jaccard, 2022), and autoencoders based on traffic images are used to detect traffic better, including encrypted traffic (Neri and Baldoni, 2025). Resource-intensive approaches such as deep vision (CNN) and deep temporal (RNN) models have the benefit of modelling intricate spatial-temporal patterns, but cannot be deployed in constrained settings in a great many cases (Kumari *et al.,* 2024; Wang and Thing, 2023). Federated learning (FL) can be seen as an improvement of the supervised and unsupervised methods, as it allows decentralized training without exchange of raw data, which benefits the privacy-disclosed fields, like IoT and healthcare (Nguyen *et al.,* 2021). FL already has an accuracy of 97.21% by using less training time and has high performance regardless of partial participation. Data bias is alleviated by aggregation methods of FL such as FedAvg-P and FedNova, which enhance generalization. In practice, frameworks such as DInfoT record 95.6 percent detection and no false positive rates (Nguyen *et al.*, 2019). All methods face the problem of class imbalance and tend to skew measurements. Imbalance can be resolved using such methods as SMOTE and cost-sensitive learning, whereas precision, recall, and F1-score are better metrics in place of accuracy (Liu *et al.,* 2024). The scalability and adaptability of hybrid learning models consisting of supervised, unsupervised, and reinforcement learning are better, but are constrained by the use of outdated data such as KDD Cup 99 (Miguel-Diez *et al.,* 2025). Formulation of new benchmarks is crucial in proper performance analysis.

2.7 Evaluating the Efficiency of ML in Resource-Constrained Ambience

Auto ML applied to automotive-embedded controls improves safety, reliability, and performance. Real-time Decision Trees and especially Random Forests can reach the accuracy of 98.3% (Becker *et al.,* 2023). Models with low weight (and even quantized) are effective to optimize edge deployment and do not hurt too much. An IDS optimized with DL in the fog computing assessment is rated 94.3 percent accurate, working with 512MB RAM and > 1GHz processors (Ji *et al.,* 2024). Yet the integration process of decision trees in the microcontrollers needs pruning and quantization, which decreases the accuracy by 15-20% in TinyML. Convergence takes longer when we are reducing neural network layers in order to fit IoT memory constraints (Kashif, 2024). The MLPerf benchmarks benchmark the ML application to newer data sets; the energy consumption of ML-powered IoT devices fell by about 40 percent with hardware-accelerated inference. The training of adaptive IoT models uses up 5-20mb of memory. The energy savings utilised by the Nordic Semiconductor Thingy 52 device cut latency by half (Kashif, 2024). Federated learning (FL) in smart grids minimizes bandwidth consumption (Jithish *et al.,* 2023), yet encounters a 15 to 20 percent drop in performance when communications are delayed in rural/mobile scenarios (Adhikari *et al.,* 2024). Naive Bayes performs well with low memory requirements (Ness *et al.,* 2025), scoring 0.85 on test cases on KDDCup99 (Ness *et al.,* 2025), but it is overly sensitive, thereby creating false positives within complex traffic.

2.8 Adaptability of ML Models for Real-World Network Traffic Scenarios

A multiscale CNN-LSTM autoencoder performs better in encrypted and high-variance network settings and gives improved accuracy with fewer false positives in comparison to conventional approaches (Singh *et al.,* 2022). LSTM models excel at predicting sequences and recognizing patterns (Agbor *et al.,* 2025), yet models with deep structures can no longer

be successfully applied to real implementations because of the differences in the traffic patterns and an absence of contextual information (Adhikari *et al.,* 2024). Learning improves by using hybrid models and reinforcement learning (RL) to be more adaptable. Mixing Deep Q-Learning and analytical search, GuidedLight allows making real-time decisions and demands a theoretical maximum of doubling the RAM and computing time of the static algorithms. The use of AI-native networking in Juniper, a supervised system, resolves VLAN problems with a 99 percent accuracy, enhancing performance and security (Sezar, 2025). Real-life implementations such as Datadog at Toyota Motor North America decreased the mean time to resolution by 90 percent and made the operations proactive to reactive through Watchdog (Sezar, 2025). Microcontroller lightweight models, such as SiFive, enable customers to shorten deployment schedules by 3-6 months, even when toolchains are restricted. Model F1 scores obtained with publicly available datasets training were 92% with a drop in training efficiency by 40 percent (Kumari *et al.,* 2024). Nevertheless, the domain shifts may decrease the effectiveness of models (Chevrot, 2022).

2.9 Optimising Computational Resource Usage in ML-Based Anomaly Detection

Optimal resource utilization is essential to implementing the ML-based systems of anomaly detection. Isolation Forests can handle big datasets using little memory and in a short time, yet find complex data difficult, so they may miss up to one-fifth of advanced attacks and even more false positives (Xu *et al.,* 2023). Distributed Isolation Forests have the potential to reduce analysis time by an average of 60 percent, although there is an overhead caused by data partitioning and synchronization (Gjannaoula *et al.,* 2021). In smart grid detection, federated learning saves 28 percent of energy consumption and 21 percent of training time, but unreliable network connections may require 15 to 20 percent more time to converge (Jithish *et al.*, 2023). The use of feature selection is also used to optimize the performance; training time is also improved by 40 percent when only the top 10 percent of features are selected (Khorasgani *et al.,* 2024). Nevertheless, the accuracy of detection may decrease by 510% due to aggressive reduction (Chen *et al.,* 2016). A median strategy on NSL-KDD attained 99.66 percent detection and 0.70 percent false positives. Efficiency is also enhanced through hardware/software optimizations. The improvement of autocorrelation functions improved Intel by a 66% drop-in run time and a reduction of the monthly cost of 165 dollars. However, it required an expensive, huge investment in accelerators and libraries (Intel, 2025). Finding the optimum trade-off between the performance, cost, and efficiency of computation is a critical challenge to real scalability.

2.10 Efficiency Calculation of ML Models Using Key Metrics

Accuracy is not a suitable indicator of performance in imbalanced datasets, as the dominance of the majority class causes a false negative, especially in medical, fraud detection, and cybersecurity tasks (Altalhan *et al.,* 2025). The preferred metrics include precision, recall, and F1-score based on the confusion matrix, where precision shows the percentage of true positive results, including positive predictions, whereas recall shows the percentage of real positives identified (Sofaer *et al.,* 2019). The F1-score balances the two parameters, and it is crucial for minority-class detection. Three predominant solutions deal with imbalance, namely data-level, algorithm-based, and hybrid. At the data level, SMOTE and its extensions (D-SMOTE, BP-SMOTE) can create a variety of diverse synthetic samples of minority instances, which in combination with the deep model, such as CNN and RNN, show better recall and F1-scores (Altalhan *et al.,* 2025). Cost-sensitive learning and the adaptive versions represent a form of algorithm-based approach where the misclassification of minority classes is penalized, while the cost of adaptation is either determined dynamically as required to achieve convergence and recall, or pre-determined in advance (Simpson, 2003).

Ensemble algorithms such as Random Forest, Gradient Boosting, and AdaBoost are also more effective in minority class detection, but AdaBoost will tend to overfit noise. Hybrid

techniques use a combination of the two strategies with better outcomes. Particle Swarm Optimization with SMOTE and MetroCost augmented the sensitivity and F1 of datasets in health care. Likewise, the combination of BIRCH clustering, Borderline SMOTE, and Tomek Links provided 85.2% F1 in detecting fraud. Smart minorities that were further treated by SMOTE-variants in combination with CNN and RNN architecture enhanced minority recognition in multimodal complex medical data (Sowjanya and Mrudula, 2023), and GANs created synthetic images in image-based imbalance applications. In general, ML models in imbalance situations should be evaluated using more subtle measures than just accuracy, with specific sampling strategies, model tuning, and the use of hybrid frameworks that improve recall, precision, and F1-score without overfitting.

2.11 Theoretical Frameworks

2.11.1 Information Theory

Claude Shannon introduced information theory, which constitutes a fundamental aspect of machine learning and examines the measurement and transmission of information in the form of entropy (Vanam, 2017). Entropy quantifies the uncertainty of a random variable- the greater the value of entropy, the more uncertain it becomes. This framework, in anomaly detection, measures uncertainty in network traffic data. Differences in anticipation lead to a corresponding increase in entropy, which is an indication that something is awry. This aids ML models in identifying the patterns that are predictable or unpredictable so that they can detect patterns with more accuracy.

2.11.2 Computational Learning Theory (CLT)

CLT is a subdiscipline of theoretical machine learning that studies the efficiency and complexity of algorithms, including their effectiveness and reliability with respect to learning (Saxena, 2025). PAC learning, VC dimension, and the bias-variance tradeoff are some key concepts. CLT supports applications in AI model optimisation, NLP, data science, and fraud detection as it informs the choice of models and feature engineering. In anomaly detection, CLT reveals how models interpret past traffic data in attempting to learn unusual behaviour. Algorithms such as LightGBM and XGBoost are observed to have a high accuracy and low variance in anomaly rates.

2.11.3 Pattern Recognition Theory

The theory of pattern recognition is concerned with the interaction of machines in identifying patterns and regularities in data (Boesch, 2024). It deals with algorithms that recognise and classify such patterns and is thus the core application in network traffic anomaly detection. Such models as SVM, Random Forest, and CNNs are trained to classify between normal and abnormal traffic flows. Other hybrid models like CNN-LSTM had high precisions and recalls and thus have been effective in detecting subtle changes in the network with high levels of sensitivity in complex networks.

2.12 Literature Gap

Recent ML studies on anomaly detection tend to use old datasets that cannot be related to real-life cases. There still exist gaps when it comes to scaling, interpretable, efficient hybrid models, zero-day persistence adaptation, and effective class imbalance solutions. Balanced detection of attackers on diverse, evolving network threats requires more generalizable methods.

2.13 Summary

This overview addresses unsupervised, supervised, and hybrid ML methods to detect anomalies, raising such challenges as considering class imbalance, old datasets, and zero-day attack detection. Although accuracy increases with techniques such as SMOTE, deep learning, and federated learning, real-world scalability, flexibility, and explainability are also essential requirements due to the changing network environments.

Chapter 3: Methodology

3.1 Introduction

This chapter describes the approach taken to investigate the way ML models represent their reactions to the network anomalies. It specifies the philosophy, approach, and design of the research and argues the usage of an inductive approach and the descriptive design. In the chapter, the procedure of secondary qualitative data collection, referring to reliable sources of information, tools, and datasets, is outlined, as well as the methods of analysis. It also seeks to address the steps to achieve validity, reliability, and ethical standards compliance, which offers a clear guideline with which the research study can be carried out and assessed.

3.2 Research Philosophy

The research philosophy used is positivist, which is applicable to studies that have empirical and quantifiable outcomes. Positivism focuses on objectivity, statistical measurements, and reproducibility of results, which are important parameters in assessing ML methods in the field of cybersecurity. The philosophy allows outlining patterns in network traffic and measuring ML performance based on observable data and metrics, making this philosophy practical in anomaly detection in the real world.

3.3 Research Approach

To derive generalisations and theories based on single observations, an inductive approach was selected (Sauce & Matzel, 2022). Lacking the limitations of already established frameworks, the study was able to interpret the network traffic and anomaly information to identify new patterns. Such flexibility becomes critical in cybersecurity, where threats keep evolving and the detection system must be adaptive. By using the inductive process, it was possible to identify trends that can be used to provide a self-improving ML-based approach to anomaly detection.

3.4 Research Design

The study applies a descriptive research design to investigate the behaviour of network traffic and how ML models respond systematically. This design explains patterns, descriptions, and anomalies without the manipulation of a variable. It enables a cleaner overlay of traffic flow characteristics to ML detection performance, and the ability to present evidence-based results in a neat manner. The method also establishes a starting point in future work on context-sensitive, adaptive detection methods.

3.5 Data Collection Methods

The research involved a secondary qualitative analysis based on reliable scholarly materials, such as Google Scholar, IEEE Xplore, Springer, ScienceDirect, and ACM Digital Library, on ML algorithms to detect anomalies. The inclusion criteria were publications of 2015 and above, the datasets that were relevant, such as CICIDS2017, NSL-KDD, and UNSW-NB15, and empirical permission of the models. Exclusion criteria did not include non-peer-reviewed and irrelevant pieces of work. Moreover, it employed practical datasets on Kaggle, allowing training and testing of ML models, and open-source libraries, PyTorch and TensorFlow, allowed the scaling, resilience, and repeatability of the research framework.

3.6 Research Tools

This study employs libraries like PyOD and NetML, which are targeted at detecting anomalies in network traffic. PyOD (Python Outlier Detection) is an extensive Python library that facilitates a broad number of algorithms for outlier detection, from traditional to neural network models. NetML provides machine learning-based methods optimized for network anomaly detection and is appropriate for use in cybersecurity.

3.7 Data Analysis Technique

Data analysis was carried out by pre-processing a KDD Cup 99 dataset by encoding the categorical variables, eliminating duplicates, and normalising numerical values. Stratified random sampling was used to deal with issues of imbalances brought about by classes, and an

80-20 training-testing split was carried out with the training set validation. Random Forest, SVM, k-NN, and Deep Neural Networks were trained and tested based on accuracy, precision, recall, F1-score, and ROC-AUC so as to offer fairly and effectively comparable performance amongst the models.

3.8 Validity and Reliability

The validity of the research was established through the usage of established datasets, such as CICIDS2017 and NSL-KDD, that modelled different anomalies and plausible network traffic. The reliability was ensured by the fact that there was uniformity in the preprocessing, model training, and model testing using such tools as PyOD and NetML. The uniformity of methods in the experiments gave it internal reliability, whereas the existence of datasets in the public and open-source libraries provided the external reliability of the measure due to its replicability. These steps made the results of this research precise, reliable, and generalizable to various settings of anomaly detection.

3.9 Ethical Considerations

Anonymised, publicly accessible data was utilised in the research to help ensure privacy and avoid disclosure of personal data. The possibility of ML bias was overcome with a variety of data and equitable training practices. Cybersecurity research accountability, integrity, and data protection standards included transparent documentation and following the ethical guidelines at the university level.

3.10 Summary

The methodology utilizes a positivist philosophy, an inductive design, and descriptive design, and secondary qualitative data from quality resources. Validity and reliability are attained because of open-source ML tools and standardized datasets. Ethical considerations such as anonymized data and mitigation of bias direct the research towards the creation of accurate, flexible, and explainable ML-based anomaly detection systems.

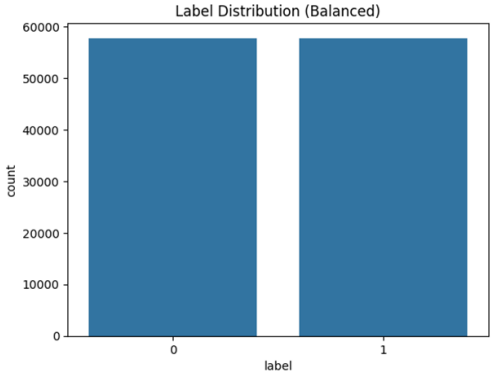
Chapter 4: Design Specification

4.1 Methodology

This paper has proposed a comprehensive ML and DL pipeline to detect network traffic anomalies and intelligently detect cyber intrusions by conducting effective preprocessing, balancing, and comparative model testing. It used the KDD Cup 1999 benchmark dataset since it has been broadly accepted in intrusion detection studies.

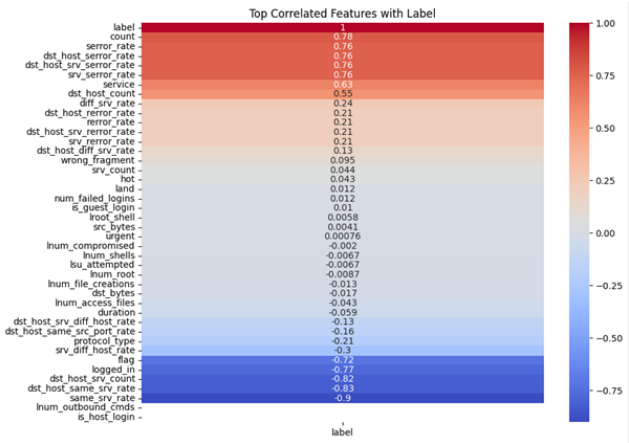
4.2 Data Collection and Preparation

The dataset included 42 features, 494,202 entries, some of which were categorical, and some were numeric. There were no missing data. The label encoder was used to encode categorical features, and 0 (normal) and 1 (anomaly) were used to encode the target variable. Duplication of records was eliminated to avoid bias.



4.3 Balancing the Dataset

To balance the classes, 57,753 examples of each class were randomly sampled, resulting in a balanced training set of 115,506 examples to allow models to learn normal and malicious patterns.



4.4 Exploratory Data Analysis (EDA)

EDA verified the balance between classes and found correlations between features through a heatmap, which helped to single out the most influential predictors.

4.5 Data Preprocessing

Input variables (X) and target (y) were divided. Standard Scaler denormalized features and made distance- and gradient-based algorithms perform better. The training, validation, and test data (60/20/20) were obtained through stratified sampling to maintain balance in the classes.

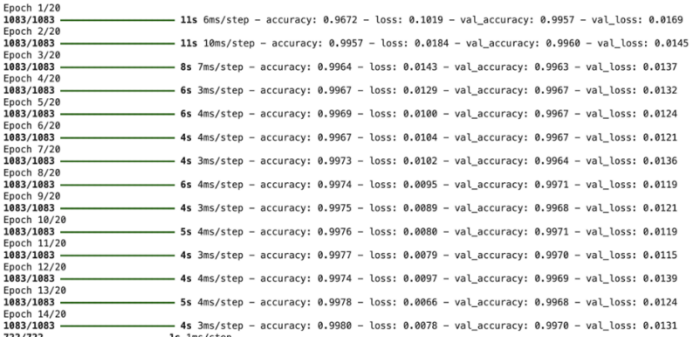
4.6 Model Building and Evaluation

A comparison was made among 3 ML (Random Forest, SVM, and KNN) and a Sequential DL model. Evaluation measures were Accuracy, Precision, Recall, F1-Score, and ROC-AUC.

- **Random Forest**: 99.88% accuracy, best precision and recall.
- **SVM**: 99.66% accuracy, slightly lower recall.
- **KNN**: 99.74% accuracy, balanced precision/recall.
- **Deep Learning**: Two hidden layers (128 & 64 neurons), dropout, sigmoid output; trained with Adam optimizer and early stopping, achieving 99.70% accuracy.

```
Classification Report for Random Forest:
              precision    recall  f1-score   support

           0       1.00      1.00      1.00     11551
           1       1.00      1.00      1.00     11550

    accuracy                           1.00     23101
   macro avg       1.00      1.00      1.00     23101
weighted avg       1.00      1.00      1.00     23101


Classification Report for SVM:
              precision    recall  f1-score   support

           0       1.00      1.00      1.00     11551
           1       1.00      1.00      1.00     11550

    accuracy                           1.00     23101
   macro avg       1.00      1.00      1.00     23101
weighted avg       1.00      1.00      1.00     23101


Classification Report for KNN:
              precision    recall  f1-score   support

           0       1.00      1.00      1.00     11551
           1       1.00      1.00      1.00     11550

    accuracy                           1.00     23101
   macro avg       1.00      1.00      1.00     23101
weighted avg       1.00      1.00      1.00     23101
```

4.7 Training and Validation



The deep learning model trained over 12 epochs, demonstrating steady improvement in accuracy and loss. The final validation accuracy reached 99.70%. Evaluation metrics on the validation set revealed precision and recall values of approximately 99.70%, with an F1-score and ROC-AUC score closely aligned with traditional models.

```
Classification Report for Deep Learning:
              precision    recall  f1-score   support

           0       1.00      1.00      1.00     11551
           1       1.00      1.00      1.00     11550

    accuracy                           1.00     23101
   macro avg       1.00      1.00      1.00     23101
weighted avg       1.00      1.00      1.00     23101
```

4.8 Evaluation Summary

Model	Accuracy	Precision	Recall	F1-Score	ROC-AUC
Random Forest	0.998875	0.999393	0.998355	0.998874	0.998874
SVM	0.996667	0.997485	0.995844	0.996664	0.996667
KNN	0.997403	0.997747	0.997056	0.997402	0.997403
Deep Learning	0.996840	0.997055	0.996623	0.996839	0.996840

All models achieved excellent results with very high accuracy, indicating the robustness of the chosen features and the effectiveness of the balancing strategy. Random Forest slightly outperformed the others in all metrics, though the differences are marginal.

4.9 Result Analysis

The results show that traditional ML models can still be very effective in anomaly detection tasks if combined with good data preprocessing and balancing. Random Forest, particularly, outperformed other models due to its ensemble character and the capacity to discover nonlinear correlations. The deep learning model also performed well, demonstrating good generalization power and the possibility for further improvement by changing architecture, applying regularization, or selecting features. Nonetheless, given the slightly higher computational cost, its benefit compared to Random Forest is not crucial in this situation. The SVM and KNN models gave results that were close, thus indicating that linear and distance-based classifiers are still possible choices for binary classification problems with structured network traffic data.

Chapter 5: Implementation

In this chapter, the findings of the study are interpreted with respect to the study's purpose and literature by considering the ML model's performance, imbalanced data, scalability, and sustainable deployment of cybersecurity.

5.1 Enhancing Detection Accuracy

The experiment established that improving ML models is more effective than traditional signature-based IDS in detecting anomalies. The most accurate model (99.88%) was Random Forest, with KNN, SVM, and the deep learning model scoring close behind (99.70%). The findings are consistent with Azam *et al.* (2023) and Kumari *et al.* (2024). According to them, ensemble and boosting strategies should be applied when the goal is to find nonlinear patterns. CNNs and LSTMs have also been noted to have high temporal-spatial analytical abilities with deep learning approaches. Although it is yet to be widely used practically, interpretability and heavy resource requirements are a resistance to extensive usage (Sowmya and Anita, 2023).

5.2 Supervised vs. Unsupervised Models

Standardized models, which were taught using labelled files such as KDD Cup 99, delivered better precision and accuracy, as reported by Adhikari *et al.* (2024). However, they may perform poorly in the case of challenges from zero-day threats. Literature has revealed that unsupervised approaches like autoencoders and clustering are flexible to work in unlabelled settings (Singh and Jang-Jaccard, 2022) but come with the danger of increased false positives. The combination has illustrated an increment in adaptability and proportional precision that hybrid approaches have shown (Pillai and Sharma, 2023).

5.3 Model Sustainability in Low-Resource Environments

Deep learning models worked well, yet needed much computational power and hence were not scalable in the use of IoT or edge networks. Lightweight models (Naive Bayes, Decision Trees) and efficiency methods of pruning or quantization are emphasized in the literature (Ji *et al.,* 2024). Federated Learning (FL) provides decentralized training and privacy advantages (Jithish *et al*., 2023), but requires reliable connections.

5.4 Managing Class Imbalance

Detection may be biased by class imbalance toward majority traffic. The accuracy of the balanced dataset used in this research was high, as demonstrated by its F1-scores (as high as 0.998), justifying the use of precision-recall-oriented evaluation (Altalhan *et al.,* 2025). Previous research suggests the use of SMOTE, EasyEnsemble, or cost-sensitive learning, but overfitting is still possible (Mirsadeghi *et al.,* 2023). The minority attack detection could be further enhanced by advanced methods such as focal loss and BP-SMOTE.

5.5 Scalable Solutions

In real-time, multi-protocol networks, scalability is essential. Zero-day detection is improved, and false positives are minimized by hybrid models (Pillai and Sharma, 2023). Edge computing and FL minimize the central bottleneck and preserve accuracy. The use of concrete explainable AI tools such as SHAP and LIME enhances trust among analysts and compliance with regulations (Gummadi *et al.,* 2024). To ensure robustness, it is preferable to use the recent dataset, such as CICIDS2017 or UNSW-NB15, instead of an older one, such as KDD Cup 99.

On the whole, the research reaffirms the idea that both ensemble and deep learning models may be used to increase accuracy in structured problems, but that hybrid, resource-aware, and scalable models would be necessary for real-world cybersecurity.

Chapter 6: Conclusion and Future work

6.1 Summary

This paper discussed the role of sophisticated machine learning (ML) algorithms in improving the anomaly detection of network traffic so as to mitigate the drawbacks of signature-based intrusion detection systems (IDS). Experiments on the KDD Cup 99 dataset demonstrated the performance of supervised models: Random Forest, SVM, and KNN, being compared with one based on deep learning. Class imbalance was dealt with by the balance of the dataset, whereas the precision, recall, and F1-score were used as the evaluation metrics.

The accurate performances were obtained by the Random Forest (99.88%), followed by the deep learning (99.70%), whilst the SVM and KNN performed well. The findings established that ML training may achieve up to 118.55 percent gross improvement on detection, flexibility, and scale, yet problems lying in readability, resiliency, and implementation in low-resource settings were identified.

6.2 Linking with Objectives

All objectives were achieved:

- Verified the relevance of ML using experiments and literature, where Random Forest and deep learning performed the best.
- Compared supervised to unsupervised methods with mention of the potential of hybrids in changing threat environments.
- Considered sustainability in low-resource environments in terms of discussing the lightweight models.
- Balanced datasets and applicable measures to address the control of the imbalance of class-specific data.
- Outlined scalable solutions such as federated learning, edge computing, and explainable AI.

6.3 Recommendations

- ***Adopting Ensemble Methods:*** Using Random Forest for balanced accuracy and interpretability.
- ***Applying Advanced Metrics:*** Including F1-score, precision, and recall for imbalanced datasets.
- ***Using Lightweight Models in IoT:*** Deploying optimized decision trees or federated learning in resource-constrained settings.
- ***Integrating Hybrid Models:*** Combining supervised and unsupervised methods for zero-day attack detection.
- ***Upgrading Datasets:*** Using recent datasets like CICIDS2017 or UNSW-NB15.
- ***Investing in Explainable AI:*** Applying SHAP or LIME for transparency.

6.4 Limitations

The analysis used existing databases such as KDD Cup 99, which do not necessarily represent the recent state of network traffic, being encrypted and multi-protocol in most cases. The models were tested under static, offline conditions, not in real-time conditions, so there are not many practical applications. Also, the analysis was limited to particular ML models, and no advanced techniques were introduced into the study, such as transformers or federated learning on a large scale. There were also computational constraints to testing high-resource models in resource-limited settings, which are becoming critical in IoT and edge computing applications.

6.5 Future Scope

Future investigation must aim at using machine learning models within the context of real-time, dynamic network environments to evaluate their adaptive nature in the face of changing threats. To become more relevant, experiments should be expanded to newer datasets, including CICIDS2023. Combining innovative models, such as transformers and reinforcement learning, and federated, as well as edge computing technologies, will enhance scalability and efficiency. Additionally, the implementation of explainable AI (XAI) would lead to increased trust and interpretability, which is vital to critical infrastructure applications. It is also possible to use continuous learning systems to respond to zero-day threats.

References

Abdullah, M.M., Khan, H., Farhan, M. and Khadim, F., 2024. An Advance Machine Learning (ML) Approaches for Anomaly Detection based on Network Traffic. Spectrum of engineering sciences, 2(3), pp.502-527.

Adhikari, D., Jiang, W., Zhan, J., Rawat, D.B. and Bhattarai, A., 2024. Recent advances in anomaly detection in Internet of Things: Status, challenges, and perspectives. Computer Science Review, 54, p.100665.

Alsaleh, N., Alnanih, R. and Alowidi, N., 2025. Hybrid Deep Learning Approach for Automating App Review Classification: Advancing Usability Metrics Classification with an Aspect-Based Sentiment Analysis Framework. *Computers, Materials & Continua*, *82*(1).

Altalhan, M., Algarni, A. and Alouane, M.T.H., 2025. Imbalanced Data problem in Machine Learning: A review. *IEEE Access*.

Azam, Z., Islam, M.M. and Huda, M.N., 2023. Comparative analysis of intrusion detection systems and machine learning-based model analysis through decision tree. IEEE Access, 11, pp.80348-80391.

Becker, T., Rousseau, A.J., Geubbelmans, M., Burzykowski, T. and Valkenborg, D., 2023. Decision trees and random forests. American Journal of Orthodontics and Dentofacial Orthopedics, 164(6), pp.894-897.

Boesch, G., (2024) What is Pattern Recognition? A Gentle Introduction. https://viso.ai/deep-learning/pattern-recognition/#:~:text=Pattern%20recognition%20is%20the%20ability,artificial%20intelligence%20(AI)%20systems.

Chen, S., Huang, Z., Zuo, Z. and Guo, X., 2016, October. A feature selection method for anomaly detection based on improved genetic algorithm. In 2016 4th International Conference on Mechanical Materials and Manufacturing Engineering (pp. 186-189). Atlantis Press.

Chevrot, A., 2022. Detection of contextual anomalies in air traffic data using neural network models (Doctoral dissertation, Université Bourgogne Franche-Comté).

Gopalsamy, M., 2022. Scalable Anomaly Detection Frameworks for Network Traffic Analysis in cybersecurity using Machine Learning Approaches. Int. J. Curr. Eng. Technol, 12(06), pp.549-556.

Gummadi, A.N., Napier, J.C. and Abdallah, M., 2024. XAI-IoT: an explainable AI framework for enhancing anomaly detection in IoT systems. *IEEE Access*, *12*, pp.71024-71054.

Intel., (2025) Accelerate Real-Time Machine Learning Based Anomaly Detection and Forecasting at Scale. https://www.intel.com/content/dam/www/central-libraries/us/en/documents/2022-11/ai-anodot-white-paper-111822.pdf

Ji, R., Padha, D., Singh, Y. and Sharma, S., 2024. Review of intrusion detection system in cyber-physical system based networks: Characteristics, industrial protocols, attacks, data sets and challenges. Transactions on Emerging Telecommunications Technologies, 35(9), p.e5029.

Jithish, J., Alangot, B., Mahalingam, N. and Yeo, K.S., 2023. Distributed anomaly detection in smart grids: a federated learning-based approach. IEEE Access, 11, pp.7157-7179.

Kashif, M., 2024. Deploying Machine Learning Models on Resource Constrained Devices:A Case Study. https://essay.utwente.nl/101088/1/Kashif_BA_EEMCS.pdf

Keerthana, P., Devith, M.D., Hariharan, G. and Muthukumar, S., 2024, March. Machine Learning-Based Anomaly Detection for Imbalanced Network Traffic. In *2024 2nd International Conference on Artificial Intelligence and Machine Learning Applications Theme: Healthcare and Internet of Things (AIMLA)* (pp. 1-5). IEEE.

Khorasgani, A.T., Shirani, P. and Majumdar, S., 2024, September. An Empirical Study on Learning Models and Data Augmentation for IoT Anomaly Detection. In 2024 IEEE Conference on Communications and Network Security (CNS) (pp. 1-9). IEEE.
Kumari, S., Prabha, C., Karim, A., Hassan, M.M. and Azam, S., 2024. A Comprehensive Investigation of Anomaly Detection Methods in Deep Learning and Machine Learning: 2019–2023. IET Information Security, 2024(1), p.8821891.
Liu, R., Shi, J., Chen, X. and Lu, C., 2024. Network anomaly detection and security defense technology based on machine learning: A review. Computers and Electrical Engineering, 119, p.109581.
Miguel-Diez, A., Campazas-Vega, A., Álvarez-Aparicio, C., Esteban-Costales, G. and Guerrero-Higueras, Á.M., 2025. A systematic literature review of unsupervised learning algorithms for anomalous traffic detection based on flows. arXiv preprint arXiv:2503.08293.
Mirsadeghi, S.M.H., Bahsi, H., Vaarandi, R. and Inoubli, W., 2023. Learning From Few Cyber-Attacks: Addressing the Class Imbalance Problem in Machine Learning-Based Intrusion Detection in Software-Defined Networking. IEEE Access, 11, pp.140428-140442.
Nassif, A.B., Talib, M.A., Nasir, Q. and Dakalbab, F.M., 2021. Machine learning for anomaly detection: A systematic review. *Ieee Access*, *9*, pp.78658-78700.
Neri, M. and Baldoni, S., 2025. Unsupervised Network Anomaly Detection with Autoencoders and Traffic Images. arXiv preprint arXiv:2505.16650.
Ness, S., Eswarakrishnan, V., Sridharan, H., Shinde, V., Janapareddy, N.V.P. and Dhanawat, V., 2025. Anomaly Detection in Network Traffic using Advanced Machine Learning Techniques. IEEE Access.
Nguyen, D.C., Ding, M., Pathirana, P.N., Seneviratne, A., Li, J. and Poor, H.V., 2021. Federated learning for internet of things: A comprehensive survey. *IEEE Communications Surveys & Tutorials*, *23*(3), pp.1622-1658.
Nguyen, T.D., Marchal, S., Miettinen, M., Fereidooni, H., Asokan, N. and Sadeghi, A.R., 2019, July. DÏoT: A federated self-learning anomaly detection system for IoT. In *2019 IEEE 39th International conference on distributed computing systems (ICDCS)* (pp. 756-767). IEEE.
Ogah, M.D., Essien, J., Ogharandukun, M. and Abdullahi, M., 2024. Machine Learning Models for Heterogenous Network Security Anomaly Detection. Journal of Computer and Communications, 12(6), pp.38-58.
Oyewola, D.O., Akinwunmi, S.A. and Omotehinwa, T.O., 2024. Deep LSTM and LSTM-Attention Q-learning based reinforcement learning in oil and gas sector prediction. *Knowledge-Based Systems*, *284*, p.111290.
Pervin, N. and Mokhtar, M., 2022. The interpretivist research paradigm: A subjective notion of a social context. *International Journal of Academic Research in Progressive Education and Development*, *11*(2), pp.419-428.
Pillai, S. and Sharma, A., 2023. Hybrid unsupervised web-attack detection and classification–A deep learning approach. *Computer Standards & Interfaces*, *86*, p.103738.
Sauce, B. and Matzel, L.D., 2022. Inductive reasoning. In *Encyclopedia of animal cognition and behavior* (pp. 3414-3421). Cham: Springer International Publishing.
Saxena, A., (2025) Computational Learning Theory in Machine Learning.https://www.appliedaicourse.com/blog/computational-learning-theory-in-machine-learning/
Sezar, S., (2025) Top 5 AI Network Monitoring Use Cases and Real Life Examples.https://research.aimultiple.com/ai-network-monitoring/
Shanmugam, V., Razavi-Far, R. and Hallaji, E., 2024. Addressing Class Imbalance in Intrusion Detection: A Comprehensive Evaluation of Machine Learning Approaches. *Electronics*, *14*(1), p.69.

Sharif, F., 2024. The role of ensemble learning in strengthening intrusion detection systems: A machine learning perspective. Int. J. Comput. Eng. Technol.
Singh, A. and Jang-Jaccard, J., 2022. Autoencoder-based unsupervised intrusion detection using multi-scale convolutional recurrent networks. arXiv preprint arXiv:2204.03779.
Singh, P.K., Jyoti, J.S. and Singh, M., 2025. Real-Time Network Traffic Anomaly Detection Using Unsupervised Machine. *Computer Vision and Robotics: Proceedings of CVR 2024*, p.333.
Sofaer, H.R., Hoeting, J.A. and Jarnevich, C.S., 2019. The area under the precision-recall curve as a performance metric for rare binary events. *Methods in Ecology and Evolution*, *10*(4), pp.565-577.
Sowjanya, A.M. and Mrudula, O., 2023. Effective treatment of imbalanced datasets in health care using modified SMOTE coupled with stacked deep learning algorithms. *Applied Nanoscience*, *13*(3), pp.1829-1840.
Sowmya, T. and Anita, E.M., 2023. A comprehensive review of AI based intrusion detection system. Measurement: Sensors, 28, p.100827.
Taik, A., Nour, B. and Cherkaoui, S., 2022. Empowering prosumer communities in smart grid with wireless communications and federated edge learning. *IEEE Wireless Communications*, *28*(6), pp.26-33.
Thwaini, M.H., 2022. Anomaly detection in network traffic using machine learning for early threat detection. Data and Metadata, 1, pp.34-34.
Ullah, F., Ullah, S., Srivastava, G. and Lin, J.C.W., 2024. IDS-INT: Intrusion detection system using transformer-based transfer learning for imbalanced network traffic. *Digital Communications and Networks*, *10*(1), pp.190-204.
Ullah, I., Liu, K., Yamamoto, T., Zahid, M. and Jamal, A., 2022. Prediction of electric vehicle charging duration time using ensemble machine learning algorithm and Shapley additive explanations. *International Journal of Energy Research*, *46*(11), pp.15211-15230.
Vanam, L., (2017) Part 2: Information Theory | Statistics for Deep Learning.https://medium.com/machine-learning-bootcamp/demystifying-information-theory-e21f3af09455
Vu, T.T.N., 2021. Understanding validity and reliability from qualitative and quantitative research traditions. *VNU Journal of Foreign Studies*, *37*(3).
Xu, H., Pang, G., Wang, Y. and Wang, Y., 2023. Deep isolation forest for anomaly detection. IEEE Transactions on Knowledge and Data Engineering, 35(12), pp.12591-12604.
Xu, W., Jang-Jaccard, J., Singh, A., Wei, Y. and Sabrina, F., 2021. Improving performance of autoencoder-based network anomaly detection on nsl-kdd dataset. *IEEE Access*, *9*, pp.140136-140146.
Yang, Y.M., Chang, K.C. and Luo, J.N., 2025. Hybrid Neural Network-Based Intrusion Detection System: Leveraging LightGBM and MobileNetV2 for IoT Security. *Symmetry*, *17*(3), p.314.
Zhang, C., Wang, N., Hou, Y.T. and Lou, W., Machine Learning-Based Intrusion Detection Systems: Capabilities, Methodologies, and Open Research Challenges. Authorea Preprints.