T.R.
SAKARYA UNIVERSITY
GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

ANOMALY DETECTION IN INTERNET OF MEDICAL THINGS USING DEEP LEARNING

MSc THESIS

Ayşe Betül BÜKEN

Department of Software Engineering

Software Engineering Program

MAY 2025

T.R.
SAKARYA UNIVERSITY
GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

ANOMALY DETECTION IN INTERNET OF MEDICAL THINGS USING DEEP LEARNING

MSc THESIS

Ayşe Betül BÜKEN

Department of Software Engineering

Software Engineering Program

Thesis Advisor: Prof. Dr. Devrim AKGÜN

MAY 2025

The thesis work titled “Anomaly Detection in Internet of Medical Things Using Deep Learning” prepared by Ayşe Betül Büken was accepted by the following jury on 19/06/2025 unanimously as an MSc in Sakarya University Graduate School of Natural and Applied Sciences, Department of Software Engineering, Software Engineering program.

Thesis Jury

Head of Jury :	**Prof. Dr. Ünal ÇAVUŞOĞLU** Sakarya University
Jury Member :	**Prof. Dr. Devrim AKGÜN** (Advisor) Sakarya University
Jury Member :	**Doç. Dr. Zehra KARAPINAR ŞENTÜRK** Duzce University

STATEMENT OF COMPLIANCE WITH THE ETHICAL PRINCIPLES AND RULES

I declare that the thesis work titled " ANOMALY DETECTION IN INTERNET OF MEDICAL THINGS USING DEEP LEARNING", which I have prepared in accordance with Sakarya University Graduate School of Natural and Applied Sciences regulations and Higher Education Institutions Scientific Research and Publication Ethics Directive, belongs to me, is an original work, I have acted in accordance with the regulations and directives mentioned above at all stages of my study, I did not get the innovations and results contained in the thesis from anywhere else, I duly cited the references for the works I used in my thesis, I did not submit this thesis to another scientific committee for academic purposes and to obtain a title, in accordance with the articles 9/2 and 22/2 of the Sakarya University Graduate Education and Training Regulation published in the Official Gazette dated 20.04.2016, a report was received in accordance with the criteria determined by the graduate school using the plagiarism software program to which Sakarya University is a subscriber, I accept all kinds of legal responsibility that may arise in case of a situation contrary to this statement.

18/07/2025

Ayşe Betül BÜKEN

To my 28th year

ACKNOWLEDGMENTS

The completion of this master's thesis represents a meaningful achievement in my academic path and I extend my sincere gratitude to all those who contributed to its realization.

I am especially grateful to my advisor, Professor Dr. Devrim AKGÜN, whose steadfast support, valuable insights and continuous motivation were instrumental throughout every stage of this research.

Ayşe Betül BÜKEN

TABLE OF CONTENTS

Page

ABBREVIATIONS

AdaBoost : Adaptive Boosting
ARP : Address Resolution Protocol
BLE : Bluetooth Low Energy
CNN : Convolutional Neural Network
CoAP : Constrained Application Protocol
DL : Deep Learning
DNN : Deep Neural Network
DoS : Denail of Service
DDoS : Distributed Denail of Service
HTTP : Hypertext Transfer Protocol
ICMP : Internet Control Message Protocol
IoT : Internet of Things
IoMT : Internet of Medical Things
LIME : Local Interpretable Model-agnostic Explanations
LR : Logistic Regression
LSTM : Long Short-Term Memory
MITM : Man In The Middle
ML : Machine Learning
MQTT : Message Queuing Telemetry Transport
PCAP : Packet Capture
RF : Random Forest
RNN : Recurrent Neural Network
SimpleRNN : Simple Recurrent Neural Network
SHAP : Shapley Additive Explanations
SMOTE : Synthetic Minority Over-sampling Technique
SYN : Synchronize
TCP : Transmission Control Protocol
UDP : User Datagram Protocol
XAI : Explainable Artificial Intelligence

LIST OF TABLES

Page

LIST OF FIGURES

Page

ANOMALY DETECTION IN INTERNET OF MEDICAL THINGS USING DEEP LEARNING

SUMMARY

The increasing adoption of the Internet of Medical Things (IoMT) in healthcare systems offers transformative benefits in patient monitoring, real-time diagnostics and automated medical services. However, this growing integration of interconnected medical devices into healthcare infrastructures simultaneously opens up critical cybersecurity vulnerabilities. IoMT environments are particularly vulnerable due to their device heterogeneity, constrained computational capabilities and the high-stakes nature of patient data and services. Traditional signature-based intrusion detection systems (IDS) fall short in handling modern and evolving network threats, particularly zero-day attacks and subtle anomalous behavior that may not match known malicious patterns. Consequently, there is a pressing need for intelligent anomaly detection solutions that are capable of learning complex attack patterns from network data in IoMT contexts.

This thesis presents a comprehensive investigation into machine learning (ML) and deep learning (DL) approaches for network-based anomaly detection using the CICIoMT2024 dataset, one of the most comprehensive benchmark datasets in this field. The goal is to evaluate and compare the effectiveness of different algorithms in classifying network traffic into normal and various anomalous categories, thereby providing a strong baseline for future IoMT security frameworks. The dataset includes traffic from diverse IoMT devices and comprises 18 distinct attack classes in addition to benign data. Our experiments are structured around three classification tasks: binary classification (benign vs. anomalous), 6-class classification (aggregated attack categories) and 19-class classification (fine-grained attacks).

In the baseline evaluation, traditional machine learning models such as Logistic Regression, Random Forest and AdaBoost were trained using standard pipeline techniques involving data cleaning, class balancing, and performance metrics evaluation. For the binary classification task, Random Forest performed exceptionally well, achieving an accuracy of 0.98 along with high precision and recall, demonstrating that classical models are still highly effective for simple anomaly detection tasks where the decision boundaries are more clearly separable.

To address more complex classification tasks, deep learning architectures were introduced. These included Dense Neural Networks (DNN), 1D Convolutional Neural Networks (Conv1D), Recurrent Neural Networks (RNN), Long Short-Term Memory (LSTM) networks and Attention mechanisms. Furthermore, we proposed hybrid deep learning models that combine RNN layers with LSTM units and Attention mechanisms. The hybrid model architecture, specifically the RNN + LSTM + Attention stack, demonstrated superior performance in handling multivariate time series data with spatial and temporal dependencies. This model achieved an accuracy of 0.99 in binary class task, 0.82 in the 6-class task and 0.66 in the 19-class task, significantly outperforming the standalone models.

The attention mechanism enhanced the model's capacity to focus on the most informative time steps, particularly in sequences where anomalous behavior manifests intermittently or subtly. This capacity to dynamically prioritize relevant temporal features proved crucial in distinguishing between closely related attack types, such as DoS vs. DDoS or Spoofing vs. Benign traffic. Additionally, the inclusion of stacked layers in the hybrid models allowed for multi-level abstraction, enabling the model to learn both short-term patterns (through RNN layers) and long-term dependencies (through LSTM layers).

Despite these promising results, the research identified several limitations. First, the reliance on a single dataset, albeit comprehensive, may limit generalizability. Second, deep models entail considerable computational overhead, rendering them less feasible for real-time or edge deployment without optimization. Lastly, class imbalance, even after downsampling, remains a challenge in training fair and unbiased models, especially when detecting underrepresented attack types.

To overcome these limitations and further advance the field, several recommendations for future work are proposed. One major direction involves optimizing models for real-time deployment on edge devices using model compression techniques like weight sparsification, reduced-precision formats and teacher-student training frameworks. Another crucial area is explainability—employing tools like SHAP, LIME, or attention heatmaps can help visualize decision-making processes and provide insights into model behavior, thereby increasing trust in AI-powered intrusion detection systems. Moreover, expanding the dataset to include traffic from newer IoMT devices and communication protocols would enhance generalization and adaptability to real-world conditions.

Feature engineering is another avenue for improvement. Re-extracting and enriching feature sets from raw pcap files could uncover new, more discriminative patterns—especially those that are currently indistinguishable using the existing features. This could also improve classification performance for subtle or overlapping attack types. Furthermore, adaptability to dynamic conditions in IoMT networks, which frequently experience changes in traffic volumes, device states, and operational contexts, is critical. Online learning, domain adaptation, and continual training approaches could help ensure models remain effective over time.

Lastly, cross-dataset generalization and benchmarking are necessary to test the robustness and transferability of proposed models across different network environments. Evaluating the trained models on other benchmark datasets or real-world IoMT traffic can reveal the degree to which these systems can handle diverse operational scenarios and novel attack strategies.

In conclusion, this thesis provides a thorough and structured comparison of various machine learning and deep learning methods for anomaly detection in IoMT networks using a large, realistic dataset. The findings underscore the potential of hybrid deep learning models with attention mechanisms to outperform traditional techniques in complex classification tasks. Nevertheless, realizing practical and generalizable solutions requires further efforts in dataset diversification, model explainability, optimization for edge devices, and adaptive learning strategies. As healthcare continues to evolve into a more connected and data-driven domain, robust and intelligent anomaly detection systems will play a crucial role in protecting patient data and ensuring the continuity and security of critical medical services.

DERİN ÖĞRENME KULLANARAK MEDİKAL NESNELERİN İNTERNETİ'NDE ANOMALİ TESPİTİ

ÖZET

Sağlık hizmetlerinin dijital dönüşümü, özellikle Tıbbi Nesnelerin İnterneti (IoMT) teknolojilerinin yaygınlaşmasıyla birlikte önemli bir ivme kazanmıştır. IoMT, sensörler, giyilebilir cihazlar, implantlar, akıllı monitörler ve hasta takibi yapan tıbbi sistemlerin internete bağlanarak veri toplaması, işlemesi ve paylaşmasını mümkün kılan bir altyapıdır. Bu sistemler; uzaktan teşhis, gerçek zamanlı hasta izleme, ilaç takibi, acil müdahale sistemleri ve ameliyat sonrası bakım süreçlerinde kullanılmakta, hasta güvenliğini ve sağlık hizmetlerinin verimliliğini artırmaktadır. IoMT sayesinde doktorlar, hastalarla fiziksel temasa gerek kalmadan sürekli veri akışı sağlayabilirken, erken teşhis ve hızlı müdahale olanağı da sağlanmaktadır.

Ancak bu dijitalleşme süreci, beraberinde ciddi güvenlik açıklarını ve siber tehditleri de getirmektedir. IoMT sistemleri, hem bireysel hem kurumsal düzeyde büyük miktarda hassas veriyi işler. Bu veriler; hastanın kimlik bilgileri, tıbbi geçmişi, anlık sağlık durumu ve tedavi süreci gibi kritik içerikler barındırır. Bu tür bilgiler siber saldırganlar için oldukça değerli olup, fidye yazılımları, veri sızıntıları ve kimlik hırsızlığı gibi saldırılara davetiye çıkarmaktadır. Ayrıca, IoMT cihazlarının çoğu sınırlı işlem gücüne ve enerji kapasitesine sahip olduğundan, geleneksel güvenlik protokollerini uygulamakta zorlanmakta ve bu durum onları daha savunmasız hale getirmektedir.

IoMT sistemlerinde karşılaşılan başlıca tehditler arasında veri gizliliği ihlalleri, hizmet kesintileri, yetkisiz erişim, cihaz sahteciliği, yazılım açıklıklarının kötüye kullanılması, kötü amaçlı kod enjeksiyonu, fiziksel güvenlik açıkları ve tedarik zinciri saldırıları yer almaktadır. Ayrıca, bu cihazların birçok farklı üretici tarafından geliştirilmesi ve genellikle merkezi olmayan yapılara sahip olması, güvenlik yönetimini daha karmaşık hale getirmektedir.

Tıbbi Nesnelerin İnterneti (IoMT) ağlarında, trafik hem iyi huylu (benign) hem de kötü niyetli (malicous) faaliyetlerden oluşabilir. İyi huylu trafik, herhangi bir zararlı amaç olmaksızın rutin veri toplama ve iletimi gibi cihazlar arasındaki normal iletişimi ifade eder. Ancak, çeşitli siber saldırı türleri bu ağları bozabilir. Bir Hizmet Reddi (DoS) saldırısı, kaynaklarını tüketmek ve kullanılamaz hale getirmek için çok sayıda istekle tek bir cihazı veya sunucuyu hedef alır. Bunun daha ciddi bir versiyonu, birden fazla tehlikeye atılmış cihazın aynı anda hedef sistemi sular altında bırakarak kurtarmayı daha da zorlaştırdığı Dağıtılmış Hizmet Reddi (DDoS) saldırısıdır. Sahtecilik, bir saldırganın yetkisiz erişim elde etmek veya iletişimi engellemek için IP veya MAC adresleri gibi verileri tahrif ederek kendisini güvenilir bir cihaz gibi gizlemesini içerir. Keşif (Recon) saldırıları, ağda güvenlik açıklarını araştırmak, hassas bilgileri toplamak ve gelecekteki istismar için sistemin mimarisini haritalamak için kullanılır. MQTT gibi hafif iletişim protokollerinin yaygın olduğu IoMT'de saldırganlar, kötü amaçlı yükler enjekte etmek, veri akışlarını ele geçirmek veya cihazlar ile sunucular arasındaki mesajlaşmayı bozmak için MQTT tabanlı güvenlik açıklarından yararlanabilir. Bu

çeşitli tehditleri anlamak ve tespit etmek, IoMT sistemlerinin bütünlüğünü, gizliliğini ve kullanılabilirliğini korumak için önemlidir.

IoMT sistemlerinin sunduğu avantajlardan en üst düzeyde yararlanabilmek için, bu sistemlerin karşı karşıya olduğu siber güvenlik tehditlerine karşı gelişmiş koruma stratejilerinin geliştirilmesi ve uygulanması büyük önem taşımaktadır. Bu yüksek lisans tezinde, IoMT ağlarında ortaya çıkan siber güvenlik açıklarının tespiti amacıyla gelişmiş anomali tespit yaklaşımları araştırılmış ve derinlemesine incelenmiştir. Araştırmanın temel amacı, modern ağ trafiği içinde normal ve anormal davranışları ayırt edebilen, öğrenme yeteneğine sahip makine öğrenmesi (ML) ve derin öğrenme (DL) tabanlı çözümler geliştirmek ve değerlendirmektir. Bu doğrultuda, günümüzün en kapsamlı ve gerçekçi veri kümelerinden biri olan CICIoMT2024 veri kümesi kullanılmıştır. Bu veri kümesi, çeşitli IoMT cihazlarından toplanan ağ trafiği verilerini içermekte olup 18 farklı saldırı türünün yanı sıra normal (benign) trafiği de kapsamaktadır.

Tez çalışması üç temel sınıflandırma görevi üzerine yapılandırılmıştır: ikili sınıflandırma (zararsız vs. saldırı), 6 sınıflı sınıflandırma (saldırı türlerinin kategori bazında gruplanması) ve 19 sınıflı ayrıntılı saldırı sınıflandırması. İlk aşamada, temel makine öğrenmesi modelleri olan Lojistik Regresyon, Rastgele Orman (Random Forest) ve AdaBoost algoritmaları uygulanmıştır. Bu algoritmalar, standart ön işleme adımlarından geçirilen veri üzerinde eğitilmiş; veri temizleme, sınıf dengesizliğinin giderilmesi ve temel başarı metriklerinin (doğruluk, kesinlik, duyarlılık, F1 skoru) hesaplanması süreçleri takip edilmiştir. İkili sınıflandırma görevinde Random Forest modeli %98 doğruluk oranı ile oldukça başarılı sonuçlar vermiştir. Bu durum, karar sınırlarının belirgin olduğu basit görevlerde klasik modellerin hâlâ etkili olduğunu göstermektedir.

Daha karmaşık sınıflandırma görevleri için derin öğrenme modelleri tercih edilmiştir. Bu kapsamda, Derin Sinir Ağları (DNN), 1-Boyutlu Konvolüsyonel Sinir Ağları (Conv1D), Yinelemeli Sinir Ağları (RNN) ve Uzun-Kısa Süreli Bellek (LSTM) gibi farklı DL mimarileri denenmiştir. Bu modellere ek olarak, hibrid yapıda olan ve RNN ile LSTM katmanlarının üzerine Dikkat Mekanizması (Attention) entegre edilen özel mimariler tasarlanmıştır. Bu hibrid model, zaman serisi tabanlı ve çok değişkenli verileri işleyebilme kabiliyeti sayesinde daha karmaşık görevlerde yüksek başarı sağlamıştır. Özellikle Yinelemeli Sinir Ağları, Uzun-Kısa Süreli Bellek ve Dikkat Mekanizması ile oluşturulan mimari, 2 sınıflı görevde %99, 6 sınıflı görevde %82 ve 19 sınıflı görevde %66 doğruluk oranı elde etmiştir. Bu oranlar, tek başına kullanılan DL mimarilerine kıyasla önemli bir gelişme ortaya koymuştur.

Dikkat Mekanizması (attention), modelin giriş verilerindeki en anlamlı zaman adımlarına odaklanmasını sağlayarak, özellikle saldırı davranışlarının aralıklı veya gizli biçimde ortaya çıktığı durumlarda önemli bir avantaj sağlamıştır. Bu sayede, normal trafik ile benzer özellikler taşıyan DoS ve DDoS gibi saldırı türlerinin ayrımı daha başarılı bir şekilde gerçekleştirilmiştir. Ayrıca, hibrid modellerin katmanlı yapısı, verideki hem kısa vadeli (RNN katmanları ile) hem de uzun vadeli bağımlılıkları (LSTM katmanları ile) yakalayabilme imkânı sunmuştur.

Bu başarılı sonuçlara rağmen çalışmada bazı kısıtlar da tespit edilmiştir. Öncelikle, çalışmanın tek bir veri kümesi ile sınırlı olması modelin genellenebilirliği açısından kısıtlayıcıdır. Ayrıca, derin öğrenme modelleri yüksek hesaplama maliyeti gerektirdiğinden, gerçek zamanlı veya uç nokta (edge) cihazlarda uygulanabilirlik açısından dezavantaj oluşturabilmektedir. Sınıf dengesizliği, her ne kadar örnekleme

teknikleri ile azaltılmış olsa da, az temsil edilen saldırı türlerinde model başarısında düşüş yaşanmasına neden olmuştur.

Bu kısıtların üstesinden gelmek ve alana katkı sağlamak amacıyla tezde çeşitli gelecek çalışmalara yönelik önerilerde bulunulmuştur. Öncelikli olarak, geliştirilen modellerin kenar cihazlarda (edge devices) çalıştırılabilmesi için model küçültme yöntemleri (örneğin pruning, quantization ve bilgi damıtma - knowledge distillation) ile optimize edilmesi önerilmektedir. Bununla birlikte, modelin karar mekanizmasının anlaşılabilirliğini artırmak adına açıklanabilir yapay zekâ (XAI) yaklaşımlarına da yer verilmelidir. SHAP, LIME veya dikkat ısı haritaları gibi araçlarla modelin hangi özelliklere göre karar verdiği görselleştirilebilir ve kullanıcı güveni artırılabilir.

Ayrıca, veri kümesinin zenginleştirilmesi ve farklı kaynaklardan gelen IoMT trafiğinin modele dâhil edilmesi, genellenebilirliği artıracaktır. Özellikle farklı protokoller, cihaz türleri ve kullanım senaryoları içeren yeni veriler, modelin daha geniş kullanım alanlarında uygulanmasını sağlayacaktır. Öznitelik mühendisliği (feature engineering) alanında yapılacak geliştirmeler, modelin daha ayrıntılı saldırı türlerini ayırt edebilmesini kolaylaştırabilir. Özellikle mevcut özniteliklerin yetersiz kaldığı durumlarda, pcap dosyalarından yeniden ve daha ayrıntılı öznitelikler çıkarılması model başarısını artıracaktır.

Modelin değişen trafik koşullarına ve çevresel değişkenlere uyum sağlayabilmesi için çevrim içi öğrenme (online learning), sürekli öğrenme (continual learning) ve alan uyumlaması (domain adaptation) gibi teknikler de önerilmektedir. IoMT sistemleri dinamik yapılara sahip olduğu için bu tür teknikler modelin zamana bağlı değişikliklere karşı dayanıklı kalmasını sağlayacaktır. Bununla birlikte, geliştirilen modellerin diğer veri kümeleri üzerinde test edilmesi, çapraz veri kümesi genelleme (cross-dataset generalization) konusunda önemli bilgiler sunacaktır. Böylelikle, modellerin farklı sistemlerdeki saldırılara karşı ne derece etkili olduğu değerlendirilebilecektir.

Sonuç olarak, bu tez çalışması, IoMT ağlarında anomali tespiti için geliştirilen klasik makine öğrenmesi ve modern derin öğrenme tabanlı yaklaşımların kapsamlı bir karşılaştırmasını sunmaktadır. CICIoMT2024 veri kümesi üzerinde gerçekleştirilen deneyler sonucunda, özellikle hibrid DL modellerin çok sınıflı görevlerde öne çıktığı gözlemlenmiştir. Dikkat mekanizması, zamana duyarlı ve değişken yapılı trafiklerde karar verme sürecini önemli ölçüde iyileştirmiştir. Ancak bu başarıyı gerçek dünya koşullarına taşıyabilmek için açıklanabilirlik, model küçültme, veri kümesi çeşitlendirme ve adaptif öğrenme alanlarında yapılacak çalışmalar gerekmektedir. Sağlık hizmetlerinin daha fazla dijitalleştiği bu çağda, güvenli, ölçeklenebilir ve akıllı saldırı tespit sistemlerinin önemi her geçen gün artmaktadır. Bu çalışma, gelecekte geliştirilecek güvenlik altyapıları için önemli bir temel teşkil etmektedir.

1. INTRODUCTION

Recent years have witnessed significant advances in the Internet of Things (IoT) with increasingly larger quantities of Internet of Things devices. One of the most useful applications of such an influence may be health, where Internet of Medical Things (IoMT) is quick to become the new norm. IoMT devices make remote diagnosis, chronic disease care and real-time monitoring of health possible in healthcare settings as well as at home in patients [1, 2]. They can also possibly decrease the healthcare network's workload and increase patient care [3]. But since most of the IoMT devices work in unforgiving environments and are susceptible to handling sensitive patient data, this increased connectivity also presents critical security and privacy concerns.

Unlike traditional IT systems, IoMT systems consist of a wide range of devices, protocols and vendors. The systems are not necessarily configured with cybersecurity capabilities and may be difficult to harden or lock down once deployed. Owing to this reality, they become viewed as vulnerabilities in a network [4][5]. They are often targeted by attackers for the intent of interfering with the provision of healthcare or stealing personal information [6]. Security threats like denial-of-service attacks (DoS and DDoS), spoofing, reconnaissance and protocol abuse have been on the rise in IoT-based systems and the healthcare sector is no exception [7].

As the complexity of such systems increases, traditional rule-based security products such as firewalls or signature-based intrusion detection systems are no longer adequate. They lag behind new and unexpected attacks and can produce an overwhelming rate of false alarms [8]. For these limitations, researchers have increasingly resorted to using machine learning (ML) and deep learning (DL) to create more intelligent, adaptive detection systems. These models can learn about real network traffic and identify patterns that suggest unusual activity, even if there has not been a previous incident of this kind of attack [9, 10].

Though the promise of ML and DL is great, they are contingent upon high-quality, realistic data. Most that presently exist for IoT security research are stale, targeted, or unrepresentative of the activity of actual medical devices. The CICIoMT2024 dataset

was created to help address that issue. It contains traffic from 40 devices (25 real and 15 simulated) and combines regular traffic with traffic belonging to 18 different attacks. The attacks range across several categories, including MQTT-based, spoofing, scanning attacks, DoS and DDoS. It was created using a controlled testbed that tried to emulate real-world environments. For example, the researchers used traffic collectors and Faraday cages in order to preserve data integrity [7]. This makes CICIoMT2024 a good platform to test various cybersecurity solutions in medical IoT.

In this paper, we use CICIoMT2024 to explore how well different machine learning and deep learning models can detect and classify attacks on IoMT devices. We compare several well-known models, including Logistic Regression (LR), Random Forest (RF), AdaBoost (AdaBoost) and Deep Neural Network (DNN), alongside more complex deep learning models like 1D Convolutional Neural Networks (Conv1D), Recurrent Neural Networks (RNNs), Long Short-Term Memory networks (LSTM) and Attention mechanisms. Along with these advanced algorithms, hybrid model architectures designed and tested These hybrid models are built to recognize spatial and temporal patterns in network traffic, with the goal of improving the detection of both common and subtle threats.

The dataset supports multiple classification options which range from binary (attack vs. benign) to more detailed schemes such as six main attack classes or up to nineteen sub-classes. Experiments focus on all of the 3 class groups indicating that some forms of attacks are harder to differentiate from one another than others. For example, DoS and DDoS traffic are often almost inseparable from one another, even to advanced models and are regularly mislabeled. Likewise, spoofing and normal traffic can share characteristics in common which confound the classifier. These challenges highlight the need for more targeted experiments in future studies, maybe on individual pairs of attack classes.

To help improve model performance, several data preparation strategies were applied. These included removing duplicate entries and applying sampling techniques to reduce class imbalance. Because some attack types had millions of entries while others had only a few thousand, models tended to overfit to the dominant classes. Balancing the dataset helped reduce this issue and made the results more stable. Layered model architectures combined with thoughtful preprocessing led to better results, especially when backed by clear sampling strategies and class balancing. Model performance was

evaluated using standard metrics such as Accuracy, Precision, Recall and F1-score. The initial dataset was split into 80% training and 20% test sets. 10% of the training set is used for validation. This made the models get tested in a fair and sound way. The major contributions of this research are:

- A detailed comparison of conventional machine learning algorithms and deep learning models using a real-world Internet of Medical Things (IoMT) dataset.
- The development and experimental testing of a range of model architectures.
- An examination of problems in IoMT anomaly detection, such as distinguishing between similar types of attacks.

The structure of this paper is as follows: Section 2 provides an overview of relevant work and practices of IoT and IoMT anomaly detection. Section 3 introduces the CICIoMT2024 dataset and explains preprocessing steps. Section 4 introduces the models and methods of experiments. Section 5 explains experiments and proposes a model architecture. Section 6 reports results and evaluation from experiments. Lastly, Section 7 concludes the paper and proposes potential avenues of future work.

2. LITERATURE REVIEW

In this section, the existing literature on anomaly detection in the Internet of Medical Things domain will be reviewed. In particular, the scope of network-based cybersecurity risks and datasets generated in Internet of Medical Devices domain will be evaluated.

2.1. Internet of Things (IoT) in Healthcare

The Internet of Things (IoT) has significantly changed many industries and healthcare is one of the most impacted. The growing use of connected devices has led to the development of the Internet of Medical Things (IoMT), which refers to medical devices and systems that are connected through the internet or local networks [11]. **Figure 2.1** represents the journey of sensor data within IoMT. These machines are employed to collect information on health, support medical choice and increase the standard of care especially in remote or ongoing patient monitoring [12].

Figure 2.1. Internet of medical things (IoMT) representation.

IoMT includes a wide range of technologies from wearable fitness trackers to hospital-based monitoring systems. Some examples of such devices are smart watches, smart inhalers, pulse oximeters, sensors for blood pressure, heart rate and glucose levels. Such devices can potentially monitor patients' conditions in real-time and notify medical professionals allowing them to respond faster and create more individualized treatment plans [13].

The CICIoMT2024 dataset, which we use in this study, includes network traffic from 40 different IoMT devices—25 real and 15 simulated. These devices represent typical smart healthcare technologies used in hospitals and at home. Among the real devices

in the dataset are the iHealth Smart Gluco-Monitoring System, the EMAY portable ECG monitor and the iHealth Blood Pressure Monitor, which help patients and doctors manage conditions like diabetes, cardiovascular diseases and hypertension [7]. Additionally, they have featured smart home assistants for general health, such as Google Home and Amazon Echo, which are occasionally included in health monitoring systems. The collection contains high-risk, high-value target smart beds for cyberattack in a hospital setting, as well as virtual insulin pumps and ECG monitors.

These networked appliances bring considerable advantages: they reduce the necessity for hospitalization, allow permanent monitoring of one's condition and even make it possible to forecast upcoming significant medical events. For example, an ECG monitor can detect abnormal rhythms beforehand, while a smart blood sugar monitor can alert the patient in case of dangerous levels of blood sugar. During the COVID-19 pandemic era, remote health monitoring usage increased very rapidly, showing the potential for such IoMT devices to deliver care outside of hospitals [14].

Most IoMT devices are low in computational resources and memory, also many are not security design-based. They have outdated software, no encryption, or poor passwords [15]. Once these devices become part of the overall hospital or home network, they are vulnerable to being compromised by attackers. The data they handle, e.g., medical history, vital signs, or personal identity, are highly sensitive and any breach can have serious consequences to the patients as well as providers [16].

In this context, ensuring the security and reliability of IoMT systems is just as important as their medical accuracy. It is no longer enough to rely only on firewalls or basic authentication. As threats become more advanced, healthcare networks must adopt intelligent and adaptive defense systems. This includes anomaly detection models that can spot suspicious activity in real time and flag potential attacks before they cause harm [17].

Overall, IoMT is transforming healthcare delivery by becoming patient-focused, evidence-based and anticipatory. But where there is such revolution, there also arises a growing need to secure the devices and information that make it possible. As the number and variety of IoMT devices grow, future research must focus on formulating

intelligent, dynamic and scalable solutions that can make their use secure in both clinical and home environments [18].

2.2. Security Challenges and Network Threats in IoMT Systems

2.2.1. Security challenges in iomt environment

IoMT systems will consist of highly heterogeneous devices that utilize lightweight communication protocols such as MQTT and CoAP. These protocols are optimized for low power usage and efficiency, occasionally at the expense of built-in security features such as encryption or strong authentication. This implies that a wide range of field-deployed IoMT platforms are relatively easy targets for attackers looking to take advantage of unsecured ports, default passwords or firmware vulnerabilities. [19].

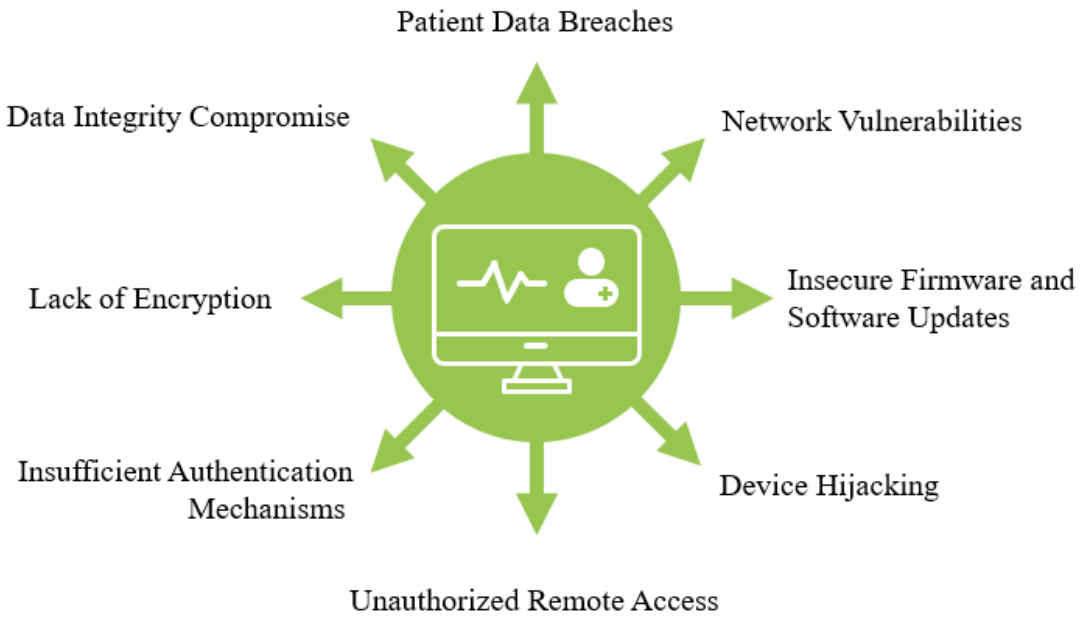

Figure 2.2. Security challenges in IoMT environment.

In hospitals, where device availability is a top priority, even routine security patches are deferred to avoid service downtime. This leaves IoMT devices vulnerable to zero-day attacks and focused exploitation. In addition, vendors' non-standardized security policies render centralized monitoring and patching useless, leading to broken defense profiles [20]. **Figure 2.2** indicates comprehensive security challenges in IoMT environment.

2.2.2. Network-based cyber threats in IoMT traffic

Communication within IoT ecosystems typically occurs between devices, edge gateways and cloud platforms. This communication follows regular patterns and if these patterns change, it could be a sign of suspicious activity. For example, unusual frequency of heartbeat packet sends or greater-than-average payloads which could indicate malicious activity.

There are several types of cyber threats specifically target IoMT network traffic:

- Denial of Service (DoS) and Distributed Denial of Service (DDoS): These include attacks where the target server or device is flooded with unwanted requests which causes denial of real access. In health scenarios, this can disrupt life-critical services [21].
- Spoofing and Replay Attacks: Adversaries impersonate trusted nodes or resend previously captured packets to mislead systems, for example, altering patient monitoring results or issuing false commands [22].
- Port Scanning and Reconnaissance: Attackers perform scans to identify open ports and services running on medical devices. These activities often precede targeted intrusions or malware delivery [19].
- Protocol Abuses and Payload Manipulations: Flaws specific to a protocol are typically exploited. For example, MQTT's lightweight authentication can be circumvented to inject harmful publish-subscribe payloads, which congest the network or lead device states astray [22].

Detecting such threats is difficult using only traditional rule-based security systems, which often fail to identify unknown or subtle threats. As highlighted in recent studies, anomaly detection based on traffic features, such as packet size distribution, inter-arrival time and connection patterns, has become an essential technique for uncovering deviations in network behavior that could point to cyber attacks [23].

2.3. Deep Learning for Anomaly Detection in IoMT Network Traffic

Anomaly detection is the method of identifying abnormal data patterns that are not the same as the normal. In IoMT network traffic, abnormalities can indicate security incidents such as unauthorized access, denial-of-service attacks or exfiltration of data

[24]. Anomaly detection methods are capable of identifying unknown or novel threats in contrast to traditional rule-based intrusion detection systems since they have the capability to learn and identify network behavior patterns over time [25].

Recent advances in deep learning have significantly enhanced the performance of anomaly detection systems in IoMT settings. These models are able to learn to extract hierarchical features from raw data automatically and this enables more accurate and scalable detection of network intrusions [26, 27].

Some of the most common deep learning architectures being exploited for IoMT anomaly detection are as follows:

- Convolutional Neural Networks (CNNs) are generally used to learn spatial relationships among traffic matrices [28].
- Recurrent Neural Networks (RNNs) and LSTM networks are generally used to learn sequential dependencies in traffic flows [29].
- Autoencoders are utilized for learning compressed representations of normal traffic and identifying deviations from these through reconstruction error [8].
- Transformers are utilized to learn multi-scale and long-range dependencies in network pattern tendencies [30].

For instance, [31] introduced FusionNet, an ensemble model combining Random Forest, K-Nearest Neighbors, Support Vector Machine, and Multi-Layer Perceptron. The study found that FusionNet outperformed traditional single-model approaches, achieving up to 99.5% accuracy, demonstrating how combining diverse ML algorithms enhances anomaly detection in IoT environments.

Similarly, [32] proposed a multi-model autoencoder + sequence-based residual learning framework to detect anomalies in IoMT-blockchain environments. This hybrid approach demonstrated strong performance in identifying complex abnormal traffic in secure medical IoT systems.

In one of their recent studies, [33] applied attention-driven deep learning model combining CNNs with self-attention to improve anomaly detection in IoT network traffic. This approach effectively captures important features and dependencies, resulting in more accurate identification of anomalies compared to traditional methods.

Furthermore, [34] proposed a hybrid swarm-neural architecture for healthcare IoMT intrusion detection. In integrating swarm intelligence with deep learning, where PSO

is utilized in feature selection and autoencoders for representation, the system achieved competitive accuracy, demonstrating the potential of hybrid AI approaches in securing low-resources medical networks.

Lastly, [35] introduced HIDS-IoMT, a Host-based Intrusion Detection System tailored for medical IoT environments. A hybrid CNN–LSTM model for edge-level intrusion detection in IoMT environments. Deployed on Raspberry Pi within a fog architecture and tested on IoTID20 and Edge-IIoTset datasets, the system achieved 99.92% accuracy, 99.91% precision, 99.99% recall and 99.95% F1-score, demonstrating great performance in detecting DDoS and other attacks.

These developments highlight the growing reliability of deep learning as a foundation technology for securing IoMT infrastructures. With the rising threats and rising complexity in data, research demands the maximization of interpretability of models, reducing computational loads and making the architectures compatible with the constraints of actual healthcare systems.

2.3.1. Dataset challenges and benchmarking standards

A persistent bottleneck in the advancement of deep learning models for IoMT anomaly detection is the availability and adequacy of benchmark datasets. IoMT traffic datasets must not only reflect the real-world behavior of medical devices but also include a variety of attack types, device categories and communication protocols. However, most available datasets fail to meet all these criteria simultaneously. They are either generated in controlled environments or lack detailed attack labels, leading to models that perform well in theory but underperform in clinical settings.

One of the principal limitations in current datasets is the lack of heterogeneity. Most IoMT environments involve a mix of devices from wearables and bedside monitors to smart beds, each with its own traffic signatures. Datasets that only cover a small subset of this range lack the diversity to support training generalizable models. This was discussed critically in [7], where it was mentioned that some existing datasets, such as IoT-23 and Bot-IoT, are not suitable for multi-protocol analysis or healthcare-specific network behavior.

Furthermore, protocol diversity and data realism are essential. IoT healthcare ecosystems utilize protocols such as MQTT, CoAP, HTTP and proprietary formats, even simultaneously. A dataset with poor protocol diversity can train a model to detect

anomalies in MQTT traffic but make it blind to threats sent over HTTP or BLE. The CICIoMT2024 dataset directly addresses this challenge with multi-protocol traffic from 40 real and emulated IoMT devices, including high-target-value hospital devices such as insulin pumps and ECG monitors.

A second root challenge is data annotation and labeling. Manual labeling is time-consuming and can be error-prone, especially for stealthy or subtle attacks such as spoofing or low-rate DDoS. Mislabeling will distort a model's performance or cause overfitting. The majority of public datasets also involve binary labeling (attack or normal), a reduction of heterogeneity in threat behavior. [36] bypass this limitation in their publication of IoMT traffic data where multi-class labeling was employed to more readily support heterogeneous attack vectors.

Benchmarking conventions are also shattered and non-uniform. No standard test procedure exists for testing intrusion detection systems on IoMT datasets. Different train-test splits, feature engineering processes and preprocessing pipelines employed by researchers make it difficult to compare studies. Some use exclusively synthetic traffic, while others employ real patient data, which has privacy and model performance implications. [37] affirm that there is an urgent need for standard evaluation measures (e.g., F1-score, AUC, latency) and benchmarking platforms that will enable reproducible and comparable studies.

2.3.1.1. Comparative overview of prominent iomt datasets

The following is a compiled comparison table (**Table 2.1**) of some of the most well-known and applicable IoMT traffic datasets, with relevant attributes for anomaly detection research highlighted.

Table 2.1. Recent datasets of internet of medical devices (IoMT).

Name of Dataset	Year	Attack Types
CICIoMT2024 [7]	2024	ARP spoofing, Ping Sweep, Recon VulScan, OS Scan, Port Scan, MQTT Malformed Data, MQTT DoS Connect flood, MQTT DoS Publish flood, MQTT DDoS Connect flood, MQTT DDoS Publish flood, DoS TCP, DoS ICMP, DoS SYN, DoS UDP, DDoS TCP, DDoS ICMP, DDoS SYN, DDoS UDP
BlueTack [38]	2022	DDoS, DoS, Bluesmack and MITM
IEC [39]	2021	Traffic Sniffing, DoS, Unauthorized Access and MITM
ICU [40]	2021	DDoS MQTT, MQTT Publish Flood, SlowITE and Brute Force
ECU-IoHT [41]	2021	Script Injection, ARP Spoofing, Smurf, DoS and Network Scan
WUSTL EHMS [42]	2020	Spoofing and Data Alteration

3. DATASET DESCRIPTION

To support anomaly detection in healthcare-focused Internet of Things (IoT) environments, this study utilizes the CICIoMT2024 dataset—one of the most recent and comprehensive public benchmarks tailored to the Internet of Medical Things (IoMT) domain. Developed by the Canadian Institute for Cybersecurity at the University of New Brunswick, this dataset is specifically designed to emulate the network behavior of diverse medical IoT devices under both normal and adversarial conditions.

3.1. CICIoMT2024 Dataset Overview

The CICIoMT2024 dataset supports three levels of granularity to classify in order to quantify the performance of anomaly detection systems: binary classification, 6-class classification and 19-class fine-grained classification. The multi-level structure allows researchers to design general-purpose detectors as well as very domain-specific classifiers for IoMT security. The distribution of each class group is illustrated in **Table 3.1**.

CICIoMT2024 dataset is provided as a training set (80%) and a test set (20%), each comprising both packet capture (PCAP) files and pre-processed CSV files. Both the PCAP and CSV files are organized according to network protocols (e.g., TCP, UDP, MQTT), packet-level flags (e.g., SYN) and specific attack types (e.g., DDoS, spoofing, reconnaissance).

Even though the CICIoMT2024 dataset provides rich coverage to a wide range of attacks across a wide range of classification levels, it does have some inherent class distribution imbalance-related issues. In the case of binary classification, attack samples are far more in number compared to benign samples and may potentially induce model bias toward the majority class when model training occurs. This pattern continues in the 6-class classification also where DoS and DDoS classes constitute the majority of the dataset and classes such as MQTT-based attacks, reconnaissance, and spoofing attacks appear with comparatively lower frequency.

The detailed attributes and their corresponding descriptions used in this study are presented in **Table 3.2**, which outlines the features extracted from the CICIoMT2024 dataset.

Table 3.1. Structure of CICIoMT2024 dataset according to classes.

Binary Classification	6 Class Classification	19 Class Classfication
Benign	Benign	Benign
Attack	DDoS	SYN
		TCP
		ICMP
		UDP
	DoS	SYN
		TCP
		ICMP
		UDP
	MQTT	DDoS Connect Flood
		DDoS Publish Flood
		DoS Connect Flood
		DDoS Publish Flood
		Malformed Data
	Recon	OS Scan
		Ping Sweep
		Port Scan
		VulScan
	Spoofing	ARP

This gets even more pronounced in the 19-class fine-grained classification where some high-frequency attack types are strongly represented, but others are comparatively underrepresented. These inherent variations in data distribution present the importance of adopting mitigation strategies such as class weighting, resampling or synthetic data augmentation. Addressing these issues with caution can help in improving model generalization and efficient detection of common as well as rare threats in IoMT

environments. **Figure 3.1**, **Figure 3.2** and **Figure 3.3** are show class distributions for the 2-class, 6-class, and 19-class classification tasks, respectively.

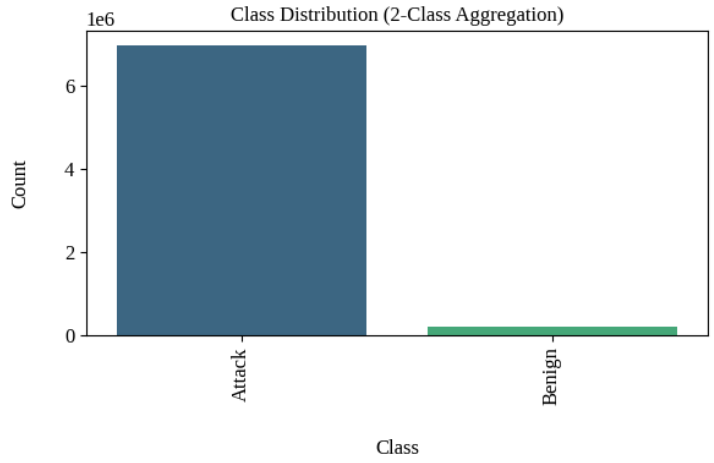

Figure 3.1. Class distribution (2-class aggregation).

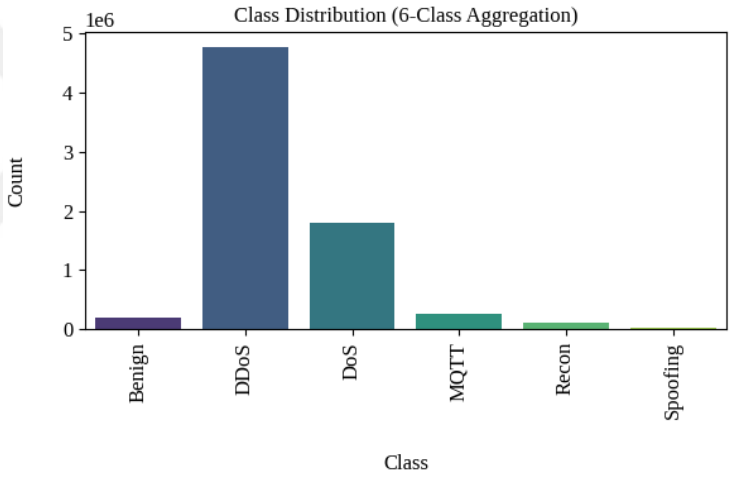

Figure 3.2. Class distribution (6-class aggregation).

a

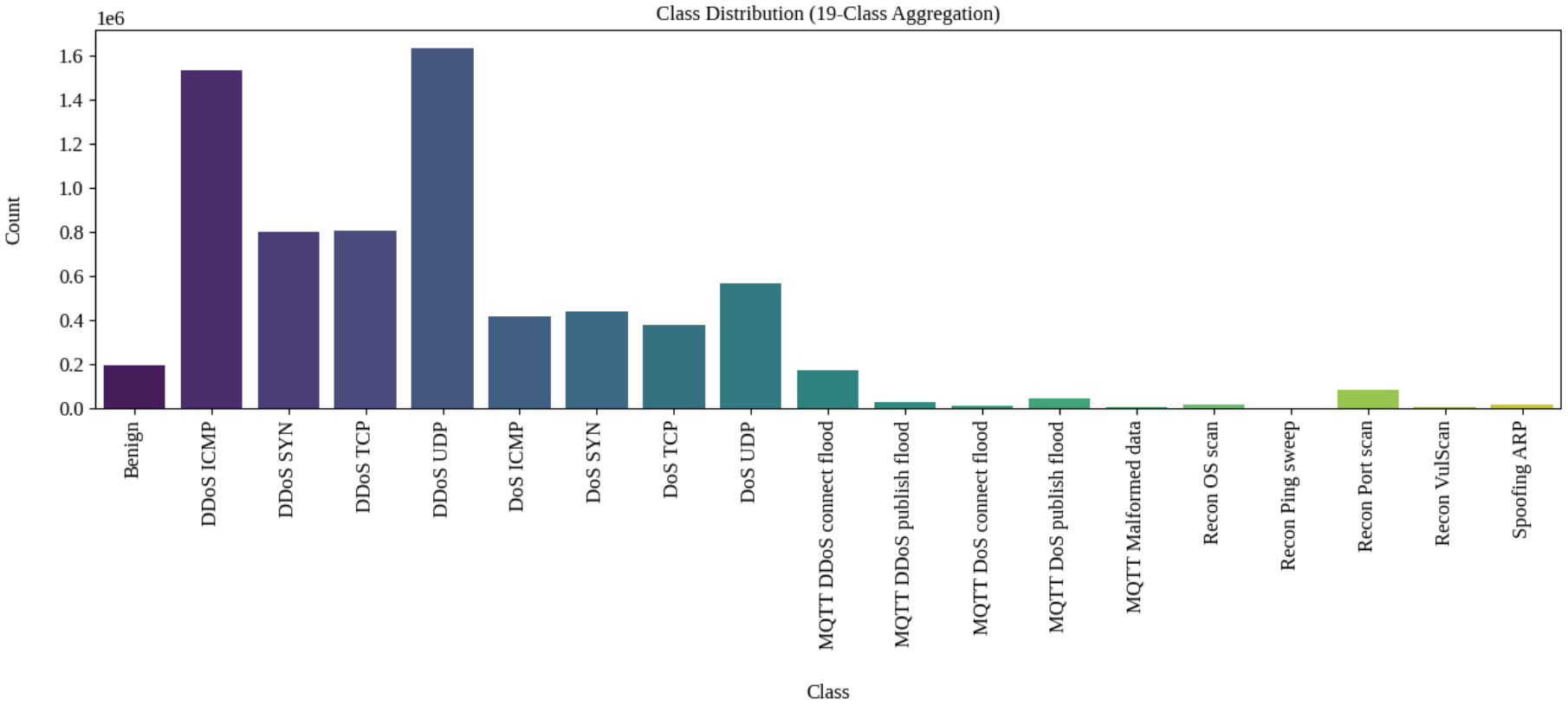

Figure 3.3. Class distribution (19-class aggregation).

Table 3.2. Features of CICIoMT2024 dataset.

Index	Feature	Description
1	Header Length	Mean of the header lengths
2	Protocol Type	Indicates the type of network protocol
3	Duration	Time duration of a network connection or flow
4	Rate	Overall data transfer rate within a connection
5	Srate	Packet sending rate from the source host
6	Drate	Packet receiving rate at the destination host
7	FIN Flag Number	Number of packets with the FIN flag set
8	SYN Flag Number	Number of packets with the SYN flag set
9	RST Flag Number	Number of packets with the RST flag set
10	PSH Flag Number	Number of packets with the PSH flag set
11	ACK Flag Number	Number of packets with the ACK flag set
12	ECE Flag Number	Number of packets with the ECE flag set
13	CWR Flag Number	Number of packets with the CWR flag set
14	ACK Count	Total count of acknowledgment packets
15	SYN Count	Total count of packets initiating connections
16	FIN Count	Total count of packets terminating connections
17	RST Count	Total count of connection reset packets
18	HTTP	Indicator or count of HTTP protocol packets
19	HTTPS	Indicator or count of HTTPS protocol packets
20	DNS	Indicator or count of DNS traffic
21	Telnet	Indicator or count of Telnet protocol packets
22	SMTP	Indicator or count of SMTP protocol packets
23	SSH	Indicator or count of SSH protocol packets
24	IRC	Indicator or count of IRC protocol packets
25	TCP	Indicator or count of TCP protocol traffic
26	UDP	Indicator or count of UDP protocol traffic
27	DHCP	Indicator or count of DHCP protocol packets
28	ARP	Indicator or count of ARP protocol packets
29	ICMP	Indicator or count of ICMP protocol packets
30	IGMP	Indicator or count of IGMP protocol packets
31	IPv	Indicator or count of IP versioned traffic
32	LLC	Indicator or count of LLC protocol frames
33	Total Sum	Sum of packet sizes in a connection
34	Min	Minimum packet size observed in the flow
35	Max	Maximum packet size observed in the flow
36	Average	Average packet size in the connection
37	Standard Deviation	Standard deviation of packet sizes
38	Total Size	Aggregate size of all packets in a flow
39	IAT	Inter-arrival time between packets
40	Number	Total number of packets or frames
41	Magnitude	Computed measure representing packet strength or impact
42	Radius	Statistical radius from a central feature metric
43	Covariance	Measure of how two flow features vary together
44	Variance	Dispersion of a feature around the mean
45	Weight	Weight score, possibly derived from traffic intensity or frequency

3.2. Preprocessing

Since each file is linked with a specific type of attack or benign traffic, this structure was leveraged to label samples into three schemes of classification: binary (attack or benign), 6-class (e.g., DDoS, DoS, MQTT) and 19-class (fine-grained attack types). Following labeling, the individual files were concatenated to form two large datasets, one for training and another for testing.

Before feeding the data into the machine learning algorithms, several preprocessing steps were applied. Initial experiments included the use of the SMOTE (Synthetic Minority Over-sampling Technique) algorithm. However, SMOTE did not lead to significant improvements in classification performance as it did not improve the metrics, and therefore, undersampling was retained as the primary balancing method. Class imbalance was addressed by controlled downsampling of the too dominant classes, thus ensuring higher balanced exposure in training without losing representational diversity. After downsampling, train-test ratios varied between 76%-85% and 15%-24% respectively. Exact number of samples per class can be seen in **Table 3.3**.

Secondly, feature value scaling was done by utilizing standard normalization techniques so that equal uniformity to the input space is provided, which is a major concern for gradient-based learning algorithms.

Additionally, feature reduction was employed to eliminate irrelevant or redundant input features that tend to hinder learning. The variance threshold method was employed to eliminate features with variance lower than a specified threshold. The idea behind this is that features having very small variance are nearly always constant across samples and hence offer little discrimination capability [43]. Mathematically, the variance of a feature X across n samples is given by

$$Var(X) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2 \tag{3.1}$$

where μ is the mean of feature X and x_i represents each sample value. Features with variance below the threshold (empirically determined and set to 0.01) were discarded.

As a result, the feature set was reduced from 45 to 30 dimensions, significantly improving model efficiency without sacrificing performance.

Table 3.3. Number of data samples after downsampling.

			Number of Samples	
	Class Name		Train	Test
2 Class	Benign		192732	37697
	Attack		300000	60000
		Total	492732	97697
6 Class	Benign		27000	5000
	DDoS		27000	5000
	DoS		27000	5000
	MQTT		27000	5000
	Recon		27000	5000
	Spoofing		16047	1744
		Total	151047	26744
19 Class	Benign		10000	3200
	DDoS ICMP		10000	3200
	DDoS SYN		10000	3200
	DDoS TCP		10000	3200
	DDoS UDP		10000	3200
	DoS ICMP		10000	3200
	DoS SYN		10000	3200
	DoS TCP		10000	3200
	DoS UDP		10000	3200
	MQTT DDoS Connect Flood		10000	3200
	MQTT DDoS Publish Flood		10000	3200
	MQTT DoS Publish Flood		10000	3200
	Recon OS Scan		10000	3200
	Recon Port Scan		10000	3200
	MQTT DoS Connect Flood		10000	3131
	Spoofing ARP		10000	1744
	MQTT Malformed Data		5130	1747
	Recon VulScan		2173	1034
	Recon Ping Sweep		740	186
		Total	168043	52642

4. METHODOLOGY

This section describes the modelling methodologies used to identify IoMT traffic anomalies using the CICIoMT2024 dataset. Our approach is structured in three tiers: standard machine learning classifiers, deep models and presented hybrid models integrating convolutional, recurrent and attention-based layers. This three-tier evolution facilitates us to contrast performance, efficacy and genericity among different model classes and levels of complexity. Both model types are trained on the same preprocessed data and evaluated on the same metrics to ensure fairness. We also employ well-defined classification metrics such as accuracy, precision, recall and F1-score to extensively measure the effectiveness of each approach to detect frequent as well as rare network threats.

4.1. Baseline Machine Learning Models

In order to establish baseline reference points for IoMT network traffic anomaly detection, we employed a set of conventional machine learning algorithms known to do well on structured tabular data. The models are interpretable, lightweight references for quantifying the performance improvements delivered by deep and hybrid architectures. Each algorithm was implemented using consistent feature preprocessing, class encoding and evaluation protocols across binary, 6-class and 19-class classification tasks.

4.1.1. Logistic regression (LR)

Logistic Regression (LR) is a linear classifier and can be used for both binary and multiclass problems, particularly when the input feature-target relationship is almost linear. Its architecture is illustrated in **Figure 4.1**. As it is simple, interpretable and has less computation involved, it is a great candidate to attempt for the first time in security datasets [44] [45]. However, it may not perform well with non-linear decision boundaries and complex feature interactions common in network intrusion data [46].

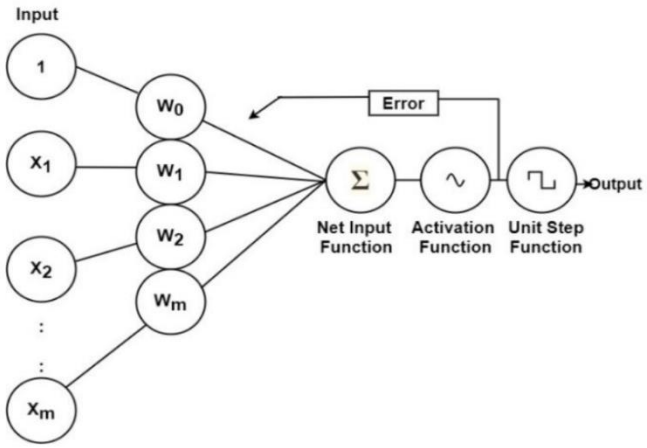

Figure 4.1. Architecture of logistic regression [47].

4.1.2. Random forest

Random Forest (RF) is a type of ensemble learning method that builds a large number of decision trees, as shown in **Figure 4.2**, and aggregates their predictions for improved generalization and stability. RF is particularly well-prepared for high-dimensional and imbalanced settings since it is able to detect non-linear feature interactions and overfitting-free [48]. It has been extensively used in intrusion detection systems (IDS) due to its great performance in various types of network attacks [49].

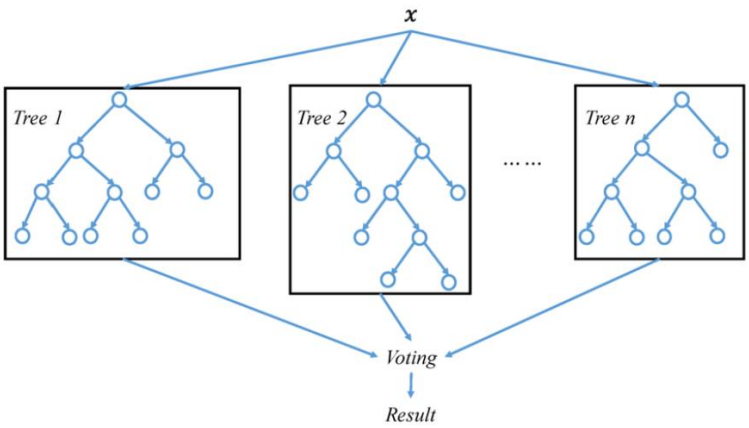

Figure 4.2. Architecture of random forest [50].

4.1.3. Adaptive boosting (AdaBoost)

Adaptive Boosting (AdaBoost) is another ensemble method that iteratively combines weak learners, typically decision stumps, to build a strong classifier by giving more importance to misclassified points in each round. AdaBoost has been found effective to detect subtle anomalies and minority class patterns, which are crucial in the IoMT

domain [51]. But it is likely to be sensitive to outliers and noise, which is one such factor in real network traffic data.

4.1.4. Fully connected dense network

A fully connected dense network is a feedforward neural model where a neuron from one layer is connected to all neurons of a subsequent layer. These architectures can model rich non-linear interactions among input features using layered transformations by using non-linear activation functions. Dense networks perform extremely well in high-dimensional tabular data environments where feature interdependencies matter and are difficult to model using linear methods [52].

4.2. Deep Learning Models

We explored a suite of deep learning architectures to effectively model the dataset's nature and feature correlations of IoMT network traffic data. These models are particularly adept at learning complex temporal dependencies and latent representations, which are critical for detecting subtle and evolving attack patterns.

4.2.1. 1D convolutional neural networks (1D-CNN)

One-dimensional convolutional neural networks (1D-CNNs) are designed to learn local feature patterns in sequential data with sliding filters. An example of CNN mechanism is illustrated in **Figure 4.3.** 1D-CNNs are effective discriminative spatial representation learners for time-ordered input and are particularly suited for traffic-based intrusion detection use cases. Earlier research demonstrated 1D-CNNs are effective for classifying encrypted or obfuscated traffic [53].

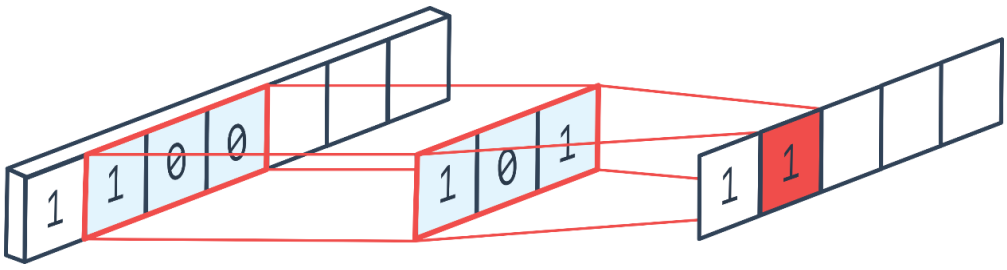

Figure 4.3. Representation of 1D-CNN [54].

4.2.2. Recurrent neural networks (RNN)

RNNs are a category of neural networks that operate on input sequences sequentially with the capability of remembering previous information in their hidden state. Each RNN cell gets an input at the current time step along with previous step's hidden state, which allows it to remember contextual information over time, such a recurrent setup is shown in **Figure 4.4**. RNNs are applied particularly to modeling time-varying relations but regular RNNs in general fail to handle long dependencies due to vanishing gradients [55]. To address this, modifications like LSTM and GRU came into being that incorporate the mechanisms of gates to retain and forget information longer across longer sequences.

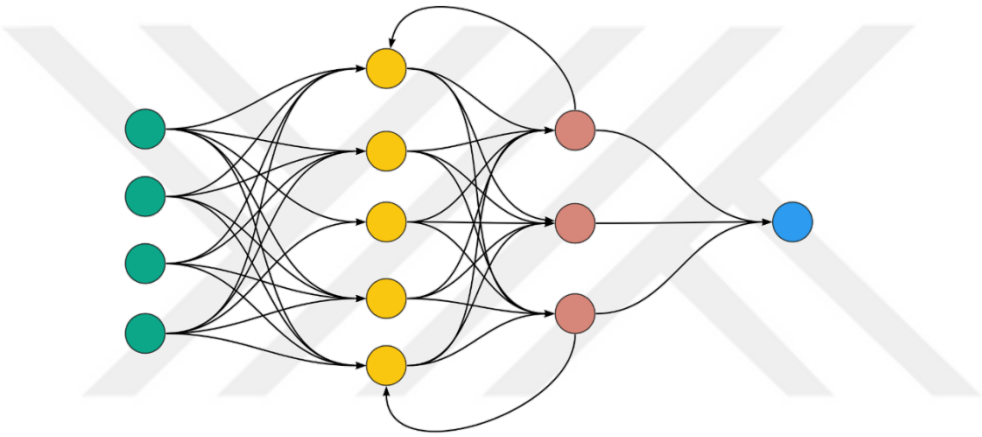

Figure 4.4. Representation of RNN [56].

RNNs have been applied in sequence modeling of behavior and identification of differences across time, notably in system logs and traffic streams in anomaly detection. They are a potential pick for identifying minor changes in sequential data patterns correlated with malicious activity because they support temporal dependencies.

4.2.3. Long short-term memory networks (LSTM)

LSTMs are a highly advanced type of RNN architecture with gated structures for regulating information flow. The general structure of an LSTM cell contains three gates (input gate, forget gate and output gate) that regulate how information is added to, removed from, or permitted to flow out of the cell; **Figure 4.5** illustrates this design. This allows them to recall important information in long sequences and forget less important parts which cancels out the vanilla RNN disadvantages [57]. With a more

long-term memory state across time steps, LSTMs can learn to capture longer-term dependencies in sequence data.

LSTM networks are optimally used in problems with multiple temporal evolution understanding tasks, such as differentiating between bursty normal traffic and persistent anomalous activity [58]. Their capacity to capture both short- and long-term patterns makes them particularly effective in detecting complex behaviors in cybersecurity scenarios.

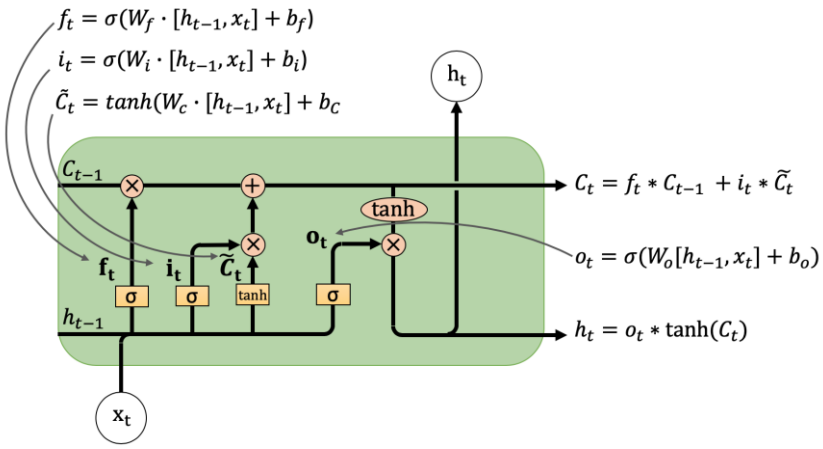

Figure 4.5. Representation of LSTM [59].

4.2.4. Self-attention mechanism

The self-attention mechanism, first introduced by Vaswani et al. in 2017 [60], is the core of the Transformer model, revolutionizing deep learning for computer vision and natural language processing. The mechanism enables models to give relative importance to elements within a sequence of inputs, enabling a richer contextual understanding.

Self-attention brings direct interaction between all elements in a sequence, unlike standard recurrent or convolutional networks. The internal mechanism of self-attention is illustrated in **Figure 4.6**. This inherent parallelizability significantly improves training and inference time computational efficiency, together with consistently capturing long-range dependencies, which is a persistent deficiency of past models. With dynamic token influence weighting based on semantic relations, self-attention generates extremely expressive, context-aware representations with state-of-the-art performance on a variety of tasks.

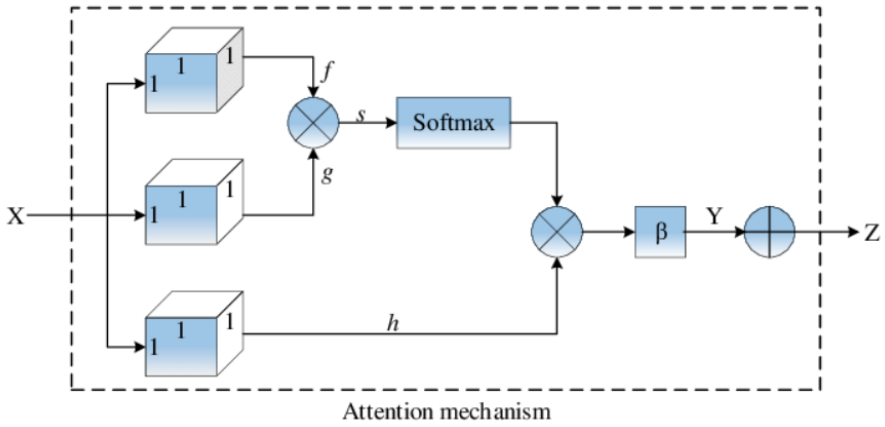

Figure 4.6. Structure of self-attention mechanism [61].

At its core, self-attention computes Query (Q), Key (K), and Value (V) matrices from an input embedding sequence $X \in R^{n \times dmodel}$, where n is sequence length and d_{model} is embedding dimension. These are derived via linear transformations:

$$Q = XW^{Q} \tag{4.1}$$

$$K = XW^{K} \tag{4.2}$$

$$V = XW^{V} \tag{4.3}$$

Here, W^{Q}, W^{K}, and W^{V} are learned weight matrices. Specifically, W^{Q} and W^{K} have dimensions $d_{model} \times d_k$, while W^{V} has dimensions $d_{model} \times d_v$. Attention scores are then computed as the scaled dot product between Q and K, normalized by softmax to yield attention weights:

$$Attention(Q, K, V) = softmax\left(\frac{QK^{T}}{\sqrt{d_k}}\right)V \tag{4.4}$$

The scaling factor d_k (dimension of key vectors) stabilizes gradients. This process produces an output that is a weighted sum of Value vectors, with weights reflecting semantic similarity, enabling each element to attend to relevant information across the sequence.

4.3. Evaluation Metrics

Evaluating the performance of anomaly detection models, especially when handling highly imbalanced datasets like CICIoMT2024, requires more than the simple accuracy measure. In this research work, we adhere to four widely used classification performance measures: Accuracy, Precision, Recall and F1-Score, each providing distinct insights into model performance under different error conditions

4.3.1. Accuracy

One definition of accuracy is the ratio of accurately predicted samples to all samples:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{4.5}$$

While commonly used, accuracy can be misleading in imbalanced datasets because it may reflect a high value even if the model performs poorly on minority classes [62].

4.3.2. Precision

The percentage of genuine positives among all predicted positives is known as precision. It reflects how many selected items are relevant:

$$Precision = \frac{TP}{TP + FP} \tag{4.6}$$

High precision is crucial in intrusion detection tasks to minimize false alarms, which can lead to alert fatigue in operational systems [63].

4.3.3. Recall

The percentage of true positives that the model accurately detects is measured by recall:

$$Recall = \frac{TP}{TP + FN} \tag{4.7}$$

High recall is particularly important in security scenarios to ensure that malicious behaviors are not overlooked [64].

4.3.4. F1-score

A single metric that balances the trade-off between precision and recall is the F1-score, which is the harmonic mean of the two:

$$F1 - Score = 2\,x\,\frac{Precision\,x\,Recall}{Precision + Recall} \tag{4.8}$$

This metric is highly informative when dealing with uneven class distributions, as it penalizes extreme imbalances between precision and recall [65].

5. EXPERIMENTAL SETUP

5.1. Hyperparameter Tuning

Several hybrid deep learning models were created to increase detection capacity and take the advantage of both local and temporal patterns in IoMT traffic. The models were designed to capitalize on the strengths of convolutional, recurrent and attention-based models to offer robust and scalable anomaly detection.

To optimize these hybrid models, multiple hyperparameter configurations were systematically explored and evaluated during the experimental phase to determine the best configuration for identifying anomalous activity in IoMT network traffic.

The development of the models was carried out incrementally and to determine the depth of best models, one to four stacked layers were experimented with in control. The isolated 1-D CNN and RNN models were initially tested with variable numbers of hidden units (32 to 512). They were followed by adding a single LSTM layer with 32 to 128 units for long temporal dependency capturing. Attention mechanism was eventually introduced to the hybrid models.

All the models were trained using the Adam optimizer with the learning rate from 0.001 to 0.01. All configurations of the models were trained for a maximum of 200 epochs with early stopping used to prevent overfitting. Batch sizes between 32 and 512 and dropout rates between 0.3 and 0.5 were tested. To calculate loss, Sparse Categorical Crossentropy and Sparse Categorical Focal Loss functions tested.

Among those models that were experimented with, one hybrid model using a Simple Recurrent Neural Network (SimpleRNN) layer incorporating Long Short-Term Memory (LSTM) units with an Attention mechanism consistently outperformed hybrid alternatives and it is followed by another hybrid model consisting of a 1-D Convolutional Neural Network (1-D CNN) layer, Long Short-Term Memory (LSTM) units and a Simple Recurrent Neural Network (SimpleRNN) layer. After experiments, selected hyperparameters for these two models can be seen in **Table 5.1** and **Table 5.2** respectively.

Table 5.1. Hyperparameters of RNN + LSTM + Self-Attention model.

Hyperparameters	SimpleRNN	LSTM	Self-Attention
Units	64	64	Uses LSTM output shape
Activation	Default (tanh)	Default (tanh)	N/A (dot-product mechanism)
Dropout Rate	0.3	0.3	0.3
Optimizer	Adam		
Learning Rate	0.001		
Batch Size	128		
Epochs	50		
Loss Function	Sparse Categorical Crossentropy		

Table 5.2. Hyperparameters of 1-D CNN + LSTM + RNN model.

Hyperparameter	Conv1D	LSTM	SimpleRNN
Units / Filters	64 filters	64 units	32 units
Kernel Size	3	N/A	N/A
Activation	ReLU	Default (tanh)	Default (tanh)
Padding	Same	N/A	N/A
Dropout Rate	0.3	0.3	0.3
Optimizer	Adam		
Learning Rate	0.001		
Batch Size	64		
Epochs	50		
Loss Function	Sparse Categorical Crossentropy		

5.2. Proposed Model Architecture

To effectively capture both the temporal dependencies and contextual patterns in sequential IoT traffic data, we propose a hybrid deep learning model combining a Simple Recurrent Neural Network (SimpleRNN), Long Short-Term Memory (LSTM)

units and a self-attention mechanism, which illustrated in **Figure 5.1**. The difficulties of conventional RNNs in learning long-term dependencies are addressed by this architecture, which also incorporates selective attention to temporal patterns of significance.

The input is first passed through a SimpleRNN with 64 hidden units. SimpleRNN units are capable of learning short-term dependencies by maintaining a hidden state across time steps, which allows them to handle sequential temporal signals. Because they do not retain long-range information very well, they are typically used for light feature extraction or as the first temporal encoders.

The second layer, a 64-unit Long Short-Term Memory (LSTM) layer, gets around this restriction and makes it possible to learn long-term dependencies. With the employment of memory cells and gate units (input, forget and output gates) LSTM design improves the network's capacity for learning. These gates control the information flow such that the network can recall or forget internal states between time steps, making it ideal for intricate time-series modelling. The dropout after the recurrent layers is used as a regularizer to combat overfitting.

A self-attention mechanism is included through the Attention layer following the LSTM layer, which calculates attention scores between each time step in the LSTM output sequence. By learning a weighted sum of features, this method enables the model to give distinct time steps in the series varied levels of priority, which improves the model's capacity to concentrate on important temporal periods rather than treating all prior data equally. This attention matrix is applied symmetrically (i.e., attention over itself), a technique known as self-attention.

Raw LSTM output and attention-enhanced representation are concatenated, enriching the feature space with both learned temporal dynamics and selectively emphasized signal segments. To reduce the dimension and pool information across time steps, we employ a GlobalAveragePooling1D layer, which computes the mean over all time steps for every feature, generating a fixed-size vector. This pooled representation is then Layer Normalized to improve training stability and convergence by reducing internal covariate shift. A dropout layer is used afterwards to regularize the representation, followed by a Dense output layer with softmax activation, which outputs class probabilities over the number of output categories specified.

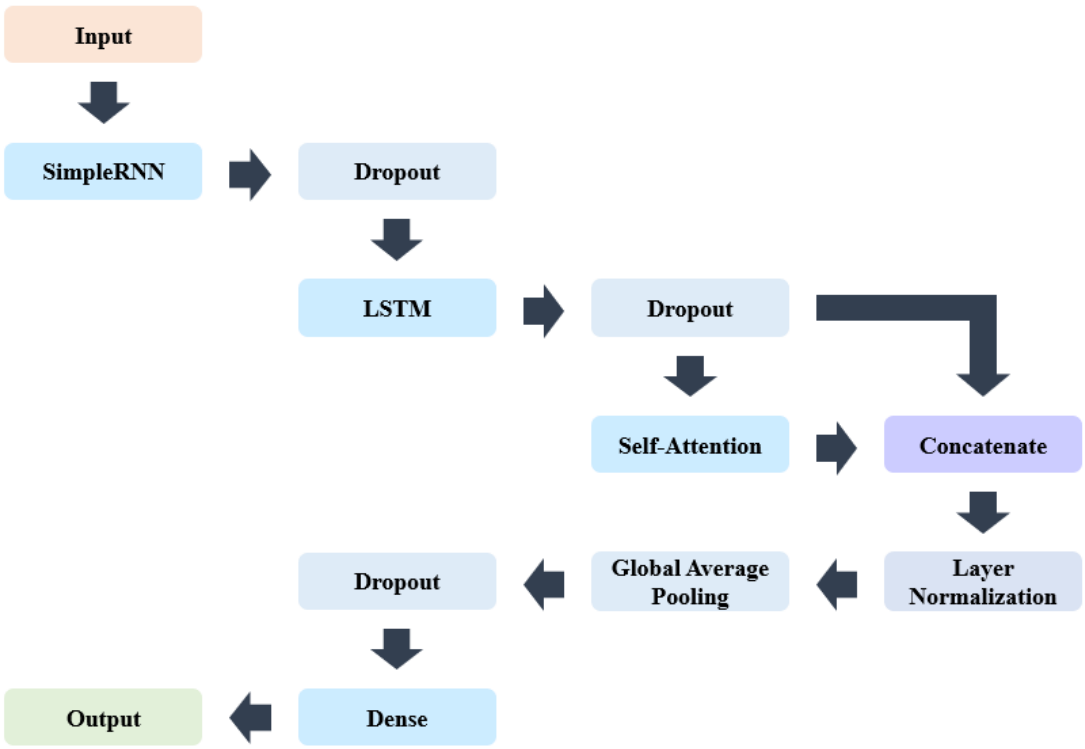

Figure 5.1. Proposed model architecture.

This architecture of the model, suggested herein, not only learns the hierarchical temporal pattern in input sequences but also dynamically pays attention to relevant information, thereby being very much suitable for anomaly detection tasks in noisy and complex IoMT environments.

5.3. Hardware and Frameworks

The experiments of this study were performed on the basis of cloud computing platforms, that is, Google Colab and Kaggle Kernels. Both these platforms provide access to GPU resources that are required to train deep learning models efficiently. Google Colab gave access to NVIDIA Tesla T4 and P100 GPUs, whereas Kaggle provided access to NVIDIA Tesla P100 GPUs. Both the platforms provided the possibility of running large-sized datasets and complex model structures without any need for local specific hardware.

On the software end, the project utilized the TensorFlow 2.x library, which used its Keras API to facilitate quick development and experimentation with neural network models. Traditional machine learning algorithms and data preprocessing were

implemented using popular Python libraries such as Pandas, NumPy and scikit-learn. This ensured effective data management and stable model testing.

Use of these cloud systems allowed flexible, scalable experimentation over homogeneous environments, allowing reproducibility while minimizing hardware constraints normally encountered in local environments.

6. RESULTS AND DISCUSSION

This section presents the performance of various baseline and proposed models across three classification settings: binary (2-class), medium-granularity (6-class) and fine-grained (19-class) target scenarios. The evaluation metrics used include accuracy, recall, precision and F1-score. **Table 6.1** presents the evaluation metrics for both the baseline and deep models.

6.1. Baseline Model Performance

Among the classical machine learning models, Random Forest (RF) consistently performed better than Logistic Regression (LR) and AdaBoost for all classification levels. In binary classification, RF yielded a great accuracy of 98.00%, which was considerably greater than LR (95.45% accuracy) and AdaBoost (97.63% accuracy). This pattern was even stronger for the 6-class and 19-class tasks, where RF scored 97.43% accuracy (6-class) and 93.38% accuracy (19-class), significantly outperforming the other classical methods.

In contrast, AdaBoost performed well on the 6-class scenario (72.60% accuracy) but was not stable for the 19-class task, where its accuracy dropped to 26.64%.

6.2. Deep Model Comparison

Deep learning models demonstrated more stable performance across tasks, with increasing benefits observed in higher-class scenarios. The simple deep neural network (DNN) and Conv1D + Dense models yielded moderate results in the 6-class task (75.80% and 78.35% accuracy, respectively) but lagged behind more complex architectures in the 19-class setting.

Notably, the hybrid model of Conv1D + LSTM + RNN outperformed all others in the binary case with a perfect accuracy of 99.74% and remained strong in the 6-class (80.50% accuracy) and 19-class (63.84% accuracy) cases. And the RNN + LSTM + Attention model was the most stable performer on all the tasks and had the highest

accuracy for the deep models in the 19-class task (66.40% accuracy) and had fairly balanced performance in 6-class (82.01% accuracy) and binary classification (99.23% accuracy). This indicates that attention improved the ability of the model to differentiate between fine-grained patterns in multi-class tasks**.**

Table 6.1. Comparison of model evaluation scores of baseline and deep models.

		Baseline Models			Deep Models				
		Logistic Regression	Random Forest	AdaBoost	DNN	Conv1D + Dense	RNN + Dense	RNN + LSTM	Conv1D + LSTM + RNN	RNN + LSTM + Attention
2 Classes	Accuracy	0.9545	0.9800	0.9767	0.9564	0.9555	0.9481	0.9509	**0.9974**	0.9923
	Recall	0.9545	0.9800	0.9767	0.9564	0.9555	0.9481	0.9509	0.9974	0.9923
	Presicion	0.9538	0.9800	0.9766	0.9563	0.9548	0.9496	0.9514	0.9974	0.9923
	F1-Score	0.9538	0.9798	0.9763	0.9549	0.9546	0.9452	0.9486	0.9974	0.9923
6 Classes	Accuracy	0.7705	0.9736	0.7260	0.7580	0.7835	0.7718	0.7872	0.8050	**0.8201**
	Recall	0.7705	0.9736	0.7260	0.7580	0.7835	0.7718	0.7872	0.8050	0.8201
	Presicion	0.7782	0.9761	0.7910	0.8122	0.8310	0.7846	0.8105	0.8127	0.8469
	F1-Score	0.7686	0.9743	0.7299	0.7139	0.7566	0.7659	0.7929	0.8074	0.8140
19 Classes	Accuracy	0.7289	0.9398	0.2664	0.5707	0.5779	0.6243	0.6304	0.6384	**0.6640**
	Recall	0.7289	0.9398	0.2664	0.5707	0.5779	0.6243	0.6304	0.6384	0.6640
	Presicion	0.7332	0.9514	0.2462	0.6070	0.5724	0.6905	0.7078	0.6924	0.7030
	F1-Score	0.6749	0.9330	0.1564	0.4623	0.4694	0.5717	0.5822	0.5709	0.6219

6.2.1. Confusion matrix and error analysis

A detailed error analysis revealed that most misclassifications in the 6-class and 19-class tasks occurred between conceptually similar or temporally overlapping attack types, particularly DoS vs. DDoS and Benign vs. Spoofing. These types often share overlapping features in short-term sequence windows, which challenges even deep models without attention mechanisms. The integration of self-attention in the RNN + LSTM + Attention model appeared to mitigate some of this confusion which enables the model to focus more effectively on distinguishing temporal nuances in network behavior.

6.2.1.1. Binary classification

In the binary classification task, the model performs well overall, with strong discriminative capability between attack and benign traffic. It correctly labels a huge majority of attack samples and benign samples, with hardly any misclassifications either way. As shown in **Table 6.2**, the Attack class had a precision of 0.9915 and recall of 0.9960, which translated to an F1-score of 0.9937, i.e., the model precisely classified the majority of attacks with barely any false positives. The Benign class also fared well, with 0.9935 precision and 0.9863 recall, showing barely more false negatives but again superb overall performance. These figures indicate the model is radically effective at distinguishing normal traffic from malicious traffic.

Table 6.2. Classification metrics for binary classification attack categories.

Attack Categories	Precision	Recall	F1-score	Number of Test Samples
Attack	0.9915	0.9960	0.9937	60000
Benign	0.9935	0.9863	0.9899	37607

In **Figure 6.1**, it can be observed that false negatives (attacks classified as benign) and the false positives (benign classified as attack) are relatively low compared to the true positives and true negatives. This suggests that the model is good and well-proportioned for general anomaly detection tasks.

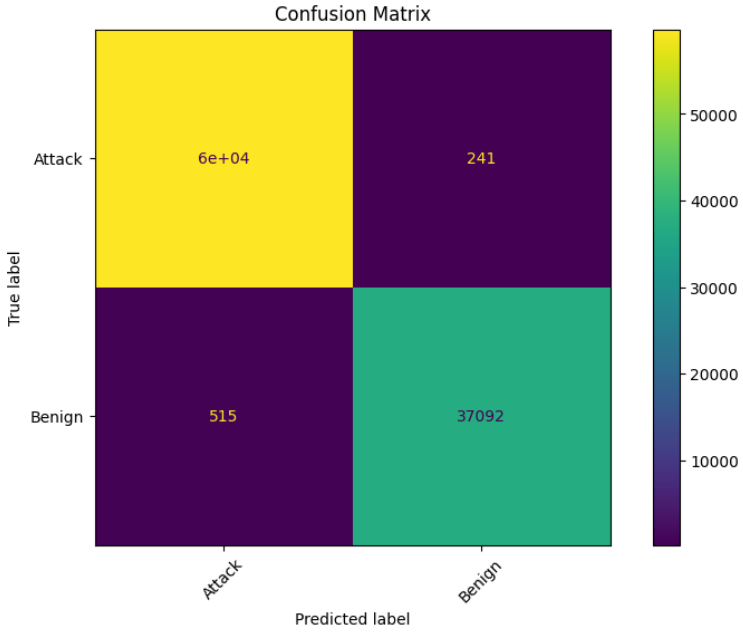

Figure 6.1. Confusion matrix of binary classification.

6.2.1.2. 6-class classification

In the 6-class classification task, the performance varied between classes. It can be observed in **Table 6.3** that MQTT and Recon classes were classified very well, achieving F1-scores of 0.9900 and 0.9593, respectively. Benign traffic also achieved good performance with F1-score of 0.8735. On the other hand, DDoS and Spoofing had unbalanced performance. DDoS achieved good recall score but its precision was lower, which indicates it has more false positives.

Table 6.3. Classification metrics for 6-class attack categories.

Attack Categories	Precision	Recall	F1-score	Number of Test Samples
Benign	0.8813	0.8658	0.8735	5000
DDoS	0.6118	0.9162	0.7337	5000
DoS	0.8338	0.4174	0.5563	5000
MQTT	0.9970	0.9832	0.9900	5000
Recon	0.9902	0.9302	0.9593	5000
Spoofing	0.6186	0.7850	0.6919	1744

In contrast, DoS had high precision but low recall, which means many attacks were not detected. This shows the model has difficulty in consistently distinguishing between overlapping or similar traffic patterns in DDoS-DoS and Spoofing-Benign

classes. This challenge often arises due to shared protocol behaviors and similar packet structures which makes it harder for the model to extract discriminative features.

As demonstrated in **Figure 6.2**, MQTT and Benign samples are well-separated, whereas DoS and DDoS attacks are confused with each other to some degree. However, Spoofing class show more noticeable misclassifications. Nevertheless, the model maintains reasonable class separation and correctly classifies the majority of samples in each category.

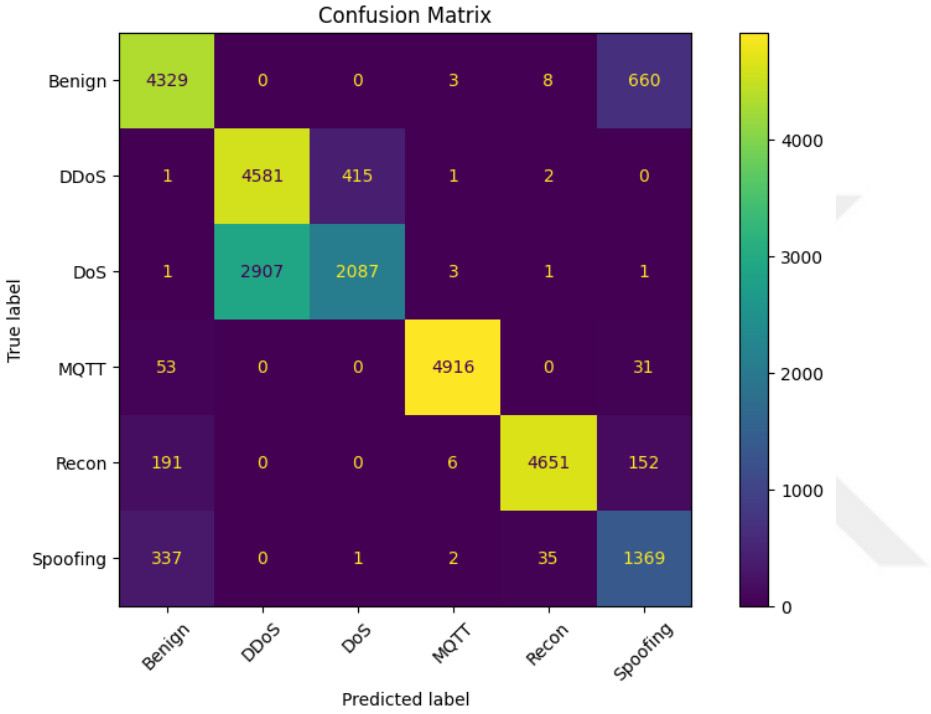

Figure 6.2. Confusion matrix of 6-class classification.

6.2.1.3. 19-class classification

In the 19-class task, the model faced greater complexity and for many subclasses the performance decreased. As **Table 6.4** indicates, classes like MQTT DDoS Connect Flood and MQTT DoS Connect Flood registered precise and accurate detection performances with F1-scores of around 0.9896 and 0.9894. Others like DDoS TCP and Recon VulScan were poorly detected while Spoofing ARP was moderately successful with a scope for improvement.

As in the 6-class classification, DoS and DDoS attacks are confused with each other mostly. It is observed in **Figure 6.3**, that this confusion increases with the separation of the MQTT main class into DoS and DDoS subclasses. These observations are a

representation of the problem of imbalanced data and the learning difficulty of identifying subtle differences in traffic at a fine-grained level.

Table 6.4. Classification metrics for 19-class attack categories.

Attack Categories	Precision	Recall	F1-score	Number of Test Samples
Benign	0.8690	0.7897	0.8274	3200
DDoS ICMP	0.6598	0.3200	0.4310	3200
DDoS SYN	0.7779	0.4991	0.6080	3200
DDoS TCP	0.6085	0.1928	0.2928	3200
DDoS UDP	0.5745	0.5603	0.5673	3200
DoS ICMP	0.5507	0.8334	0.6632	3200
DoS SYN	0.6323	0.8556	0.7272	3200
DoS TCP	0.5203	0.8756	0.6528	3200
DoS UDP	0.5706	0.5850	0.5777	3200
MQTT DDoS Connect Flood	0.9848	0.9944	0.9896	3200
MQTT DDoS Publish Flood	0.9792	0.1175	0.2098	3200
MQTT DoS Connect Flood	0.9932	0.9856	0.9894	3131
MQTT DoS Publish Flood	0.5335	0.9984	0.6954	3200
MQTT Malformed Data	0.7680	0.8867	0.8231	1747
Recon OS Scan	0.5673	0.3475	0.4310	3200
Recon Ping Sweep	0.8742	0.7097	0.7834	186
Recon Port Scan	0.5763	0.7884	0.6659	3200
Recon VulScan	0.3257	0.1654	0.2194	1034
Spoofing ARP	0.5342	0.7391	0.6202	1744

In conclusion, this study contrasts and compares classic machine learning algorithms with a number of deep learning models on the CICIoMT2024 dataset. A range of model architectures were developed and experimentally tested with consideration placed on the trade-off between performance and computational complexity. For this, feature selection through variance thresholding was used, reducing input features and enhancing training efficiency without loss of classification quality.

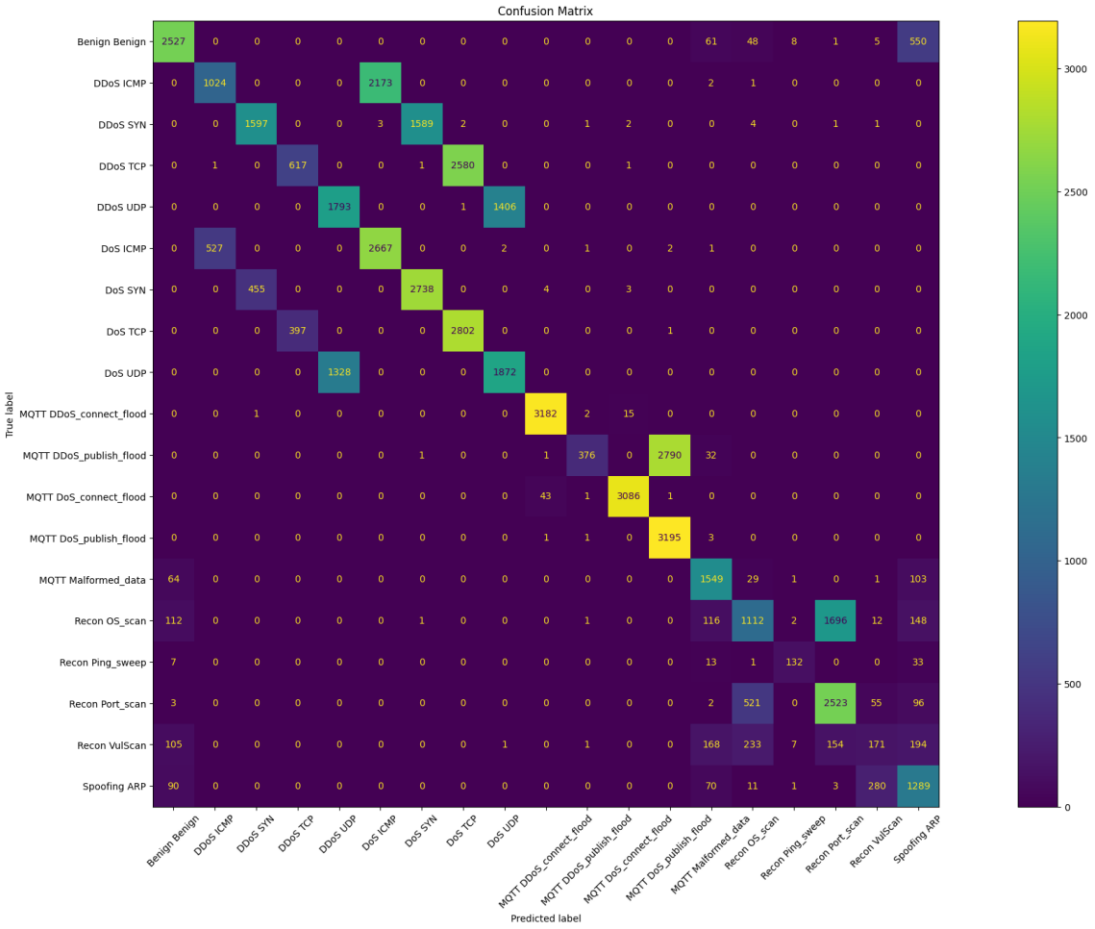

Figure 6.3. Confusion matrix of 19-class classification.

The research also investigated fundamental challenges in IoMT anomaly detection, e.g., the inappropriateness of discrimination between highly similar attack types like DoS and DDoS. Finally, a number of hybrid architectures were tried out and the best-performing model (combining RNN, LSTM and self-attention layers) achieved improved detection capability through the ability to capture temporal relationships as well as subtle patterns in traffic data. These studies aim at designing more accurate and efficient anomaly detection systems for medical IoT networks.

6.3. Runtime Efficiency Comparison Between Hybrid Architectures

In addition to evaluating classification performance, this study also examined the computational efficiency of the proposed deep learning model along with the second best model. Runtime efficiency is particularly significant in IoMT contexts where real-time anomaly detection is desired and computational resources may be constrained, especially in edge devices.

To assess this, training and prediction times were recorded for two of the most complex hybrid architectures: Conv1D + LSTM + RNN (shown as 1 in Model column in **Table 6.5**) and RNN + LSTM + Attention (shown as 2 in Model column in **Table 6.5**). These measurements were taken for both 6-class and 19-class classification tasks using feature sets of 30 and 45 dimensions. The results are summarized in **Table 6.5**.

Table 6.5. Runtime efficiency comparison.

Classification Task	Feature Count	Model	Training Time (s)	Prediction Time (s)
6-class	30	1	680,3628	9,1652
		2	606,6737	9,7036
	45	1	983,3850	12,6763
		2	812,5085	12,7913
19-class	30	1	823,4000	17,6594
		2	1.202,6254	23,7467
	45	1	1.126,7298	24,2481
		2	1.828,0146	36,4287

Reducing the feature set from 45 to 30 significantly improved runtime efficiency across both hybrid architectures, especially for the more complex RNN + LSTM + Attention model. In 19-class classification, for example, the training time dropped from approximately 1828 seconds to 1202 seconds and prediction time dropped from 36.43 seconds to 23.75 seconds when using 30 features instead of 45. This demonstrates that feature reduction not only simplifies the model input but also reduces the computational load, which is crucial in time-sensitive IoMT environments. While the attention-based architecture inherently demands more computation due to its internal operations, the impact of feature size remains a major factor in optimizing model performance and efficiency.

7. CONCLUSION

Growing utilization of Internet of Medical Things (IoMT) devices in healthcare systems comes with advantages as well as significant cybersecurity threats. As these networked medical devices are becoming integral to patient care, security and integrity of the network infrastructure are paramount. Anomaly detection is especially important in this context when traditional signature-based intrusion detection systems cannot keep up with new and unknown threats.

In this paper, we proposed and compared a range of machine learning and deep learning models for network-based anomaly detection using the CICIoMT2024 dataset. The CICIoMT2024 dataset is a realistic IoMT traffic in different devices and attack scenarios and can be utilized for comparing the effectiveness of different detection techniques.

7.1. Evaluation Across Models

Experiments were conducted with both traditional machine learning algorithms (Logistic Regression, Random Forest and AdaBoost) and also deep learning models such as Dense Neural Networks (DNN), RNNs, LSTMs, 1D Convolutional layers and hybrid models with attention mechanisms. The models were tested on three classification setups: binary classification (normal vs. anomalous), 6-class classification (big attack categories) and 19-class fine-grained attack detection.

The findings were clear performance variation with the model type based on the complexity of the tasks. In the binary scenario, the traditional models performed well with Random Forest at 0.98 accuracy level and comparable scores on precision, recall and F1-score. This indicates that for less complex detection tasks where feature boundaries are clearer, traditional tree-based models are still a good option.

7.1.1. Role of hybrid deep learning

As the classification complexity increased to 6 and 19 classes, deep learning models—especially those capturing temporal and sequential relationships—became increasingly dominant. The RNN + LSTM + Attention model consistently

outperformed others, achieving an accuracy of 0.99 in binary class task, 0.82 in the 6-class task and 0.66 in the 19-class task. These results validate the hypothesis that temporal attention can significantly enhance the model's capacity to distinguish nuanced attack patterns, which are often difficult to capture using static feature sets alone.

Interestingly, while DNNs and standalone Conv1D models showed moderate performance, stacking architectures such as Conv1D + LSTM + RNN yielded better generalization, likely due to their layered ability to extract both spatial and temporal dependencies. Attention mechanisms further improved model discrimination by allowing the network to dynamically focus on the most essential time steps, which is particularly valuable in complex traffic patterns typical of IoMT environments.

7.2. Limitations of the Study

Despite these promising outcomes, the current work has some limitations. First, all experiments were conducted on a single dataset. While CICIoMT2024 is comprehensive, future studies should validate the model's generalizability across other datasets and real-world deployments. Second, the models, particularly the deep architectures, require substantial computational resources and hyperparameter tuning, which may hinder their immediate deployment on edge devices with constrained processing power. Additionally, the class distribution in the dataset remains skewed, even after downsampling techniques, which may bias some performance metrics, particularly in underrepresented classes.

7.3. Future Work Recommendations

Future work can explore real-time anomaly detection frameworks that enable on-device processing in edge environments. This would involve lightweight and optimized models that can operate under strict latency and energy constraints. Model compression techniques like weight sparsification, reduced-precision formats, and teacher-student training frameworks could help in this transition.

Also interpretability is crucial, particularly in healthcare-related systems where security decisions must be transparent and explainable. Future efforts should incorporate explainable AI (XAI) techniques such as SHAP values or LIME to

visualize model decisions. This could enhance trust and allow system administrators or healthcare professionals to make informed judgments based on AI predictions.

IoMT environments are inherently dynamic, with frequent changes in traffic volume, device roles, and contextual network behavior. Therefore, future research should explore adaptive learning mechanisms capable of responding to real-time distribution shifts and concept drift. Online learning, continual training, and domain adaptation methods may help models remain effective even when faced with evolving usage patterns or novel attack vectors.

REFERENCES

[1] M. Mohsin and A. I. Jony, “A comparative analysis of medical IoT device attacks using machine learning models,” Malaysian Journal of Science and Advanced Technology, pp. 429–439, Sep. 2024, doi: 10.56532/mjsat.v4i4.318.

[2] H. Naeem, A. Alsirhani, F. M. Alserhani, F. Ullah, and O. Krejcar, “Augmenting Internet of Medical Things Security: deep ensemble integration and methodological fusion,” Computer Modeling in Engineering & Sciences, vol. 141, no. 3, pp. 2185–2223, Jan. 2024, doi: 10.32604/cmes.2024.056308.

[3] A. Misbah, A. Sebbar, and I. Hafidi, “SmartEdge: An Adaptive Framework for Efficient and Secure IoMT Data Processing with Real-Time Cyber Attack Detection,” in Proc. Third ICMDS'24: Machine Learning, Inverse Problems and Related Fields, A. Laghrib and A. Ghazdali, Eds., Lecture Notes in Networks and Systems, vol. 1466, Cham, Switzerland: Springer, 2025, pp. [insert page numbers if available]. doi: 10.1007/978-3-031-94802-2_2

[4] J. Doménech, O. León, M. S. Siddiqui, and J. Pegueroles, “Evaluating and enhancing intrusion detection systems in IoMT: The importance of domain-specific datasets,” Internet of Things, p. 101631, May 2025, doi: 10.1016/j.iot.2025.101631.

[5] K. Ramesh, N. C. Miller, A. Faridi, F. Aloul, I. Zualkernan and A. R. Sajun, "Efficient Machine Learning Frameworks for Strengthening Cybersecurity in Internet of Medical Things (IoMT) Ecosystems," 2024 IEEE International Conference on Internet of Things and Intelligence Systems (IoTaIS), Bali, Indonesia, 2024, pp. 92-98, doi: 10.1109/IoTaIS64014.2024.10799438.

[6] F. G. Abdiwi, “Hybrid machine learning and blockchain technology for early detection of cyberattacks in healthcare systems,” International Journal of Safety and Security Engineering, vol. 14, no. 6, pp. 1883–1893, Dec. 2024, doi: 10.18280/ijsse.140622.

[7] S. Dadkhah, E. C. P. Neto, R. Ferreira, R. C. Molokwu, S. Sadeghi, and A. A. Ghorbani, “CICIoMT2024: A benchmark dataset for multi-protocol security assessment in IoMT,” Internet of Things, vol. 28, p. 101351, Aug. 2024, doi: 10.1016/j.iot.2024.101351.

[8] A. S. Mirkhail and Z. Xinyou, “Deep learning for anomaly detection in IoT healthcare systems,” International Research Journal of Multidisciplinary Scope, vol. 06, no. 02, pp. 1480–1494, Jan. 2025, doi: 10.47857/irjms.2025.v06i02.03768.

[9] A. Mohammadi, H. Ghahramani, S. A. Asghari and M. Aminian, "Securing Healthcare with Deep Learning: A CNN-Based Model for Medical IoT Threat Detection," 2024 19th Iranian Conference on Intelligent Systems (ICIS), Sirjan, Iran, Islamic Republic of, 2024, pp. 168-173, doi: 10.1109/ICIS64839.2024.10887510.

[10] A. Misbah, A. Sebbar, and I. Hafidi, “Innovative Federated Learning Approach to Secure Internet of Medical Things,” in Innovative Technologies on Electrical Power Systems for Smart Cities Infrastructure, I. Aboudrar, F. Ilahi Bakhsh, A. Nayyar, and I. Ouachtouk, Eds., Sustainable Civil Infrastructures, ICESST 2024, Cham, Switzerland: Springer, 2025, pp. 220-228. doi: 10.1007/978-3-031-86705-7_21

[11] G. R. Pradyumna, R. B. Hegde, K. B. Bommegowda, T. Jan and G. R. Naik, "Empowering Healthcare With IoMT: Evolution, Machine Learning Integration, Security, and Interoperability Challenges," in IEEE Access, vol. 12, pp. 20603-20623, 2024, doi: 10.1109/ACCESS.2024.3362239

[12] K. S. Bughio, D. M. Cook, and S. A. A. Shah, “Developing a novel ontology for cybersecurity in internet of Medical Things-Enabled Remote Patient Monitoring,” Sensors, vol. 24, no. 9, p. 2804, Apr. 2024, doi: 10.3390/s24092804.

[13] A. H. Ameen, M. A. Mohammed, and A. N. Rashid, “Enhancing security in IOMT: a Blockchain-Based cybersecurity framework for Machine Learning-Driven ECG signal classification,” Fusion Practice and Applications, vol. 14, no. 1, pp. 221–251, Dec. 2023, doi: 10.54216/fpa.140117.

[14] S. Ksibi, F. Jaidi, and A. Bouhoula, “MLRA-Sec: an adaptive and intelligent cyber-security-assessment model for internet of medical things (IoMT),” International Journal of Information Security, vol. 24, no. 1, Nov. 2024, doi: 10.1007/s10207-024-00923-y.

[15] N. K. Jabbar, M. Naderan, M. S. Taha, “HybridIOMT: a Dual-Phase machine learning framework for robust cybersecurity in Internet of Medical Things,” International Journal of Intelligent Engineering and Systems, vol. 18, no. 4, pp. 307–321, Apr. 2025, doi: 10.22266/ijies2025.0531.20.

[16] G. Lazrek, K. Chetioui, Y. Balboul, S. Mazer, and M. E. Bekkali, “An RFE/Ridge-ML/DL based anomaly intrusion detection approach for securing IoMT system,” Results in Engineering, vol. 23, p. 102659, Aug. 2024, doi: 10.1016/j.rineng.2024.102659.

[17] A. A. Khan et al., “BDLT-IoMT—a novel architecture: SVM machine learning for robust and secure data processing in Internet of Medical Things with blockchain cybersecurity,” The Journal of Supercomputing, vol. 81, no. 1, Dec. 2024, doi: 10.1007/s11227-024-06782-7.

[18] G. R. Pradyumna, R. B. Hegde, K. B. Bommegowda, T. Jan, and G. R. Naik, “Empowering healthcare with IOMT: evolution, machine learning integration, security, and interoperability challenges,” IEEE Access, vol. 12, pp. 20603–20623, Jan. 2024, doi: 10.1109/access.2024.3362239.

[19] C. Tagliaro et al., “Large-Scale security analysis of Real-World backend deployments speaking IoT-Focused protocols,” *The 27th International Symposium on Research in Attacks, Intrusions and Defenses (RAID 2024)*, p. 18, 2024, [Online]. https://doi.org/10.1145/3678890.3678899

[20] Z. Laaroussi and O. Novo, "A Performance Analysis of the Security Communication in CoAP and MQTT," 2021 IEEE 18th Annual Consumer Communications & Networking Conference (CCNC), Las Vegas, NV, USA, 2021, pp. 1-6, doi: 10.1109/CCNC49032.2021.9369565.

[21] A. Djenna, S. Harous, and D. E. Saidouni, “Internet of Things meet Internet of Threats: New concern Cyber security Issues of critical cyber infrastructure,” *Applied Sciences*, vol. 11, no. 10, p. 4580, May 2021, doi: 10.3390/app11104580.

[22] D. Koutras, G. Stergiopoulos, T. Dasaklis, P. Kotzanikolaou, Dimitris Glynos, and C. Douligeris, “Security in IOMT Communications: a survey,” journal-article, Aug. 2020. doi: 10.3390/s20174828.

[23] M. Sarhan, S. Layeghy, N. Moustafa, and M. Portmann, “NetFlow datasets for Machine Learning-Based network Intrusion detection systems,” in *Springer eBooks*, 2021, pp. 117–135. doi: 10.1007/978-3-030-72802-1_9.

[24] A. Naghib, F. S. Gharehchopogh, and A. Zamanifar, “A comprehensive and systematic literature review on intrusion detection systems in the internet of medical things: current status, challenges, and opportunities,” Artificial Intelligence Review, vol. 58, no. 4, Jan. 2025, doi: 10.1007/s10462-024-11101-w.

[25] G. Akar, S. Sahmoud, M. Onat, Ü. Cavusoglu, and E. Malondo, “L2D2: A novel LSTM model for multi-class intrusion detection systems in the era of IOMT,” IEEE Access, p. 1, Jan. 2025, doi: 10.1109/access.2025.3526883.

[26] J. Al. Shaikh, C. Wang, N. Saifullah, M. W. U. Sima, M. Arshad, and W. U. A. Rathore, “Memory feedback transformer based intrusion detection system for IoMT healthcare networks,” Internet of Things, p. 101597, May 2025, doi: 10.1016/j.iot.2025.101597.

[27] N. Faruqui et al., “SafetyMed: A novel IOMT intrusion detection system using CNN-LSTM hybridization,” Electronics, vol. 12, no. 17, p. 3541, Aug. 2023, doi: 10.3390/electronics12173541.

[28] A. K. Rai, D. K. Verma, and R. K. Dwivedi, “RTAD-HIS: Regulated Transformer Architecture based Anomaly Detection Framework Towards Security in Healthcare IoT Systems,” Applied Soft Computing, vol. 177, p. 113209, Apr. 2025, doi: 10.1016/j.asoc.2025.113209.

[29] S. Loganathan, J. R. F. Raj, R. S. Krishnan, A. A. Blessie, V. V. Kumar, and S. Kavitha, An AI-Driven Cybersecurity Framework for the Internet of Medical Things: A Hybrid LSTM-CNN-GAN Approach. 2025, pp. 1883–1890. doi: 10.1109/icmlas64557.2025.10967601.

[30] N. Ying, H. Li, Z. Zhang, Y. Zhou, H. Chen, and M. Yang, “VITDFNN: a vision transformer enabled deep fuzzy neural network for detecting sleep Apnea-Hypopnea syndrome in the internet of medical things,” IEEE Transactions on Fuzzy Systems, pp. 1–12, Jan. 2024, doi: 10.1109/tfuzz.2024.3388091.

[31] D. Alsalman, “A comparative study of anomaly detection Techniques for IoT Security using Adaptive Machine Learning for IoT threats,” *IEEE Access*, vol. 12, pp. 14719–14730, Jan. 2024, doi: 10.1109/access.2024.3359033.

[32] J. Wang, H. Jin, J. Chen, J. Tan, and K. Zhong, “Anomaly detection in Internet of medical Things with Blockchain from the perspective of deep neural network,” *Information Sciences*, vol. 617, pp. 133–149, Oct. 2022, doi: 10.1016/j.ins.2022.10.060.

[33] M. L. Hernandez-Jaimes, A. Martinez-Cruz, K. A. Ramírez-Gutiérrez, and A. Morales-Reyes, “Network traffic inspection to enhance anomaly detection in the Internet of Things using attention-driven Deep Learning,” *Integration*, p. 102398, Mar. 2025, doi: 10.1016/j.vlsi.2025.102398.

[34] J. B. Awotunde, K. M. Abiodun, E. A. Adeniyi, S. O. Folorunso, and R. G. Jimoh, “A Deep Learning-Based intrusion detection technique for a secured IOMT system,” in *Communications in computer and information science*, 2022, pp. 50–62. doi: 10.1007/978-3-030-95630-1_4.

[35] A. Berguiga, A. Harchay, and A. Massaoudi, “HIDS-IOMT: a Deep Learning-Based Intelligent Intrusion Detection System for the Internet of Medical Things,” IEEE Access, p. 1, Jan. 2025, doi: 10.1109/access.2025.3543127.

[36] J. Areia, I. A. Bispo, L. Santos, and R. L. De C Costa, “IOMT-TrafficData: Dataset and Tools for benchmarking intrusion detection in Internet of Medical Things,” *IEEE Access*, vol. 12, pp. 115370–115385, Jan. 2024, doi: 10.1109/access.2024.3437214.

[37] G. Zachos, G. Mantas, K. Porfyrakis, J. M. C. S. Bastos, and J. Rodriguez, “Anomaly-Based Intrusion Detection for IOMT Networks: design, implementation, dataset generation and ML Algorithms evaluation,” *IEEE Access*, p. 1, Jan. 2025, doi: 10.1109/access.2025.3547572.

[38] M. Zubair et al., “Secure Bluetooth communication in smart healthcare systems: a novel community dataset and intrusion detection system,” Sensors, vol. 22, no. 21, p. 8280, Oct. 2022, doi: 10.3390/s22218280.

[39] P. Radoglou-Grammatikis et al., “Modeling, detecting, and mitigating threats against industrial healthcare systems: a combined software defined networking and reinforcement learning approach,” IEEE Transactions on Industrial Informatics, vol. 18, no. 3, pp. 2041–2052, Jul. 2021, doi: 10.1109/tii.2021.3093905.

[40] F. Hussain et al., “A framework for malicious traffic detection in IoT healthcare environment,” Sensors, vol. 21, no. 9, p. 3025, Apr. 2021, doi: 10.3390/s21093025.

[41] M. Ahmed, S. Byreddy, A. Nutakki, L. F. Sikos, and P. Haskell-Dowland, “ECU-IoHT: A dataset for analyzing cyberattacks in Internet of Health Things,” Ad Hoc Networks, vol. 122, p. 102621, Jul. 2021, doi: 10.1016/j.adhoc.2021.102621.

[42] A. A. Hady, A. Ghubaish, T. Salman, D. Unal, and R. Jain, “Intrusion Detection System for healthcare systems using medical and network Data: A comparison study,” IEEE Access, vol. 8, pp. 106576–106584, Jan. 2020, doi: 10.1109/access.2020.3000421.

[43] I. Guyon and A. Elisseeff, “An introduction to variable and feature selection,” *Journal of Machine Learning Research*, Mar. 2003, doi: 10.5555/944919.944968.

[44] D. Jurafsky and J. H. Martin, “Speech and Language Processing: Chapter 5 Logistic Regression,” *Stanford University*, Jan. 12, 2025, Accessed on: May 19, 2025. [Online]. Available: https://web.stanford.edu/~jurafsky/slp3/5.pdf

[45] T. Laird, “Partially ordered logistic regression,” MSc. Dissertation, Dept. Of Statistics and Sata Science, University of Arizona, 2024.

[46] S. Kotsiantis, Supervised Machine Learning: A Review of Classification Techniques. *Informatica*, 31. 249-268, 2007.

[47] S. Biswas, S. Ghosh, S. Roy, R. Bose, and S. Soni, “A Study of Stock Market Prediction through Sentiment Analysis,” *Mapana Journal of Sciences*, vol. 22, no. 1, Feb. 2023, doi: 10.12723/mjs.64.6.

[48] L. Breiman and Statistics Department, University of California, “*Random forests”*. 2001, p. 4, Accessed on: May 19, 2025. [Online]. Available: https://www.stat.berkeley.edu/~breiman/randomforest2001.pdf

[49] S. J. Sam and S. B. Xavier, “Intrusion Detection System for Industrial IoT,” *IEEE Access*, pp. 1–6, Jul. 2024, doi: 10.1109/icait61638.2024.10690722.

[50] Y. Wang, Z. Pan, J. Zheng, L. Qian, and M. Li, “A hybrid ensemble method for pulsar candidate classification,” *Astrophysics and Space Science*, vol. 364, no. 8, Aug. 2019, doi: 10.1007/s10509-019-3602-4.

[51] Y. Freund and R. E. Schapire, “A Decision-Theoretic generalization of On-Line learning and an application to boosting,” *Journal of Computer and System Sciences*, vol. 55, no. 1, pp. 119–139, Aug. 1997, doi: 10.1006/jcss.1997.1504.

[52] A. M. Javid, S. Das, M. Skoglund, and S. Chatterjee, “A RELU dense layer to improve the performance of neural networks,” *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 2810–2814, May 2021, doi: 10.1109/icassp39728.2021.9414269.

[53] W. Wang et al., “End-to-end Encrypted Traffic Classification with One-dimensional Convolution Neural Networks,” *IEEE International Conference on Intelligence and Security Informatics (ISI)*, pp. 43-48, 2017 doi: 10.1109/ISI.2017.8004872.

[54] “Size of output of a Conv1D layer in Keras”. [Online]. Available: https://stackoverflow.com/questions/65006011/size-of-output-of-a-conv1d-layer-in-keras

[55] Y. Bengio, P. Simard, and P. Frasconi, “Learning long-term dependencies with gradient descent is difficult,” IEEE Transactions on Neural Networks, vol. 5, no. 2, pp. 157–166, Mar. 1994, doi: 10.1109/72.279181.

[56] “How Recurrent Neural Network (RNN) Works”. [Online]. Available: https://dataaspirant.com/how-recurrent-neural-network-rnn-works/

[57] G. Kim, H. Yi, J. Lee, Y. Paek, S. Yoon, and Seoul National University, “LSTM-based system-call language modeling and robust ensemble method for designing host-based intrusion detection systems,” *arXiv Preprint*, 2016.

[58] R. Vinayakumar, K.P. Soman, P. Prabaharan, “Applying deep learning approaches for network traffic prediction,” *International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, 2017, doi: 10.1109/icacci.2017.8126198

[59] R. Kizito, P. Scruggs, X. Li, M. Devinney, J. Jansen, and R. Kress, “Long Short-Term memory networks for facility infrastructure failure and remaining useful life prediction,” *IEEE Access*, vol. 9, pp. 67585–67594, Jan. 2021, doi: 10.1109/access.2021.3077192.

[60] A. Vaswani et al., “Attention is all you need,” *31st Conference on Neural Information Processing Systems (NIPS 2017),* Accessed on: May 19, 2025. [Online]. Available: https://arxiv.org/pdf/1706.03762.pdf

[61] S. R. Fahim, Y. Sarker, S. K. Sarker, Md. R. I. Sheikh, and S. K. Das, “Self attention convolutional neural network with time series imaging based feature extraction for transmission line fault detection and classification,” *Electric Power Systems Research*, vol. 187, p. 106437, Jun. 2020, doi: 10.1016/j.epsr.2020.106437.

[62] T. Brownlee, “Classification Accuracy is Not Enough: More Performance Measures You Can Use,” *Machine Learning Mastery*, 2020, Accessed on: May 19, 2025. [Online]. Available: https://machinelearningmastery.com/classification-accuracy-is-not-enough-more-performance-measures-you-can-use/

[63] P. Huilgol, “Precision and Recall in Machine Learning”. *Analytics Vidhya*, 2024, Accessed on: May 19, 2025. [Online]. Available: https://www.analyticsvidhya.com/articles/precision-and-recall-in-machine-learning/

[64] H. B. McMahan, “Metrics for Classification Problems – Why Accuracy Can Be Misleading,” *Google AI Blog*, 2019. Accessed on: May 19, 2025. [Online]. Available: https://developers.google.com/machine-learning/crash-course/classification/precision-and-recall

[65] S. N. Sruthy, “Precision-Recall Trade-Off: Navigating the Delicate Balance in Imbalanced Datasets,” Medium, Jan. 30, 2024. Accessed on: May 19, 2025. [Online]. Available: https://medium.com/@sruthy.sn91/precision-recall-trade-off-navigating-the-delicate-balance-in-imbalanced-datasets-66ad175ccf1e

CURRICULUM VITAE

Name Surname: Ayşe Betül BÜKEN

EDUCATION:

- **Undergraduate** : 2018, Sakarya University, Engineering Faculty, Department of Industrial Engineering
- **Graduate** : Ongoing, Sakarya University, Institute of Science and Technology, Department of Software Engineering

PROFESSIONAL EXPERIENCE:

- She worked as a design engineer at the KOLARC Makine between 2018-2019.
- She works at BMC Savunma since 2019.

PUBLICATIONS, PRESENTATIONS AND PATENTS ON THE THESIS:

- Buken A. B., Akgun D. (2025, 23-25 May). Anomaly Detection in IoT Device Traffic Data Using Deep Learning. *21th International Istanbul Congress,* Istanbul, Turkiye.