THÈSE POUR OBTENIR LE GRADE DE DOCTEUR DE L'INSTITUT MINES-TÉLÉCOM (IMT) – ÉCOLE NATIONALE SUPÉRIEURE DES MINES D'ALÈS (IMT MINES ALÈS)

1

En Informatique

École doctorale ED 583 Risques et Sociétés

Unité de recherche Laboratoire des Sciences des Risques

A MULTILAYERED CYBERSECURITY APPROACH FOR THE HEALTHCARE SECTOR BASED ON DATABASE TECHNOLOGY, AI/ML, AND RISK ASSESSMENT

Présentée par Emmanuel Song Shombot

Le 25 mars 2025

Sous la direction de Gilles DUSSERRE et Robert BESTAK

Devant le jury composé de

Gouenou COATRIEUX, Professeur, IMT Atlantique	Rapporteur
Didier Van CAILLIE, Professeur, University of Liège	Rapporteur
Gilles DUSSERRE, Professeur, University of Nîmes	Directeur de Thèse
Robert BESTAK, Professeur, CTU Prague	Co-Directeur de Thèse
Stéphane LECOEUCHE, Professeur, IMT Mines Alès	Encadrant
Johanne VINCENT, Maître de Conférences, IMT Atlantique	Examinateur
Jean-Yves LEFRANT, Professeur, CHU de Nîmes	Examinateur, Pr. Du jury
caroline de RUBIANA, Mme, Cyber'Occ, France	Invitè

Collège Doctoral Languedoc-Roussillon

IMT Mines Alès
École Mines-Télécom

ABSTRACT

In recent years, there has been a growing frequency and sophistication in cyberattacks, which have made cybersecurity in the healthcare sector a pressing global concern. Some of the factors responsible for this fragility include overdependence of healthcare facilities on linked digital infrastructures, sensitive patient information, and the use of antiquated security systems. This thesis explores a multilayered cybersecurity approach in healthcare using database technology, artificial intelligence (AI), machine learning (ML), and risk assessment methodologies. We begin by assessing the current threat landscape through a web application that leverages publicly available cyber incident records mainly from the Center for International and Security Studies Maryland (CISSM), which contains over 1,500 healthcare-related cases, and the Office for Civil Rights (OCR), which has documented over 800 investigated incidents. The developed platform enables users to add, delete, search, view, and download cyber-related incidents. To address cybersecurity vulnerabilities, we introduce the Baze University Hospital Cyber Risk Assessment Framework (BUHCRAF), applied at Baze University Hospital as a case study. This framework identifies over 50 security risks, with some linked to known threats documented in the Common Vulnerabilities and Exposures (CVE) database. Key risks include unpatched software, unsecured medical IoT devices, and weak access controls. Based on these insights, AI/ML-driven cybersecurity solutions are proposed. The study includes several case studies demonstrating AI/ML applications in cybersecurity. The first case focuses on anomaly detection in the Internet of Health Things (IoHT), using the ECU-IoHT dataset containing 111,000 records and five attack classes (Normal, Nmap, DoS, ARP, Smurf). We compare four ML models—Random Forest (RF), Gradient Boosting (GB), Decision Trees (DT), and Multilayer Perceptron (MLP)—where GB achieves the highest accuracy of 99.53%. The second case study applies AI/ML to phishing detection, analyzing features such as request URL, domain age, HTTPS/SSL status, website traffic, and pop-ups. Support Vector Machines (SVM) were used with two kernel functions—Radial Basis Function (RBF) and Polynomial. The Polynomial kernel outperformed RBF, achieving 84.5% accuracy compared to 82.6%. The third case study explores AI/ML in intrusion detection by proposing a Cross-ML framework using three public datasets: UNSW-NB15, X-IIoTID, and ToN-IoT. Individual models implies models trained on their respective datasets achieved F1-scores ranging from 92.0% to 98.1%, while the combined dataset approach recorded an accuracy of 90.9%, recall of 84.0%, and an F1-score of 84.9%. Cross-dataset evaluations showed that the X-IIoTID model tested on ToN-IoT achieved the best performance, reaching 75% accuracy and a 74% F1-score. This research highlights the importance of a multilayered cybersecurity approach that integrates database-driven approach, AI/ML-enhanced security applications, and structured risk assessment frameworks to strengthen cybersecurity resilience in healthcare organizations.

Keywords: *Cybersecurity, Artificial Intelligence, Machine Learning, Databases, Risk Assessment*

ACKNOWLEDGMENTS

Truly, each PhD program is a journey that requires the participation of both institutions and individuals; therefore, I am immensely grateful to everyone who has supported this project since its inception.

First and foremost, I would like to thank IMT Mines Alès, the University of Nîmes, Campus France, the Petroleum Technology Development Fund (PTDF), and the Federal University Lafia for their collaborative roles in providing an enabling atmosphere and the necessary funding for the successful completion of this PhD research. I also want to thank the Barrande fellowship program, a collaboration between the government of France and the Czech Republic, for affording me the opportunity to conduct part of my research at the Czech Technical University in Prague.

I am immensely grateful for the constant support of my supervisory team and the staff throughout this study. My heartfelt gratitude goes to my thesis directors, Gilles Dusserre and Robert Bestak, whose enthusiasm, trust, and constructive criticism gave me the confidence to navigate the roller-coaster ride of this doctoral program. I would also like to thank Nasir Baba-Ahmed and Stéphane Lecoeuche for their invaluable guidance and contributions during the research process.

I extend my sincere appreciation to all members of the jury and manuscript reviewers for taking the time to assess this work and provide significant professional feedback. I am also grateful to the staff at Laboratoire CERIS (LGI2P), particularly Edith Teychene, for their prompt assistance with administrative matters. Special thanks to Anne-Catherine Denni for her registration support and Kattialyn Gossiaux for the numerous tutorials on the use of scientific resources.

My heartfelt gratitude goes to Mrs. Christine Charmasson for her dedication to helping me master the French language, which was crucial for my integration and success in this program.

I am profoundly thankful to my family for their patience, love, and encouragement over the past few years. My heartfelt gratitude goes to my father, Emmanuel Shombot Dung, whose words of encouragement have been a wellspring of strength; my mother, Angela Emmanuel Dung, whose unwavering prayers have been my anchor; and my siblings, Chuwang, Chundung, Kumbo, Mafeng, and Kaweng, for their love and constant support.

To my alma mater, the University of East London (UEL), I extend my special thanks to Ameer Al-Nemrat, Amin Karami, Mashal Alqudah and Marwan Alshar'e for their encouragement, advice, and professional insights, which have had a lasting impact on my academic journey.

Finally, I would like to thank my friends and colleagues, including Hamza Bayd, Karim Radouane, Nanlok Nimlang, Henry Ukwuoma, Nimisingha Amakama, Isaac Akawu, Aaron Yusuf, and Ajala Anang, for sharing this journey with me. Your friendship and support have been invaluable throughout this process.

DEDICATION

“Firstly, I want to dedicate this thesis to God. His infinite mercy has kept me in good health and in sound mind to complete this thesis. My short prayers during this challenging journey have always been Matthew 7:7: Ask, and it shall be given you; seek, and ye shall find; knock, and it shall be opened unto you”

“I also want to dedicate this thesis to my beloved family, whose unwavering love, support, and sacrifices kept me going through this incredible journey.

To my parents, your guidance has been my compass through every challenge.”

TABLE OF CONTENTS

ABSTRACT 2
ACKNOWLEDGMENTS 3
DEDICATION 5
LIST OF FIGURES 11
LIST OF TABLES 13
LIST OF ABBREVIATIONS 14
RESEARCH ACTIVITIES 15
RÉSUMÉ DE LA THÈSE 17
CHAPTER 1 – INTRODUCTION 29
1.1 Introduction 29
1.2 Background 29
1.3 Statement of the problem 31
1.4 Aim 33
1.5 Research questions 33
1.6 Objectives 34
1.7 Artificial Intelligence (AI) and Machine Learning (ML) synopsis 35
1.7.1 Application of AI/ML in cybersecurity 37
1.7.1.1 Supervised approaches 37
1.7.1.2 Unsupervised approaches 39
1.7.1.3 Semi-supervised approaches 39
1.7.1.4 Reinforcement approaches 40
1.8 Structure of the document 40
1.9 Conclusion 41
CHAPTER 2 – STATE-OF-THE-ART 43
2.1 Introduction 43
2.2 Cybersecurity in the healthcare sector 43
2.2.1 Network and IT facilities in hospitals 43
2.2.2 Common cyberthreats 45
2.2.3 Current threat landscape 48
2.3 Regulatory and compliance requirements 51
2.3.1 General Data Protection Regulation (GDPR) 51
2.3.2 Health Insurance Portability and Accountability Act (HIPAA) 52

2.3.3 Personal Information Protection and Electronic Documents Act (PIPEDA) 53
2.3.4 General Data Protection Law (LGDP) 54
2.4 Multilayered cybersecurity approaches 54
2.4.1 Defense in depth 57
2.4.2 Zero trust architecture 59
2.5 Database technologies 60
2.5.1 Application in other sectors 60
2.6 AI/ML related work and use-cases 65
2.6.1 AI/ML in phishing detection 65
2.6.2 AI/ML in Network Intrusion Detection Systems (NIDS) 67
2.6.3 AI/ML in malware detection 69
2.7 Cybersecurity risk assessment frameworks 70
2.7.1 ISO/IEC 27000 series 70
2.7.2 NIST CSF 72
2.8 Resilience considerations 74
2.8.1 Resilience implications in this work 79
2.9 Gap in research 80
2.10 Conclusion 81
CHAPTER 3 – METHODOLOGY 83
3.1 Introduction 83
3.2 Research methodology 83
3.3 Research approach and justification 83
3.3.1 Research category: quantitative vs. qualitative 84
3.3.2 Nature of study: descriptive vs analytical 85
3.3.3 Purpose of study: applied vs. fundamental 85
3.3.4 Research design: exploratory vs. conclusive 85
3.3.5 Data type: primary vs. secondary 86
3.4 Methodology 87
3.5 Tools and frameworks 91
3.5.1 Visual studio code 91
3.5.2 Anaconda distribution environment 91
3.5.3 MySQL 93
3.5.4 Flask 93

3.6 Principles of AI/ML algorithms used....93
3.7 Interaction between tools....95
3.8 Mapping resilience phases to control, impact, and organizational factors....96
3.9 Conclusion....97
CHAPTER 4 – SYSTEM DESIGN....98
4.1 Introduction....98
4.2 Analysis and design....98
4.3 System Development Life Cycle (SDLC) methodology....99
4.3.1 SDLC Methodology....99
4.3.2 Throwaway prototyping....100
4.3.3 Justification for methodology selection....101
4.4 Requirements engineering....102
4.4.1 Object oriented analysis and design....102
4.4.2 Functional requirements....103
4.4.3 Use case modeling....103
4.4.4 Sequence diagram modeling....105
4.5 Database design....110
4.5.1 Conceptual clarification (database vs dataset)....110
4.5.2 Data sources....112
4.5.2.1 Office for Civil Rights (OCR)....112
4.5.2.2 Center for International and Security Studies Maryland (CISSM)....113
4.5.2.3 The European Repository of Cyber Incidents (EuRepoC)....114
4.5.2.4 Codification of cyberattack incidents....115
4.6 Outcomes and visual analysis from developed web application and proposed database 116
4.7 Conclusion....120
CHAPTER 5 – CASE STUDIES AND APPLICATIONS, RESULTS, DISCUSSIONS AND ANALYSIS 121
5.1 Introduction....121
5.2 Case study 1: Risk assessment (Baze University Hospital)....121
5.2.1 BUH cyber setup and operations....123
5.2.2 Proposed risk assessment framework....124
5.2.2.1 Scope definition....125
5.2.2.2 Assets identification....125

5.2.2.2.1 Assets description 125
5.2.2.2.2 Assets value 128
5.2.2.3 Threats identification 128
5.2.2.3.1 MITRE ATT&CK® 129
5.2.2.3.2 Threat rating 129
5.2.2.4 Vulnerability identification 130
5.2.2.4.1 Common Vulnerabilities and Exposures (CVE) 131
5.2.2.5 Risk Determination 131
5.2.2.6 Risk ranking and control 135
5.3 Case Study 2: AI/ML in anomaly detection 141
5.3.1 Dataset description 142
5.3.2 Attack scenarios 143
5.3.3 AI/ML algorithms used for anomaly detection 144
5.3.4 Methodology and performance metrics 148
5.3.5 Parameter settings for the models used 149
5.3.6 Visualizations, results and discussions 152
5.4 Case study 3: AI/ML in website phishing detection 157
5.4.1 Dataset and description 158
5.4.2 Proposed system 160
5.4.3 Justification for algorithm used 161
5.4.4 Support Vector Machines (SVM) 162
5.4.5 Parameter settings for the experiment 163
5.4.6 Visualizations, results and discussions 164
5.5 Case Study 4: AI/ML in intrusion detection 168
5.5.1 Used datasets and description 170
5.5.2 AI/ML algorithms used for intrusion detection 173
5.5.3 The proposed cross-ML framework 173
5.5.4 Common features categorization 175
5.5.5 Encoding, normalization and split 176
5.5.6 Training and testing procedure 177
5.5.7 Parameter settings for the experiment 177
5.5.8 Results and discussions 178
5.6 Conclusion 186

CHAPTER 6 – CONCLUSION 187
6.1 Introduction 187
6.2 Contributions 187
6.3 Future work 189
6.4 Conclusion 190
REFERENCES 191
APPENDICES 212

LIST OF FIGURES

Fig. 1.1: ML approaches [16] 37
Fig. 2.1: Hospital Network Architecture [27] 44
Fig. 2.2: Average cost of data breaches worldwide by industry, 2022 vs. 2021(in dollars) [43] 48
Fig. 2.3: Multilayered Cybersecurity approach [78] 55
Fig. 2.4: Defense in depth [79] 58
Fig. 2.5: Core zero trust logical components [88] 59
Fig. 2.6: Use-case application of AI/ML in email phishing detection [124] 66
Fig. 2.7: Use-case application of AI/ML in email NIDS [118] 68
Fig. 2.8: Use-case application of AI/ML malware detection [118] 69
Fig. 2.9: ISO 27005 Risk management process 71
Fig. 2.10: NIST CSF 73
Fig. 3.1: Types of Research Methodology [172] 84
Fig. 3.2: High-level abstraction of methodology 90
Fig. 3.3: Interaction between tools and frameworks 96
Fig. 4.1: SDLC. [186] 99
Fig. 4.2: Throwaway prototyping [186] 101
Fig. 4.3: Justification for methodology selection [186] 102
Fig. 4.4: Use case diagram 104
Fig. 4.5: Predict phishing attack 105
Fig. 4.6: Search cyberattack incident 106
Fig. 4.7: Add new cyberattack incident 107
Fig. 4.8: View statistics of cyberattack incidents 108
Fig. 4.9: API of cyberattack incidents 109
Fig. 4.10: Delete cyberattack incidents 110
Fig. 4.11: Cyberattack incidents in MySQL 112
Fig. 4.12: GUI of cyberattack database 117
Fig. 4.13: Filtering parameters for database queries 118
Fig. 4.14: Dynamic filtering of results from database based on user input 119
Fig. 4.15: Distribution of cyberattack incidents 119
Fig. 5.1: BUH Cyber Operations and processes 123
Fig. 5.2: Baze University Hospital Cyber Risk Assessment Framework (BUHCRAF) 124

Fig. 5.3: Number of Assets in BUH 127
Fig. 5.4: Categorization of Assets based on ENISA 127
Fig. 5.5: Asset Values and Impacts of Incident 128
Fig. 5.6: Likelihood of Threat 130
Fig. 5.7: ECU-IoHT Testbed Architecture [11]. 143
Fig. 5.8: Proposed system 148
Fig. 5.9: 3D and 2D t-SNE EDA 153
Fig. 5.10: Comparison of Evaluation Metrics 154
Fig. 5.11: Confusion matrices 157
Fig. 5.12: Proposed System 160
Fig. 5.13: EDA result 165
Fig. 5.14: Output of phishing plot using SVM 167
Fig. 5.15: Proposed Cross-ML Framework 173
Fig. 5.16: Common features Ton-IoT, X-IIoTID and UNSW-NB15 175
Fig. 5.17: Performance metric Ton-IoT, X-IIoTID and UNSW-NB15 179
Fig. 5.18: Combined model results 180
Fig. 5.19: Cross ML performance 181
Fig. 5.21: Confusion matrices of all three datasets 184
Fig. 5.22: ROC curve of all three datasets 185

LIST OF TABLES

Table 2.1: Comparison of Incident Databases 64
Table 4.1: Dataset vs Database Differences 111
Table 4.2: Cyberattack incident types and definitions 116
Table 5.1: Department, services and description 122
Table 5.2: Assets, Categories and Description 125
Table 5.3: Risk determination 131
Table 5.4: Risk register and control 136
Table 5.5: Dataset Description 142
Table 5.6: Features and Definition 158
Table 5.7: Samples from ToN-IoT dataset 170
Table 5.8: Samples from X-IIoTID dataset 171
Table 5.9: Samples from UNSW-NB15 dataset 172
Table 5.10: Common features and description 176

LIST OF ABBREVIATIONS

ATT&CK	Adversarial Tactics, Techniques, and Common Knowledge
AI	Artificial Intelligence
ANN	Artificial Neural Networks
BUH	Baze University Hospital
BUHCRAF	Baze University Hospital Cyber Risk Assessment Framework
CISSM	Center for International and Security Studies Maryland
CNN	Convolutional Neural Networks
CVE	Common Vulnerabilities and Exposure Database
DDoS	Distributed Denial of Service
DoS	Denial of Service
DT	Decision Trees
EDA	Exploratory Data Analysis
EHR	Electronic Health Records
EMR	Electronic Medical Records
EuRepoC	European Repository of Cyber Incidents
FAA	Federal Aviation Administraion
GB	Gradient Boosting
GDPR	General Data Protection Regulation
GUI	Graphical User Interface
HIPAA	Health Insurance Portability and Accountability Act
IDE	Integrated Development Environment
ML	Machine Learning
MLP	Multilayer Perceptron
NASA	National Aeronautics and Space Administation
NIDS	Network Intrusion Detection System
NIST CSF	National Institute of Standards Cybersecurity Framework
OCR	Office for Civil Rights
OO	Object Oriented
OOA	Object Oriented Analysis
OOD	Object Oriented Design
PHI	Protected Health Information
PII	Personal Identifiable Information
PIPEDA	The Personal Information Protection and Electronic Documents Act
RF	Random Forest
SDLC	Systems Development Life Cycle
SVM	Support Vector Machines
UML	Unified Modeling Language
ZTA	Zero Trust Architecture

RESEARCH ACTIVITIES

JOURNAL ARTICLE

- Shombot, E. S., Dusserre, G., Bestak, R., & Ahmed, N. B. (2025). Maximizing healthcare security outcomes through AI/ML multi-label classification approach on IoHT devices. Health Technol. (2025). https://doi.org/10.1007/s12553-025-00963-x

- Shombot, E. S., Dusserre, G., Bestak, R., & Ahmed, N. B. (2024). An application for predicting phishing attacks: A case of implementing a support vector machine learning model. *Cyber Security and Applications*, *2*, 100036. https://doi.org/10.1016/j.csa.2024.100036

CONFERENCE PUBLICATION

- Shombot, E. S., Dusserre, G., Bestak, R., & Ahmed, N. B. (2024b). Cross ML for Io(H)T Network Traffic Classification: A New Approach Towards Standardization. *2024 8th Cyber Security in Networking Conference (CSNet)*, 282–288. https://doi.org/10.1109/CSNet64211.2024.10851757

- Shombot, E. S., Dusserre, G., Bestak, R., & Ahmed, N.-B. (2023b, September 11). *State-of-the-art Analysis of Cyberattack Trends in the Healthcare Sector*. TIEMS 2023 - The International Emergency Management Society Hybrid Annual Conference. https://imt-mines-ales.hal.science/hal-04208536

- Shombot, E. S., Dusserre, G., & Bestak, R. (2023, November 8). Prediction and evaluation of cyberattack databases with AI tools in healthcare sector. Séminaire IMT « Data Analytics & Artificial Intelligence ». https://imt-mines-ales.hal.science/hal-04316053

JOURNAL ARTICLE (CO-AUTHORSHIP)

- Thawbaan, A., Shombot, E. S., Dusserre, G., Ahmed, N. B., Babatunde, Z., Isa, L. M., & Job, D. A. (2024). Chatbots in Cybersecurity: Enhancing Security Chatbot Efficacy through Iterative Feedback Loops and User-Centric Approaches. *American Journal of Innovation in Science and Engineering*, *3*(3), Article 3. https://doi.org/10.54536/ajise.v3i3.2919

- Yusuf, M. K., Danladi, A. J., Shombot, E. S., Dusserre, G., Odey, V. A., Baba-Ahmed, N. B., Bestak, R., & Lawan, M. I. (2024). The Growing Cybersecurity Crisis in Healthcare: A Call to Action. *American Journal of Innovation in Science and Engineering*, *3*(3), Article 3. https://doi.org/10.54536/ajise.v3i3.3576

COLLOQUIUM

- Shombot, E. S., Dusserre, G., Bestak, R., & Ahmed, N.-B. (2023a, April 13). *Cyber Security in Healthcare: Utilizing Artificial Intelligence and Database Technology to Enhance Understanding of Cyberattacks*. Colloque IMT 2023 : « Sécurité et Résilience ». https://doi.org/10.5281/zenodo.7937270

- Shombot, E. S., Dusserre, G., Daclin, N., & Olivaux, M. (2022, March 31). *Artificial Intelligence in Critical Infrastructures: AI Paradigm for Enhanced Cyber Attack Detection in Critical Infrastructures*. Colloque IMT - Gestion de crise et numérique : nouvelles menaces et nouvelles solutions. https://hal-emse.ccsd.cnrs.fr/emse-03683423

RÉSUMÉ DE LA THÈSE

Cette thèse explore des méthodes innovantes pour améliorer la posture de cybersécurité des organismes de santé. Dans un contexte d'escalade des menaces, en particulier après que Covid-19 a révélé la cyber-fragilité de la plupart des établissements de santé, les incidents de cyber-attaques enregistrés ont connu un pic significatif. Le secteur de la santé reste très vulnérable aux cyberattaques. Le secteur de la santé reste très vulnérable aux cyberattaques. Les raisons en sont la dépendance excessive à l'égard de la technologie numérique pour les activités quotidiennes au caractère lucratif de l'obtention de données sensibles sur les patients. Cette étude identifie les défis majeurs, notamment les infrastructures obsolètes, le manque de préparation et la sophistication croissante des cybercriminels.

La recherche adopte une approche de cybersécurité à plusieurs niveaux qui intègre la technologie des bases de données, l'intelligence artificielle (IA), Apprentissage automatique (ML) et les méthodologies d'évaluation des risques. Les bases de données se sont révélées efficaces pour le stockage, la gestion et l'extraction des données; par conséquent, le catalogage des incidents historiques de cyberattaques permettra de reconnaître des schémas qui conduiront à l'élaboration de solutions de sécurité proactives.

Comme les modèles de menaces sont établis à partir de l'historique des incidents, des mesures concrètes doivent être prises, ce qui conduit cette recherche à présenter le cadre d'évaluation des risques cybernétiques de l'hôpital universitaire de Baze (BUHCRAF), qui combine des éléments de la série ISO/IEC 27000 et du cadre de cybersécurité du NIST afin d'identifier, d'évaluer et d'atténuer les risques dans un contexte pratique de soins de santé. Certaines de ces stratégies d'atténuation sont fondées sur des politiques, tandis que d'autres nécessitent des contrôles techniques. Pour celles qui nécessitent des contrôles techniques, ce travail a adopté une approche de cas d'utilisation pour mettre en œuvre des techniques d'IA/ML dans la détection d'anomalies, la prévention de l'hameçonnage et les systèmes de détection d'intrusion tout en prenant en compte l'hétérogénéité des données et en améliorant la généralisabilité grâce à un cadre Cross-ML proposé.

La thèse est divisée en six chapitres. Le premier chapitre est l'introduction, le deuxième chapitre est l'état de l'art, le troisième chapitre est la méthodologie, le quatrième chapitre est la conception du système, le cinquième chapitre est constitué d'études de cas et d'applications, de résultats, de discussions et d'analyses, et le sixième chapitre est la conclusion.

Dans le premier chapitre, nous avons détaillé le contexte de l'étude dans les deux sous-domaines informatiques de l'intelligence artificielle (IA), de Apprentissage automatique (ML) et de la cybersécurité dans lesquels s'inscrit notre recherche. Le chapitre commence par souligner la fréquence croissante et la sophistication des cyberattaques, en particulier dans le secteur des soins de santé qui dépend fortement des systèmes informatiques pour ses activités quotidiennes. Le chapitre montre également à quel point le secteur des soins de santé est mal équipé en raison d'une technologie obsolète, de mesures de sécurité inadéquates et d'une attitude essentiellement réactive face aux menaces émergentes. Étant donné que les problèmes auxquels est confronté le secteur des soins de santé sont multiples, l'objectif de la recherche, tel qu'il est formulé dans le chapitre, est de mettre en œuvre une approche de cybersécurité à plusieurs niveaux adaptée au secteur des soins de santé. Cette approche s'appuie sur la technologie des bases de données, l'intelligence artificielle (IA), les techniques d'apprentissage automatique et la stratégie d'évaluation des risques. Étant donné que les différentes approches de cybersécurité sont conçues pour des scénarios de menace différents et spécifiques, cette thèse tient compte de cette particularité. Cette thèse tient compte de cette particularité en proposant une approche de cybersécurité multicouche à trois niveaux. Par conséquent, le chapitre synthétise les objectifs du travail, qui impliquent l'analyse des incidents de cyberattaques existants à partir de dossiers publics qui représentent le paysage des menaces, l'évaluation et la mise en œuvre de techniques d'IA/ML pour la détection des cyberattaques, et l'incorporation de stratégies d'évaluation des risques pour renforcer la position de la cybersécurité dans les soins de santé.

Le deuxième chapitre est consacré à l'état de l'art. Il présente un examen approfondi des travaux de recherche antérieurs et présente des bases théoriques pour la thèse. L'objectif

est de s'assurer que tous les fondements théoriques pertinents et les concepts connexes autour du domaine problématique établi sont examinés en profondeur. Parmi les thèmes abordés, nous avons examiné la configuration typique de cybersécurité des hôpitaux, en mettant l'accent sur l'infrastructure informatique, les vulnérabilités et les défis inhérents à la protection des dispositifs connectés tels que les dossiers médicaux électroniques (EHR), car ils semblent être parmi les ressources les plus ciblées pour l'exploitation. L'analyse des structures des réseaux hospitaliers révèle des éléments clés : les réseaux médicaux, les patients et les réseaux de médecins privés, qui présentent tous des risques pour la cybersécurité.

Le chapitre décrit les cybermenaces courantes, notamment les logiciels malveillants, le phishing, les attaques DoS/DDoS, les ransomwares et les menaces internes, en tant que types de menaces courantes responsables de cyberattaques notables telles que le ransomware WannaCry, la violation de l'hôpital universitaire de Brno, la violation du système de suivi de Kaiser Permanente et la violation des données de HCA Healthcare, mettant ainsi en évidence le paysage croissant des menaces et les coûts associés aux violations de données.

La recherche dans le domaine des soins de santé peut parfois être limitée par les stipulations des cadres réglementaires établis en raison de la confidentialité et de la sensibilité des données. C'est pourquoi nous avons examiné certains cadres réglementaires tels que le GDPR, HIPAA, PIPEDA et LGDP, afin de nous assurer que nous avons pris en compte les exigences réglementaires dans notre thèse.

Le chapitre traite également des approches de cybersécurité multicouches en tant que procédures holistiques de cybersécurité qui impliquent la formulation de plusieurs couches d'outils et de techniques de défense pour protéger de manière tenace les composants de l'infrastructure informatique ou cybernétique d'une organisation. Cette stratégie repose sur l'idée qu'aucun mécanisme de défense n'est une solution unique pour répondre à tous les problèmes de sécurité qui se posent. Une organisation peut réduire le risque d'un point de défaillance unique et se défendre contre un plus grand nombre de menaces en utilisant

plusieurs couches de défense. La sécurité multicouche comprend les mesures de sécurité physique qui limitent l'accès aux systèmes informatiques (agents de sécurité, portes verrouillées et clôtures renforcées), les mesures de sécurité technique qui protègent les systèmes de réseau à l'aide de matériel et de logiciels spécialisés (pare-feu, protection par mot de passe et logiciel antivirus) et les mesures de sécurité administrative qui consistent en des politiques ou des procédures destinées aux employés de l'organisation (formation de sensibilisation à la cybersécurité et évaluation des risques). Toutes ces couches servent à répondre à des problèmes de sécurité spécifiques dans l'infrastructure informatique d'une organisation.

En outre, nous avons exploré des travaux connexes qui utilisent des technologies de base de données pour cataloguer les incidents, comme on l'a vu dans d'autres secteurs tels que l'aviation (base de données de la Federal Aviation Administration et de la National Aeronautics and Space Administration) et la cybersécurité (MITRE ATT&CK). Cette approche du catalogage a contribué à améliorer la sécurité, à analyser les incidents et à améliorer la réponse de la direction, une pratique qui peut être adaptée à la cybersécurité des soins de santé.

Ce chapitre traite également des cadres d'évaluation des risques en matière de cybersécurité et des différentes adaptations disponibles pour évaluer l'adéquation au contexte des soins de santé. Nous avons examiné de près la série ISO/IEC 27000, largement utilisée et communément appelée famille de normes ISMS, qui fournit les meilleures pratiques et recommandations pour la gestion de la sécurité de l'information, et le cadre de cybersécurité du National Institute of Standards and Technology (NIST NCF), conçu pour aider les organisations à améliorer leur position en matière de cybersécurité en identifiant les lacunes exploitables en matière de cybersécurité dans les infrastructures de toutes tailles.

Le chapitre s'est également penché sur les applications de l'IA/ML dans le domaine de la cybersécurité. Notre enquête a permis de déduire que les applications pratiques sont principalement des techniques supervisées avec des algorithmes tels que les random

forests, le gradient boosting, les decision trees, les support vector machines, etc. et que les cas d'utilisation spécifiques sont la détection du phishing, les systèmes de détection d'intrusion dans les réseaux (NIDS) et la détection des logiciels malveillants. Nous avons également constaté que les ensembles de données permettant d'effectuer ces tâches sont limités en raison de contraintes liées à la protection de la vie privée ou sont tout simplement trop anciens pour les menaces actuelles.

L'étude de l'état de l'art s'est avérée utile pour nous aider à identifier les lacunes de la recherche, telles que l'absence de stratégies de cybersécurité contextuelles adaptées aux soins de santé et la disponibilité limitée d'ensembles de données pour les applications d'IA/ML en général et dans les soins de santé, ce qui situe davantage notre recherche dans le contexte approprié de la littérature existante.

Le troisième chapitre détaille la méthodologie employée pour atteindre les objectifs de la recherche et répondre aux questions posées. L'objectif de ce chapitre est notamment de garantir l'objectivité et la reproductibilité des travaux de recherche. Les critères clés que nous avons pris en compte sont que la méthodologie est la plus appropriée pour atteindre les objectifs et qu'elle peut être reproduite pour des recherches similaires de même nature.

D'un point de vue fonctionnel, nous simplifions notre méthodologie en trois parties distinctes. (1) Sources/fournisseurs de données : Cette partie a permis d'identifier une source primaire de données pour la base de données proposée, qui s'est appuyée sur des bases de données existantes et des rapports d'actualité sur les cyberincidents dans le secteur des soins de santé, que nous avons pu identifier auprès du Center for International and Security Studies Maryland (CISSM) avec plus de 1500 incidents liés aux soins de santé et la base de données de l'Office for Civil Rights (OCR) avec plus de 800 incidents signalés et ayant fait l'objet d'une enquête. La source de données secondaires utilisée dans cette recherche est plus spécifique et adaptée aux tâches d'IA/ML. Nous avons utilisé des données primaires pour développer une application web qui peut prendre en charge les opérations d'ajout de nouveaux enregistrements, de recherche d'enregistrements, de suppression d'enregistrements et de visualisation des enregistrements des cyberincidents,

ainsi qu'une API ouverte pour rendre public le contenu de la base de données. (2) Méthodes et techniques d'IA/ML : À partir des sources de données secondaires, nous avons utilisé diverses techniques d'IA/ML. Nous avons utilisé l'ensemble de données sur l'hameçonnage pour classer les attaques d'hameçonnage, l'ensemble de données ECU-IoHT pour la détection des anomalies, l'ensemble de données Ton-IoT, l'ensemble de données UNSW-NB15 et l'ensemble de données X-IIoTID pour la détection des intrusions dans le réseau. (3) Évaluation des risques : Au-delà de la construction d'une base de données qui permet des manipulations fonctionnelles pour présenter le paysage des menaces des cyberincidents, les résultats de la base de données sont utilisés dans le cadre de ce travail avec les résultats de l'application des techniques d'IA/ML pour alimenter un modèle d'évaluation des risques. Pour ce faire, nous nous appuyons sur des cadres largement utilisés tels que le cadre NIST, la série ISO 27000 et le cadre ENISA pour la catégorisation des cyberbiens dans le secteur des soins de santé. Notre méthodologie est d'extraire les éléments applicables de ces cadres et propose un cadre adapté à l'hôpital universitaire de Baze, qui est l'établissement de soins de santé adopté comme cas d'utilisation.

La méthodologie de recherche de cette étude utilise divers outils et cadres majeurs pour faciliter le processus de recherche. Ces outils contribuent à la collecte et à l'analyse des données, ainsi qu'à la réalisation des objectifs généraux de la recherche. Les justifications nécessaires sont fournies pour expliquer pourquoi certains outils sont préférés à d'autres, et la déclaration de ces outils dans ce chapitre est importante pour la reproductibilité de l'expérimentation.

Le chapitre quatre porte sur la conception du système et couvre l'application et la base de données permettant de cataloguer les incidents de cyberattaque que nous obtenons auprès de nos principaux fournisseurs de données (CISSM et OCR). Pour développer l'application, nous avons utilisé l'approche du cycle de vie du développement des systèmes (SDLC). Bien que le SDLC comprenne diverses méthodologies telles que la chute d'eau, le modèle en V et le développement agile, cette recherche adopte l'approche du prototypage jetable en raison de sa flexibilité, des boucles de rétroaction itératives et de sa capacité à répondre à des exigences floues. Nous avons formulé des exigences fonctionnelles afin

d'établir la fonctionnalité de base nécessaire pour qu'un système soit utile, et nous avons pu dériver six exigences fonctionnelles de l'application développée. (1) Le système doit permettre à un utilisateur de prévoir les attaques d'ingénierie sociale (phishing). (2) Le système doit permettre à un utilisateur de rechercher une attaque dans la base de données. (3) Le système doit permettre à un utilisateur d'ajouter de nouvelles attaques à la base de données. (4) Le système doit permettre à un utilisateur de consulter les statistiques relatives à la diffusion des attaques. (5) Le système doit permettre à un utilisateur de consulter et d'obtenir des enregistrements de la base de données via une API publique. (6) Le système doit permettre à un utilisateur de supprimer les enregistrements de cyberattaques de la base de données. Nous avons utilisé le langage de modélisation unifié (UML) pour élaborer un modèle de cas d'utilisation qui montre comment un utilisateur interagit avec le système et le diagramme de séquence, qui est un modèle comportemental qui capture les séquences temporelles entre les objets et le processus impliqué dans la mise en œuvre d'une fonctionnalité donnée. En conclusion, ce chapitre démontre la réussite de la conception et de la mise en œuvre d'une base de données et d'une application web adaptées au secteur des soins de santé, intégrant des ensembles de données publiques et des méthodologies orientées objet afin de fournir des informations précieuses sur les modèles de cyberattaques.

Le cinquième chapitre est consacré aux études de cas, aux applications, aux résultats, aux discussions et à l'analyse. Dans ce chapitre, nous avons présenté l'approche de cybersécurité multicouche proposée pour le secteur des soins de santé avec ses applications pratiques et ses études de cas, étayées par des discussions détaillées sur les résultats et l'analyse. Plus précisément, nous nous sommes concentrés sur quatre études de cas importantes pour montrer l'application pratique dans les établissements de santé modernes.

La première étude de cas est l'évaluation des risques de l'infrastructure cybernétique de l'hôpital universitaire de Baze (BUH), qui est basée sur l'intégration des composantes des lignes directrices ISO27000 et NIST CSF. Nous avons utilisé le cadre de l'Enisa pour catégoriser les actifs informatiques dans les soins de santé, et notre visite à l'établissement

de santé de l'université de Baze a révélé que leurs actifs informatiques sont constitués à 29 % d'équipements de réseau, à 7 % de bâtiments et d'installations, à 7 % de données, à 21 % de systèmes d'information clinique interconnectés, à 7 % d'appareils clients mobiles et à 29 % d'appareils médicaux en réseau. Cette approche nous a permis d'identifier les actifs, les menaces et les vulnérabilités. Les vulnérabilités sont des points faibles sur lesquels les cybercriminels s'appuient généralement pour perpétrer des attaques. Comme nous avons réussi à catégoriser leurs cyberactifs, nous avons ensuite proposé le cadre d'évaluation des cyberrisques de l'hôpital universitaire de Baze (BUHCRAF), qui se compose de cinq étapes. (1) Définition du champ d'application pour établir les limites de ce que l'évaluation des risques couvrira. Nous avons limité l'évaluation des risques à la cyberinfrastructure de l'hôpital. Tous les autres actifs, tels que les bâtiments, les automobiles, les fauteuils roulants et les autres actifs physiques, ont été exclus. (2) L'étape d'identification des actifs, que nous définissons comme le matériel, les logiciels, les réseaux ou les substances ayant une valeur matérielle pour l'hôpital universitaire de Baze. Dans cette phase, nous avons également utilisé une matrice en 5 points pour déterminer la valeur des actifs. (3) L'étape d'identification des menaces, que nous définissons comme des événements potentiels ou des agents agissant de l'extérieur qui peuvent endommager ou compromettre les actifs. (4) La phase d'identification des vulnérabilités est définie comme des failles ou des faiblesses inhérentes aux actifs où les menaces peuvent être exploitées pour compromettre la confidentialité, la disponibilité et l'intégrité de l'actif. Nous avons également utilisé la base de données CVE (Common Vulnerability and Exposure database) pour contextualiser notre évaluation de la vulnérabilité identifiée pour l'hôpital universitaire de Baze (BUH) avec les évaluations déjà signalées et leur gravité correspondante. (5) L'étape de détermination du risque, au cours de laquelle nous avons utilisé la feuille de calcul Menaces-vulnérabilité-actifs (TVA) qui présente logiquement la liste des actifs, leurs menaces et les vulnérabilités correspondantes, ainsi qu'un registre des risques qui catégorise le risque d'élevé à faible, calculé comme suit : Risque = Actif x Score de menace x Score de vulnérabilité.

Nous avons ensuite présenté les mécanismes techniques qui peuvent être utilisés pour atténuer ces risques. Bien que, d'après le registre des risques, différents contrôles

physiques, techniques et administratifs puissent être adoptés pour faire face à certains des risques élevés, nous en avons sélectionné quelques-uns qui nécessitent des contrôles techniques.

La deuxième étude de cas porte sur l'application de l'IA/ML à la détection d'anomalies dans l'environnement de l'Internet des objets de santé (IoHT). L'IoHT est une adaptation spécifique du concept de l'IdO, spécialement conçue pour le secteur des soins de santé, où l'interconnectivité des objets sensoriels a pour seul objectif l'administration des services de santé aux patients. Bien que cela améliore la prestation des soins de santé, cela expose également les dispositifs IoHT à diverses menaces de cybersécurité. La détection des anomalies est utilisée pour identifier les écarts dans le comportement du système qui indiquent des attaques potentielles ou des irrégularités.

Nous avons utilisé l'ensemble de données ECU-IoHT, qui se concentre sur les cyberattaques dans le secteur des soins de santé, en ciblant spécifiquement les appareils IoHT. L'ensemble de données contient 111 000 enregistrements et 9 caractéristiques, collectés entre février et juin 2020. L'ensemble de données comprend du trafic normal et du trafic d'attaque (classes d'attaque Nmap, DoS, ARP et smurf).

L'environnement de l'expérience comprend (1) un ordinateur portable, HP EliteBook Pro, x360 8th Gen, 500GB SSD, Intel Core i7, CPU 2.11GHz, et RAM 16GB. (2) Anaconda Distribution Version 3 (Jupyter Lab-6.4.12 fonctionnant sur localhost 8888, Python 3.9, pandas, SKlearn et matplotlib).

Nous avons expérimenté et comparé quatre algorithmes de ML qui n'avaient pas été pris en compte à l'origine dans la recherche menée. Les algorithmes sélectionnés sont Random Forest (RF), Gradient Boosting (GB), Decision Trees (DT) et Multilayer Perceptron (MLP). En utilisant les principales mesures de performance et d'évaluation telles que l'exactitude, la précision, le rappel et le score F1, le gradient boosting a obtenu de meilleurs résultats que les autres algorithmes avec plus de 99,53 %. Les résultats de la deuxième étude de cas nous ont permis de démontrer l'applicabilité de l'IA/ML dans les environnements IoHT, étant

donné que les IoHT font partie des actifs qui peuvent être compromis par des criminels pour accéder aux établissements de soins de santé.

La troisième étude de cas porte sur l'application de l'IA/ML à la détection de l'hameçonnage. L'étude de cas s'appuie sur les travaux d'Abdelhamid et al. (2014), qui ont rassemblé un ensemble de données de 1353 sites web provenant de diverses sources et identifié plusieurs caractéristiques associées aux sites légitimes et aux sites d'hameçonnage. Les sites de phishing proviennent des archives de données Phishtank (www.phishtank.com), un site web communautaire qui permet aux utilisateurs de soumettre, de vérifier, de suivre et de partager des données sur le phishing. Les sites web légitimes proviennent de Yahoo et des annuaires de points de départ. En intégrant le script PHP à un navigateur, les auteurs ont pu rassembler 548 sites web valides sur un total de 1353. L'ensemble de données comprenait 702 URL d'hameçonnage et 103 URL suspectes. Les caractéristiques de l'ensemble de données comprennent l'URL de la requête, l'âge du domaine, HTTPS et SSL, le trafic du site web, l'URL longue, SFH, la fenêtre pop-up, l'URL d'ancrage, l'URL de redirection et l'adresse IP. Nous avons utilisé le même environnement d'expérimentation que pour la deuxième étude de cas. L'étude a utilisé des machines à vecteurs de support (SVM) pour la classification, en s'appuyant sur deux fonctions de noyau : la fonction de base radiale (RBF) et les noyaux polynomiaux pour classer les attaques. Le noyau RBF cartographie les données d'entrée dans un espace à plus haute dimension pour améliorer la séparabilité, tandis que le noyau polynomial transforme les données d'entrée à l'aide de fonctions polynomiales pour faire la distinction entre les classes. Les deux noyaux ont été appliqués pour classer l'ensemble de données dans les catégories hameçonnage, suspect et légitime.

Bien que mineurs, nos résultats ont montré que le noyau polynomial était plus performant que le noyau RBF, avec une précision de classification de 84,5 % contre 82,6 %. Les deux noyaux ont démontré leur capacité à différencier efficacement les cas d'hameçonnage, le noyau polynomial excellant dans la gestion des modèles complexes de l'ensemble de données.

Notre dernière étude de cas porte sur l'IA/ML dans la détection des intrusions. L'objectif de cette étude de cas est d'explorer l'apprentissage inter-machines (Cross-Machine Learning - Cross-ML), qui consiste essentiellement à utiliser des données provenant de diverses sources pour généraliser le trafic réseau. Par conséquent, cette étude de cas présente un cadre clair pour le Cross-ML et montre comment l'utiliser sur trois ensembles de données accessibles au public: UNSW-NB15, X-IIoTID et ToN-IoT. Le concept de Cross-ML, tel qu'il est présenté dans l'étude de cas, fait référence à une méthode claire et reproductible pour les phases d'entraînement et de test des modèles de ML. L'idée qui sous-tend ce concept est que les modèles de ML ont tendance à dériver lorsqu'ils sont déployés en production. La formation croisée consiste donc à faire travailler l'ensemble du pipeline de ML sur différents ensembles de données, par opposition à la formation standard, qui se concentre sur un seul, avec la probabilité d'améliorer la capacité de généralisation du modèle. L'évaluation de la capacité de généralisation et des performances d'un modèle sur de nouvelles données plus représentatives des données examinées dans les scénarios du monde réel.

Comme condition préalable à la ML croisée, les caractéristiques de l'ensemble de données doivent être similaires. Pour ce faire, ce travail a comparé les caractéristiques communes à tous les ensembles de données. Pour garantir la cohérence et avoir des caractéristiques communes, nous avons codé et normalisé les valeurs des trois ensembles de données. La procédure de formation et de test pour le cadre cross-ML que nous proposons est la suivante. (1) ToN-IoT entraîné a testé ToN-IoT, X-IIoTID entraîné a testé X-IIoTID, et UNSW-NB15 entraîné a testé UNSW-NB15. D'après les expériences, nos résultats sont satisfaisants ; les résultats de la classification ont enregistré des scores très élevés, allant de 92,0 % à 98,1 % en termes de score F1 sur les trois ensembles de données.

(2) L'ensemble de données combiné entraîné a été testé sur l'ensemble de données de test combiné, l'ensemble de données combiné entraîné a été testé sur ToN-IoT, l'ensemble de données combiné entraîné a été testé sur X-IIoTID, et l'ensemble de données combiné entraîné a été testé sur UNSW-NB15. Le modèle combiné sur les trois ensembles de données dans leur ensemble et l'évaluation des résultats montrent une précision, un rappel

et un score F1 de 90,9 %, 84,0 % et 84,9 %, respectivement. Ensuite, le modèle combiné entraîné est testé sur les ensembles de données de tests individuels. Comme on pouvait s'y attendre, les performances du modèle diminuent considérablement par rapport à celles des modèles individuels ; toutefois, les scores de précision restent prometteurs, allant de 82,4 % à 97,8 %.

(3) ToN-IoT entraîné a testé X-IIoTID ; ToN-IoT entraîné a testé UNSW-NB15 ; X-IIoTID entraîné a testé ToN-IoT ; X-IIoTID entraîné a testé UNSW-NB15 ; et UNSW-NB15 entraîné a testé ToN-IoT ; UNSW-NB15 entraîné a testé X-IIoTID. Les résultats les plus remarquables de l'expérience cross-ML montrent que le modèle X-IIoTID sur ToN-IoT a obtenu les mesures les plus élevées parmi les évaluations cross-dataset, atteignant plus de 75 % de précision et un score F1 de 74 %.

La principale contribution de cette étude de cas est le cadre reproductible proposé pour la formation et le test des modèles ML dans divers ensembles de données, ce qui garantit une applicabilité et une résilience plus larges dans les environnements de réseau dynamiques.

Le chapitre six est la conclusion de cette thèse. Nous avons résumé la manière dont notre travail a jeté des bases très solides pour la question urgente de la cybersécurité dans le secteur des soins de santé. La recherche s'est appuyée sur des données historiques qui alimentent un cadre d'évaluation des risques sur mesure tout en utilisant des méthodes d'IA/ML pour traiter les menaces courantes telles que le phishing, la détection d'anomalies et la détection d'intrusions dans le réseau. La nature des cybermenaces étant dynamique, une innovation et une collaboration continues seront nécessaires. Nous espérons que le présent travail, en plus de satisfaire un intérêt académique curieux pour les méthodologies d'investigation exceptionnelles, pourra également devenir une incitation à lancer de nouvelles recherches et de nouveaux efforts pour protéger l'infrastructure numérique du secteur de la santé pour la sécurité et le bien-être non seulement des patients, mais aussi des professionnels de la santé.

CHAPTER 1 – INTRODUCTION

1.1 Introduction

This is the opening chapter of this work and provides a comprehensive overview of two important computing subfields of artificial intelligence (AI), machine learning (ML) and cybersecurity. The research examines how AI/ML tools can be applied to healthcare cyber security context to mitigate risk that often affect critical infrastructure such as hospitals. The chapter also entails the background of the research, the problem statement, aims and objectives, the research questions, and the overall structure of the thesis.

1.2 Background

Cybersecurity has become more delicate to administer for experts because of the exponential growth in volume and complexity of cyberattacks. This is further exacerbated by the fact that there is increasing dependence on digital information in hospitals today and, by consequence, poses the imperativeness of protecting digital assets from cyber attackers. According to [1], there are over 3 billion internet users today seeking to access different services including healthcare services. Also, most cultural, social, commercial and governmental activities carried out by individuals, corporations, governmental and non-governmental organizations are implemented on cyberspace. Furthermore, the internet has provided an enabling economic environment for billions of dollars to be exchanged annually and by this reality, many cybercriminals consider cyberspace viable for wreaking havoc and carting away with financial benefits.

The criticality of the security of the healthcare sector cannot be overemphasized in the event of a compromise in the security architecture of a typical hospital. Consequences are far-reaching because of the sensitivity of equipment, existential threat to lives, and disruptions in routine operations. As practitioners continually make efforts to combat

cybercrime in the global space, there are sufficient reasons to infer that the healthcare sector is underprepared in terms of capacity to deal with and respond to cybercriminals. The most recent pandemic of our lives; the COVID-19 provided a full-blown manifestation of such under preparedness in the healthcare sector as over 500 healthcare companies reported a variety of cyber breaches that compelled a shutdown in operations. This along with many other data security incidents provides reasonable grounds to assume that the healthcare sector needs to do more in terms of preparation, security, and response to cyberattacks [2].

The obscurity in the operations of a hospital's cyber infrastructure, justifiably so, resulting from the sensitivity of the kind of services they routinely provide, inherently erects a barrier for genuine security professionals to have a clear understanding of the actual state of their cybersecurity needs. Therefore, carrying out precise issue-based research to address specific issues in the context of emerging cyber threats in the healthcare sector becomes a daunting endeavor. Furthermore, security solutions have the reputation of being reactive: researchers at cybersecurity companies are saddled with the task of finding new cyber threat samples, analyzing them, and adding them to their threat lists. This certainly accounts for the inefficiency in dealing with new threat variants in healthcare cybersecurity architecture fortification. Although the industry continues to employ this strategy, it is now operating more proactively, particularly considering concerns from social engineering-based attacks.

Given all these realities, a revised approach from a reactive to a proactive strategy is crucial in the healthcare sector. One way of leveraging technology to achieve this expectation is the adoption of AI/ML. In fact, an author suggests [3] that AI/ML software should have the ability to learn from the consequences of past events in order to help predict and identify cybersecurity threats.

AI/ML has seen far more usability in areas that are heavily data reliant such as e-commerce and marketing because they have proven to be tremendously helpful for quickly automating decision-making processes and identifying patterns from incomplete or altered data. Even though they are not a one-stop solution for all cybersecurity issues, when a miniscule margin of error can be achieved with their deployment, it is conceivable that they will become indispensable for cybersecurity at scale. These algorithms function by initially learning from real-world information, such as current security risks, false positives, and the most recent dangers discovered by researchers around the world.

Banafa [3], emphasize that currently, cybersecurity is reactive implying that in most cases, it helps alert IT staff about data breaches, identity theft, suspicious applications, and suspicious activities however, conventional prevention methods lack the decisiveness of proactivity in halting attacks which can be a critical aspect of AI adoption. if they can possess the capacity of stopping Cyberattacks even before they happen. Another critical talking points the article emphasizes is how the industry is notoriously unbalanced, with the bad actors getting to pick from thousands of vulnerabilities to launch their attacks, along with deploying an ever-increasing collection of tools to evade detection once they have breached a system. While they only must be successful once, the security experts tasked with defending a system have to stop every attack, every time.

1.3 Statement of the problem

Hospitals provide essential services, but cyber-attackers pose considerable threat to optimum safety of their operations. Since the Corona virus pandemic, the Federal Bureau of Investigation (FBI) reported that there was an increase in cybercrime on medical facilities by over 400% [4]. According to another report by [5], the number of health organizations compromised by cyberattacks increased by 90% in the three months leading up to June 30 in comparison to the first three months of 2022. The 3,200 events from all industries that were reported to the consultant over the previous 12 months served as the basis for this

study. It is also worth mentioning that the National Health Service (NHS) suffered two major incidents in 5 years. The First was WannaCry Ransomware in 2017 costing over £92 million and led to the cancellation of over 19,000 patient appointment [6]. The second attack as reported by [7] happened on August 4th, 2022 and It impacted patient referrals, emergency medications, ambulance dispatch, after-hours appointment scheduling, and mental health services. Even though most hospitals will want to exercise discretion in publicizing every time their IT infrastructure is compromised, it will be a fair assertion to ascribe to that, from several security reports out there in the public domain, there is evidence to suggest how much of a growing cyberattack has being in the healthcare sector.

As hospitals continue to rely on data driven technology for carrying out their routine services efficiently, digitizing medical assets will create the right avenue for cyber criminals to exploit. This is because, there is over 430 million connected medical devices worldwide therefore, there is an expansion in the attack surface susceptible to criminal exploitation [8]. This issue is further compounded because the healthcare sector has built a reputation of projecting a lackadaisical posture when dealing with issues of cybersecurity. A study by [9] shows that more than 80% of medical imaging devices run on outdated operating systems that can easily be compromised by hackers due to inherent vulnerabilities of the Operating System. In fact, the security of patient information and the entirety of the health care cyber infrastructure as articulated by [10] cannot just be reduced to an ethical theme but also a first level national security priority.

Even though it can be argued that considerable progress has been made so far, however, measures of dealing with cybersecurity incidents in hospitals is still widely regarded as inadequate [11]. This research considers a different approach to the problem by gathering publicly accessible cyberattack incidents that occurred in the past, in the healthcare sector and building it into a comprehensive database. The reasoning for this approach is elaborated sufficiently in Chapter 2. is essentially because cybersecurity incidents in hospitals are mostly treated as isolated.

Again, the vast attack surface that results from the high number of connected devices, compounded by the fact that there are different variants of attack vectors and huge sensitive data to protect has certainly moved beyond a human scale problem. Therefore, having a broad holistic view of the aggregated incidents will provide a good understanding of the patterns of cyberattack incidents in hospitals and with the data, employ AI/ML tools and techniques to make inferences, identify patterns so that measures can be recommended for critical stakeholders to tackle the problem of cybersecurity breaches in hospitals.

1.4 Aim

The aim of this research is to implement a multilayered cybersecurity approach tailored to the healthcare sector. This approach is shaped based on database technology, artificial intelligence (AI), machine learning (ML) techniques, and risk assessment strategy. Given that different cybersecurity approaches are designed for different and specific threat scenarios. This thesis accounts for this peculiarity by proposing a three-layer multilayered cybersecurity approach.

1.5 Research questions

From our aim, we articulate a clear, focused, and concise inquiry that will guide the direction of our investigation. Given that we found a niche of research interest corroborated by an identified knowledge gap that requires a solution, we formulated a set of research questions to guide our inquiry.

RQ1: How can structured, database-driven incident reporting systems enhance understanding of cyber threats in the healthcare sector, and what insights can be derived from public incident repositories (e.g., OCR, CISSM)?

RQ2: What is the comparative performance of selected AI/ML algorithms (e.g., Gradient Boosting, Random Forest, SVM) in detecting anomalies, phishing, and

network intrusions within healthcare-specific datasets (e.g., ECU-IoHT, Phishing, ToN-IoT)?

RQ3: How can a risk assessment framework, combining ISO 27000/NIST-CSF standards, improve cybersecurity risk prioritization in a healthcare context (e.g., Baze University Hospital)?

1.6 Objectives

From the research question posed in the previous subsection (subsection 1.5), this research articulates targeted objectives to address critical cybersecurity challenges in healthcare facilities. The process of formulating these objectives includes analysis of publicly available records or repositories that are healthcare-related to identify recurring patterns, common attack vectors, and vulnerabilities within the sector. With this analysis, the objective will therefore be to develop a detailed cyber incident database, which offers a valuable resource for tracking trends in cyber threats and informing prevention strategies.

Additionally, another objective of this research is to investigate applicable AI/ML techniques that are usable in the domain of cybersecurity. Given that different methods are applicable to different problems, this objective will explore different ML models and customize them where necessary to the specific challenges faced in healthcare settings. By adapting these tools to the sector's unique context, the research will seek to establish their usability in improving both the accuracy and efficiency of cybersecurity systems. Furthermore, the study also addressed the issue of data heterogeneity and generalizability in intrusion detection systems by introducing a novel Cross-ML framework. This approach enhances the ability of detection models to generalize across different datasets, ensuring more robust performance in diverse environments.

The final objective of this study is to develop a tailored risk assessment framework known as the Baze University Hospital Cyber Risk Assessment Framework (BUHCRAF). This

objective closely ties to the first objective, given that identified patterns of compromise will likely imply vulnerabilities of key assets in a typical healthcare cyber infrastructure. Also, in developing the BUHCRAF framework, we will draw on elements from ISO27000 and the NIST Cybersecurity Framework to provide a structured approach to assessing risks, identifying system vulnerabilities, and prioritizing mitigation strategies.

1.7 Artificial Intelligence (AI) and Machine Learning (ML) synopsis

Artificial intelligence (AI) has expanded the capabilities of what Computers can be used to achieve [12]. The ability of machines to simulate human cognition to solve real-time problems with even greater precision is a novel innovation that is positively altering the dynamics of how people live today [13]. The phrase Artificial Intelligence (AI) was originally used in 1956 at Dartmouth College when computer scientists from IBM, Carnegie Mellon University, and Massachusetts Institute of Technology introduced ideas for teaching computers specific methods. The foundation of many industry sectors that have adopted new information technologies is artificial intelligence (AI). Although the origins of AI may be found in earlier times, there is no denying the crucial role those modern intelligent computers with the capacity for learning, reasoning, and adaptability play.

The study of artificial intelligence (AI) has grown in popularity due to claims of success in symbolic approaches, rule-based systems, expert systems, etc. for addressing problems. The years 1956 through 1974 are regarded as the "golden years of AI" due to the numerous projects that the Defense Advanced Research Projects Agency (DARPA) conducted. Research during this period concentrated on natural language processing, logic, and optimization. After extensive investment but no significant innovations, some organizations began to reduce their research budgets, and criticism began to mount. AI went through its first "winter" (1974–1980) due to a lack of processing, storage, and computational power [12].

Subsequently, the technology advanced, big data became popular, Moore's law continued to advance, computing power increased, and AI began to reemerge. This marked the revival period of the technology. Recently, AI/ML has been employed in fields including human capital, robotics, vision, gaming, medical research, fraud detection, and predictive analytics. Across all industries, AI is beginning to spread. A strong "prediction tool" that may be applied in practically any situation where human judgment or decision-making is necessary is the ability to reason and make decisions based on data.

According to Morning Consult and IBM, AI may be the largest economic opportunity of the 21st century so far, estimated to contribute around $16 trillion to GDP by 2030. Recent advances in a range of technologies from data prep, to bias detection, as well as skills training and re-training efforts is causing the rate of AI exploration and adoption to take off faster than some have predicted [14], [15].

Machine Learning (ML) is widely considered to be a subset of AI that has considerable parts of its foundation in computational statistics, data mining, data science and analytics. The prior idea is targeted at making computers learn from past data [16]. The experience therefore formulates the basis of future actions without the need for specific reprogramming of the computing devices. Machine Learning considerably blurs the line of distinction between work that will be carried out by humans and that which will be carried out by machines and that precisely defines the reason why it is considered a standing pillar of the industry 4.0 technology [17]. To uncover interesting data patterns, recognize sequences, or forecast behavior, machine learning models often consist of a collection of rules, procedures, or complicated functions and equations. These capabilities might be significant in the field of cybersecurity. In this thesis, the author will go through several approaches to solving ML problems and how they can potentially relate to cybersecurity improvement. Fig. 1.1. shows the relationship between AI and ML as it is currently understood in the literature.

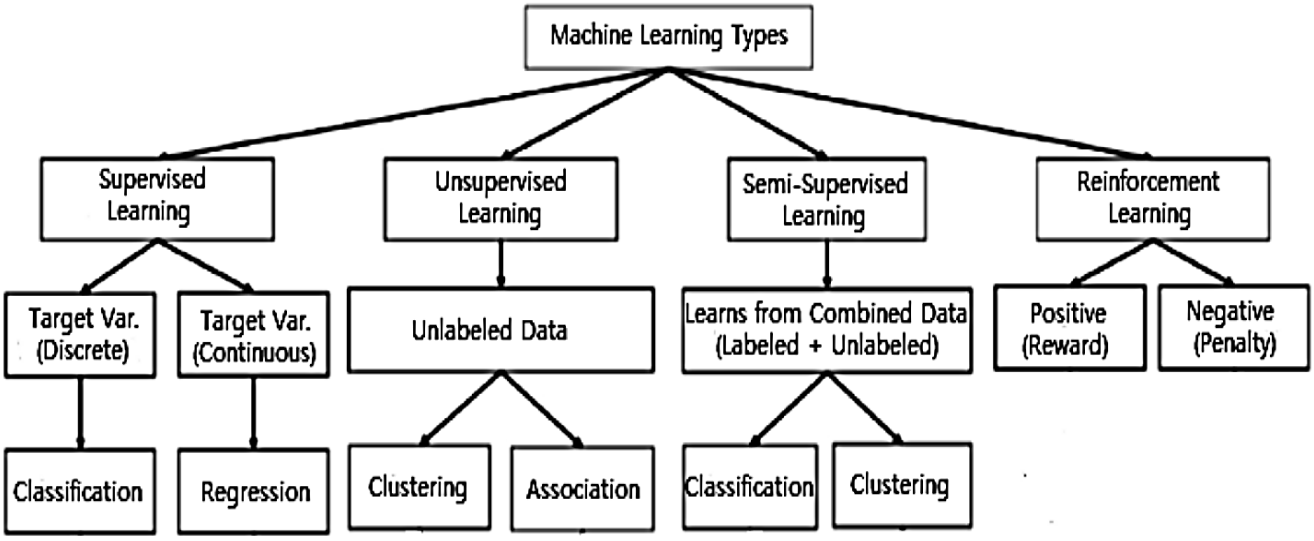


Fig. 1.1: ML approaches [16]

1.7.1 Application of AI/ML in cybersecurity

Models are the simplification of reality through implicit and explicit assumptions with the intention to solve practical problems. Because of the nature of models, perfection is not feasible to achieve. The well-known statistician George Box's articulates that "all models are flawed, but some are useful" relates to this notion [18]. Therefore, how good a model is can only be attributed to its usefulness. For ML, there are several types of models that fall broadly into two categories of supervised and unsupervised machine learning models. Recently, there are models such as semi-supervised and reinforcement models as shown in Fig. 1.1. However, they are still under researched in cybersecurity applications.

1.7.1.1 Supervised approaches

Supervised ML attempts to approximate a target by using the input provided. The algorithm is given a collection of training data which is labeled. The algorithm will learn a rule that it will use to predict the labels for subsequent observations based on large set of data. In other words, supervised learning algorithms are fed historical data and asked to identify the greatest prediction link. A simplistic way of looking at supervised machine

learning as put forward by [19] is supervised model contain input variables (x) and output variables (y). From the input variable (x) the algorithm can learn mapping functions to the output (y) which can be mathematically depicted as y = f(x). Generally, when certain objectives are determined to be achieved from a specific set of inputs, or when a task-driven method is used, supervised learning is generally idealistic [16].

When specified targets are established to achieve from a particular set of inputs, or when using a task-driven approach, supervised learning is carried out. The most widely used supervised learning methods in the field of machine learning are referred to as classification and regression approaches [20]. These methods are frequently used to categorize or forecast the future of a specific security concern. For instance, classification techniques can be used in the cybersecurity domain to predict denial-of-service attacks (yes, no), or to identify various classes of network attacks like scanning and spoofing. The well-known classification methods are ZeroR, OneR, Navies Bayes, Decision Tree, K-nearest Neighbors, Support Vector Machines, Adaptive Boosting, and Logistic Regression [20].

Regression approaches, on the other hand, can be used to predict continuous or quantitative values, such as the total number of phishing activities over a specific time or the parameters of network packets. The underlying reasons of fraud and other types of crime can be found using regression analyses. The two most widely used regression techniques are linear regression and support vector regression. The primary distinction between classification and regression is that the former's anticipated output is numerical or continuous, whilst the latter's is categorical or discrete. As an extension of supervised learning, ensemble learning combines various simple models, such as Random Forest learning, which creates multiple decision trees to address a specific security issue [20].

1.7.1.2 Unsupervised approaches

Unsupervised learning, which is a data-driven process, analyzes unlabeled datasets without the requirement for human intervention. This is frequently used for generative feature extraction, relevant trend and structure identification, result groupings, and experimental reasons. Clustering, density estimation, feature learning, dimensionality reduction, association rule discovery, anomaly detection, etc. are some of the most popular unsupervised learning techniques [16].

Cyberattacks like malware in the field of cybersecurity hide in some instances, changing their behavior constantly and autonomously to evade detection. Unsupervised learning methods like clustering can be used to extract hidden patterns and structures from datasets and find clues to such complex attacks. Like this, clustering approaches can be helpful in locating anomalies, determining policy violations, and locating and removing noisy occurrences in data. Popular partitioned clustering algorithms include K-means and K-medoids, and well-known hierarchical clustering algorithms utilized in a variety of application sectors include single linkage or complete linkage [20].

In events of security incidents, further analysis may be of benefit for feature engineering tasks like optimal feature selection or extraction relevant to a specific security issue which is usually very hard to determine. An article written by [21], stipulated an approach for selecting security features based on their importance score values and they were derived from the popularly known conventional methods of dimensionality reduction such as non-negative matrix factorization, Pearson correlation analysis, linear discriminant analysis and principal component analysis.

1.7.1.3 Semi-supervised approaches

Semi-supervised learning is the hybridization of the supervised and unsupervised methods, as it operates on both labeled and unlabeled data. With so much appreciation to

the peculiarity of the real world, where labeled data could be such a rarity in several contexts, and unlabeled data seem to be more readily available, the semi-supervised model tend to be the most applicable [22]. A semi-supervised learning model's main objective is to create predictions that are superior to those made using just the labeled data from the model. Semi-supervised learning is utilized in a variety of applications, including text categorization, fraud detection, machine translation, and fraud detection [16].

1.7.1.4 Reinforcement approaches

Reinforcement approaches are a form of ML that enables agents and machines to automatically assess the ideal behavior in a specific context or environment to increase its efficiency. The goal of this incentive-based or penalty-based learning approach is to use the knowledge gained from environmental activists to take steps that will either raise the benefit or decrease the risk. However, it is not recommended to use it for resolving simple or elementary issues. It is a strong tool for training models that can increase automation or maximize the operational efficiency of complex systems like robotics, autonomous driving tasks, manufacturing, and supply chain logistics [16].

1.8 Structure of the document

This research work is structured into six chapters. Chapter one is the introduction, chapter two is state-of-the-art, chapter three is methodology, chapter four is system design, chapter five is case studies and application, results, discussion and analysis, chapter six is the conclusion. Each chapter is opened by an introduction of what the chapter covers and ends with a conclusion of all that was detailed.

Chapter one is the introduction of the document detailing the background of the research domain, a statement of the problem from literature to rationalize the imperative for conducting the study, the questions raised that the research seeks to address then finally, the aim and objectives of the research.

Chapter two is the state-of-the-art description of the studied subject by making comparison with related research in the subject area for the purpose of avoiding repetition, understanding what the knowledge gap is, and proffering solutions that will become novel contribution to the existing body of knowledge.

Chapter three of the thesis is the methodology, and it critically presents the research processes employed to address the research objectives. The chapter provides a coherent framework of procedures, techniques, and tools meticulously selected to gather and interpret data in a systematic and organized manner.

Chapter four is system design, and it details the analysis and design principles used in developing the web application and database of cyberattack incidents. This chapter also defines the functionality that the system will perform and the actors that will interact with the system.

Chapter five of the thesis is case studies and applications, results, discussions and analysis. This chapter focuses on the implementation of the proposed solution in a typical modern healthcare network and the various applications of AI/ML methods. All set up, additional data, refined specific methodology, and outcomes are detailed. This provides practical relevance to the results in real-world healthcare cybersecurity settings.

Chapter six is the conclusion, the chapter serves as a concise summary of the research progress up to this point, highlighting the significant findings and contributions. Furthermore, it offers valuable insights and perspectives into future research work.

1.9 Conclusion

To assert that the healthcare sector is not well equipped to deal with cyber threat is demonstrably stating the least and this assertion is well supported by [23] highlighting that,

healthcare organizations are more vulnerable to cyberattacks because of the industry's major cybersecurity gaps, staff members' low levels of digital literacy, out-of-date software, and a lack of rules and enforcements. This research seeks to capitalize on technology such as AI/ML to provide new ways of looking at the problem of cybersecurity in hospitals and effectively reduce the persistent criminality perpetrated in the sector.

CHAPTER 2 – STATE-OF-THE-ART

2.1 Introduction

In this chapter, a comprehensive discussion of similar previously researched work is investigated to ensure that this thesis is solidly grounded in literature. In order to comprehend how devices connect in a typical hospital's network, a look at the typical cybersecurity setup of a hospital is first examined and broken down into precise components. Secondly, a focus on database technology, AI/ML and risk assessment frameworks were discussed. This review is important in understanding how a multilayered cybersecurity method can be specifically used for the healthcare sector to address inherent cybersecurity challenges.

2.2 Cybersecurity in the healthcare sector

Technologies in healthcare have the power to prolong, save, and improve lives. Electronic Health Records (EHRs) or electronic medical records (EMR) usually rely on technology for storing data. Other technologies include wearable devices, general-purpose devices, and in some cases, technology inserted into the human body [24]. All these can be achieved using digital health systems and a well-coordinated IT infrastructure [25]. Telemedicine technology has advanced to the point where care can be provided remotely, even across international borders. Patients are increasingly using their own mobile applications, and the medical Internet of Things can now integrate these apps with telemedicine and telehealth.

2.2.1 Network and IT facilities in hospitals

Hospitals are health care facilities that provide nursing care, patient treatment (both medical and surgical), using specialized health science and medical equipment. For hospitals to carry out their operations efficiently, they must rely on the combination of personnel, IT infrastructure, operational technology, medical equipment and networks. Not all hospitals are the same, therefore, the decomposition of the routine processes should be proportionate to their needs. This research is not tailored to the specificity of an isolated hospital but rather to address the generality of predominant cyber threats that put the delivery of their services at risk. However, regardless of the hospital or the nomenclature used as a matter of convenience, there are always three key network structures in use. Medical networks for doctors and nurses to access medical records, inventories and information; patient networks utilized by patients and visitors of the facility; and private networks that leases out connectivity from the hospital to private physicians who can also have access to the information systems and records on the internal network [26].

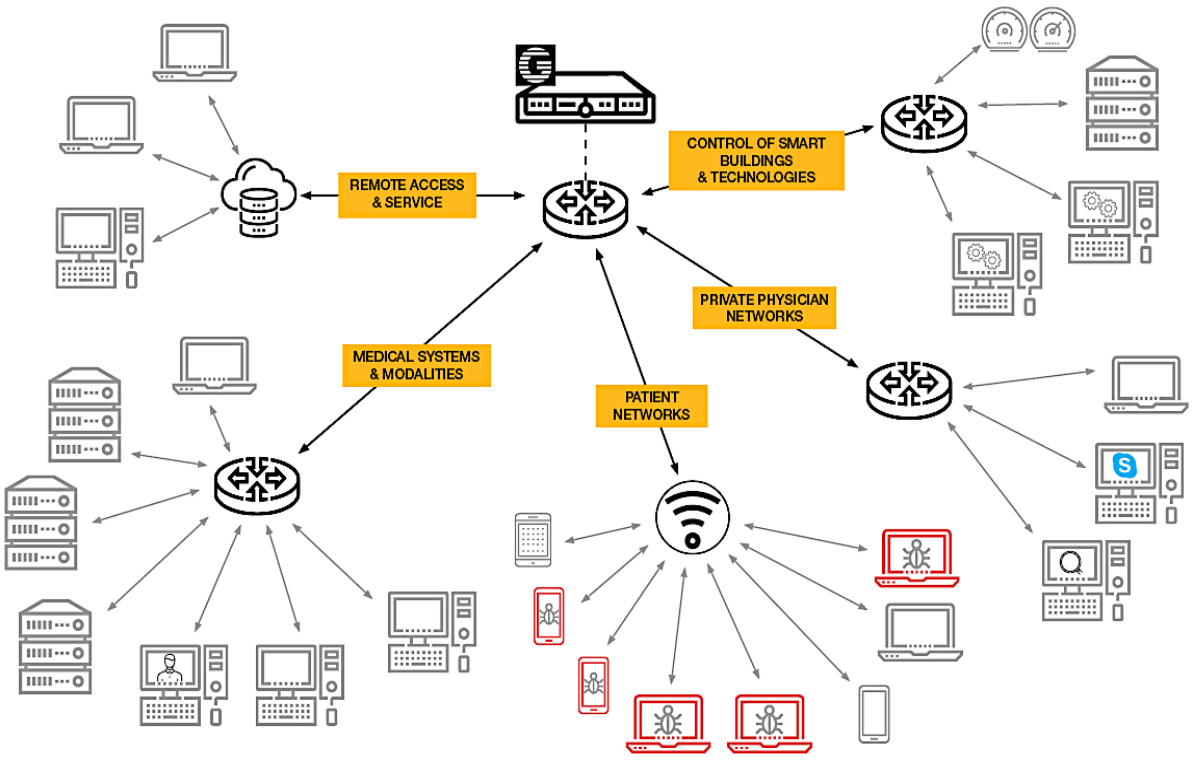


Fig. 2.1: Hospital Network Architecture [27]

The hospital's network architecture, as represented in Fig. 2.1, simplifies an interconnected network infrastructure within typical healthcare systems with cyber attackers infiltrating the patient networks as a vulnerable access point which is signified by the red malware infested mobile phones and personal computers. Also, it shows various subsystems such as a network hub can be seen at the center, which connects medical systems and modalities, remote access and service, the private physician network, smart buildings and technologies control and patient networks. This interconnectedness introduces significant cybersecurity risks where cyber attackers can remotely compromise any of the networks for illicit activities. Therefore, as key equipment is interwoven to provide essential services over the network, securing this networked equipment is as pivotal to the hospital as the services the equipment provides. Cybersecurity, as interpreted by most people, is an all-encompassing term that can mean measures taken by cybersecurity firms to protect users when accessing internet services. A more descriptive definition, as put forward by [28] is, cybersecurity is the protection of internet-connected systems such as hardware, software and data from cyberthreats.

2.2.2 Common cyberthreats

There have been numerous cyberattacks or threats that have plagued the healthcare sector in the past. Some are well-documented and described in the literature, while others are not. Generally, these threats and cyberattacks against healthcare facilities can broadly take two forms depending on the effects they have on the organization. Primary effects are the direct effects of a cyber-attack that can cause problems or use targeted devices. Secondary effects are the indirect effects created by cyber threats. Secondary effects often proceed from primary attack scenarios, and they can include regulatory and legal implications [29], [30]. Most cyberthreats usually come in the form of unauthorize access to computerized systems by cybercriminals. To protect any facility from cybercriminals, it is vital to have a good understanding of the different types of cyberthreats and the mechanics of their exploitative tendencies. The different types of cyberthreats include

1. **Malware**: A malware is any software that is purposefully created to disrupt a computer, server, client, or computer network, leak confidential information, obtain unauthorized access to data or systems, prevent access to information, or inadvertently compromise a user's privacy and security online. Malware gets its name from the combination of “malicious” and “software”, and it is generally an umbrella term used to describe viruses, trojans, spyware, ransomware, botnets and adware. Malware can either spread from one device to another or remain in place, only impacting its host device [31]. A malware attack requires the target device to have the program installed. The user must act to preempt the attack life cycle. Users should therefore be informed on the kinds of software to avoid, the kinds of links they should examine before clicking, and the emails and attachments they should avoid in addition to employing firewalls that can identify malware [32], [33].
2. **Phishing:** Phishing is a technique for exploiting social engineering that is widely used to steal sensitive data from users, including usernames, passwords, and login information. The goal is to persuade users to reveal their account information, login credentials, and other sensitive information. As mentioned by [34], the simplest way for hackers to infect devices with malware is through phishing. Phishing tactics lure victims into opening emails or clicking on links that appear to be from reliable or legitimate sources. The link could drive users to a malicious website that directly infects user PCs with malware or a fake website that requests personal information. As a result, one should not click a link if they are unsure about it [35].
3. **DoS and DDoS Attacks:** A denial-of-service (DoS) attack aims to overburden a system's resources to the point that it becomes unable to respond to valid service requests. Like a denial-of-service (DoS) attack, a DDoS attack aims to deplete a system's resources when a massive number of host computers under the attacker's control that are infected with malware target a network. Due to the victim site's inability to serve users that request access, these are known as "denial of service" attacks. DoS and DDoS attacks are distinct from other cyberattacks that provide

hackers more access to a system or the ability to gain access to it more easily. These assaults give the attacker a quick return on their investment. On the other hand, the goal of DoS and DDoS network assaults is to simply prevent the target's service from being effective. If the attacker is employed by a rival company, they might make money off of their actions [36], [37].

4. **Ransomware:** Ransomware, as the name implies, is a malicious software designed by criminals to prevent users from accessing electronic services unless they pay a fee. During the COVID-19 pandemic, ransomware attacks multiplied as more people worked remotely [23]. Ransomware is becoming more and more complex. In addition to encryption, other technologies are now being incorporated into ransomware's arsenal. The first demonstration of ransomware by Joseph Popp in 1989 used symmetric-key encryption to hijack victims' hard drives and demanded payment for decryption [38] [39].

5. **Insider threats:** The most dangerous individuals can occasionally be found within an organization. People working inside a company's walls are particularly dangerous because they frequently have access to a variety of systems and, in certain cases, administrative rights that give them the ability to make important adjustments to the system's security standards. Additionally, employees frequently have a thorough awareness of the organization's cybersecurity architecture as well as how the company responds to attacks. This information can be used to enter restricted places, alter security settings, or determine the most advantageous moment to launch an attack [40]. Limiting employee access to critical systems to just those who require them to accomplish their jobs is one of the best strategies to stop insider threats in organizations. Additionally, for the limited few who require access, employ Multi-Factor Authentication (MFA), which will necessitate that they utilize both a physical item they possess and at least one thing they know in order to access a critical system. The user might be required to input a USB device and enter a password, for instance. In some setups, the user must log in to a mobile device that

generates an access number. If both the password and the number are accurate, the user can only enter the secure area [41], [42].

2.2.3 Current threat landscape

The current threat landscape in the healthcare sector does not leave much to be desired as there has been a steady increase in reported cases over the years. As seen in Fig. 2.2, the healthcare sector appears to be the most affected amongst other sectors in terms of the average cost of data breaches worldwide [43]. This indicates its significant vulnerability as the picture shows a significant increase from $9.23 million in 2021 to $10.10 million in 2022.

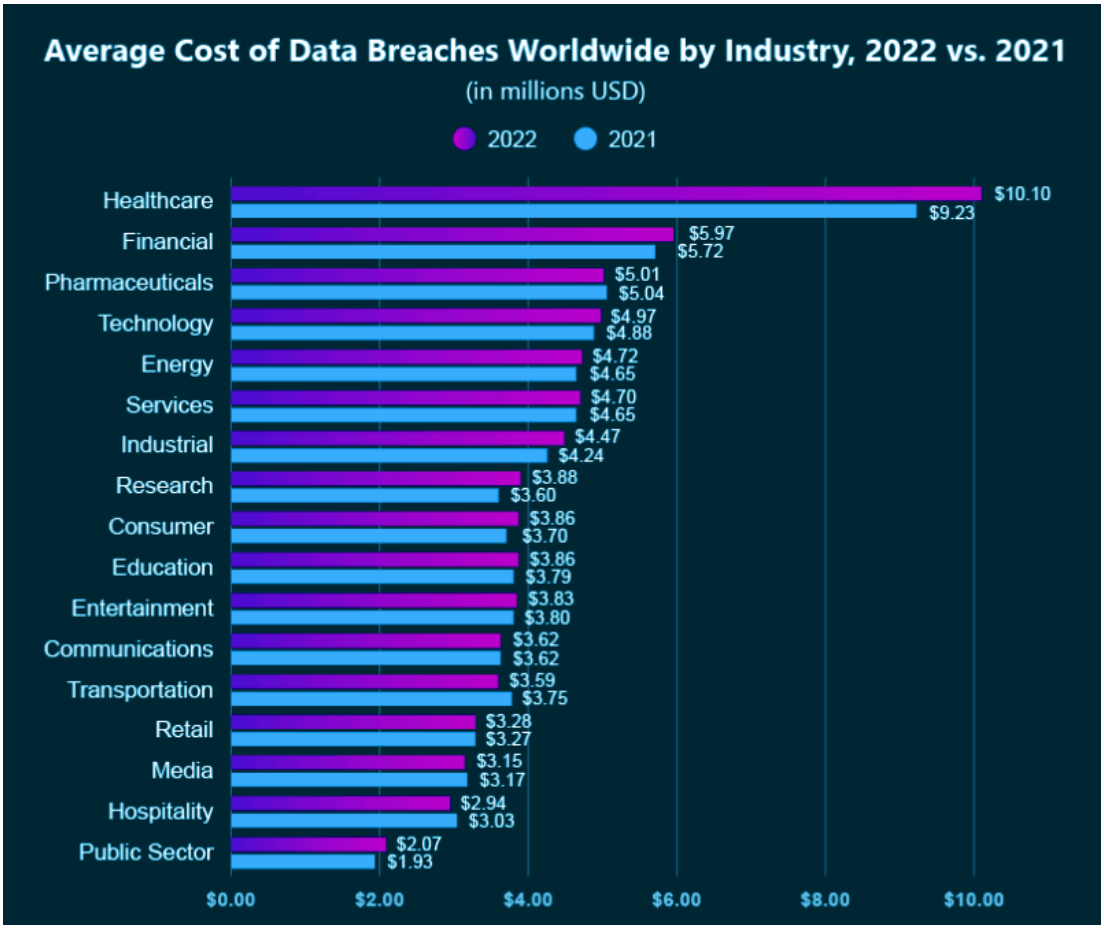


Fig. 2.2: Average cost of data breaches worldwide by industry, 2022 vs. 2021(in dollars) [43]

There have been many scientific literature works that have reported notable cyberattacks in the healthcare sector [44], [45], [46], [47], [48], [49], [50], [51], [52], [1], [53], [54], [55], [56], [57]. These studies not only documented acts of violence against hospitals, but they also demonstrated that the health care industry is one of the top three worldwide targets of ransomware attacks. This is a hugely important issue because hospitals all over the world are adopting network systems, cloud storage services, and connected medical equipment. These days, hospital information systems play a major role in providing medical, financial, and administrative support. It is commonly acknowledged that innovation must balance privacy and security with usefulness and effectiveness; however, technology is evolving faster in terms of patterns and complexity than the typical security measures or safeguards put in place to protect the healthcare industry.

The WannaCry ransomware is perhaps one of the most catastrophic incidents in recent years where over 200 000 computer systems in 150 countries have been impacted by a cyberattack that began in more than 60 trusts within the National Health Service (NHS) of the United Kingdom. The WannaCry ransomware first attacked the National Health Service (NHS) on May 12, 2017 and has since spread quickly, impacting numerous companies worldwide, including FedEx [58]. When someone opens an attachment or clicks on a link in an email, the virus is downloaded to their computer. After that, the user must pay a ransom to unlock the encrypted files on the computer. The WannaCry virus's creators demanded US$300 in bitcoin to unlock each impacted computer. The fee doubled after three days, and if payment was not made within a week, all data would be lost. Experts on cybercrime advise against making these kinds of payments since there is no assurance that doing so will allow users to access their data again. Several practitioners have promptly attributed the attack's prevention to the NHS's lack of investment in preventive technology. While it does not seem that the WannaCry ransomware specifically targeted the NHS, the organization was exposed to attack because a few of its Windows operating systems were over 15 years old and no longer received updates or support from Microsoft [59].

On the wake of COVID-19, a cyberattack targeted the Brno University Hospital in Brno, Czech Republic [60]. The hospital is the second largest facility in the country and the attack has become widespread in central Europe [61]. The specifics of the security breach have not been disclosed by hospital officials, but it was considered serious enough to delay critical surgical procedures and divert newly admitted patients in critical condition to St. Anne's University Hospital. Two other hospital branches, the Children's Hospital and the Maternity Hospital, were also impacted by the incident, which forced the hospital to shut down its entire IT network [62]. The primary asset of target is all Microsoft Windows-based systems and subsequent reports indicated that ransomware was likely used by the attackers to encrypt the hospital's data after they had successfully gained access to the network via spear-phishing [63].

In 2024, 13.4 million people were affected by the Kaiser Permanente website tracker breach. Kaiser Foundation Health Plan Inc. is alerting 13.4 million people that certain tracking technologies on its websites and applications have exposed some of patient’s personal information to outside entities, including Microsoft (Bing), Google, and X (Twitter). This is the biggest confirmed healthcare data breach that involves website tracking technologies to date, as well as the largest reported healthcare data breach of 2024. According to Kaiser, the tracking technologies have been eliminated from its websites and mobile applications after being discovered during a voluntary internal investigation. To avoid such incidents in the future, more precautions have been taken. Out of an abundance of caution, notifications are being sent to everyone who may have been impacted, including users of Kaiser's websites and mobile apps as well as current and former health plan members across all the markets in which it operates. In May 2024, notifications are anticipated to be sent out. Names, IP addresses, and sign-in statuses were among the categories of information that might have been given to tech corporations [64], [65].

Healthcare Corporation of America (HCA), the largest healthcare system in the United States with more than 180 hospitals and 2,300 healthcare locations, is based in Nashville, Tennessee. It was recently revealed that patient-protected health information was obtained by unauthorized parties. According to HCA Healthcare's initial report, over 11 million records were affected. According to the HHS Office for Civil Rights, 11,270,000 individuals were currently affected by the third-largest healthcare data breach reported by a HIPAA-regulated company. HCA Healthcare revealed on July 10, 2023, that hackers had obtained access to an external storage location that was being used to format emails automatically, including those informing patients about HCA Healthcare programs and services and reserving appointments for patients. The compromised data lists contained 27 million rows of information, including the protected health information of about 11 million patients who received care at HCA hospitals and doctor's offices in 20 U.S. states, though the investigation into the data breach is still ongoing [66].

2.3 Regulatory and compliance requirements

All sectors must protect patient data, but the healthcare industry must take extra precautions because patient data is sensitive. To ensure patient privacy and data protection, healthcare providers must take all reasonable steps. Patients can suffer numerous personal and professional complications due to inadequate protection of confidentiality, but the provider can also suffer serious consequences, including possible reputational damage [67]. In the context of health services, there are several guidelines and regulations for the storage, sharing and use of personal data stipulated by several countries

2.3.1 General Data Protection Regulation (GDPR)

For data protection purposes, the General Data Protection Regulation (GDPR) defines health data and recognizes it as a distinct type of data[68]. The GDPR is European Union legislation that came into force in May 2018. The main aim of the regulation is, first

and foremost, to strengthen and consolidate the protection of European Union citizens and, furthermore, to regulate the way in which organizations handle this data [69]. Consent is a pivotal aspect of the GDPR that ensures the expression of will by the data subject through an unequivocal positive declaration that enables the data managers to treat such data according to the regulation. The right to portability also ensures technical feasibility of transmitting data to individuals in a machine-readable format and to also transmitting the data between organizations without unnecessary interference. In the era of digitalization where organizations are plagued with not exercising due diligence in treating patient data, these rights serve to place substantial control in the hands of the individuals in respect to the treatment of their personal data. In the event of a data breach, the legislation also requires that organizations inform the individuals affected and all regulatory authorities including a comprehensive description of the breach, the likely consequences and remediation steps taken to address it. Furthermore, there are penalties attached to culpable organizations who do not adhere to the regulations or are complicit in their implementation of the data protection rules. Depending on the seriousness of the breach, fines can reach over 20 million euros or 4% of the total revenue accrued by the organization and in less serious cases, the fines can be up to 10 million euros or 2% of the organization global turnover [69].

2.3.2 Health Insurance Portability and Accountability Act (HIPAA)

In the United states of America, the Health Insurance Portability and Accountability Act (HIPAA) establishes national standards for safeguarding individual's medical records and other relevant personal health information [70], [71]. The need to ensure adherence to standards while enabling essential information flow for care coordination gets more difficult as digital technologies are ingrained in healthcare procedures [72]. Therefore, achieving high-quality healthcare requires care coordination, which is described as the purposeful planning of patient care activities to enable the proper delivery of health services and that is what HIPAA delivers [73]. The Act applies to providers, organizations, plans and companies

that can process health data on behalf of the entities, setting clear rules about who can access Protected Health Information (PHI). In the stipulation of the HIPAA Act, patients must provide express permission in writing before their health data can be transmitted to third parties. Furthermore, the legislation addresses the security of health data detailing physical, technical and administrative safeguards that can be employed to curtail unauthorized access, misuse and disclosure. Administrative sanctions alongside civil fines can be imposed on culpable organizations and the assessment of such fines is carried out by the U.S Department of Health and Human Services (HHS). The severity of the violation is also considered, and legal liability can range between $100 to $50,000 for every instance of a violation. The maximum limit is capped at $1,500,000 per type of violation, however, when severe intentional violation or gross negligence of the HIPAA provisions can be established, then criminal proceedings can be opened and referred to the U.S Department of justice for investigation. Notable examples of healthcare organizations paying penalties for non-compliance with HIPAA law include Anthem Inc.'s breach, which resulted in the insurer paying $16 million in 2018 after a data breach put 79 million people at risk had; Massachusetts General Hospital paid $1 million for the loss of 192 patients' medical records and Cignet Health paid $4.3 million for refusing to provide medical records to patients upon request [69].

2.3.3 Personal Information Protection and Electronic Documents Act (PIPEDA)

The Personal Information Protection and Electronic Documents Act (PIPEDA) is the Canadian law that regulates how private sector businesses gather, use, and disclose personal data. It went into effect in 2001 and outlines policies for gathering and using personal data, defines conditions for using that data for purposes, and grants individuals the ability to update their personal data. Furthermore, to safeguard personal data from loss, theft, and unauthorized access, PIPEDIA mandates the implementation of appropriate security measures. In cases of non-compliance, the commissioner may issue correction

orders and pursue fines through the federal court of Canada. Penalties include investigations, hearings, and consent agreements conducted by the Privacy Commissioner of Canada [69].

2.3.4 General Data Protection Law (LGDP)

Brazil has a legislation named the General Data Protection Law (LGDP) that came into force in 2020. The legislation aims to protect citizens' privacy and ensure their control over personal information by prohibiting businesses and organizations from processing personal data, including health data. The GDPR serves as an inspiration for most of the laws, which set forth precise rules for the use, storage, and exchange of personal data [69].

2.4 Multilayered cybersecurity approaches

Multilayered cybersecurity approaches are holistic approaches to cybersecurity that involves formulating multiple layers or procedures of defense tools and techniques to tenaciously defend vulnerable components of an organization’s cyber or IT infrastructure [74]. This strategy is predicated on the idea that no defense mechanism is a one-stop shop for addressing all security concerns [75]. An organization can reduce the risk of a single point of failure and defend itself against a greater range of threats by utilizing multiple layers of defense [76]. Multilayered security consists of the physical, technical and administrative layers and they all function to address specific security concerns in the IT infrastructure of an organization as shown in Fig. 2.3 [77].

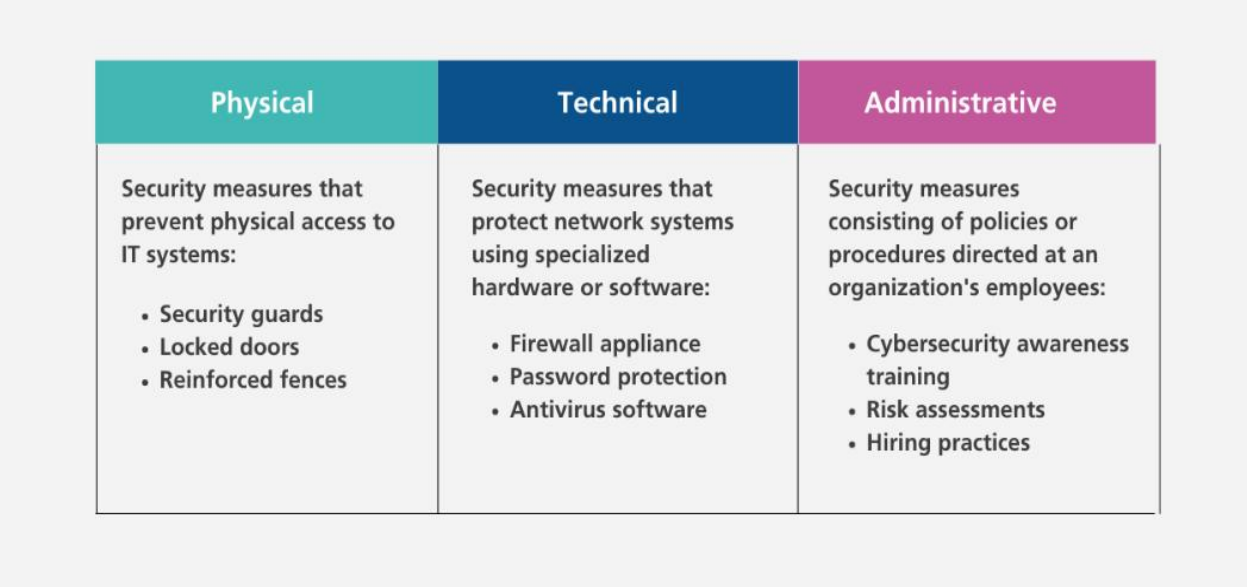


Fig. 2.3: Multilayered Cybersecurity approach [78]

There are several scholarly works that investigate the concept of multilayered (defense in depth) approaches applied to different contexts to improve cybersecurity and resilience outcomes. Some of these researches are given in the proceeding paragraphs;

The work of [79], provide a multilayered cybersecurity framework for addressing cybersecurity concerns in SCADA systems through the implementation of granular access controls, network micro-segmentation, anomaly detection, encrypted communications, and legacy system updates. Their technology was used in an electronics manufacturing facility for access control, zoning, monitoring, and encryption applications. The multilayered defense-in-depth (DID) strategy integrates policies, processes, and technologies to combat growing vulnerabilities. Furthermore, their findings reveal increased security, with 57.4% fewer unauthorized access incidents, 41.2% faster threat containment, and 79.2% fewer hacker attempts.

The work of [80] offer a "Multi-Layer Security Scheme for Mitigating Smart Grid Vulnerability against Faults and Cyber-Attacks" to address outages by identifying network weak areas caused by physical faults and cyber-attacks. Their method consists of two parts.

The first layer is a cyber-security-based blockchain approach that serves a generic purpose. The second layer employs a cyber-security-based reinforcement-learning method to assist susceptible points by monitoring data. Their simulation results show that using such a comprehensive method can significantly reduce the probability of blackouts.

Additionally, the work of [81] used the concept of multilayered approaches (defense in-depth) for industrial control systems (ICS) to develop a cyberattack detection system developed using network traffic data, host system data, and measured process parameters. The data used to demonstrate the proposed detection method come from a real-time ICS testbed. Five attacks, including man in the middle (MITM), denial of service (DoS), data exfiltration, data manipulation, and false data injection, are used to replicate the effects of a cyber attack and create data for data-driven detection models. Their research examined four classical classification models based on network and host system data, including k-nearest neighbor (KNN), decision tree, bootstrap aggregating (bagging), and random forest (RF), to provide a secondary line of defense for cyber-attack detection in the event that the intrusion prevention layer fails.

Furthermore, [82] presented a "Multilayer Resilience Paradigm Against Cyber Attacks in DC Microgrids" Their research focuses on man-in-the-middle (MITM) attacks, which employ a multilayer-based event-driven technique to remove signals by establishing numerous mitigation layers. Depending on whether the authentication signal for each neighboring agent is True or False, the data flow between the multilayer cyber network is restricted. Finally, they conducted simulations and tests to verify and validate their suggested robust system.

Finally, the work of [83] offer a multilayered framework consisting of two elements: a three-layered logical model and reference architecture for cyber-physical systems, and a meta-model of cyber-physical system attacks known as the cyber-physical system kill-chain. The layered reference architecture provides a methodical foundation for investigating

how the causal chain associated with cyber disturbances may be traced back to physical perturbations. The framework is assessed against a simulated unmanned aerial system. The long-term goal is to use the framework to deduce cyber-physical system security attributes and to list the principles for constructing cyber-attack-resistant systems.

2.4.1 Defense in depth

In practice, the term "defense in depth" is sometimes used interchangeably with the layered cybersecurity approach because it requires the use of multiple cybersecurity products or solutions at different levels of control, as well as adaptable practices and protections to secure an organization's network, properties and resources [84], [85]. It is a comprehensive strategy that addresses security on all fronts by considering internal linkages within an organization and making use of resources and technologies already in place [78]. Fig. 2.4 shows a defense in depth architecture and several security mechanisms that can be employed at each layer or endpoint. For example, the systems and network security layer (the third layer on the right-hand side of Fig. 2.4) of the defense in depth architecture is subdivided into host-based security and network-based security. For network-based security, the defense in depth architecture identifies intrusion detection systems and anomaly detection systems as a technique for securing that endpoint. Therefore, in our work, we proposed the AI/ML driven anomaly detection system using gradient boosting, random forest, decision trees and multilayer perceptron. Additionally, to secure the intrusion detection systems endpoint, we proposed the cross-ML approach using decision trees. Furthermore, for the web and application security layer (the sixth layer on the right-hand side of Fig. 2.4), the defense in depth architecture identifies web application defense, application security, and web defense as an endpoint; therefore, in our work we propose web phishing detection using SVM to secure that endpoint.

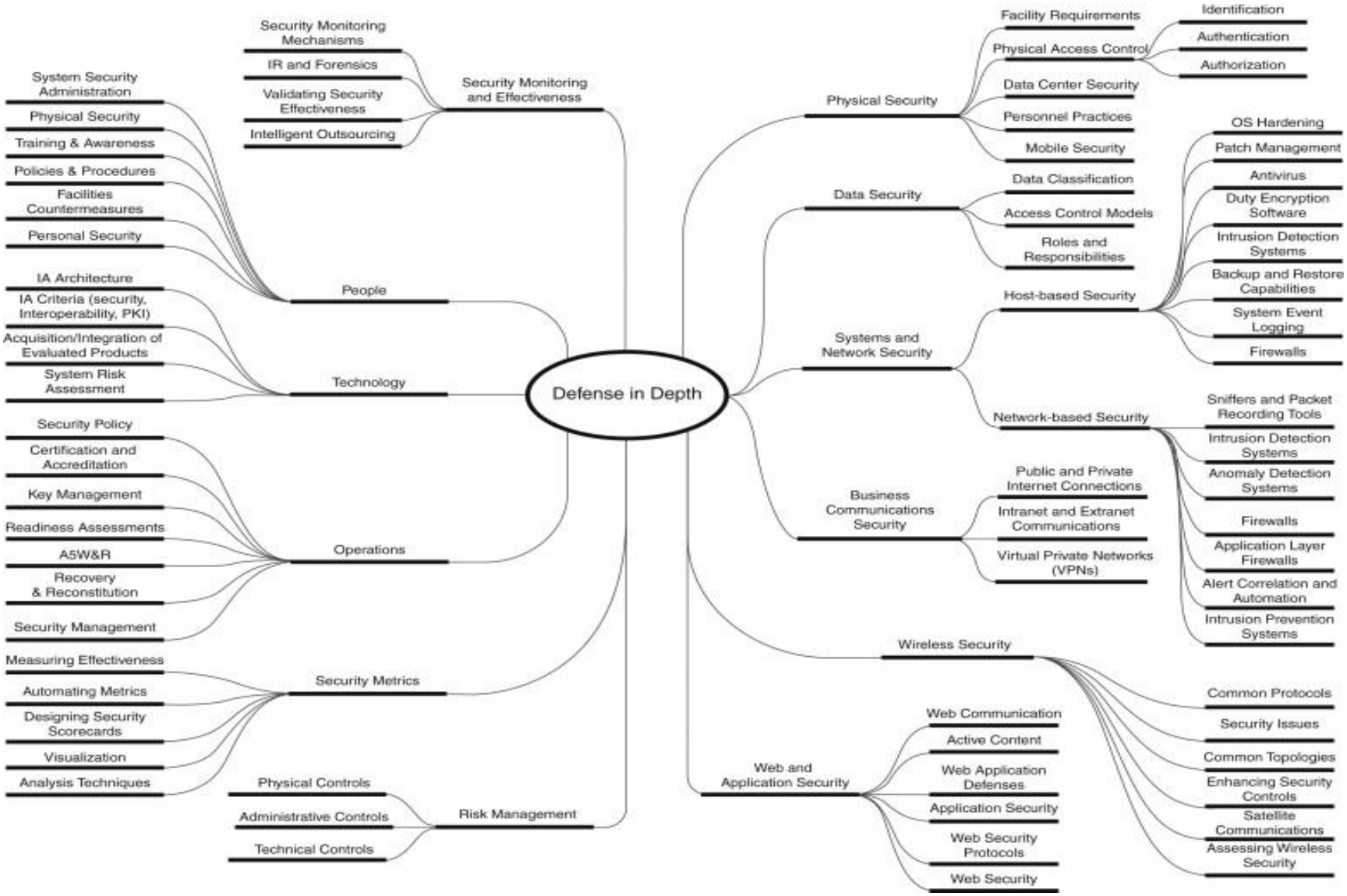


Fig. 2.4: Defense in depth [84]

2.4.2 Zero trust architecture

The term "zero trust" (ZT) refers to a set of evolving cybersecurity paradigms that center defenses on users, assets, and resources instead of static, network-based perimeters. The infrastructure and workflows of an industrial and enterprise are planned using zero trust principles rely on access policy based on context and continuous user and device authentication and verification [86], [87]. Zero trust presupposes that assets or user accounts are not implicitly trusted based only on their physical or network location so, zero trust architecture (ZTA) is based on the notions of least privilege, granular access control, and dynamic, stringent policy enforcement where no user or device—regardless of size or location—is implicitly trusted [88], [89], [90], [91].

According to the Zero Trust model seen in Fig 2.5, every network request should be handled as though the system has already been breached, and even the most basic request should be considered as possible threats. Before a session can begin or access is allowed in the first place, multi-factor authentication and authorization are needed. Furthermore, credentials must be verified once more for every authenticated user requesting access to a new resource. This method facilitates the quick detection, identification, and neutralization of threats—whether they originate from the outside or within the network—and helps to limit lateral movement once a threat has entered the system [92].

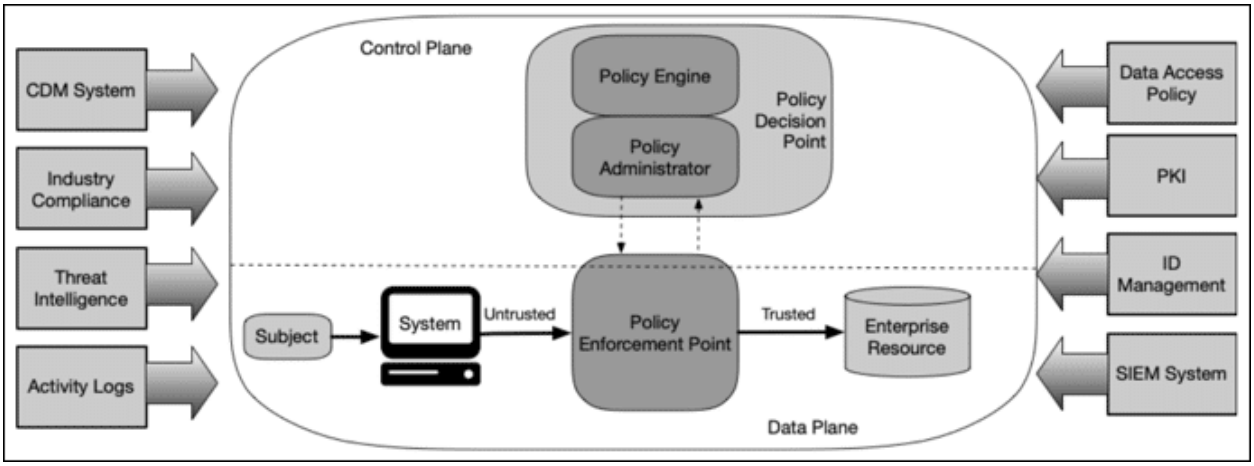


Fig. 2.5: Core zero trust logical components [93]

2.5 Database technologies

Database technology has been around since the 1950s [94]. However, their application has evolved tremendously. In principle, databases are structured collections of data that are organized and stored in a systematic manner and are very important for the management of an organization [95], [96]. They are designed to efficiently store, retrieve, manage, and manipulate large amounts of data. Databases are widely used in various applications, ranging from small-scale personal projects to large-scale enterprise systems.

Database systems play a pivotal role in the successful implementation of systems that can be used for a variety of purposes [97]. Database technology can be effective in cataloging cyberattack incidents that have occurred in the healthcare sector so that lessons can be learned from indicators such as the frequency of incidents, attack types, dates, attack vectors used, and the number of persons affected because of a successful compromise. These collected records can be better situated to provide an outlook to help develop and improve the cybersecurity posture of healthcare organizations.

2.5.1 Application in other sectors

Other sectors have made significant progress beyond the healthcare sector in terms of adopting an incident database for specific reasons, and it has helped them address some of their organizational peculiarities.

In the context of safety and security, McGregor [98] elucidates how intelligent systems are being increasingly utilized by governments and corporations. However, the deployment of these systems can potentially carry significant risks to people's lives and well-being. Unfortunately, designers of these systems rarely anticipate the potential problems that may arise in real-world scenarios because of using these systems. To further Compound this issue, there is the lack of a formal system within the artificial intelligence community to help practitioners learn from past mistakes. While there are commendable

efforts by individuals in various fields to collect information on AI failures, such as through Google Docs and GitHub, the existing lists are challenging to navigate, lack comprehensiveness in most cases, and represent subjective viewpoints on real-world AI failures [98]. To prevent recurring failures in AI, it is imperative to share information about past failures with AI practitioners. Thus, the work of McGregor [98], presents a structured compilation of instances where intelligent systems have caused real-world issues related to safety. The collection, known as the AI Incident Database (AIID), aims to address the question of what can go wrong when deploying a specific system. The work offers an infrastructure that supports industry's best practices in artificial intelligence, a dataset consisting of over a thousand incident reports, and an architecture for developing research based on the expanding collection of incidents.

The Federal Aviation Administration [99], and National Aeronautics and Space Administration [100] keep a database of past incidents and accidents. The Federal Aviation Administration (FAA) maintains a database of preliminary accidents and incidents that are reported to the Office of Accident Investigation and Prevention through Aviation Safety Information Analysis and Sharing (ASIAS) [101]. Likewise, the National Aeronautics and Space Administration (NASA) [102] maintains a database of accidents and incidents under the Aviation Safety Reporting System (ASRS) [100]. The ASRS database serves as the largest global collection of confidential safety information voluntarily shared by frontline personnel in aviation, including pilots, flight attendants, mechanics, controllers, and dispatchers. It serves as an essential resource for the development of targeted products and subsequent research aimed at addressing diverse aviation safety concerns. The reason why the aviation industry appears to be the safest sector is because of this database archiving system, and by systematically analyzing and archiving previous accidents and incidents within a shared database, valuable insights can be gained. In the aviation industry, accidents refer to cases involving significant damage or loss of life, while incidents represent situations where the risk of an accident significantly escalates. For instance, if a small fire is swiftly extinguished in the cockpit, it is classified as an "incident," but if the fire causes harm to the crew during

the extinguishing process, it becomes an "accident." The aviation database organizes flight log data and expert investigations to facilitate comprehensive evaluations of both technological and human factors. This continuous self-examination has contributed to making air travel one of the safest modes of transportation since its inception over a century ago. The safety of commercial airlines has undergone remarkable advancements during these years. The rates of fatal accidents in large scheduled jet airlines have significantly decreased to a point where, across multiple aspects, aviation has emerged as the safest mode of commercial transportation [103]. Aviation incidents are infrequent occurrences and out of 32.2 million flights in 2022, only about five accidents with fatal impact were recorded making air travel one of the safest activities in which individuals can participate in [104].

Another database used by the cybersecurity community is Adversarial Tactics, Techniques, and Common Knowledge (ATT&CK), which was developed by the Massachusetts Institute of Technology Research and Engineering (MITRE). They operate as a not-for-profit organization focused on advancing technical expertise for the benefit of society, especially in cybersecurity. MITRE ATT&CK® is an expansive and globally-accessible knowledge base that offers a comprehensive collection of adversary tactics and techniques [105]. Developed based on real-world observations and empirical data, the ATT&CK knowledge base serves as an invaluable resource for understanding and countering malicious activities across various cyber threat landscapes [106]. This dynamic repository acts as a foundational framework for the development of specific threat models and methodologies within multiple sectors, including the private industry, government agencies, and the broader cybersecurity product and service community. By leveraging the insights provided by ATT&CK, organizations can better comprehend the tactics employed by adversaries and proactively fortify their defenses against emerging threats [107]. With the creation and ongoing maintenance of ATT&CK, MITRE since 2013, the organization actively fulfills its overarching mission of solving critical problems and promoting a safer world. This initiative fosters collaboration, bringing together diverse communities and stakeholders to

collectively enhance cybersecurity practices and responses. By providing an open platform, ATT&CK encourages inclusivity, enabling any individual or organization to leverage its vast resources and contribute to the advancement of cybersecurity without any financial barriers or restrictions [108]. The accessibility and broad utilization of ATT&CK contribute to the collective efforts of the cybersecurity community, fostering a stronger defense against evolving threats and ensuring the continual growth and improvement of global cybersecurity practices [109]. Overall, the ATT&CK framework helps cyber security administrators to understand the evolving threat landscape, focus on specific threat actors and the tactics they employ, provide real-time threat intelligence and incident response and a more robust foundation for risk quantification [110].

In recent times, the United States Food and Drug Administration (FDA) has introduced a database into its operations. This database plays a crucial role in monitoring and evaluating the safety and effectiveness of drugs and therapeutic products in the market. [111]. The database is made available via the Adverse Events Reporting System (FAERS) public dashboard [112]. The database collects and stores information on adverse events and medication errors related to drugs and therapeutic biologic products. The FAERS database serves as a valuable resource for monitoring the safety and effectiveness of medications and identifying potential risks or issues associated with specific drugs. Healthcare professionals, consumers, and manufacturers can report adverse events to FAERS, and the data within the database are used for ongoing surveillance, analysis, and research in the field of pharmacovigilance [113]. Since its establishment, millions of cases have been reported to the FAERS by manufacturers, healthcare professionals, and consumers, contributing to a comprehensive and dynamic database that strengthens the FDA's ability to ensure the safety of drugs in the market [114].

Table 2.1 compares other known structured incident reporting systems across different domains that have contributed to enhancing safety and security. Each row highlights a

database or reporting system, its source, function, the need it addresses and its overall purpose.

Table 2.1: Comparison of Incident Databases

Title	Authors/sources	Function	Need	Purpose
Artificial Intelligence Incident Database (AIID)	Sean McGregor 2020	presents a structured compilation of instances where intelligent systems have caused real-world issues related to safety	Lack of a formal system within the artificial intelligence community to help practitioners learn from past mistakes	Infrastructure that supports industry best practices in AI by sharing information about past failures with practitioners
The Federal Aviation Administration (FAA)	Aviation Safety Information Analysis and Sharing (ASIAS)	The ASRS database serves as the largest global collection of confidential safety information	Systematically analyze and archive previous accidents and incidents within a shared database	Provide resources for the development of targeted products and subsequent research aimed at addressing diverse aviation safety concerns
National Aeronautics and Space Administration (NASA)	Aviation Safety Reporting System (ASRS)	The aviation database organizes flight log data and expert investigations for both technological and human factors.	Systematically analyze and archive previous accidents and incidents within a shared database	Provide resources for the development of targeted products and subsequent research aimed at addressing diverse aviation safety concerns
Adversarial Tactics, Techniques, and Common Knowledge (ATT&CK)	Massachusetts Institute of Technology Research and Engineering (MITRE)	Dynamic repository acts as a foundational framework for the development of specific threat models	The need to ensure the continual growth and improvement of global cybersecurity practices	Organizations can better comprehend the tactics employed by adversaries and proactively fortify their defenses against emerging threats
United States Food and Drug Administration (FDA)	Adverse Events Reporting System (FAERS)	The database collects and stores information on adverse events and medication errors	The need to have a valuable resource for monitoring the safety and	Identify potential risks or issues associated with specific drugs

effectiveness of medications

2.6 AI/ML related work and use-cases

With the advent and increasing popularity of ML, numerous works proposed ML solutions for diverse cybersecurity tasks, resulting in hundreds of research papers. Such abundance inspired many literature surveys that aggregate or summarize the state of the art. However, most of such studies may provide a detailed analysis but on a single application, such as cyber risk assessment [115] or IoT security [116]. Others may focus on a specific cyber detection problem, e.g., malware [117], [118], [119] spam detection [120], [121], or intrusion detection [19], [122]. Most scholarly works proposed by academicians on the application of AI/ML in cybersecurity are typically theoretical, devoid of real use-case application. The problem with pedaling such narrative is the unintended consequence of assuming that AI/ML approaches can solve all cybersecurity challenges and that is not quite the case. To achieve water-tight cybersecurity posture in any organization requires the right interwovenness of several tools and techniques. Whilst AI/ML is predominantly data driven; therefore, the extent of research will be constrained by the quality and availability of data in the public domain that are suited for AI/ML algorithms or models to solve. Therefore, it is imperative to sieve through literatures that are futuristic and theoretical and narrow down to works that actualizes the application of AI/ML in cybersecurity [123]. So far, most applications of AI/ML in cybersecurity are mostly supervised approaches and the specific use-cases are.

2.6.1 AI/ML in phishing detection

Phishing represents one of the most common vectors for penetrating a target network and is still a rampant threat in the cybersecurity landscape. Phishing is a social engineering attack where attackers manipulate their victims with the intent of retrieving important personal details [124]. Phishing is one of the most common attack vectors that

has led to very prominent cybersecurity incidents. Even big tech giants such as Facebook and Google has been compromised by phishing attacks leading to financial losses of over 90 million Euros [125]. The application of AI/ML for detecting phishing attacks usually takes two forms. Website phishing attacks which can either be based on uniform resource locator (URL) or website features [126], [127], [128]. The other form of use-case application of AI/ML is email phishing attacks as elaborated in this work [129], [130], [131]. Fig. 2.6 shows a diagrammatic representation of use-case application of AI/ML in real-time phishing detection where, the pipeline begins with preprocessing, model training that utilizes the one-dimensional convolutional neural network phishing detection (1D-CNNPD) and an advanced 1D-CNNPD model for classification to predict if an email is either "phishing" or "legitimate".

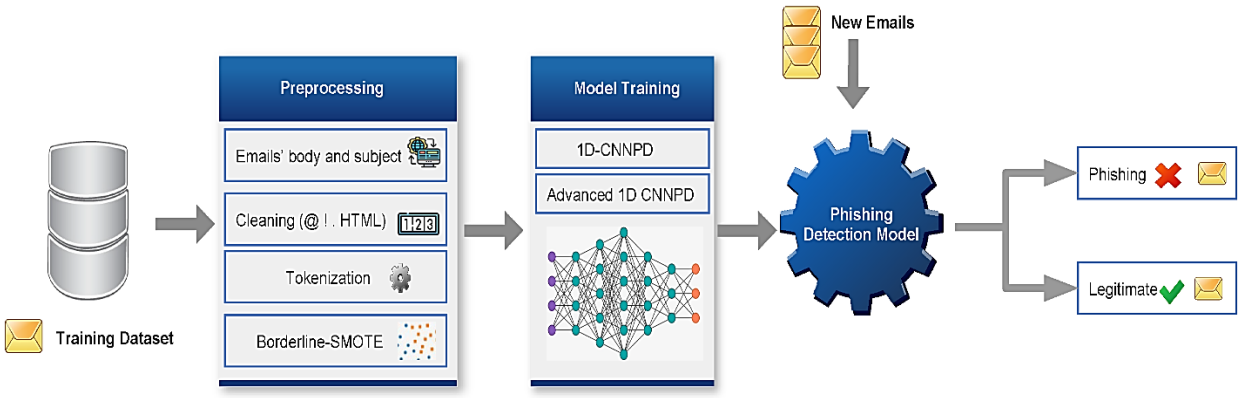


Fig. 2.6: Use-case application of AI/ML in email phishing detection [129]

The authors in [132] used affinity graph-based semi-supervised learning framework that leverages four categories of email profiling feature that consist of origin, text, attachment, and recipient to detect known and unknown spear phishing campaigns. Their work utilized a dataset of 1,467 spear phishing emails from 8 campaigns. Their campaign attribution used only 25 labeled emails, constituting less than 1% percent of the data to simulate real-time traffic. Their proposed system provided a 0.9 F1 score and a 0.01 false positive rate.

The authors in [133] emphasize the growing sophistication of phishing attacks. Their work evaluated the performance of XGBoost and random forest on a dataset with 6,157 legitimate and 4,898 phishing website features. From their results, XGBoost showed high accuracy of 98.32%, and F1 score of 97.68% which was slightly higher than the random forest with an accuracy of 97.27% and an F1 score of 97.56%. They also experimented with decision trees, logistic regression, and AdaBoost but showed significantly lower results in comparison with XGBoost and random forest.

2.6.2 AI/ML in Network Intrusion Detection Systems (NIDS)

An integral component of network security research is the Network Intrusion Detection System (NIDS). Using active defense technology, it detects intrusion behaviors and responds to threats by alerting users and stopping the intrusion [134]. Network intrusions systems are either host based which involves a single machine/endpoint or network based that sieve through network traffic to identify anomalous data packets as shown in Fig. 2.7. Before now signature-based approaches were the common technical method for building NIDS. Signature-based NIDS defines attack patterns and stores it in a signature database or dictionary with the intent of capturing any network traffic that matches the signature and flagging it as an intrusion or attack [135]. The problem with signature-based NIDS is novel attacks that are yet to be captured in the intrusion database will not be flagged by the intrusion detection system. This is the major limitation of the existing approach that paved the way for several research on ML based techniques for NIDS [136], [137], [138], [139]. When it comes to detecting threats against a variety of targets, including cloud, IoT, endpoint devices, NIDS can be deployed anywhere in a network environment and leverage ML for detection purposes. A typical NIDS deployment that makes use of ML, can analyze a variety of data kinds, such as DNS records, complete packet captures (PCAP), network flows7 (NetFlows), Simple Network Management Protocol (SMNP), and so on [123], [140].

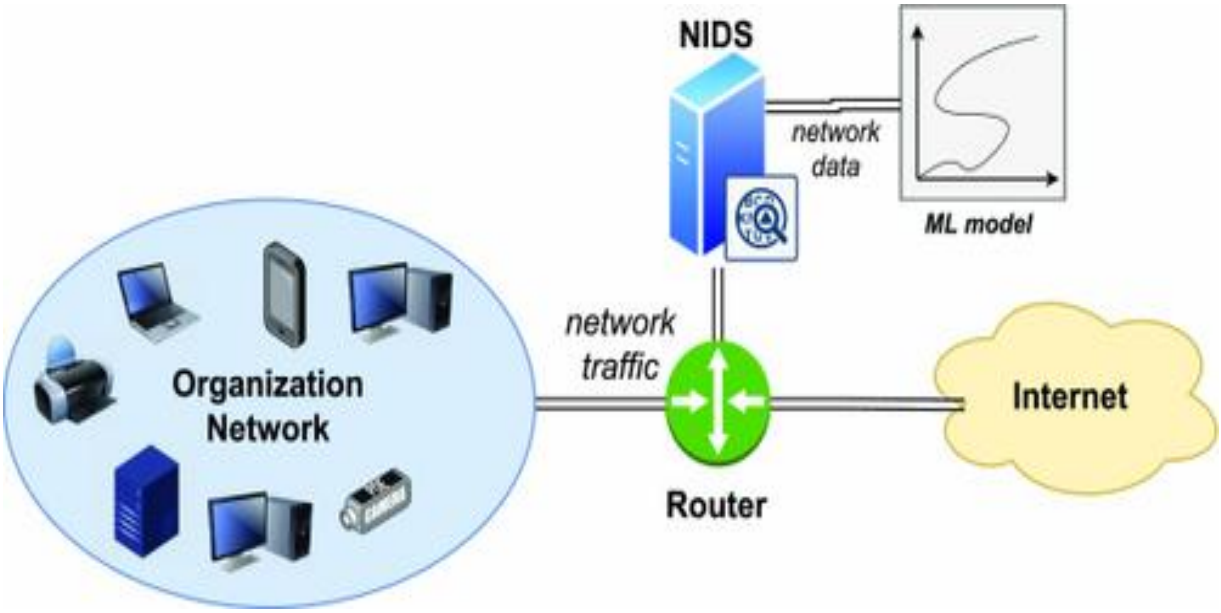


Fig. 2.7: Use-case application of AI/ML in email NIDS [123]

The work of [141], "Classification Analysis of Intrusion Detection on NSL-KDD Using Machine Learning Algorithms" employed three ML models (artificial neural networks, support vector machines, and decision trees) to establish the optimal ML technique among the models. Their experiment used 60% and 70% of the NSL-KDD dataset for training and the rest of the dataset for testing. The result of their experiment shows that the decision tree model performed better than the other two models with an accuracy score of 98.84%. Although, their work is a binary classification problem, they also did not include the accuracy score of the artificial neural networks and support vector machines.

Another study by [142], proposed an AI-based intrusion detection system for IoT architectures that is based on two stages. The first stage used the Bat algorithm with swarm division to acquire optimal features from the traffic network data. In the second stage they applied the random forest algorithm for a multiclass classification. The result of their method shows 96.42% accuracy and a 0.98% false positive rate.

2.6.3 AI/ML in malware detection

The fight against malware is one of the most emblematic challenges of cybersecurity. Because malware affects a specific device, its detection is performed by analyzing data at the host level. Malware is the root cause of many security breach incidents, such as data loss and denial of service attacks [143]. In its simplest terms, malware is defined as a malicious software. Malware can be classified into various categories according to their nature of dissemination and the charade of security breach activities they pose. A worm is a form of malware that can multiply and leave the device infected to infect further ones. Ransomware disables the victim's access to the data until the attackers pay the ransom. Fileless malware makes changes in the native OS files. Spyware collects user data without the user's knowledge; it can even spy on the user's activity data. Trojans disguise themselves as desirable code or software. Rootkits install themselves on the victim's machines and provide the attackers with remote control on the victim's devices [143]. Majority of the use-case application of AI/ML techniques is in Android Malware detection [144], [145], [146], [147], [148]. As shown in Fig. 2.8, static and dynamic analysis are the two methods available for malware detection. The former seeks to identify malware by examining a specific file without executing any code. The latter concentrate on examining how a piece of software behaves while it is being executed, typically by putting it in a controlled environment and keeping an eye on its actions [123]. Machine learning (ML) can be useful for both static and dynamic analyses [149], [150].

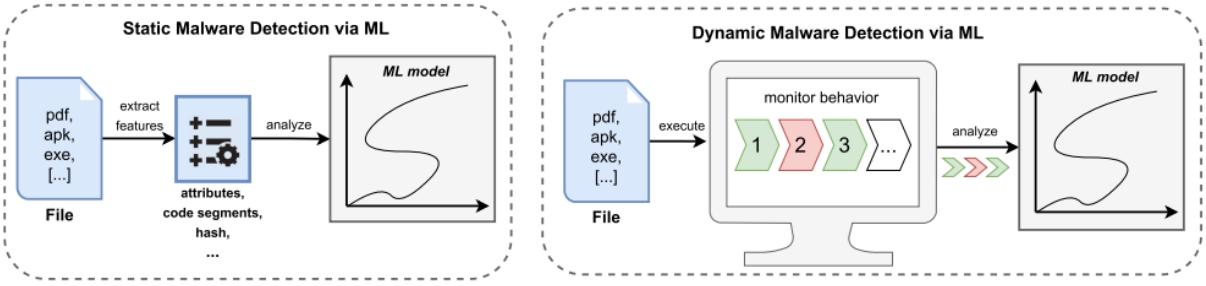


Fig. 2.8: Use-case application of AI/ML malware detection [123]

One notable work on malware detection offered by [151] proposed visual-AT (adversarial techniques) to enhance detection. Their work proposed a novel ML-based malware visualization detection method to enhance accuracy and compared it to traditional ML based methods. Their study utilized the BIG 2015 dataset with 10,678 labeled malware samples across nine classes and the Ember 2018 dataset with 2,000 malicious and 2,000 benign files. Their results show that the visual-AT algorithm shows over 28% increase in detection accuracy in comparison to random forest and SVM and CNN algorithm.

2.7 Cybersecurity risk assessment frameworks

In the domain of risk management, there are several standards and frameworks designed for cybersecurity risk assessment [152]. Some of the widely known frameworks include the International Organization for standardization (ISO 27000) series [153], the National Institute of Standards Cybersecurity Framework (NIST CSF) [154], and the European Union Agency for Cybersecurity (ENISA) [155]. In the area of research, authors have proposed frameworks such as the "Yet Another Cybersecurity Risk Assessment Framework" (Yacraf) [156], while other authors have leveraged the strengths of the Information Technology Infrastructure Library (ITIL) model tailored towards meeting specific security management objectives of a defined context [157].

It is not an uncommon practice to decompose several cybersecurity frameworks and adapting it according to the unique needs of a firm [158]. In our work, we evaluated the similarities between the NIST and ISO 27000 series and used it as a reference to develop the risk assessment solution suited for the healthcare sector.

2.7.1 ISO/IEC 27000 series

The ISO/IEC 27000 series popularly referred to as the ISMS family of standards provides the best practices and recommendations for information security management [159]. The scope of the series is very comprehensive in design and transcends beyond

confidentiality, privacy, and cybersecurity issues. For instance, the ISO/IEC 27001 standard generally stipulates the requirements for establishing, implementing, maintaining and continuous development of information security strategy [160]. The ISO/IEC 27002 stipulates a universally accepted and acknowledged benchmark restrictions tailored for specific information security risk circumstances [160]. The ISO/IEC 27005 stipulates guidance for the management of information security risks [161]. Most of the ISO/IEC standards were implemented to varying degrees to achieve the objectives of this work.

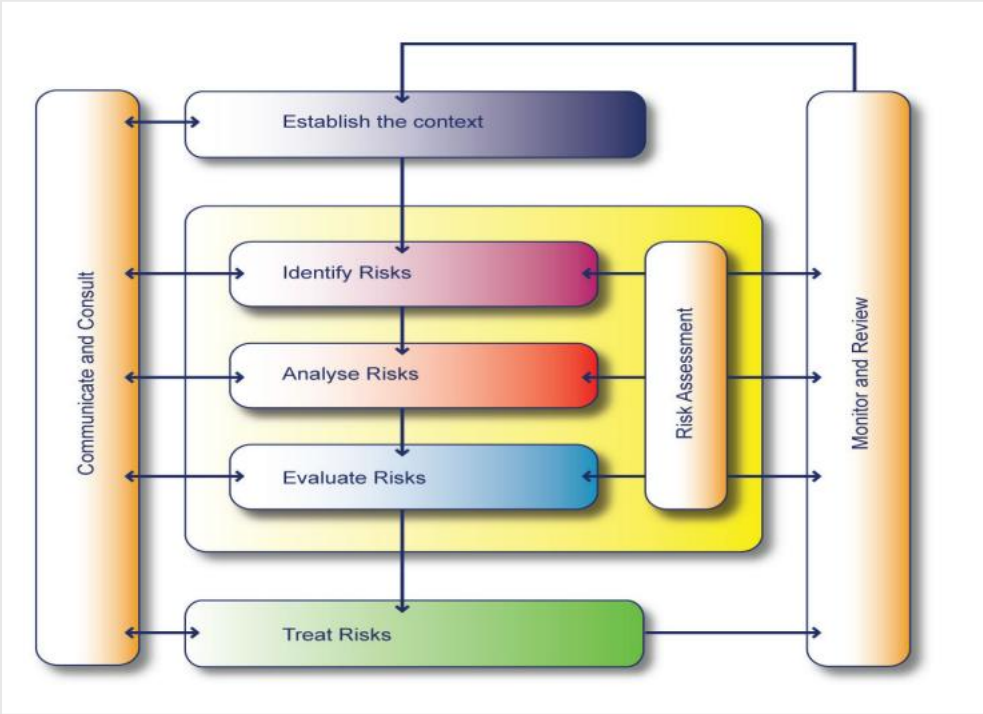


Fig. 2.9: ISO 27005 Risk management process

1 **Communicate and Consult:** This level of the risk management process opens the communication pathways of stakeholders that will be involved in the process. Ensuring there is a healthy feedback loop of information flow and decision making [161].

2 **Establish the context:** This stage involves defining the scope, context and criteria in alignment with organizational objectives. Thoroughness should be exercised in this stage as it should provide details for how risks are identified and how the impact and likelihood will be calculated [162, p. 27005].

3 **Risk Assessment (Identify Risks, Analyze Risks, Evaluate Risks):** The Risk assessment process generally, as shown in Fig. 2.9 consists of three key steps of identifying risks, analyzing risks and evaluating risks. To achieve this, the risk assessment analyst will comprehensively compile all the information assets and identify all subsisting threats and vulnerabilities applicable to the assets. Subsequently, the risks criteria are defined then followed by assigning impact and likelihood values. To complete this stage, each identified risk is evaluated against predetermined levels of acceptance so that priorities can be established on which risks should be addressed.

4 **Risk treatment**: The risk treatment is also referred to as the risk control strategy [163]. This stage stipulates the five key strategies that can be taken to treat the articulated risks. it could be defense, where measures are taken to stop the exploitation of vulnerabilities; transference, which moves the responsibility associated with the risks to other assets or organizations; mitigation, to reduce the impact of an incident; acceptance, where no action is taken to protect an asset especially when the cost of the asset is less than the safeguards employed to protect it; termination are conscientious measures taken to avoid introducing activities that expose the organization to uncontrollable risks.

5 **Monitor and Review:** cybersecurity risks are not stationary; the risk landscape can abruptly change because of changing threats; therefore, continuous monitoring is needed to identify such changing dynamics and include them in the risk management scope.

2.7.2 NIST CSF

The National Institute of Standards and Technology Cybersecurity Framework (NIST NCF) is designed to help organizations improve their cybersecurity posture by identifying exploitable cybersecurity gaps [164]. The framework facilitates understanding, managing and reducing risk associated with cybersecurity infrastructure in organizations of all sizes [165]. The NIST CSF framework is fully customizable to precise needs of an organization.

This functional adaptability gives the organization sufficient leverage to determine the specifics of the implementation framework [166]. The framework is made up of three major component: core, implementation tiers and profiles [158]. The core consists of five activities designed to deliver specific cybersecurity outcomes. These activities as shown in Fig: 2.10 are identify, protect, detect respond and recover [167], [168].

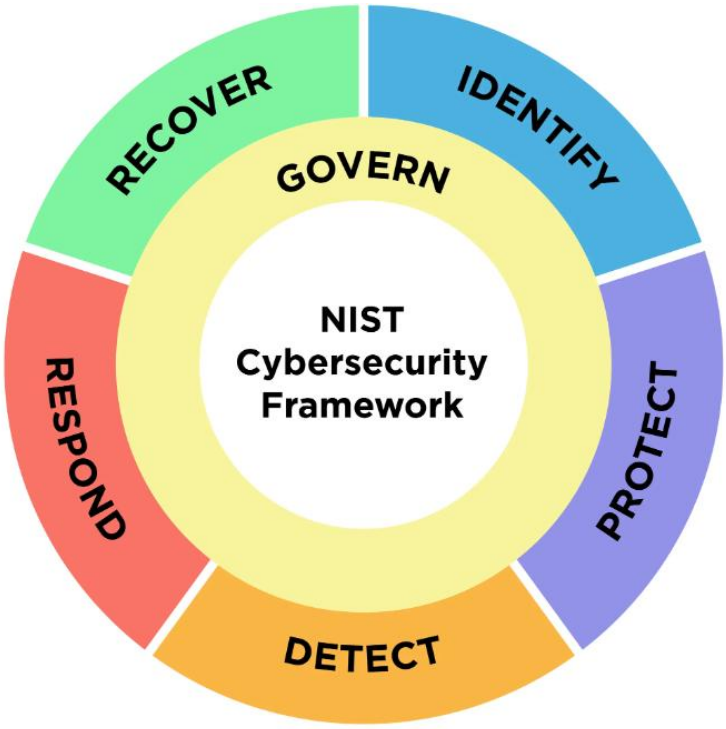


Fig. 2.10: NIST CSF

1. **Govern**: In the governance layer, the organizational context is established and planning a programme that aligns with the mission, goals and stakeholders' expectations. Roles are also established and assigned to individuals to define hierarchy and responsibility.
2. **Identify**: This stage constitutes the identification of assets owned by the organization. These assets are typically data, hardware, software, systems, facilities and people. Additionally, the risks to these assets are also identified and classified.

Another crucial activity carried out in this stage is Risk Analysis (RA) which involves attack surface management (ASM) and vulnerability analysis (VA).

3. **Protect:** These are the measures or safeguards put in place to ensure the confidentiality, integrity and availability (CIA) of the assets already identified. Some of the measures used in protecting assets could also be cryptography, multifactor authentication, passkeys and backup or recovery systems.
4. **Detect:** This layer of the framework ensures timely discovery of cyberattacks. Usually, this is achieved with specialized tools such as a monitoring dashboard, Network intrusion detection systems (NIDS) and anomaly detection systems.
5. **Respond:** This level of the framework involves practical actions taken to contain the impact of the detected cybersecurity incidents. Dynamic playbooks are employed in this stage to outline a series of steps to be taken to resolve the incident.
6. **Recover:** The recovery process supports the timely restoration to normalcy in operations in the event of a successful attack exploitation. The backup systems used for protection are commonly used in recovery. Also, communication dynamics must also to manage at this level to notify regulatory bodies of the incident.

2.8 Resilience considerations

As put forward [169], "Resilience is the intrinsic ability of a system or an organization to adjust its functioning prior to, during, or following changes and disturbances, so that it can sustain required operations under both expected and unexpected conditions." Other authors such as [170], [171] defines resilience as an organization's ability to anticipate potential threats, to cope effectively with adverse events, and to adapt to changing conditions.

Resilience in critical infrastructure presupposes systems that are capable of monitoring their environment, detecting disruptions, and self-correcting under uncertainty. Although several researchers use different labels for the concept, In this section, we

considered different authors that have articulated resilience under different themes. (eg., cyber resilience, organizational resilience, critical infrastructure resilience, resilience potential, resilient organization, and institutional resilience).

The work of [170] on organizational capacities articulates resilience as a meta-capability and break it into its constituent pieces. Inspired by process-based studies, they suggest three successive resilience stages (anticipation, coping, and adaptation) and give an overview of underlying capabilities that together form organizational resilience. Based on their framework, they analyze the linkages and interactions between the various resilience stages, as well as the primary antecedents and drivers. They further organized their studies in three parts depending on how resilience is viewed, either as an outcome, a process or capabilities. To conclude, their findings imply that resilience is a core organizational ability that contributes to organizational progress. It enables businesses to withstand stress, continuously innovate, and respond rapidly to change.

The findings of [172] on improving the resilience of critical infrastructures (CI) provides insights for governance systems of the critical infrastructure management for the effective and efficient reduction of damages on CIs through improving their resilience. They proposed eight approaches that can prevent damage to CIs, hence enhancing resilience. They include clear definitions, improved understanding of CI interdependencies, vulnerabilities, and exposure, dedicated bodies with clear responsibilities, stakeholder coordination and information sharing, well-standardized data collection, knowledge enhancement through building capacities, engagement of the private sector, and integration of CI resilience strategies into standards and building codes. This thesis integrates several of these dimensions through its multilayered framework, particularly by applying structured risk registers, data-centric knowledge bases, and AI-driven threat anticipation.

The work of [173] perceives institutional resilience in healthcare systems from a slightly different perspective. They provided Checklist for assessing institutional resilience (CAIR) based on three cultural drivers of commitment, competence and cognizance. They further posit that managing system safety in health care does not equate to zero adverse events rather, the maximum degree of intrinsic resistance consistent with the organization's reasons for existence which in turn should inform and direct the institution's principles, policies, procedures, and practices.

The work of [174] considers the capacities for resilience in healthcare, where they proposed a conceptual framework in terms of contextualization and collaboration. They deemed their approach fit for purpose because it bridges the gap in literature as it contributes to theorization and guidance in tailoring operationalization efforts, although they recognized the ambiguity in terms of research themes and content of different resilience studies. Their work also relied heavily on Hollnagel potentials for resilient performance in healthcare, including anticipation, monitoring, responding, and learning [169]. They also acknowledged limitations in terms of consensus indicators for operationalizing these potentials.

In cybersecurity, resilience has a slightly different connotation. As defined by [175], Cyber resilience is the ability of an actor to resist, respond and recover from cyber incidents to ensure the actor's operational continuity. An actor in their definition means any individual or collective unit usually assumed to preserve its own cyber resilience. One interesting example they provided in their work is IoTs colossal attack surface given that everything is connected to everything therefore, the quality of cyber resilience in IoTs depends on the state of the art within the IoT. Furthermore, their work stresses that cyber resilience is embedded across levels of interacting agents which could be individuals, organizations, systems or nations. Lastly, their research also identified gaps in current practices ranging from lack of systemic and technical focus on controls, measures, and recovery mechanisms in order to account for a changing cyber landscape.

The work of [176] provides an interesting viewpoint on the fundamental concepts of cyber resilience. Their research establishes a relationship between cybersecurity resilience and risk assessment. They stated that risk assessment is the empirical quantification of a cyberattack that exploit known vulnerabilities, while cyber resilience is a system's ability to prepare for, absorb, recover from, and adapt to cyber threats. They also noted that because cyber threats might be difficult to measure and foresee in all cases, cyber resilience begins when risk management ends, as residual risk can persist even after thorough implementation of mitigation techniques. They also acknowledged the overlap between cyber risk assessment and cyber resilience concepts and why both paradigms are interdependent and should not be treated as separate entities, as that will introduce redundancy to cybersecurity culture. Furthermore, their work also posits a strategy for risk management that includes identifying all critical components of a system that are vulnerable to failure and devising measures to harden them. This approach closely aligns with the strategy we employed in this thesis, although real questions arise in terms of the practicality of identifying all known critical components.

Additionally, the work of [177] emphasizes how to identify the correct harmony between risk management and resilience approaches. Their work considered the NIST CSF framework consisting of five core elements of identify, protect, detect, respond and recover as a typical example of risk management practices. They also emphasized that implementation can vary in sophistication; a non-resilience approach concentrates on protecting the existing system design rather than going after fundamental capacity, whereas a resilience-oriented approach employs the framework's concepts to design a comprehensive resilience solution. Depending on the company, the proliferation of standards and guidelines demonstrates the variety of perspectives on how to ensure strong cybersecurity practices, which can sometimes take the form of risk management or resilience approaches.

In this thesis we followed [169] idea of resilience, where he sees resilience within complex socio-technical systems, particularly in critical infrastructures, as a dynamic capability and not just a static attribute. Drawing from principles of Resilience Engineering, organizational resilience is best understood not as something an entity possesses, but as something it performs. To that end, for an organization to be resilient, it must be capable of these four basic abilities.

1. Ability to respond: To be resilient, an organization must be able to adapt to both expected and unexpected changes. These could be threats, disruptions or opportunities. To be resilient, an organization must have a clear plan of what action to take and when to take it. This refers to the ability to confront the *actual*.
2. Ability to monitor: Another immediate aspect of resilience for an organization is the ability to monitor events and identify any significant changes that may impact existing or planned operations. A resilient organization should be capable of understanding what to focus on, at least for the duration of the current activity or process. This capability can be interpreted as the ability to address the *critical*.
3. Ability to anticipate: A resilient organization must be able to predict developments that are distant in the future, beyond the scope of current activities. It must be able to analyze potential future events, conditions, or state changes that may impact the organization's ability to function, either positively or negatively, such as technology innovation, changes in customer needs, new legislation, and so on. This refers to the ability to address *potential*.
4. Ability to learn: Lastly, a resilient organization must be capable of learning from past experiences. It is critical to comprehend what has occurred and to be equipped to draw the appropriate lessons from the event. This is the ability to address the *factual*.

2.8.1 Resilience implications in this work

Building on the work of [169], [170], and [176], resilience is increasingly viewed as a dynamic and emergent property that integrates sensing, responding, anticipating, and learning across intertwined technical and organizational domains.

Within the scope of this thesis, resilience is acknowledged as an essential capability for healthcare organizations that are inadvertently operating under heightened digital and operational uncertainty. Therefore, the proffered multilayered cybersecurity framework embeds resilience as a latent property within the layers of the framework. To this end, this thesis did not treat resilience as a central design paradigm because of the following reasons.

1. ***Defined scope and primary emphasis on detection and risk:*** This research primarily focused on proposing a framework that integrates AI-based threat detection, database driven incident analysis and structured risk assessment tailored for healthcare environments (Baze University Hospital). Although, at the most basic level, our approach inherently contributes to resilience, especially in terms of responsiveness and adaptation, however, the structure of the thesis did not explicitly adopt resilience theory to structure research questions and experimental design.
2. ***Challenges in operationalization and conceptual maturity:*** Resilience in complex socio-technical systems, as evidenced by healthcare institutions, is necessarily fragmented and, in most cases, difficult to operationalize. The triggering components are still linked to human behavior, institutional learning, and organizational culture that are poorly defined which Baze University as a case study unavoidable inherits these limitations. This posed a challenge to our studies, as it becomes challenging to evaluate resilience in a reproducible and quantifiable way; therefore, this study tilted more towards technical considerations than organizational and human aspects of resilience.

However, within the proposed multilayered framework, there is an implicit consideration of resilience although, not treated as a separate add-on, but as an emergent property built through sensing (monitoring), reacting (responding), forecasting (anticipating), and adapting

(learning) across all layers. Table 2.2: provides the specifics of how resilience is implicitly captured in our work.

Table 2.2: Resilience considerations

SN.	Resilience Ability	What it means	Thesis context
1.	Respond	Ability to act quickly to variability, threats, or disruption	Dynamic threat detection & AI-driven response (e.g., anomaly detection, phishing models, intrusion detection)
2.	Monitor	Detect internal/external changes that affect system stability	Process monitoring in network traffic using ML classifiers (Random forest, gradient boosting, decision trees, and SVMs)
3.	Anticipate	Predict long-term changes, threats, or shifts	Use of risk registers in risk assessment to prioritize and anticipate trends on vulnerable assets
4.	Learn	Adapt based on past events, both good and bad	Feedback loops provided by cyber incident database on changing trends of cyber attacks

2.9 Gap in research

As shown in subsection 1.2 and 1.3, the healthcare sector faces cybersecurity challenges of immense proportions and the COVID-19 pandemic really tested the limits of healthcare information systems [178]. Although the healthcare sector is an umbrella term that encapsulates every health institution that facilitates the provision of services to patients, it is worth noting that the peculiarities in terms of IT needs of each facility will be

different. The difference in this IT adaptation may be the reason why one healthcare facility may be more susceptible to attacks than another.

Subsection 2.5.1 has shown how database technologies have been implemented in other sectors including aviation, AIID and FDA. Although database technologies in other areas of application do not directly stop cyberattacks from happening, they play a crucial role in helping security analysts better understand the general threat landscape. In doing so, adaptation in terms of security tools and techniques are tailored to address evolving threats that are documented in the database.

Furthermore, there are limited studies that looked at the peculiarities of the healthcare cyberinfrastructure and proposed adapted cybersecurity strategy at least for a single context. Some just address recent practices and trends in a comprehensive way [179], others highlight holistic managerial perspective for management of healthcare cybersecurity [180], while others focus on generalizable cybersecurity insights for the healthcare industry in the USA [181]. Most of these approaches for tackling cybersecurity in the healthcare sector seem to provide a disjointed method of dealing with the issue. In this work, having a context of reference is integral to narrowing down tailored healthcare multilayered security approaches and grounding the proposal in concrete terms.

Finally, there are shreds of evidence in the literature that demonstrate practical applicability of AI/ML in cybersecurity practices [123], [11], [116] [182]. However, representative data that replicate real life perspective is inadequate. Therefore, research in cybersecurity security AI is often restricted to same datasets even though, the possibility of extending research from theoretical dimensions are numerous, the unavailability of data places a glass ceiling from a practical dimension [183].

2.10 Conclusion

This chapter highlighted the current gaps in the state-of-the-art, details of healthcare cybersecurity practices, the growing threat landscape, and current regulatory and compliance requirements that must be met regarding healthcare research and data. Also, multilayered cybersecurity approaches is introduced as a concept and fine-tuned to only include AI/ML, database technology and risk assessment in the right proportionality to tackle cybersecurity challenges in the healthcare sector. This chapter also reveals that technical cyber security solutions can only be implemented according to the kind of threat that is confronted by the healthcare facility.

CHAPTER 3 – METHODOLOGY

3.1 Introduction

This is the third chapter of this thesis, and it details *"how"* the research will be conducted. In this chapter a coherent structure of procedures, techniques, tools employed to gather, interpret data in a systematic and organized manner is presented. The two key components of this section are to firstly achieve objectivity by ensuring that personal biases and preconceived notions do not factor in the outcome of the study and secondly, ensure that results are consistent that it can be examined by other researchers to ensure reproducibility.

3.2 Research methodology

According to [184], methodology can be understood as the underlying framework linked to specific paradigmatic assumptions that guide a research work. As emphasized by [185], a research methodology must fulfill the following two criteria:

1. Firstly, the methodology should be the most appropriate to achieve objectives of the research.
2. Secondly, it should be made possible to replicate the methodology used in other research of the same nature.

3.3 Research approach and justification

As shown in Fig. 3.1, there are certain criteria that a research methodology should encapsulate to ensure that it efficiently addresses identified aims and objectives. These are

the general category of the research work, nature of the study, purpose of the study, research design and the data type [186].

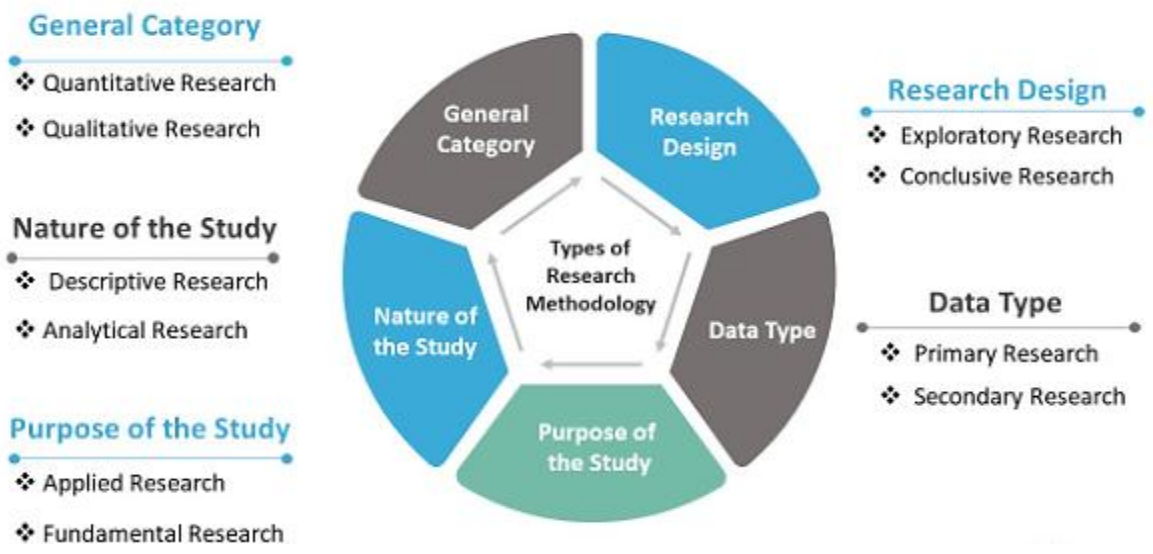


Fig. 3.1: Types of Research Methodology [186]

3.3.1 Research category: quantitative vs. qualitative

From a general category selection standpoint between quantitative or qualitative, this thesis assumes the posture of qualitative research as it focuses on exploring subjective experiences, meanings, and interpretations in a contextual manner. From a computing perspective, qualitative research allows researchers to delve into the complexities and nuances of human-computer interactions, user experiences, and social implications of technology. This research aims to generate in-depth insights, uncover patterns, and gain a comprehensive understanding of the phenomena being studied. Qualitative research helps in articulating better summaries or explanation to theories that are embedded in a given data and since this research leverage sources of data already in the literature, it seems to lean more as a semi- qualitative study [186].

3.3.2 Nature of study: descriptive vs analytical

In terms of the nature of the research, this work is more inclined as analytical research as it involves investigation to uncover supporting data to validate the body of knowledge associated with a given hypothesis. Also, there are components that fits the criteria of a descriptive study as the thesis will rely on the observations made from a lot of variables on the cyberattack datasets to describe some hidden patterns that are pertinent in the datasets [186].

3.3.3 Purpose of study: applied vs. fundamental

As for the purpose of the study, this research leans more as fundamental research because Its objective is to explore and uncover information that has broad applicability, complementing existing knowledge within a particular field or industry. For this research, it is evident that cybersecurity threats are prevalent in the healthcare sector, and this is well documented in literary sources investigations such as articles and news media. Although, this work can also be described as applied research because it entails the study of real-world problems with the objective of creating practical answers and directly addressing specific concerns or challenges plaguing the healthcare sector in terms of cybersecurity threats and the applying AI methods to address cyber threats in the sector. Fundamentally, there is a reliance on current information, theories, and processes to develop practical solutions that may be implemented in a variety of fields, sectors, or domains. Applied research seeks to bridge the theoretical and practical divides by giving actionable insights and contributing to the development of novel and useful solutions [186].

3.3.4 Research design: exploratory vs. conclusive

Based on the design of the research, this aligns more with an exploratory research design because rather than offering definitive proof, exploratory research seeks to

improve understanding of a subject. It takes a flexible and inquisitive approach that is defined by a loosely couple framework in most cases. As a result, hypotheses are not investigated, and the results may have no direct outward application. Nonetheless, the findings frequently help to enhance research in related fields. Exploratory research is typically qualitative, though a quantitative technique with a large sample size may be used in some cases. It is also known as interpretive research or the grounded theory approach due to its adaptability and open-ended nature. Furthermore, as conclusive research findings often have a specific purpose, it is critical to identify this study's research objectives and data requirements. Using a conclusive study design, the findings of exploratory studies can be validated and quantified [186].

3.3.5 Data type: primary vs. secondary

As it pertains to the data used for this study, this research categorizes as secondary because the data has already been collected and published by other researchers or organizations for their own purposes. The OCR (Office for Civil Rights) and CISSM (Center for International and Security Studies Maryland CISSM) database provide a useful data source for this research work. For secondary research datatypes, the data is collected for purposes other than the current research project and is obtained from sources such as books, journal articles, government reports, websites, databases, or previously conducted studies. On the other hand, Primary research data refers to original data collected firsthand by researchers specifically for their research project or study. This data is obtained directly from original sources through methods such as surveys, experiments, observations, interviews, or focus groups. Primary research data is customized and tailored to address the specific research objectives and questions of the study. It is considered firsthand information and is not previously published or available elsewhere [186].

3.4 Methodology

The methodology for this research is schematically depicted in Fig. 3.2. This provides a high-level abstraction of the research methodology highlighting data sources, tools, operations performed on the data and expected results. In the following chapter (Chapter 4 – System design), the high-level presentation of the methodology is further broken down into granular components as they play a critical role in addressing the research questions.

We made a high-level abstraction of the methodology used in this research work as shown in Fig. 3.2. We simplified the methodology into three subcategories (data sources/providers, AI/ML methods and techniques, risk assessment) which are made up of components and processes. Since the methodology is a high-level abstraction of the work, at this level, we conceptualized it to be generic. However, Chapter 5 provides more details of the methodology according to the case study in perspective.

1. **Data sources/Providers**: This research utilized a variety of data sources to achieve defined set objectives (primary data source and auxiliary/secondary datasets).
    - We begin by identifying healthcare cyberattack incident databases for our proposed database of cyberattack incidents. We utilized reported incidents that are documented in other databases such as CISSM and OCR. We also added the capability for user inputted record into the proposed database
    - Next, we built an application that can support the operations of adding new records, searching records, deleting records, and viewing records, and an open API to publicize the content of the proposed database. We affirm that the functionalities of the proposed database primarily provide a mechanism for cataloging already reported incidents. The patterns observed only function to inform users of the evolving threat landscape and not necessarily be useful for AI/ML operations.

- Lastly, we identify other auxiliary datasets (Phishing dataset, ECU-IoHT dataset, Ton-IoT dataset, UNSW dataset and X-IIoTID dataset for specific AI/ML tasks.

NB: To clarify, this work initially intended to use existing databases of cyberattack incidents to perform AI/ML techniques for the classification of cyberattacks. However, we found through initial experiments that data from the CISSM and OCR databases are not suitable for AI/ML operations because the data are categorical and do not conform to the principles of independent and identically distributed random variables, which are necessary to make ML models usable, as articulated by [123].

Therefore, we had to change our view of the research problem by extensively looking at existing cybersecurity research that applied AI/ML techniques, and from the plethora of the works available, we found the context of applications of AI/ML in cybersecurity to be confined to tasks such as intrusion detection, phishing detection, malware detection, etc. This helped us situate our problem better and to understand what data is applicable according to the task and what algorithms or techniques have been used to solve such problems.

2. **AI/ML methods and techniques**: A number of AI/ML techniques were used on the auxiliary datasets during experimentation to predict different scenarios.
    - We used the phishing dataset for classifying phishing attacks.
    - We used the ECU-IoHT dataset for anomaly detection.
    - We used the Ton-IoT dataset, UNSW dataset and X-IIoTID dataset for network intrusion detection.

Using these datasets where necessary because they are more suited for AI/ML tasks. So therefore, we used support vector machines, random forest, gradient boosting, multilayer perceptron, and decision trees for the tasks. A similar but unique AI/ML pipeline was used on the datasets to produce different outputs resulting from the application of the different

algorithms mentioned. Chapter 5 provides more details on the specifics of how the pipeline will be implemented.

3. **Risk assessment**: Beyond building a database that enables functional manipulation to present the threat landscape of cyber incidents, the output from the database is utilized within the scope of this work alongside the results from the application of AI/ML techniques to feed into a risk assessment model. The components that make up this phase are,
    - We proposed the BUH risk assessment model by identifying critical assets, threats to the assets and likelihood of the threats exploiting vulnerabilities in the assets.
    - We leveraged the NIST framework, ISO 27000 series and the Enisa framework as a base. Our BUH model distilled applicable components of the frameworks and proposed a tailored framework for Baze University hospital which is the adopted healthcare facility we investigated as a case study.
    - The output from this process is the risk assessment register that prioritizes the risk and their severity.
    - Lastly, we proposed controls for the identified risk.

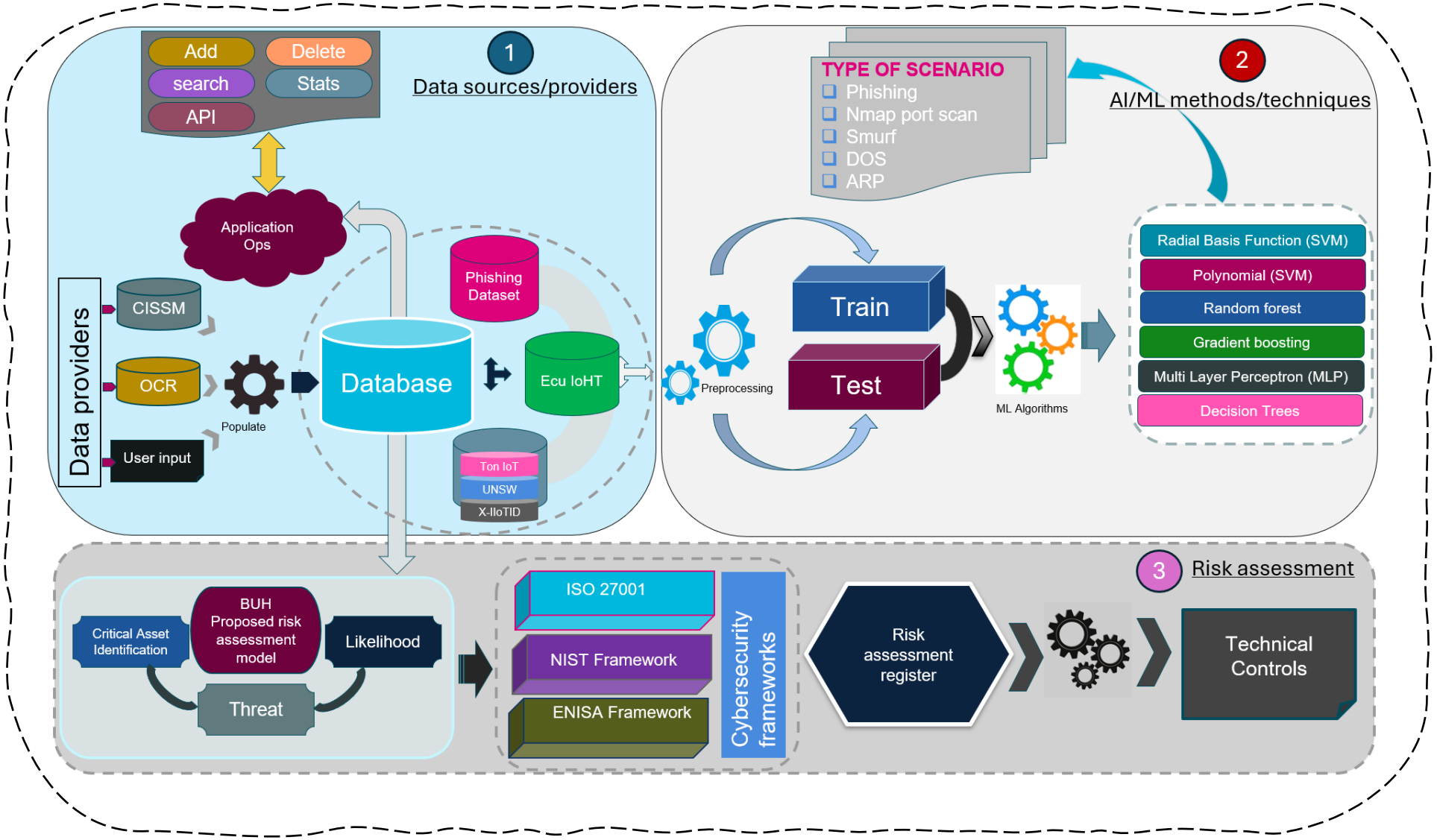


Fig. 3.2: High-level abstraction of methodology

3.5 Tools and frameworks

The research methodology in this study employs various major tools and frameworks to facilitate the research process. These tools are instrumental in gathering and analyzing data, as well as supporting the overall research objectives. Necessary justification is provided why some tools are preferred over others and, declaring these tools in this chapter is important for reproducibility of the experimentation.

3.5.1 Visual studio code

VS Code, short for Visual Studio Code, is a source-code editor developed by Microsoft using the Electron Framework. It is compatible with Windows, Linux, and macOS operating systems. Included within its features are native capabilities for JavaScript, TypeScript, and Node.js, and an extensive collection of extensions for various languages and runtimes (such as C++, C#, Java, Python, PHP, Go, and .NET) [187].

3.5.2 Anaconda distribution environment

Anaconda is a free distribution (collection of packages or libraries) software that provides a specialized toolkit tailor-made for research and scientific purposes. By installing Anaconda, you gain access to diverse environments that enable coding in either Python or R which is widely used for working on data science, machine learning applications, predictive analytics, and large-scale data processing projects. These environments, also referred to as integrated development environments (IDEs), serve as platforms or applications that greatly simplify code development. They play a similar role to text processors like Microsoft Word, Google Docs, and Pages, which are used for writing text, but IDEs offer a multitude of additional functionalities [188]. IDEs encompass a wide range of useful features for writing, editing, and debugging code, visualizing and analyzing data, storing variables, presenting results, and collaborating on projects [188].

Python is an interpreted, general purpose, versatile, interactive, object-oriented, open-source programming language. It is considered among many to be a concise and easy to read programming language. Python stands out for its remarkable versatility, offering a wide range of applications across various domains. From machine learning to web development and beyond,

NumPy is a widely adopted Python library that plays a vital role in numerical computing. This open-source package introduces a powerful array object capable of handling large datasets with multiple dimensions efficiently. Its extensive collection of mathematical functions and operations are specifically optimized for numerical computations.

Pandas is a robust and widely used open-source Python data manipulation and analysis library. It includes data structures and functions for handling and analyzing structured data, such as tabular and time-series data. Pandas allow users to easily load, manipulate, filter, and transform data making it an invaluable tool for data cleaning, exploration, and preparation tasks. Pandas is a well-liked option for data scientists, analysts, and researchers working on data-intensive projects because it integrates well with other libraries in the Python ecosystem [189].

Scikit-learn is a Python package or library for machine learning. It is an open-source software built on top of SciPy offering a comprehensive range of functionalities supporting both supervised and unsupervised learning. Scikit-learn provides very comprehensive tools for model development, selection, evaluation, and a host of other utilities including data pre-processing capabilities. The primary features of scikit-learn encompass classification, regression, clustering, dimensionality reduction, model selection, and pre-processing tasks. This library is renowned for its user-friendly nature, ensuring ease of use.

Matplotlib is an open-source Python toolkit that is frequently used for creating high-quality visualizations and plots. It has a wide range of tools and functionalities for creating various types of charts, graphs, and figures. Users can use Matplotlib to create line plots, scatter plots, bar plots, histograms, and other data visualizations to effectively convey and present data.

3.5.3 MySQL

MySQL is a relational database management system (RDBMS) that is open source and widely used for organizing and storing structured data. It offers a scalable and effective solution for organizing, accessing, and manipulating data across a wide range of applications and industries [190]. According to [191], MySQL is the most widely used, open source relational database management system owned by the Oracle Corporation. MySQL includes support for different storage engines, powerful querying and indexing algorithms, transactional support, and robust security measures [192].

3.5.4 Flask

Flask provides a minimalist yet powerful framework for building web applications, enabling developers to create efficient and scalable solutions with ease [193]. Flask is classified as a micro framework due to its ability to execute framework activities without the use of external libraries. It has its own collection of tools, technologies, and libraries to meet the needs of web application development.

3.6 Principles of AI/ML algorithms used

This research utilized five AI/ML algorithms. Each algorithm was tailored for a cybersecurity use case as illustrated in Fig. 3.2. Factors that guided the selection of the

algorithm include the nature of the dataset, the inherent complexity of the attack patterns and the desired classification performance.

- **Phishing detection:** We classified phishing attacks based on extracted web features and applied Support Vector Machines (SVM) with Radial Basis Function (RBF) and polynomial kernel.
- **Anomaly Detection in IoHT:** We used the ECU-IoHT dataset to classify network attacks in an internet of health things environment. The attack classes are Nmap port scans, ARP spoofing, DoS, and Smurf attacks. The algorithms used for this case study include Gradient Boosting (GB), Multilayer Perceptron (MLP), Decision Trees (DT), and Random Forest (RF)
- **Intrusion Detection:** We used the decision tree algorithm to classify malicious traffic in networks. We achieved this using TON-IoT, UNSW-NB15, and X-IIoTID.

Detailed justification is provided for the algorithms used in chapter 5. When implementing different case studies alongside their mathematical formulations. In this section, we limit our explanation to the operating principle of each of the algorithm.

1. **Support Vector Machines (SVMs):** A support vector machine creates a hyperplane or set of hyperplanes in a high or infinite dimensional space that can be used for classification, tasks. On the surface, a decent separation is obtained by the hyper-plane with the greatest distance to the closest training data points of any class (known as the functional margin), because the larger the margin, the lower the classifier's generalization error [194].
2. **Decision Trees (DT):** Decision Trees are a non-parametric supervised learning technique for classification and regression. The principle works by building a model that can predict the value of a target variable using basic decision rules derived from data attributes. In classification tasks the decision tree algorithm takes two arrays as input. X (n_samples, n_features) that holds the training samples, and an array Y of

integer values, shape (n_samples,) that holds the class labels for the training samples [195].

3. **Gradient Boosting (GB):** This approach constructs an additive model in a forward stage-wise manner, allowing for the optimization of arbitrary differentiable loss functions. In each level, n_classes regression trees are fitted to the negative gradient of the loss function, such as binary or multiclass log loss. Binary classification is a specific instance in which just one regression tree is generated [196].
4. **Multilayer Perceptron (MLP):** MLP is a feedforward neural network. It learns a function $f: R^m \rightarrow R^o$ by training on dataset where m indicates the number of dimensions for the input and o indicates the number of dimensions for output. $X =_{x1,\ x2,\ldots,xm}$ and a target y, MLP can learn a nonlinear function for both classification and regression. More layers can be added between the input and output layer [197].
5. **Random Forest (RF):** RF is a tree-based classifier, and it functions by adding randomization to the classifier design results. The ensemble prediction is calculated by averaging the individual classifier predictions [196].

3.7 Interaction between tools

The tools utilized in this study were designed to interact and complement each other, facilitating a seamless workflow and enhancing the research process as some objectives seem to be interwoven. These tools work synergistically, in data collection, analysis, and interpretation. A loosely coupled interaction between these tools is presented in Fig. 3.3. The anaconda environment allows for the development of ML algorithms, the VS Code server handles the logic and model execution, while the XAMPP server ensures secure data storage and retrieval.

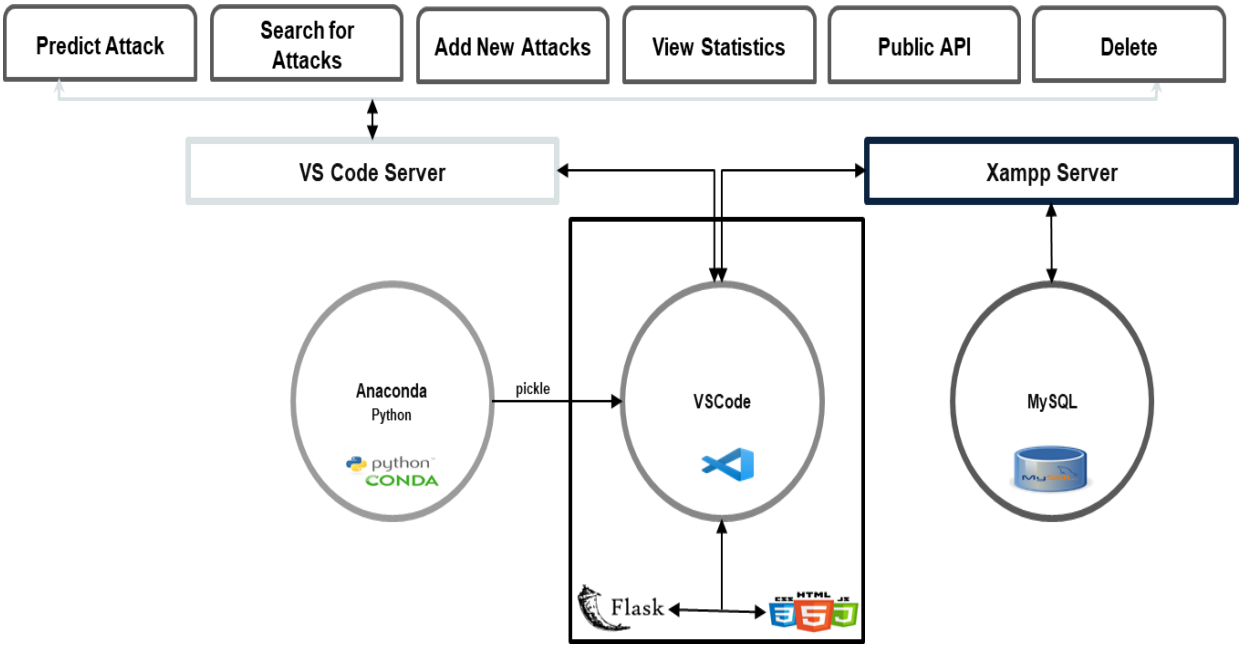


Fig. 3.3: Interaction between tools and frameworks

3.8 Mapping resilience phases to control, impact, and organizational factors

Following up on resilience considerations introduced in chapter two, which align with the four core phases of resilience, consisting of preparation, absorption, recovery, and adaptation. We account for this in our methodology section by mapping each resilience phase to relevant impact and control variables. For instance, control variables in our work would imply predefined system settings or configurations, while impact variables would mean measurable outcomes that can be quantified. However, since the scope of this research focused more on technical controls as identified by the risk register, we highlight both organizational and human factors of resilience with their integration points without explicit integration protocols. This mapping reinforces the systemic and socio-technical scope of the research, demonstrating that resilience is not only a product of technical robustness but also of staff readiness, coordination capacity, and institutional learning. Furthermore, linking these controls with behavioral and organizational enablers, the

proposed multilayered framework offers a more holistic understanding of how resilient thinking applies to the healthcare context. Table 3.1 presents an integrated view of the proposed framework's alignment with the four core phases of resilience.

Table 3.1: Resilience Phases with control, impact, and Organizational factors

	Resilience Phase	Control variables	Impact variables	Organizational /Human factor
1.	Prepare	AI model parameters, attack scenarios, dataset types, system configurations	Readiness level, risk scores, threat model coverage	Training participation, awareness readiness, protocol dissemination
2.	Absorb	Anomaly thresholds, detection pipeline architecture, access control settings	Detection accuracy, F1 score, false positives/negatives	Operator alertness, staff response behavior, escalation pathways
3.	Recover	Failover strategies, risk registers, restoration protocol designs	Time-to-response, mean-time-to-recovery, impact score	Coordination efficiency, recovery team readiness, stakeholder feedback
4.	Adapt	Incident logs, retraining frequency, system updates, feedback loops	Model improvement rate, parameter settings, risk register updates, organization policy change	Organizational learning, policy revision, system-user behavior change

3.9 Conclusion

This chapter provides the methodology adopted for the research. Including tools and frameworks that were used to achieve the research objectives. Justification is provided by the author to provide validity to the research approach used and the tools that were adopted and used throughout the research. Overall, this chapter's goal is to be able to achieve reproducibility of results when the author's methods are followed.

CHAPTER 4 – SYSTEM DESIGN

4.1 Introduction

This chapter details the analysis and design principles used in developing the web application and database of cyberattack incidents. It also defines the functionality that the system will perform and the actors that will interact with the system. Since one of the objectives of this thesis is to design a database of cyberattack incidents, it is important that the most appropriate industry standards are adhered to.

4.2 Analysis and design

Often, organizations may have reasons to redesign their organizational requirements or change processes to achieve business objectives, which may be related to security, growth and profitability. System analysis and design provides the right orientation to enable these goals to be achieved. Typically, emphasis is placed on the functioning of systems, their relationships with other subsystems, and the ability of both to achieve a specific goal. Organizations may assess how improvements might address their needs now and in the future. The process by which an organization creates a newer system or strategy to supplement or replace an older one in a suitable scenario is known as system design [198].

Although system analysis and design methodologies are very comprehensive and include renowned modules such as project selection and management, requirements engineering, process modeling, architectural design, user interface design, etc., this research only employed applicable components of system analysis and design such as System Development Life Cycle (SDLC) methodology and some components of requirements elicitation. The reason for using the system analysis and design approach to

build the web and database application in this research is to better understand the components of the system and their interrelationships. In other words, system analysis and design provide a reliable structure for understanding requirements, identifying and decomposing problems, defining system specifications, building system models, and evaluating alternatives.

4.3 System Development Life Cycle (SDLC) methodology

The system development life cycle in systems engineering, information systems and software engineering is a systematic process for creating, deploying and testing information systems. The SDLC also provides a conceptual model for project management [199]. The SDLC usually constitutes a process of planning, analysis, design, implementation and testing or maintenance as a final stage as illustrated in Fig. 4.1.

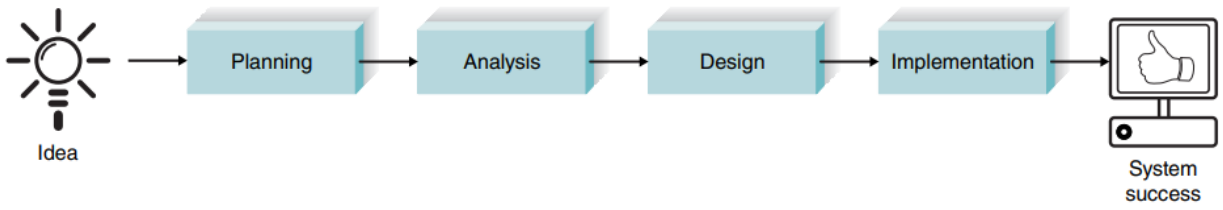


Fig. 4.1: SDLC. [200]

4.3.1 SDLC Methodology

Whilst the SDLC provides foundational paradigms or processes for building IS systems or software products, the methodology provides a formal approach of implementing the SDLC which consists of tasks steps and deliverables. Some of the popular known methodologies known to practitioners and academicians include waterfall development, parallel development, V-model, iterative development, system prototyping, throwaway prototyping and extreme programming. Depending on the characteristics of the project,

there are some factors that should be considered to select the appropriate methodology for SDLC [200].

1. **Clarity of user requirements**: This refers to the extent of the clarity of the requirements for the new system. Unclear requirements are better suited for some methodologies than others.
2. **Familiarity with technology**: This refers to the wealth of experience the developers have with the technology used in designing the new system. Having insufficient expertise with a particular technology often compromises developing systems that meet defined requirements.
3. **System complexity**: Factors such as system complexity are important when selecting a SDLC methodology. When systems are complex to the point that they must span across other organizational units, integrate with other existing systems, and provide an extensive array of features then, there are some SDLC methodologies that are more suitable for this reality than others.
4. **System Reliability:** When choosing a methodology, strong considerations should be given if there is any tolerance for some downtime, or will this system need to be extremely reliable.
5. **Short time schedules:** Short time schedules are more appropriate for some methodologies such as agile development. Other methodologies don't seem to work well with timelines that are not flexible.
6. **Schedule visibility:** This refers to the stakeholder's expectations regarding the interaction in relation to progress participation. When high engagement and visibility is required, some methods are more suitable than others.

4.3.2 Throwaway prototyping

For our proposed database application of cyberattacks and defining the functionalities that it should possess. We selected throwaway prototyping shown in Fig. 4.2

as the methodology of choice because it allows for a scaled-down version of the solution that can be tested for improvements [200].

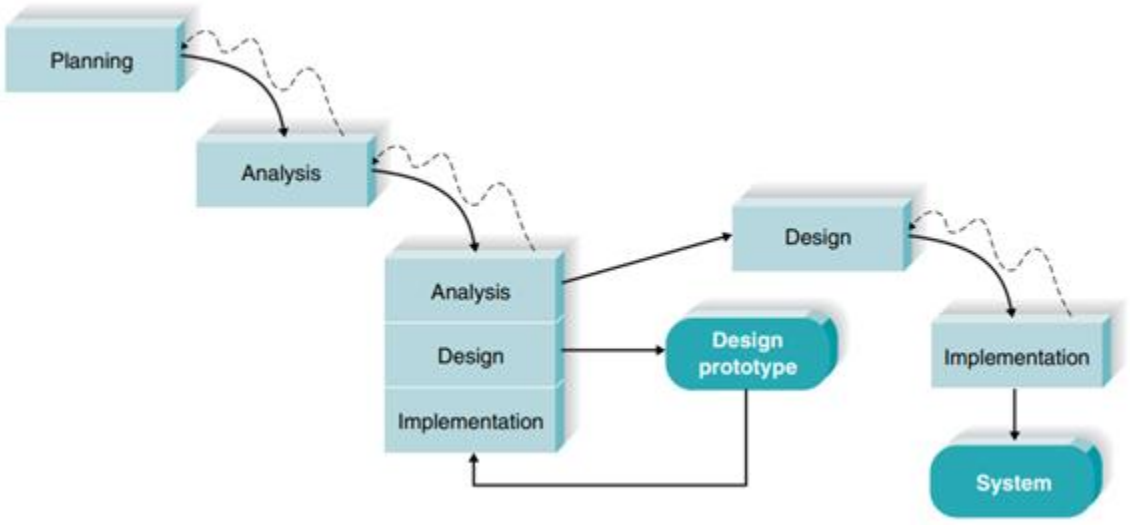


Fig. 4.2: Throwaway prototyping [200]

4.3.3 Justification for methodology selection

The use of disposable prototypes is advantageous for many reasons, as shown in Figure 4.3. It offers a lot of flexibility when working with unclear requirements as it provides the right environment for engagement and iterative feedback loops. Additionally, this methodology encourages the development of multiple prototypes in the design phase, which is helpful in anticipating likely problems that may arise, thereby significantly reducing the risk associated with building a system.

Usefulness in Developing Systems	Waterfall	Parallel	V-Model	Iterative	System Prototyping	Throwaway Prototyping	Agile Development
With unclear user requirements	Poor	Poor	Poor	Good	Excellent	Excellent	Excellent
With unfamiliar technology	Poor	Poor	Poor	Good	Poor	Excellent	Poor
That are complex	Good	Good	Good	Good	Poor	Excellent	Poor
That are reliable	Good	Good	Excellent	Good	Poor	Excellent	Good
With short time schedule	Poor	Good	Poor	Excellent	Excellent	Good	Excellent
With schedule visibility	Poor	Poor	Poor	Excellent	Excellent	Good	Good

Fig. 4.3: Justification for methodology selection [200]

4.4 Requirements engineering

Simply put, a requirement is a list of the functions or qualities that the system must possess [200]. With our proposed database, it is essential to define all system needs, functionalities, constraints and expectations. Therefore, we used the UML approach to model our web application as a set of interacting objects and define how each function is implemented using sequence diagrams. Without requirements engineering, there will be no anchor to establish if our final system sufficiently implements the functions we modelled.

4.4.1 Object oriented analysis and design

In systems/software engineering, there are majorly two methods of analysis and design. The first is structured analysis and design which focuses on disassembling a system into smaller, easier-to-manage parts. The overall goal is to show how data moves through the sub-systems, and it typically makes use of graphical tools such as Data Flow Diagrams (DFDs). The second is Object-Oriented Analysis (OOA) which models a system as a group of interacting objects. Real-world entities and their interactions are mirrored by objects, which embody both behavior and data. To promote a more modular, reusable, and intuitive design approach, OOA places a strong emphasis on the identification of classes, objects, attributes, and methods. The use of use cases, sequence and class diagrams or the

UML is usually used in OOA. The major advantage of OOA over structured analysis is that it provides a more systematic model of breaking down systems functionality that is easily reusable, reliable and maintainable [201].

4.4.2 Functional requirements

Functional requirements establish the baseline functionality necessary for a system to be useful [202]. It also encapsulates the capabilities or things a system must do to support a user in completing a task [200]. Following this definition in this thesis, we outline the key functional requirements for the system that is captured in the first objective, and they are as follows.

1. The system shall allow a user to predict social engineering (phishing) attacks.
2. The system shall allow a user to search for attack from the database.
3. The system shall allow a user to add new attacks to the database.
4. The system shall allow a user to view statistics of attack distribution.
5. The system shall allow users to view obtain database records via a public API.
6. The system shall allow a user to delete records of cyberattacks from the database.

4.4.3 Use case modeling

A common but important graphical notation in the Unified Modeling Language (UML) is the use case [203]. In its most basic form, a use case diagram shows how a user interacts with the system, illustrating how the various user types interact with the system and trigger unique functionalities. In most cases, all functional requirements transform to instantiated use cases or auxiliary use cases with *"extended"* or *"included"* dependencies.

Fig. 4.4: shows the use-case model for the proposed system with two main actors: a general user and an admin. Users can add, search, delete, and view statistics on cyberattack incidents, as well as share data through APIs. The system's core feature is

phishing prediction, which analyzes indicators such as SSL certificates, long URLs, pop-up windows, suspicious domains, and other phishing signs to determine the likelihood of a phishing attempt.

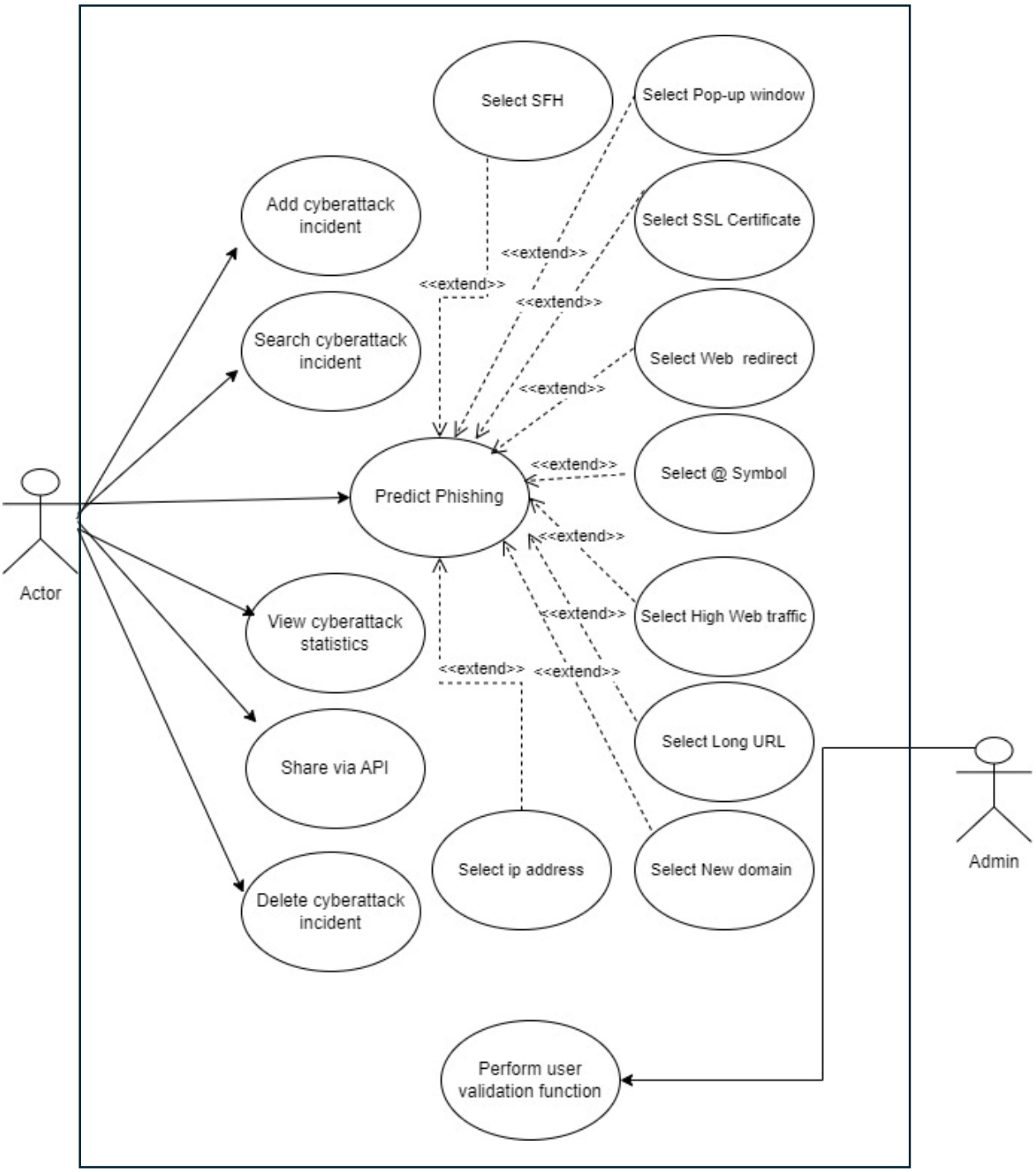


Fig. 4.4: Use case diagram

4.4.4 Sequence diagram modeling

Sequence Diagram diagrams are behavioral diagrams that depict time sequences between objects and the processes involved to implement a given functionality. Usually, it depicts a logical transformation from the use case diagram to dynamic model of the implementation. Fig. 4.5: through 4.10: shows the sequence diagram from the decomposition of the use case and functional requirements. The models typically describe how an actor interacts with the to-be system to achieve different functional requirements. For example, Fig. 4.5: shows how a user can enter parameter for phishing prediction on the web application. The system forwards the message using the parameters in the flask control to predict if the user input corresponds to either phishing, suspicious or legitimate. After that, the entered parameters will be saved in a phishing table in the MySQL database and the result page displayed to the user.

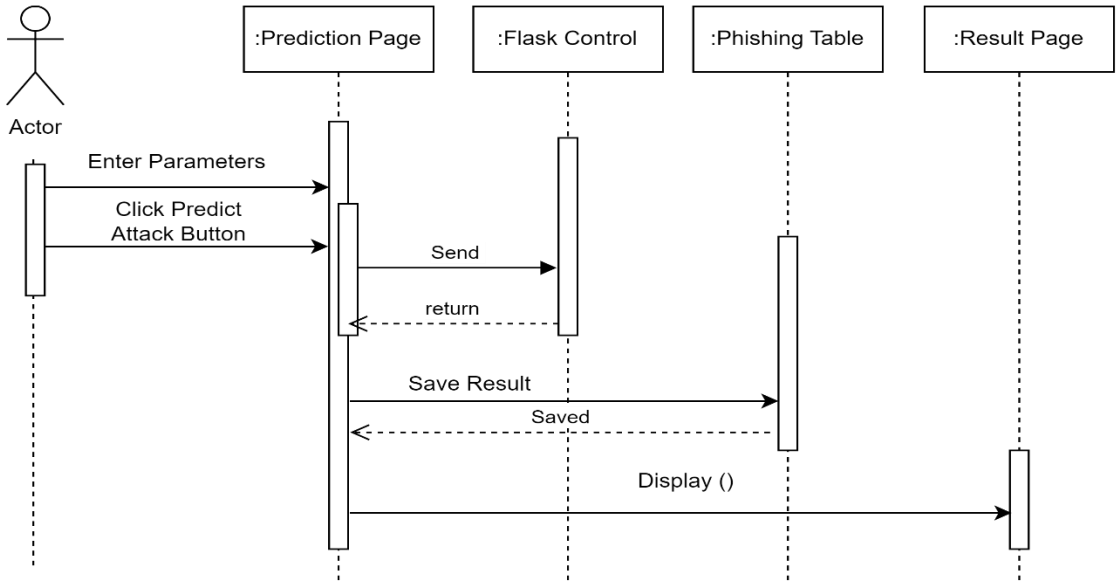


Fig. 4.5: Predict phishing attack

Fig. 4.6 shows the sequence diagram for the search function. The function begins with a user clicking the search button on the homepage. The system sends the request through the Flask control to the cyberattack table. If the record is found, it is displayed to the user on the result page. However, if the search query by the user is not found in the cyberattack table, then the system throws an exception “no cyberattack incident in database” on the homepage.

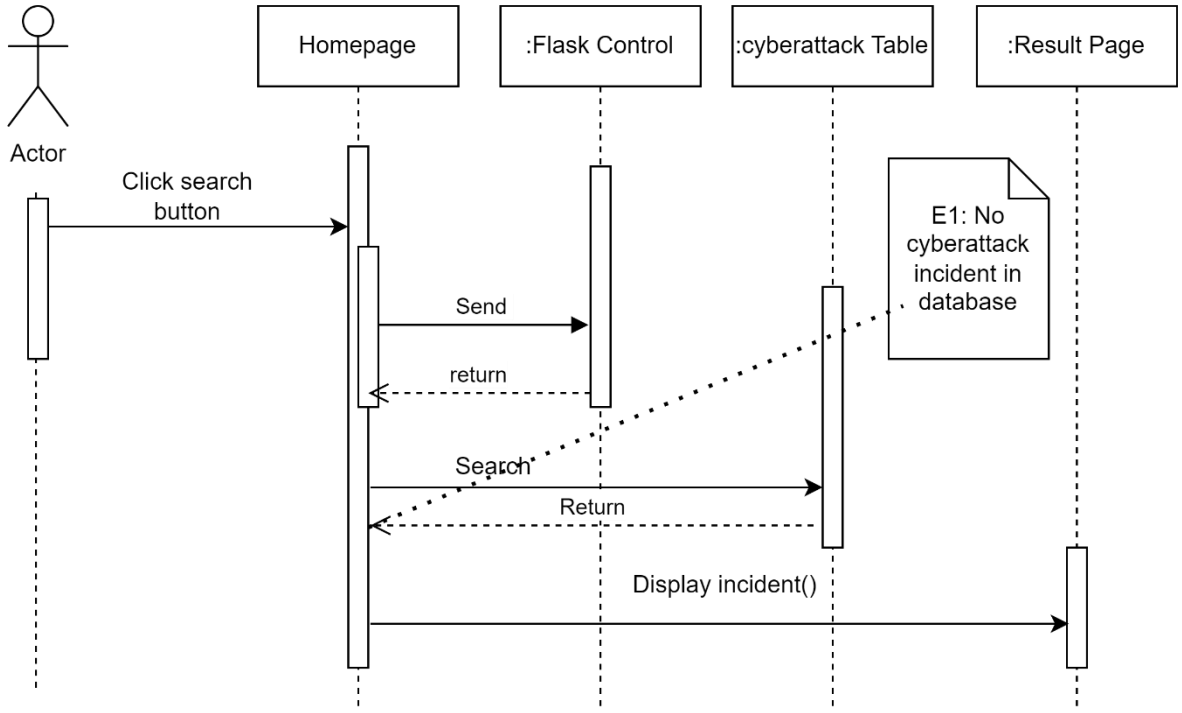


Fig. 4.6: Search cyberattack incident

Fig. 4.7 shows the sequence diagram for the add new cyberattack function. The function begins with a user clicking the add new cyberattack button on the homepage. The system prompts the user to enter parameters. After the user enters parameters, the Flask control process the data and saves it to the cyberattack table. If the record is already in the table, then the systems throw an exception “incident already exist”. If the user does not enter the complete parameters, the system throws another exception prompting the user to “enter complete parameters”. Subsequently, the user is directed to the result page.

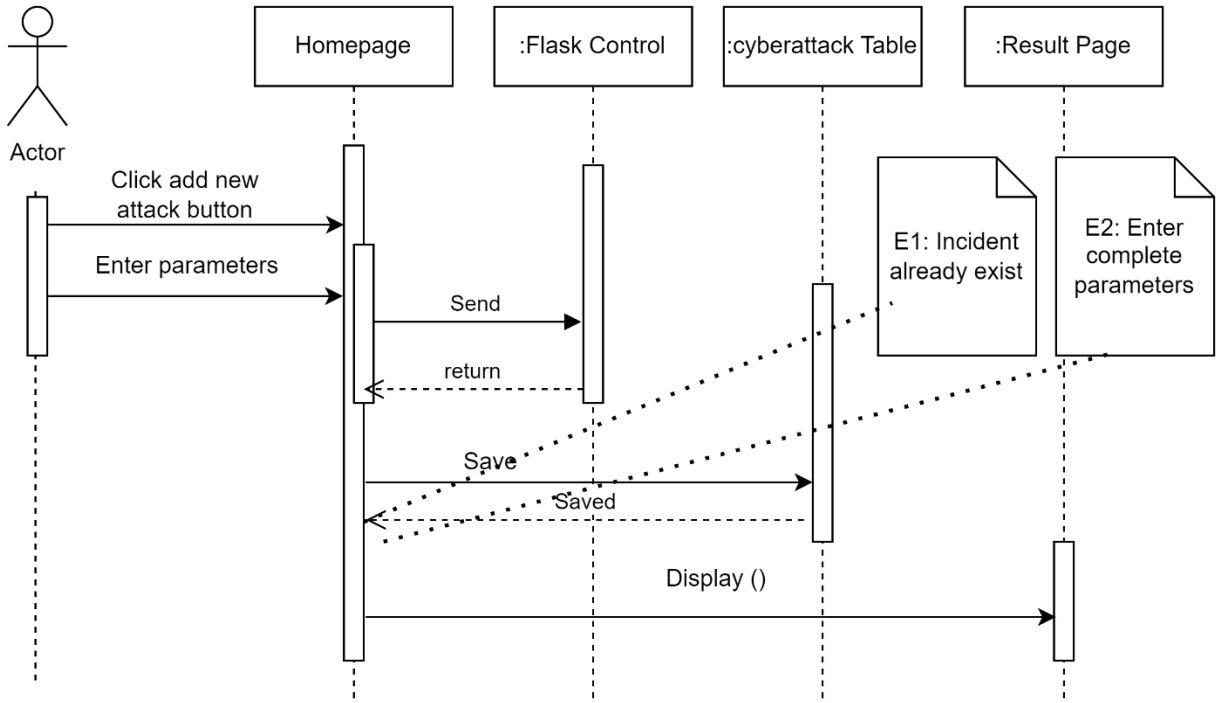


Fig. 4.7: Add new cyberattack incident

Fig. 4.8 shows the sequence diagram to view statistics of cyberattack incidents. To initiate this function, the user selects the parameters for what he wants to view (For example, view cyberattack incident by type). The Flask control queries the cyberattack table and displays the statistics on the result page.

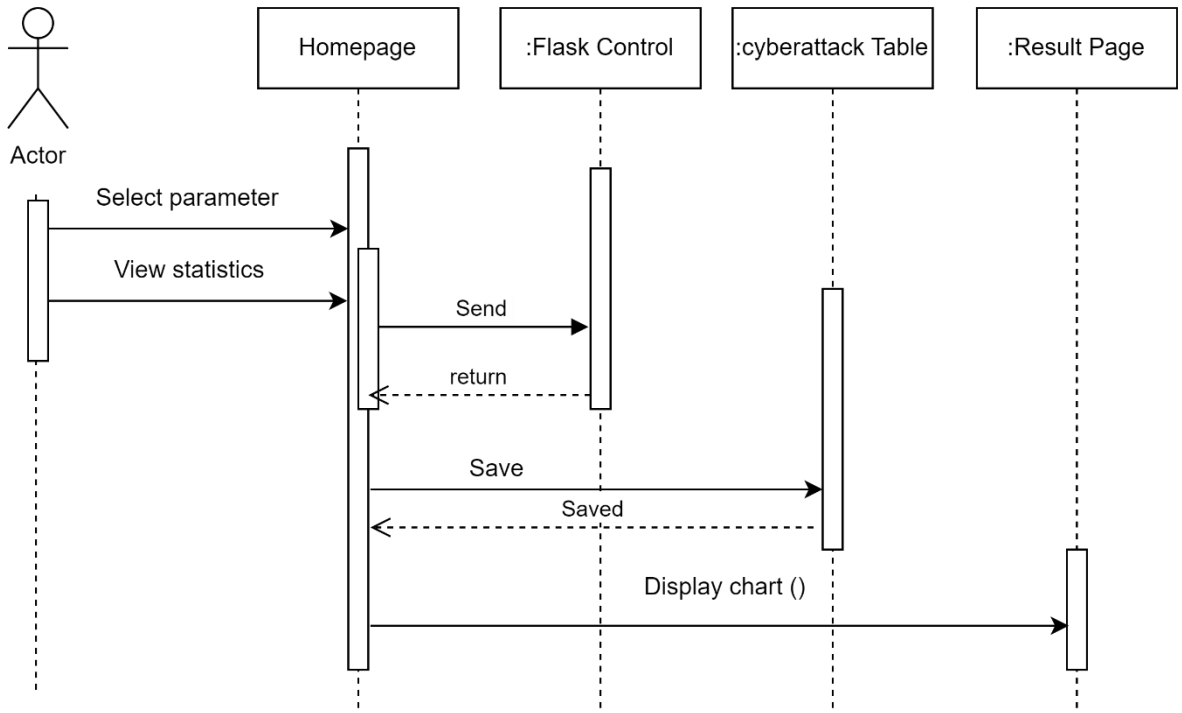


Fig. 4.8: View statistics of cyberattack incidents

Fig. 4.9 shows the sequence diagram for viewing the API for cyberattack incidents. To initiate this function, the user clicks the API button and flask control parse the data as a json or csv file for the user on the result page.

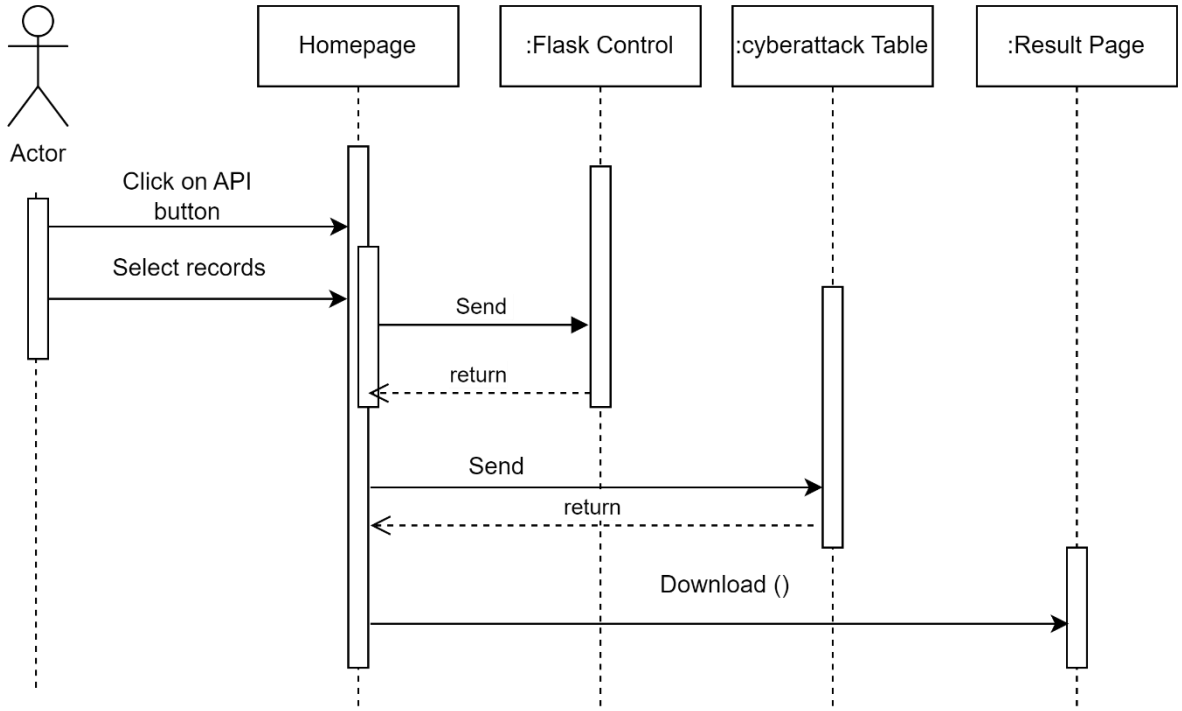


Fig. 4.9: API of cyberattack incidents

Fig. 4.10 shows the sequence diagram for deleting cyberattack incidents. To initiate this function, the user selects the record and clicks the delete button on the homepage. The flask control handles the delete operation and delete the record.

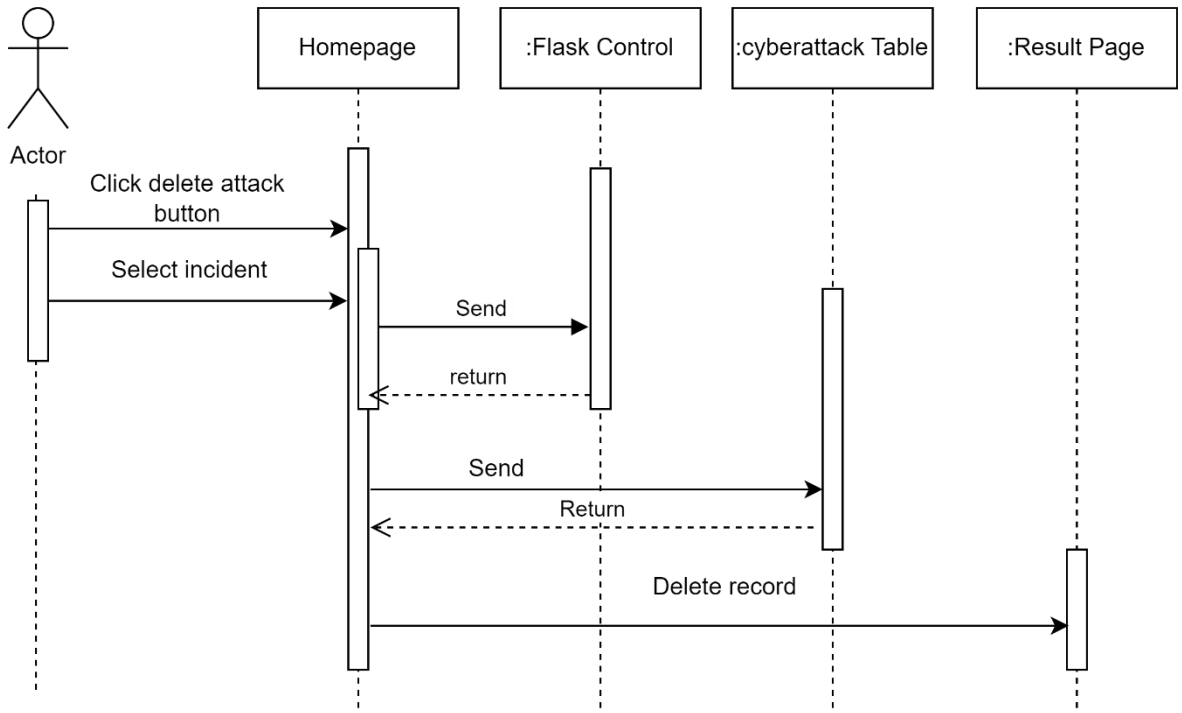


Fig. 4.10: Delete cyberattack incidents

4.5 Database design

This thesis uses the MySQL database as the repository of choice in collating cyberattack incidents or records. The key reason for this decision is its open-source nature, ability to run on a local host, integrates well with most development environments or cross platforming capabilities, ease of use and versatility. Fig. 4.11: shows the proposed database design and the tables as captured in MySQL.

4.5.1 Conceptual clarification (database vs dataset)

A database is intended to efficiently store, retrieve, and manipulate massive volumes of data. It is often hosted on a server and can accommodate numerous users at the same time. Databases are required for processes involving complicated querying, analysis, and regular

updates. They also include security and backup features to govern concurrent data access and maintain data integrity.

A dataset, on the other hand, is largely utilized for analysis and modeling, most commonly in domains such as machine learning or statistical research. It is often smaller than a database and can be saved in a variety of forms, including CSV, Excel, and JSON. Datasets are useful for specific activities like training machine learning models or doing data visualizations. They are typically utilized for research rather than long-term data management.

Table 4.1: Dataset vs Database Differences

Aspect	Dataset	Database
Data structure	Contains specific observations or measurements	Holds different types of data, allows operations like adding, deleting and updating data
Connectivity	Disconnected architecture; does not require active connections	Requires active electronic connections for data operations
Data organizations	Data can be unsorted; used mainly for analysis with statistical and ML tools	Data is well-organized in computerized files with search and cross-referencing capabilities
User interaction	Users analyze the data to understand specific aspects	Users interact with data using commands to sort, search and generate reports
Data manipulation	Limited manipulation capability; operations are mainly for hypothesis and research	Language for manipulation includes SQL to support queries and management

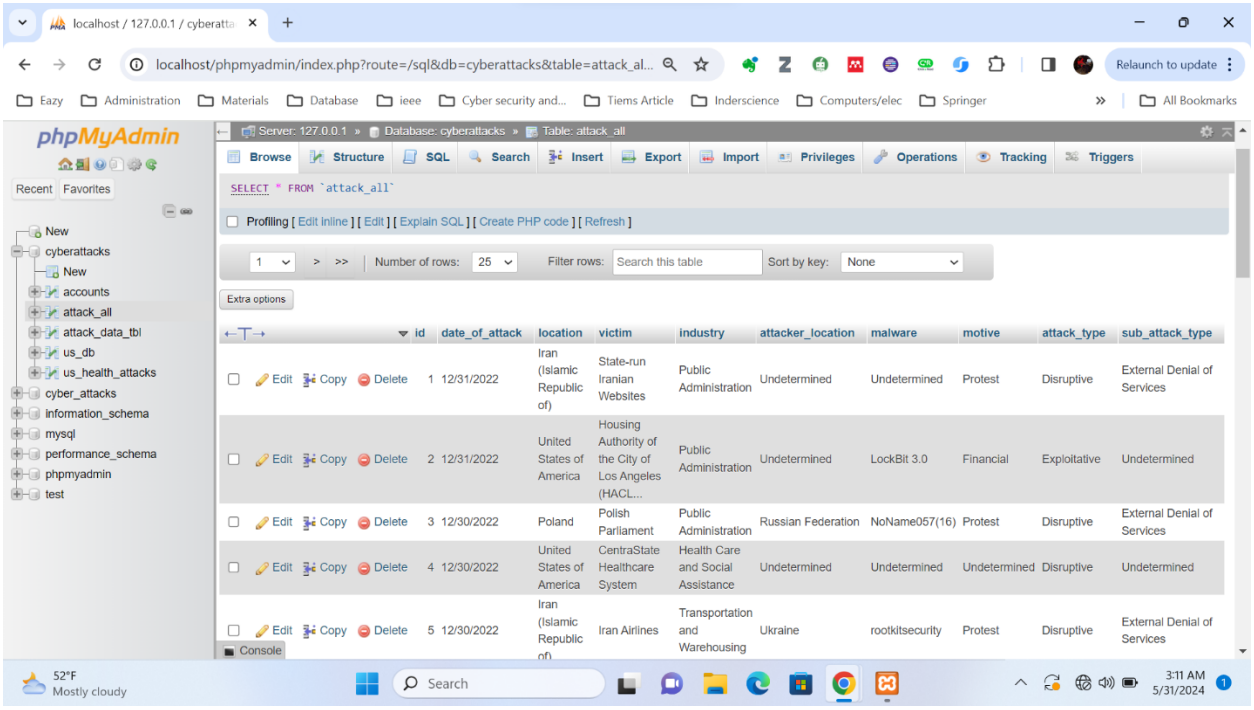


Fig. 4.11: Cyberattack incidents in MySQL

4.5.2 Data sources

The healthcare sector typically maintains some crucial databases for the execution of their operations, such as electronic health records, claims databases, disease registries, clinical trial databases, public health databases, and research databases. However, cyberattack incident databases that catalog known cyberattacks are a rarity in the sector [204]. Investigation into public records reveals two cyberattack sources that seemingly mirror cyberattacks in the healthcare sector and were adopted to populate the proposed database.

4.5.2.1 Office for Civil Rights (OCR)

The Office for Civil Rights (OCR) of the United States of America - Department of Health and Human Services (HHS) oversees enforcing and promoting civil rights legislation

that protect persons' rights in healthcare settings. The Office for Civil Rights (OCR) guarantees that healthcare companies and providers follow federal laws and regulations such as the Health Insurance Portability and Accountability Act (HIPAA) Privacy, Security, Breach Notification Rules and the Patient Safety Act and Rule. These regulations jointly protect fundamental rights like nondiscrimination, conscience, religious freedom, and the privacy of health information at organizations covered by them. In accordance with HITECH Act section 13402(e)(4), the Secretary is required to publish a compilation of events involving breaches of unsecured protected health information affecting 500 or more individuals. The Office for Civil Rights aims to preserve individuals' rights to equal access to healthcare services as well as the privacy and security of their personal health information. It examines complaints from individuals who believe their rights have been infringed and conducts compliance audits of covered businesses such as healthcare providers, health plans, and healthcare clearinghouses to ensure compliance with privacy and security standards. The OCR plays a critical role in protecting individuals' privacy and civil rights in the healthcare system through its enforcement activities, counseling, and education campaigns. It seeks to establish a compliance culture and to promote fair access to healthcare for all people, regardless of their origin or characteristics [205].

4.5.2.2 Center for International and Security Studies Maryland (CISSM)

CISSM is an abbreviation for the Center for International and Security Studies at Maryland, which is part of the school of public policy. It is a research University located at the University of Maryland, College Park. CISSM undertakes multidisciplinary research and analysis on global security concerns such as nuclear security, cybersecurity, terrorism, international relations, and strategic studies. The center's mission is to provide policy-relevant insights and expertise to help influence decision-making in international security, defense, and conflict resolution. CISSM conducts collaborative research initiatives, produces academic publications, organizes conferences and events, and provides

educational programs to create a better understanding of the world's complex security concerns [206]. CISSM makes available a cyber event database that is publicly accessible, and the motivation of the author is based on inadequate classification methods pertaining to cyber events which hinder technical personnel, organizational leaders, and policymakers from engaging in substantive and nuanced discussions regarding the challenges they encounter. The research paper by Harry and Gallagher [207] proposes a novel taxonomy of cyber effects, which is utilized to analyze a dataset comprising publicly reported cyber events. The database serves as a comprehensive collection of open-source information pertaining to a variety of publicly acknowledged cyber events targeting both private and public organizations. It encompasses events spanning from 2014 to the present, each meticulously coded to ensure standardized information on key aspects such as the threat actor, their country, motive, target, end effects, industry, and the country impacted [208]. Furthermore, the database provides source links to the respective news sources, allowing for easy access to additional information and verification reveal variations across industry sectors concerning the magnitude of events experienced, the distribution between exploitative and disruptive event types as well as the methods employed for data theft or disruption of organizational operations. These results underscore notable sector-specific distinctions and emphasize the importance of developing tailored strategies based on a comprehensive understanding of the threat landscape [209].

4.5.2.3 The European Repository of Cyber Incidents (EuRepoC)

The European Repository of Cyber Incidents (EuRepoC) is an independent research consortium that aims to improve understanding of the cyber threat environment in the European Union and beyond. The group was launched in November 2022 and the primary goals are to promote data-driven discussions and decisions in the field of cyber security, as well as to raise awareness of cyber security concerns. They accomplished this by developing an analytical framework for assessing and comparing the 'life cycle' of cyber events, with a focus on technological, political, and legal considerations. The data from their studies are

useful to a diverse variety of stakeholders, including government officials, civil society members, corporate executives, journalists, educators, students, and the public. They have a dedicated team that tracks, documents and analyses public reports and then compile it into an interdisciplinary repository that is publicly accessible [210].

Although there are some databases in other sectors such as AI safety and security [98] Federal Aviation Administration (FAA) [99], Adversarial Tactics, Techniques, and Common Knowledge (ATT&CK) [105], the United States Food and Drug Administration (FDA) [111]. In this work, we tried to focus on healthcare cyberattack incidents that relate only to the healthcare sector. The CISSM and especially the OCR database appears to possess cyber incidents that fit the scope of this research study.

4.5.2.4 Codification of cyberattack incidents

There exists a long non-exhaustive list of cyber-attack types like Malware, Phishing, DDoS, Man-in-the-Middle (MitM) Attack, SQL Injection, Social Engineering, Zero-Day Exploits, Insider Threats, Password Attacks, Advanced Persistent Threats (APTs), Ransomware, Exploit Kits etc. [211], [212], [34], [1], [213]. With the long list of cyber-attack types comes the enormous challenges of properly classifying an attack type because of how unprecise their definitions can be. For example, all phishing attacks are social engineering attacks however, not all social engineering attacks can be classified as phishing [214]. Also, a ransomware and a malware attack can be functionally similar in execution, but the express request for a fee from the victim could be the insignificant determinant that will influence the categorization of this attack type as ransomware. The lack of sophistication by individuals reporting on these cyberattack incidents without knowing the granular details escalates this seemingly infinitesimally small issue into a humongous one, especially when precision is needed in knowing an attack type so that a comprehensive database can be built to reflect the actual reality on the ground.

Table 4.2: Cyberattack incident types and definitions

	OCR Type of Breach	Definition	CISSM Type of Attack	Definition
1.	Hacking/IT Incident	An event or occurrence where unauthorized individuals gain access to computer systems, networks, or data by exploiting vulnerabilities or employing unauthorized methods. Network server and email compromise fall in this category.	Exploitive	Illicitly access or exfiltrate sensitive information such as personal identifiable information, classified information, or financial data.
2.	Unauthorized Access/Disclosure	The act of gaining entry or revealing sensitive information without proper authorization, potentially leading to privacy breaches, data leaks, or unauthorized use of resources.	Disruptive	This type of incident Impedes the target organization's normal operations.
3.	Theft	The act of unlawfully taking someone else's property or valuable information, often resulting in financial loss, data compromise, or disruption of operations.	Mixed	Incidents incorporates both disruptive and exploitative elements, such as a ransomware attack.
4.	Improper Disposal	The incorrect or careless handling of sensitive data or information, which may involve disposing of it in a manner that does not adequately protect against unauthorized access or retrieval.	Undetermined	When the type of attack cannot be determined from the limited information available
5.	Loss	The unintended or accidental misplacement, disappearance, or destruction of assets, including data or physical devices, which can result in the inability to access or recover the affected resources.		

4.6 Outcomes and visual analysis from developed web application and proposed database

For my thesis, we proposed a database of cyber incidents, with corresponding web application designed to enable implementation of the functional requirements that are elucidated in subsection 4.4.2. The outcomes of some key queries are provided from Fig. 4.12: through 4:15.

The cyberattack database implemented using the tools stipulated in section 3. However, the content of the database in terms of the attributes coincides with the OCR and CISSM database. We used this approach because the databases contain very robust structure that is adaptable and easy to integrate into our MySQL driven application. Moreover, the OCR database is based on reported incidents from health institutions while the CISSM has a research team that ensures the legitimacy of these incidents. Our approach thus helps us legitimize the reported events, thereby given us a basis for establishing cyberattack trends in the healthcare sector.

Figure 4:12: shows the Graphical User Interface (GUI) appearance of our developed database. From the interface, a user can add attack incident, download a CSV file of the incident as a CSV or excel file and print. A search and delete function can also be seen on the figure.

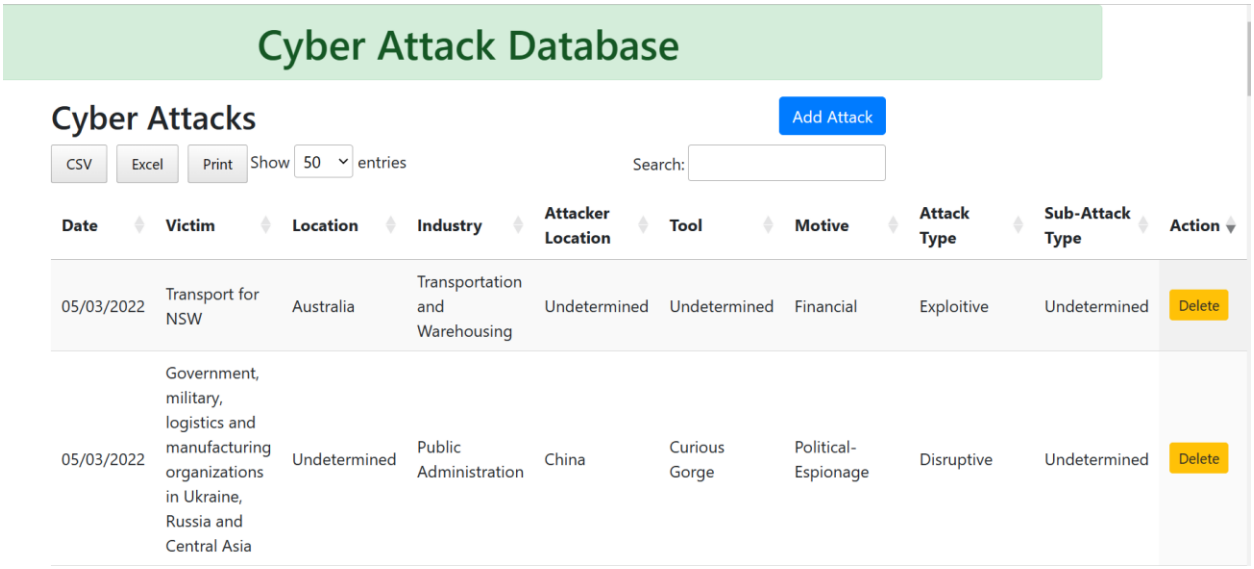


Fig. 4.12: GUI of cyberattack database

Fig. 4.13: shows the parameters a user can use to filter the results from the analysis of the database. A user can select the date of cyberattack incidents alongside the location, the healthcare provider and type of attack to view.

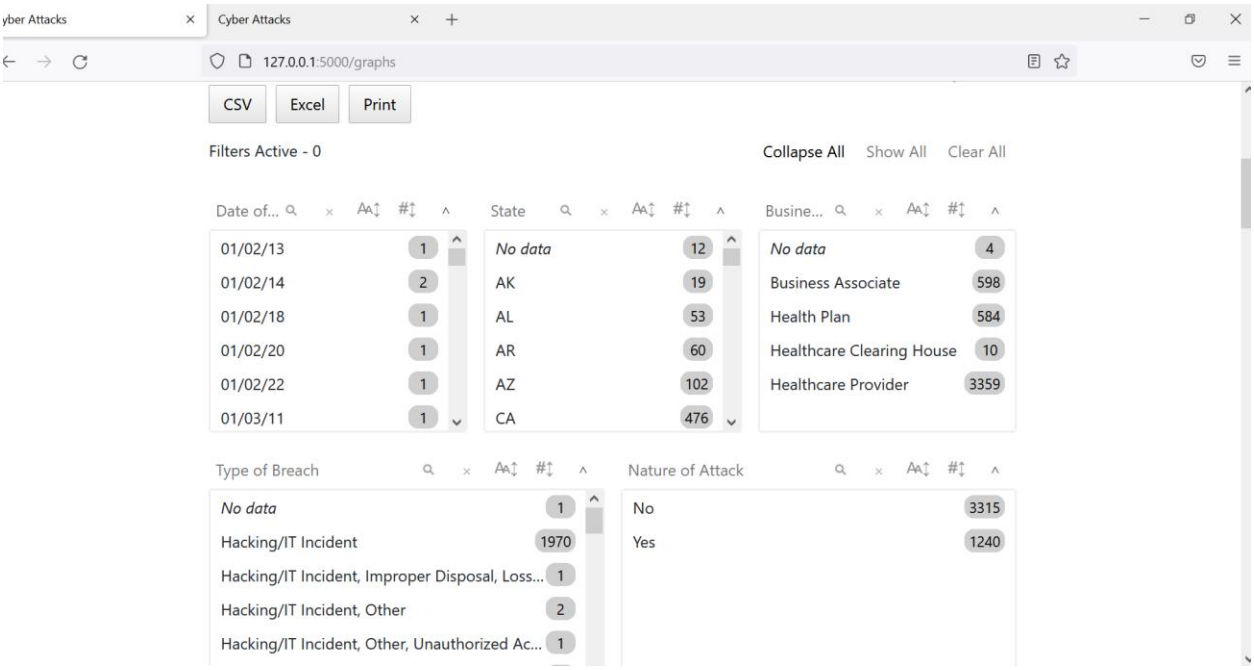


Fig. 4.13: Filtering parameters for database queries

From our developed application, Fig. 4.14, it shows the name of the healthcare facility that has been breached. A user can use different filters and combinations to view different kinds of results. This capability of manipulating the database is important to study patterns of attacks. Also, Fig. 4:15, shows the potential locations where cyber attackers initiated their attacks.

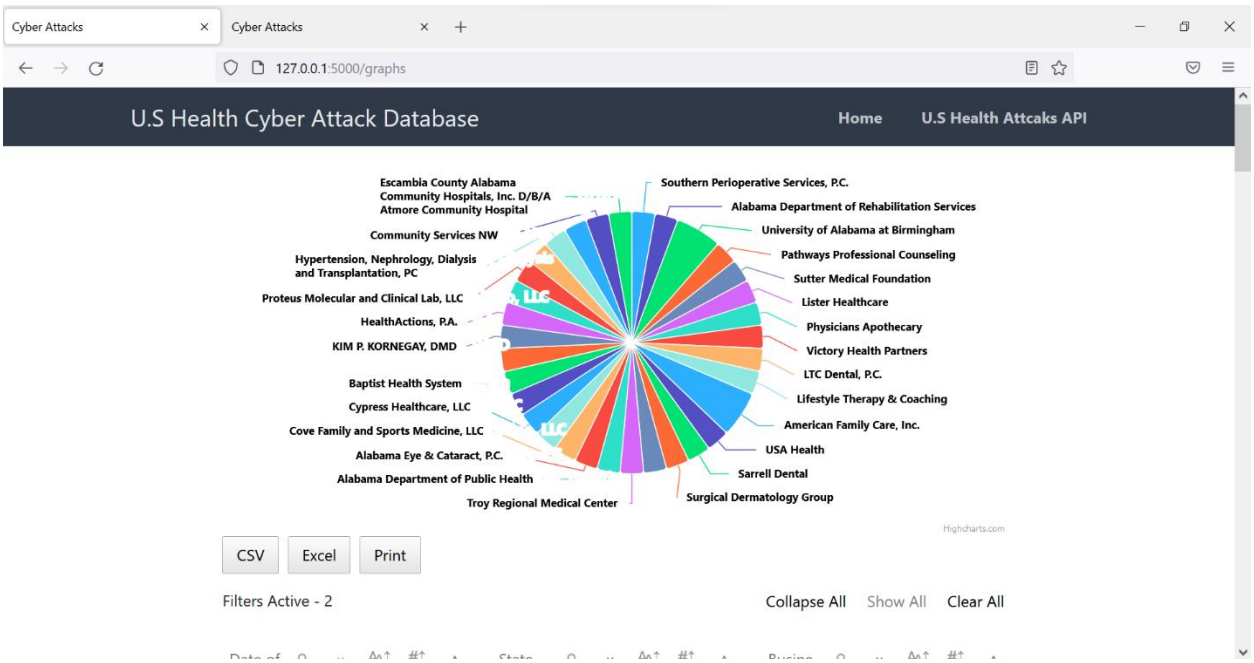


Fig. 4.14: Dynamic filtering of results from database based on user input



Fig. 4.15: Distribution of cyberattack incidents

4.7 Conclusion

This chapter narrows on building the database of cyberattacks for a typical healthcare facility leveraging public sources of cyberattacks such as CISSM, OCR and public records. A web application that runs on a local machine was developed to interact with the database to allow for dynamic manipulation of the database. In designing the application, the Object oriented (OO) approach of software and systems development was adhered to where the key functionalities are modeled using the standard notation of the unified modeling language (UML).

CHAPTER 5 – CASE STUDIES AND APPLICATIONS, RESULTS, DISCUSSIONS AND ANALYSIS

5.1 Introduction

This chapter presented the proposed multilayered cybersecurity approach for the healthcare sector with its practical applications and case studies, supported by detailed discussions of results and analysis. More precisely, four important case studies were concentrated upon to show the practical application in modern healthcare facilities. The first case study is the risk assessment of the cyber infrastructure in Baze University Hospital (BUH) that is based on integrating components of ISO27000 and NIST CSF guidelines. This approach enabled us to identify assets, threats, and vulnerabilities. Vulnerabilities are weak spots that cybercriminals typically capitalize on to perpetrate attacks. We then proceeded with technical mechanisms that can be leveraged to mitigate the exposure of these assets. Since Internet Health Things (IoHT) are some of such assets that can be compromised, we explored AI/ML techniques in anomaly detection using models such as random forest, gradient boosting, decision trees, and MLP to address IoHT asset exposure. Thirdly, we also demonstrated the use of SVMs by comparing the radial basis function and polynomial kernels for phishing website detection, which is another known threat in the healthcare sector. Lastly, we propose a cross-ML approach that is conceived to address the issue of generalizability and concept drift common in training AI models and practical deployment.

5.2 Case study 1: Risk assessment (Baze University Hospital)

For a better appreciation of cybersecurity solutions in the healthcare sector a typical case study is necessary to connect theorems, research hypothesis to real life practicability. It is well documented in the literature that there are a lot of methods and techniques by

which cybersecurity can be administered to protect IT assets. Some are more technical and tailored to very specific cyber threats, i.e., using machine learning to detect spoofing, DoS, NMap port scans, and smurf attacks in Internet of Health Things (IoHT) environment [11], while others are more policy-driven towards people and processes [215]. In most cases, the approaches used are stochastic and do not present a holistic perspective to address cybersecurity in healthcare.

Baze University Hospital (BUH) is a 200-bed hospital associated with Baze University in the Federal Capital Territory of Abuja, Nigeria. BUH was established to provide advanced healthcare services. The facility is equipped with modern facilities for the sole purpose of providing quality health services. Additionally, the hospital engages in research activities leveraging technology that are functionally and aesthetically urbane [216].

Table 5.1: Department, services and description

SN.	Services	Description
1.	Registration Kiosk	Registration portal with Data protection compliant efficient hospital management systems (HIMS).
2.	Emergency Services	Ambulance, Advanced point-of-care diagnostics, mobile X-Ray, mobile ultrasound scan, ABG machine, Monitors, Defibrillators and Mobile Ventilators.
3.	Family Medicine & General Outpatient Clinic	Routine medical checkups, health screening and wellness packages.
4.	Pharmacy Department	Counselling and drug information - high level pharmaceutical administration.
5.	Pediatric Unit	Emergency and Neonatal Intensive care unit.
6.	Ear, Nose & Throat (ENT) Department	Audiological examination, general ENT, head and neck surgery, ENT diagnosis and emergency management.
7.	Ophthalmology	Diagnosis and treatment of eye diseases, cataract surgeries and vision rehabilitation.
8.	Obstetrics and Gynecology	Assisted conception, general surgery,
9.	Laboratory Services	Microscopy, CSF, ESC, HVS, culture and sensitivity of clinical specimens.
10.	Radiology Department	Magnetic Resonance Imaging, digital radiography, computed tomography (64 slice CT) with coronary and Angiographic study options.
11.	Histopathology Department	Incisional, excisional, Exfoliative and aspiration cytology.
12.	Blood Transfusion and Haematology Department	Diagnosis and treatment of haemato-oncology disease. Leukemia and bleeding disorder.

13.	Chemical Pathology Laboratory	Cardiac marker testing for cancer screening, endocrine diagnosis, metabolic disorders, infertility evaluation and cardiovascular risk assessment.
14.	Physiotherapy and Rehabilitation Department	Hydrotherapy, laser therapy, reatherm diathermy and shockwave therapy
15.	Utility Unit	Water supply, power system supply, firefighting equipment, gas plant and Air conditioning system.
16.	Catering Services	Fully functional Catering and laundry services.

5.2.1 BUH cyber setup and operations

At BUH, all internet traffic entering and leaving the organization is filtered through a firewall. However, it is worth mentioning that Baze University Hospital emerged from Baze University, founded in 2011; Therefore, there is the main site and the permanent site of Baze University Hospital as shown in Fig. 5.1. Currently, the main site houses electronic medical records (EMR), an employee/student portal, and a website server. From the main site, traffic flows to the permanent site, where all medical equipment, switches, management systems, etc. are located.

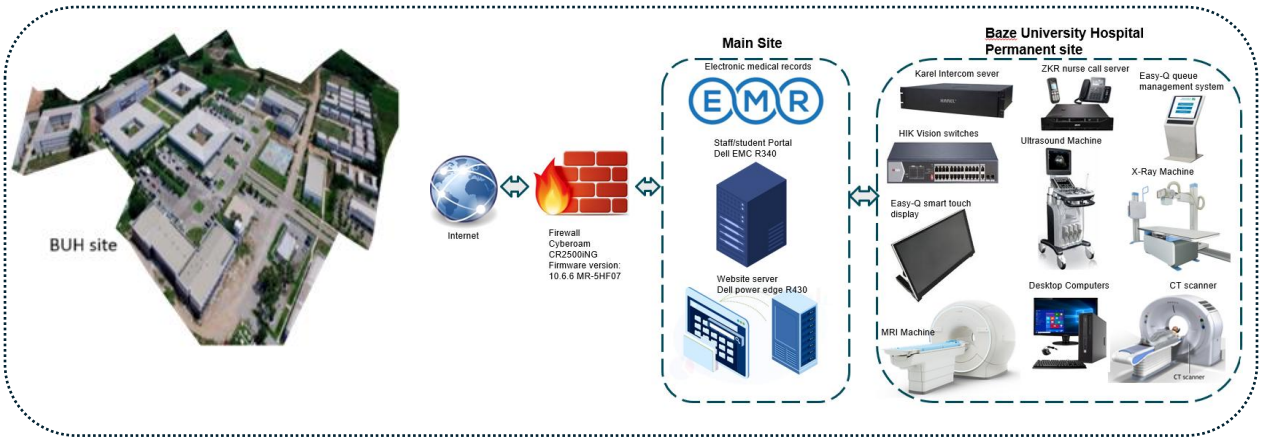


Fig. 5.1: BUH Cyber Operations and processes

5.2.2 Proposed risk assessment framework

Risk is generally perceived as a function of threat, vulnerability, and impact [156] or the probability of an external cyber threat exploiting a vulnerability in an asset and the corresponding impact of that adverse incident on an organization [217]. The proposed Baze University Hospital Cyber Risk Assessment Framework (BUHCRAF) is a streamlined and fully adapted approach based on the iterations of the ISO27000 series and NIST CSF established in chapter 2. Using the two frameworks provides flexibility in leveraging just the strong and applicable components from both in the proportion that applies to the defined context. Fig. 5.2. shows the five stages proposed in the BUHCRAF framework.

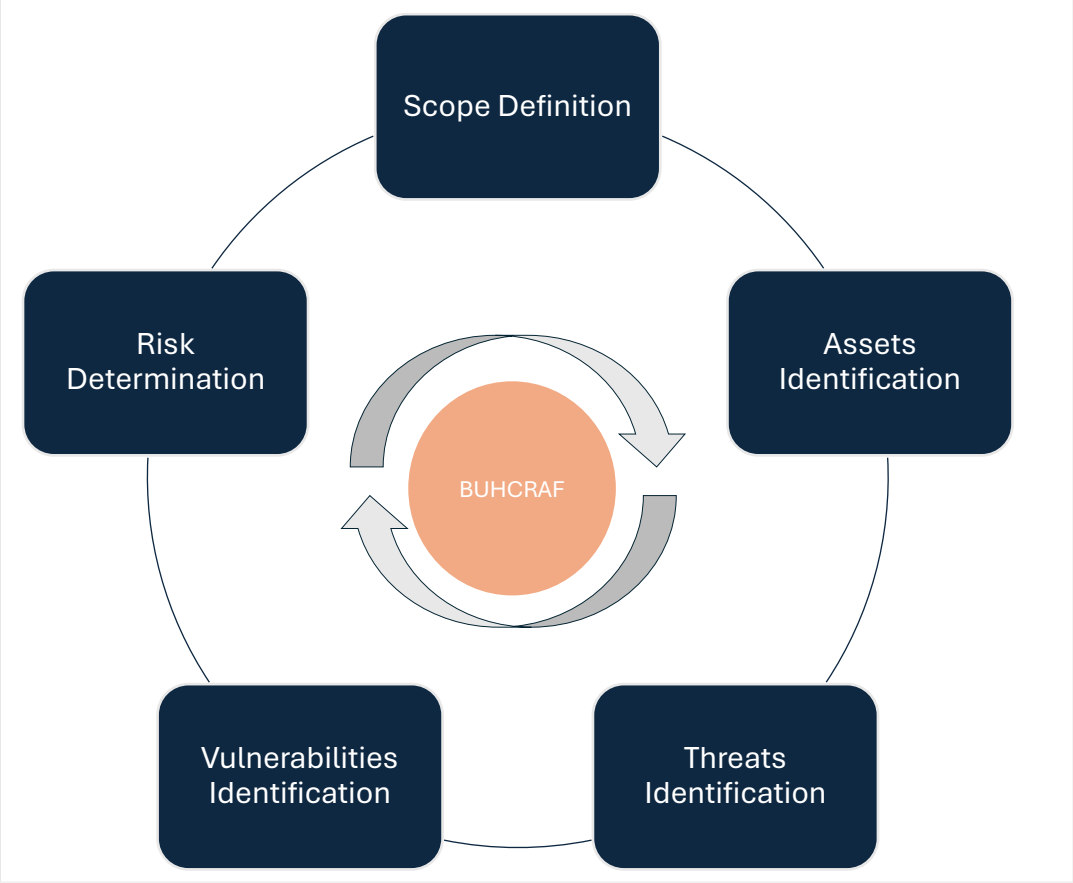


Fig. 5.2: Baze University Hospital Cyber Risk Assessment Framework (BUHCRAF)

5.2.2.1 Scope definition

In the scope definition stage, we articulated the boundaries of what the risk assessment would cover. Even though the mission and objectives of BUH are futuristic and very wide in scope, in this work, we confined the risk assessment to comprehensively cover the cyber infrastructure of the hospital. All other assets, such as buildings, automobiles, wheelchairs, and other physical assets were excluded.

5.2.2.2 Assets identification

Assets, within the context and scope definition of this work, are defined as any hardware, software, networks, or substance of material value to Baze University Hospital. Should there be any compromise in the integrity of an asset, then it will have a significant impact on the performance of the asset and negatively affect hospital operations.

5.2.2.2.1 Assets description

Hospitals generally have diverse assets. The European Union Agency for Cybersecurity (ENISA) has a loosely coupled methodology that helps categorize and group cyber resources in hospitals. We have adopted the categorization technique to identify and group relevant assets that fall within the scope of our study. Table 5.2 shows the different assets, their categories and description.

Table 5.2: Assets, Categories and Description

SN.	Asset	Type	Category	Description
1.	Switches	HIK Vision	Networking equipment	These switches are used for network distribution. There are two variations of the switches. (1) Power Over Internet (POE) (2) non-power over ethernet
2.	Ultra-sound machines	GE Logiq	Networked medical devices	Medical devices used for clinical imaging. There are two variations of the ultra-sound machines. (1) GE Logic X (2) GE Logic Q
3.	MRI machines	GE signa creator	Networked medical devices	Magnetic resonance imaging (MRI) uses non-invasive imaging technology that produces detailed anatomical images in three dimensions
4.	CT-Scan	GE Revolution maxima	Networked medical devices	Computed or computerized tomography (CT) scanner is a medical imaging technique used for obtaining detailed images of internal body parts.

5.	Advantage workstation	GE AW4.6	Networked medical devices	The advantage workstation is a networked medical imaging console dedicated to examination, review and diagnosis on film.
6.	Remote energy monitoring	Cerbo Gx	Buildings and facilities	The remote energy monitoring system helps in controlling, optimizing and managing energy consumption from a central location in a building asset.
7.	Intercom server	Karel	Networking equipment	Karel a modular and server-based IP communication platform that is designed for medium and large-scale organizations.
8.	Nurse call system	ZKR	Networking equipment	Communication system infused with medical nurse calling and queue management system.
9.	Firewall	Cyberoam CR2500iNG	Networking equipment	A hardware device for plugging into the hospital network at the perimeter serving as a gateway for traffic.
10.	Website server	Dell power edge R430	Networking equipment	Used for high performance computing, web technology and infrastructure scale out.
11.	Staff and student portal	Dell EMC R340	Data	Single socket 1U rack server designed for productivity and application that are data intensive and automation of daily tasks.
12.	Electronic medical record	EMR	Interconnected clinical information systems	The EMR is a digital patient health record for clinical care billing and reporting.
13.	Queue management system	Eazy-Q	Interconnected clinical information systems	Queue management solution that offers unique way of handling customer inflow.
14.	Computers	Hp - Windows 11	Mobile client devices	Standard computers used by staff for technical and admirative tasks all running windows 11

Fig. 5.3 shows the number of cyber assets in BUH, while Fig. 5.6 shows the distribution of cyber assets based on the ENISA framework. The distribution consists of 29% networking equipment, 7% buildings and facilities, 7% data, 21% interconnected clinical information systems, 7% mobile client devices, and 29% networked medical devices.

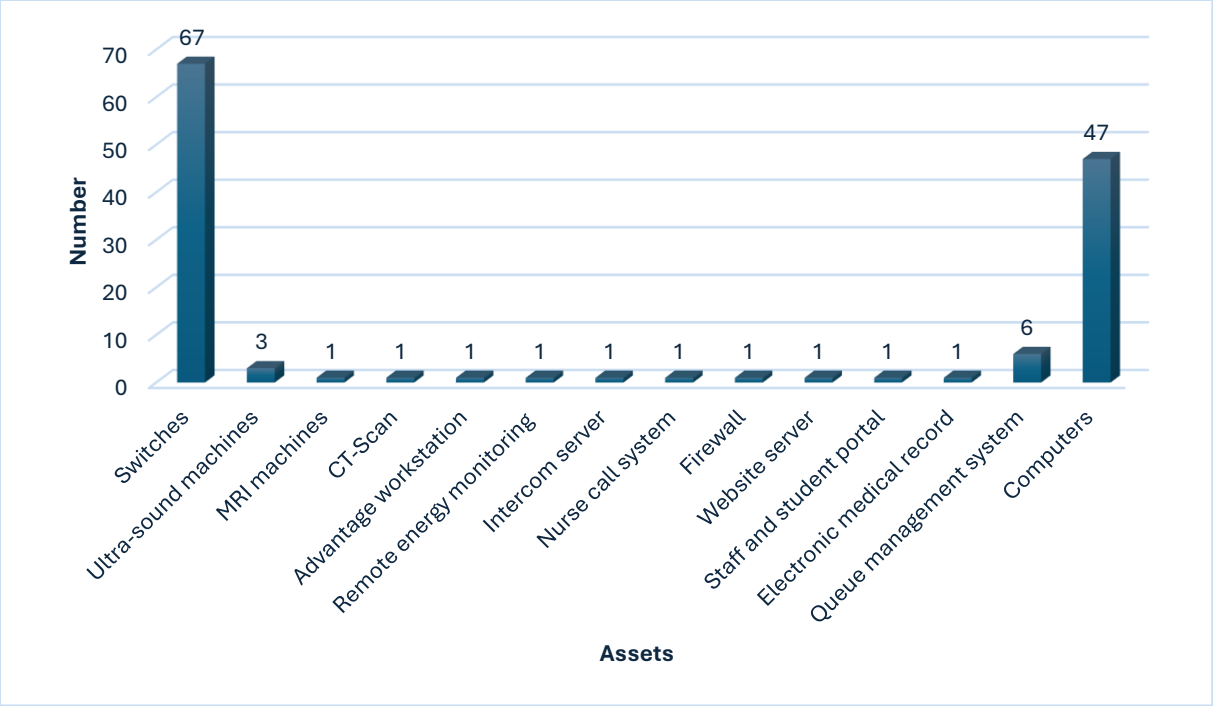


Fig. 5.3: Number of Assets in BUH

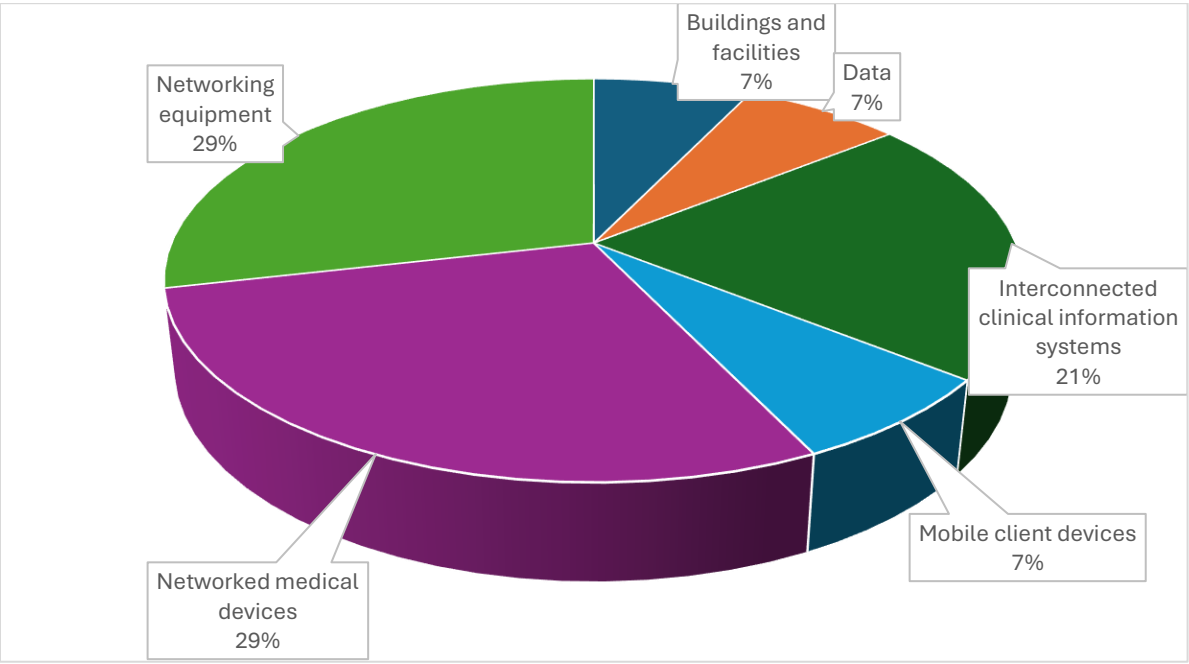


Fig. 5.4: Categorization of Assets based on ENISA

5.2.2.2.2 Assets value

When determining the hierarchy of importance of each asset to the company, it is necessary to assign a relative value that indicates the order of priority. Factors that must be considered in this assessment include assets that are critical to the success of the organization, assets that generate the highest revenue/profitability, assets with the highest replacement cost, and assets that if compromised cause the greatest embarrassment [218]. Therefore, in this work, we propose the matrix in Fig. 5.5 was used to assign relative values to each asset for the cyber assets in Baze University Hospital. In our matrix 1 means insignificant while 5 is acute.

Value	Type of Effect Level of Effect	Company Embarrassment Level	Personal Safety Implication	Personal Privacy Infringement	Failure to Meet Legal Obligations	Financial Loss (£)	Disruption to Activities (£) (Time & Effort to Recover from Incident)
1	Insignificant	Contained within Work Area at worst	Minor injury to individual	Isolated personal detail revealed	Civil suit resulting in less than £10k damages	Up to 10k	Up to 10k
2	Minor	Contained within Company at worst	Minor injury to several people	Isolated personal detail compromised	Civil suit (above £10k). Small fine (up to £1k)	10k to 100k	10k to 100k
3	Significant	Local public or Press become aware	Major injury to individual	Several personal details revealed	Large fine (above £10k)	100k to 500k	100k - 500k
4	Major	National public or Press become aware	Major injury to several people or death of individual	Several personal details compromised	Custodial sentence imposed	500k - 1000k	500k - 1000k
5	Acute	Senior Staff forced to resign or Company fails	Death of several people	All personal details revealed and/or compromised	Multiple civil or criminal suits	Above 1000k	Above 1000k

Fig. 5.5: Asset Values and Impacts of Incident

5.2.2.3 Threats identification

Threats are potential occurrences or externally acting agents that can damage or compromise assets. Threats can come from authorized and unauthorized users of assets; however, they cannot come within the asset itself. Identifying all threats to a defined asset can be a tortuous and sometimes impractical endeavor as threats are dynamic in nature.

However, a good articulation of known threats is usually recommended to situate risks in their proper perspective.

5.2.2.3.1 MITRE ATT&CK®

MITRE ATT&CK® framework is a curated globally accessible knowledge base drawn from adversarial tendencies and actions observed during cyberattacks [109]. ATT&CK® which stands for Adversarial Tactics, Techniques, and Common Knowledge primarily aims to model the behaviors with tactics, techniques, and procedures (TTP) to examine attack patterns [219]. The framework is primarily activated by most individuals/organizations to develop specific threat models however, what is often overlooked with the ATT&CK® framework is the huge repository of articles that describe threats in detail and explain how attackers were able to carry out successful cyber-attacks and compromised certain cyber assets [220]. Therefore, to efficiently identify potential threats to our assets, this research leveraged the ATT&CK® framework, articles and web sources to identify known threats to the assets in the BUHCRAF model.

5.2.2.3.2 Threat rating

Different threats will have different severities or impacts therefore, a consistent way of assessing these threats is necessary. Fig. 5.6 provides a rating system that captures several categories of threats which includes but is not limited to espionage, software attacks, theft, sabotage, technical hardware failures and information extortion [163].

Risk Likelihood	Description	Interpretation
1	**Negligible**	Once every 1000 years or less
2	**Extremely unlikely**	Once every 200 years
3	**Very unlikely**	Once every 50 years
4	**Unlikely**	Once every 20 years
5	**Feasible**	Once every 5 years
6	**Probable**	Annually
7	**Very probable**	Quarterly
8	**Expected**	Monthly
9	**Confidently expected**	Weekly
10	**Certain**	Daily

Fig. 5.6: Likelihood of Threat

5.2.2.4 Vulnerability identification

Vulnerabilities are inherent flaws or weaknesses in the assets where threats can exploit to compromise the confidentiality, availability, integrity of an asset [218]. The process of vulnerability identification usually overlaps with the threat identification stage because all known vulnerabilities associated with an asset are mapped precisely to the known potential threats that can likely exploit them. Third-party catalogues are critical to finding a great deal of documented vulnerabilities. Alongside the MITRE ATT&CK® framework, this work greatly utilized the Common Vulnerability and Exposure (CVE) database [221], and other sources to identify known vulnerabilities to the assets in the scope of the BUHCRAF model.

5.2.2.4.1 Common Vulnerabilities and Exposures (CVE)

Common Vulnerabilities and Exposure (CVE) database is a standardized system for identifying, tracking and cataloging known vulnerabilities [222]. The CVE keeps a unique identification number, a brief description and references for publicly known cyber security vulnerabilities [223]. Using the same rating system in Fig. 5.6, we assigned corresponding vulnerability rating to the threats and assets. The goal of using the CVE is simply to contextualize our identified vulnerability rating for BUH with ratings that are already reported and their corresponding severity. This helped tone down the imminent subjectivity that is unavoidable in creating risk registers.

5.2.2.5 Risk Determination

In determining the overall risks, we used a systematic approach, utilized in a form of a Threats-vulnerability-Assets (TVA) worksheet that logically presents the list of assets, their threats and corresponding vulnerabilities [163], [224]. Table 5.3 shows how we determined the risk based on the asset value and threat rating. To minimize the level of subjectivity in the rating system, we used references from the CVE database as applicable.

Table 5.3: Risk determination

SN.	Asset	Value	Threat	Rating	Vulnerability	Rating	Reference
1.	Switches	2	MAC Address Flooding	3	Use of HTTP protocol	3	[225], [226], [227], [228], [229]
	Switches	2	MAC Address Flooding	3	Use of Telnet protocol	2	
	Switches	2	MAC Address Flooding	3	Inadequate segregation of network traffic	4	
	Switches	2	MAC Address Flooding	3	Outdated firmware	4	
	Switches	2	MAC Address Flooding	3	Unimplemented port security control	4	
	Switches	2	VLAN Hopping	3	Not utilizing virtual local area networks (VLANS)	3	

	Switches	2	VLAN Hopping	3	Network segmentation deficiencies	2	
	Switches	2	Insider threats	4	Incorrect access control list (ACLs)	3	
	Switches	2	Insider threats	4	Unchanged default passwords	3	
	Switches	2	Insider threats	4	Lack of log review system	3	
	Switches	2	DoS Attacks	8	Not disabling unused ports	6	
	Switches	2	DoS Attacks	8	Not utilizing virtual local area networks (VLANS)	6	
	Switches	2	DoS Attacks	8	Lack of Network monitoring tools	6	
	Switches	2	DoS Attacks	8	Lack of regular updates	6	
2.	Ultra-sound machines	3	Insider threats	6	Protection mechanism failure	5	[230], [231], [232], [233], [234]
	Ultra-sound machines	3	Insider threats	6	Incorrect User Management	6	
	Ultra-sound machines	3	Insider threats	6	Improper input validation	5	
	Ultra-sound machines	3	Elevation of privileges	5	Misconfigured access control list	6	
	Ultra-sound machines	3	Command injection	5	Bug in operating system	5	[235]
	Ultra-sound machines	3	Path Traversal	4	Bug in operating system	5	
3.	MRI machines	4	Hacking	6	GE (remote) service user privileges	4	[236], [237], [238], [239]
	MRI machines	4	Hacking	6	Unprotected Transport of Credentials	4	[240], [241],
	MRI machines	4	Hacking	6	Exposure of Sensitive System Information to an Unauthorized Control Sphere	4	
	MRI machines	4	MDHRex	7	Vulnerability in FTP (port 21), SSH (port 22), Telnet (port 23) REXEC (port 512)	5	
	MRI machines	4	Unauthorized access	6	Insufficiently protected credentials	5	[242]
4.	CT-Scan	4	MDHRex	7	Vulnerability in FTP (port 21), SSH (port 22), Telnet (port 23) REXEC (port 512)	5	
	CT-Scan	4	Hacking	7	Unprotected Transport of Credentials	5	
	CT-Scan	4	Hacking	7	Flaws in Unix operating system	4	[243]
	CT-Scan	4	Arbitrary code execution	5	Unix OS vulnerability	3	
5.	Advantage workstation	3	Hacking	3	Stack-based Buffer Overflow	5	[244]
	Advantage workstation	3	Insider threat	3	Improper Authorization in Handler for Custom URL Scheme	5	

6.	Remote energy monitoring	3	DoS attack	5	Missing Authentication for Critical Function	4	[245]
	Remote energy monitoring	3	DoS attack	5	Unsafe Reflection	4	
	Remote energy monitoring	3	Unauthorized access	3	Remote function calling	3	
7.	Intercom server	1	Packet sniffing	3	Unencrypted packets in transit	3	[246], [247]
	Intercom server	1	SIP (Session Initiation Protocol) spoofing	2	Lack of call session monitoring state	3	[248], [249]
	Intercom server	1	Jamming attacks	2	Poorly implemented anti jamming algorithm	2	
	Intercom server	1	Black Storm Attacks	2	Susceptibility to IP manipulation	3	
	Intercom server	1	DDoS Attacks	3	Poor network segmentation and multi-tenant service provision	2	
	Intercom server	1	Vishing attacks	2	Untrained operator staff susceptible to social engineering manipulation	2	
8.	Nurse call system	2	sabotage	4	Insecure access protocol	5	
9.	Firewall	4	SQL injection threats	5	pre-authentication SQL injection vulnerability	4	[250],[251], [252]
	Firewall	4	Cross site scripting (XSS)	5	Improper sanitation of login credentials	4	
	Firewall	4	Cross site scripting (XSS)	5	Vulnerability in “LiveConnectionDetail.jsp” application	3	[253]
	Firewall	4	Shell injection	3	Web admin portal vulnerability	6	
	Firewall	4	Shell injection	3	Susceptibility to SSL VPN console manipulation	2	
	Firewall	4	Privilege escalation	6	Component access restriction	3	
	Firewall	4	Privilege escalation	6	Unknown function of “Licenseinformation.jsp” vulnerability	2	
	Firewall	4	Privilege escalation	6	Improper access control	3	
	Firewall	4	Privilege escalation	6	Improper input validation	3	
	Firewall	4	Privilege escalation	6	Incorrect Permission Assignment for Critical Resource	4	
	Firewall	4	Privilege escalation	6	Improper privilege management	3	
	Firewall	4	Privilege escalation	6	Exposure of Resource to Wrong Sphere	3	
	Firewall	4	Privilege escalation	6	Exposure of Sensitive Information to an Unauthorized Actor	5	
	Firewall	4	Command injection	3	Improper Neutralization of Special Elements used in a Command	5	
	Firewall	4	DoS Attacks	8	Use of custom signature based IPS	8	

10.	Website server	5	Sabotage	6	SMM communication buffer verification vulnerability	7	[254], [255]
	Website server	5	Data breach	8	Insufficient security and access control	6	
	Website server	5	Phishing attacks	9	Outdated Vulnerable web application components	9	
	Website server	5	DDoS Attacks	8	Loophole in Apache server	8	
	Website server	5	DDoS Attacks	8	Server resource insufficiency	7	
	Website server	5	DDoS Attacks	8	UDP overflow	6	
	Website server	5	Missing function level access control	5	Insufficient server authentication and authorization checks	5	
	Website server	5	Missing function level access control	4	Misconfigured function level protection	6	
11.	Staff and student portal	3	Sabotage	6	SMM communication buffer verification vulnerability	6	[254],[256]
	Staff and student portal	3	Component tampering	5	Physical server vulnerability	4	
	Staff and student portal	3	Malware Injection	7	Firmware vulnerability	4	
	Staff and student portal	3	Malware Injection	3	Software vulnerability	5	
	Staff and student portal	3	Unauthorized open-port attacks	5	Server management vulnerability	4	
12.	Electronic medical record	5	SQL injection	7	Unvalidated input fields	5	
	Electronic medical record	5	Data breach	9	Weak authentication	6	
	Electronic medical record	5	Cross-site request forgery	5	Exploitation of session cookie vulnerability	5	
	Electronic medical record	5	Cross-site scripting	6	Insufficient sanitation of user supplied data	6	
	Electronic medical record	5	Cross site scripting	6	Loophole in client-side script	5	
	Electronic medical record	5	Phishing attacks	9	Email compromise	9	
	Electronic medical record	5	Phishing attacks	9	Web features susceptible to phishing attacks	9	

	Electronic medical record	5	Phishing attacks	9	Redirect vulnerability	8	
	Electronic medical record	5	Ransomware attacks	8	Inadequate monitoring and regulation system	7	
	Electronic medical record	5	Malware	6	Outdated anti-virus scans	9	
	Electronic medical record	5	Malware	6	Inadequate monitoring and regulation system	8	
13.	Queue management system	2	ARP poisoning	6	Misconfigured IP to MAC address table	6	[257]
14.	Computers	3	Virus	10	Indiscriminate flash drive use on computers	9	[258], [259]
	Computers	3	Spoofing	6	Windows certificate vulnerability	4	[260],[261]
	Computers	3	Spoofing	6	read/write/execute capabilities vulnerability	3	
	Computers	3	Spoofing	6	Secure server handler compromise	4	
	Computers	3	Escalation of privilege	5	Improper input validation	7	
	Computers	3	DoS attack	7	Information disclosure via adjacent access	3	
	Computers	3	Buffer overflow	4	System bugs leading to infinite loop	2	[262]
	Computers	3	Buffer overflow	4	Arbitrary code execution	5	
	Computers	3	Buffer overflow	4	Overwriting function pointers	4	
	Computers	3	Control list DoS	4	Misconfiguration of access control list	4	
	Computers	3	Remote code execution	3	Poor defined access control list	6	
	Computers	3	Insider threats	4	Windows Security Center API Remote Code Execution Vulnerability	5	
	Computers	3	preimage attack	3	MD5 thumbprint certificate vulnerability	5	

5.2.2.6 Risk ranking and control

The Risk register is a repository for all anticipated risks ranked according to their priorities [263], [264]. The risk register may also include other associated information that guides the risk management process [265], [266]. Within the scope of this research, the relevant control as identified by the NIST framework and other relevant scholarly sources

capable of mitigating the impact of the risk are also included. As shown in Table 5.4, all identified risks that could potentially put assets at risk are assigned a value and ranked in descending order. This means that the higher the number, the more senior managers should prioritize these risks. The formula used to determine risk score is Risk = Asset x Threat Score x Vulnerability Score.

Table 5.4: Risk register and control

Asset	Value	Threat	Rating	Vulnerability	Rating	Risks value	Control
Website server	5	Phishing attacks	9	Outdated Vulnerable web application components	9	405	Regular patch updates
Electronic medical record	5	Phishing attacks	9	Email compromise	9	405	Build strong social engineering training and culture
Electronic medical record	5	Phishing attacks	9	Web features susceptible to phishing attacks	9	405	Introduce technical sophistication and AI detection techniques
Electronic medical record	5	Phishing attacks	9	Redirect vulnerability	8	360	Strong usage policies and redirect checks
Website server	5	DDoS Attacks	8	Loophole in Apache server	8	320	Regular patch updates
Website server	5	DDoS Attacks	8	Server resource insufficiency	7	280	Implement technical DDoS attack techniques
Electronic medical record	5	Ransomware attacks	8	Inadequate monitoring and regulation system	7	280	Strong usage policies and training
Electronic medical record	5	Data breach	9	Weak authentication	6	270	Secure user authentication profiles
Electronic medical record	5	Malware	6	Outdated anti-virus scans	9	270	Use most recent anti-virus software
Computers	3	Virus	10	Indiscriminate flash drive use on computers	9	270	Regulate flash drive usages for systems on networks
Firewall	4	DoS Attacks	8	Use of custom signature based IPS	8	256	Implement robust AI prevention techniques
Website server	5	Data breach	8	Insufficient security and access control	6	240	Deploy efficient traffic monitoring routine
Website server	5	DDoS Attacks	8	UDP overflow	6	240	Socket monitoring
Electronic medical record	5	Malware	6	Inadequate monitoring and regulation system	8	240	Strong usage policies and training

Website server	5	Sabotage	6	SMM communication buffer verification vulnerability	7	210	Secured component verification (SCV), Intrusion detection
MRI machines	4	MDHRex	7	Vulnerability in FTP (port 21), SSH (port 22), Telnet (port 23), REXEC (port 512)	7	196	Restrict and monitor ports, use port control mechanisms
CT-Scan	4	MDHRex	7	Vulnerability in FTP (port 21), SSH (port 22), Telnet (port 23), REXEC (port 512)	7	196	Restrict and monitor ports, use port control mechanisms
Electronic medical record	5	Cross-site scripting	6	Insufficient sanitation of user supplied data	6	180	Systems to check user supplied data
Electronic medical record	5	SQL injection	7	Unvalidated input fields	5	175	Validated form checks
Electronic medical record	5	Cross site scripting	6	Loophole in client-side script	5	150	Regular patch update
Ultra-sound machines	3	Insider threats	6	Protection mechanism failure	8	144	Strengthen physical access protocol
CT-Scan	4	Hacking	7	Unprotected Transport of Credentials	5	140	credential update, patch update
Website server	5	Missing function level access control	5	Insufficient server authentication and authorization checks	5	125	Strong culture of verification and validation checks
Electronic medical record	5	Cross-site request forgery	5	Exploitation of session cookie vulnerability	5	125	Session variable management control
MRI machines	4	Unauthorized access	6	Insufficiently protected credentials	5	120	Protect credentials LDAP, keystore
Firewall	4	Privilege escalation	6	Exposure of Sensitive Information to an Unauthorized Actor	5	120	Strong organizational policy on assets usage
Website server	5	Missing function level access control	4	Misconfigured function level protection	6	120	Due care given to configuring level access control
Switches	2	DoS Attacks	8	Not utilizing virtual local area networks (VLANS)	7	112	Proportionately deploy VLANS
CT-Scan	4	Hacking	7	Flaws in Unix operating system	4	112	Ensure Unix is patched with current OS
Ultra-sound machines	3	Insider threats	6	Incorrect User Management	6	108	Restrict access to kiosk mode functionality
Staff and student portal	3	Sabotage	6	SMM communication buffer verification vulnerability	6	108	Secured component verification (SCV), Intrusion detection
Ultra-sound machines	3	Elevation of privileges	5	Misconfigured access control list	7	105	Strengthen physical access protocol
Computers	3	Escalation of privilege	5	Improper input validation	7	105	continuous monitoring of access privileges

Firewall	4	Cross site scripting (XSS)	5	Improper sanitation of login credentials	5	100	Strong credentials assignment and monitoring
Switches	2	DoS Attacks	8	Not disabling unused ports	6	96	Closed unused ports
Switches	2	DoS Attacks	8	Lack of Network monitoring tools	6	96	Deploy Network monitoring tools i.e. Nagios XI
Switches	2	DoS Attacks	8	Lack of regular updates	6	96	Regular patch update
MRI machines	4	Hacking	6	GE (remote) service user privileges	4	96	change default login credentials
MRI machines	4	Hacking	6	Unprotected Transport of Credentials	4	96	Implement cryptographic mechanisms
MRI machines	4	Hacking	6	Exposure of Sensitive System Information to an Unauthorized Control Sphere	4	96	Implement cryptographic mechanisms
Firewall	4	Privilege escalation	6	Incorrect Permission Assignment for Critical Resource	4	96	Vetted ACLS and proper coordination of access privileges
Ultra-sound machines	3	Insider threats	6	Improper input validation	5	90	enable the "system lock" password in the Administration GUI menu if possible
Staff and student portal	3	Malware Injection	7	Firmware vulnerability	4	84	Silicon-based Root of Trust
Firewall	4	SQL injection threats	5	pre-authentication SQL injection vulnerability	4	80	Update “hotfix” patch
Ultra-sound machines	3	Command injection	5	Bug in operating system	5	75	Restrict physical access
Firewall	4	Shell injection	3	Web admin portal vulnerability	6	72	Patch update
Firewall	4	Privilege escalation	6	Component access restriction	3	72	Patch update
Firewall	4	Privilege escalation	6	Improper access control	3	72	Vetted ACLS and proper coordination of access privileges
Firewall	4	Privilege escalation	6	Improper input validation	3	72	Vetted ACLS and proper coordination of access privileges
Firewall	4	Privilege escalation	6	Improper privilege management	3	72	Vetted ACLS and proper coordination of access privileges
Firewall	4	Privilege escalation	6	Exposure of Resource to Wrong Sphere	3	72	Patch update
Queue management system	2	ARP poisoning	6	Misconfigured IP to MAC address table	6	72	Deploy secure traffic management systems

Computers	3	Spoofing	6	Windows certificate vulnerability	4	72	continuous monitoring of access privileges
Computers	3	Spoofing	6	Secure server handler compromise	4	72	Access privilege segmentation
Computers	3	DoS attack	7	Information disclosure via adjacent access	3	63	System resource management systems
Ultra-sound machines	3	Path Traversal	4	Bug in operating system	5	60	Restrict physical access
CT-Scan	4	Arbitrary code execution	5	Unix OS vulnerability	3	60	Ensure end-to-end encryption of PHI
Remote energy monitoring	3	DoS attack	5	Missing Authentication for Critical Function	4	60	use a firewall
Remote energy monitoring	3	DoS attack	5	Unsafe Reflection	4	60	Block access from untrusted networks and hosts
Firewall	4	Cross site scripting (XSS)	5	Vulnerability in “LiveConnectionDetail.jsp” application	3	60	Patch update
Firewall	4	Command injection	3	Improper Neutralization of Special Elements used in a Command	5	60	Check externally influencing input from upstream components
Staff and student portal	3	Component tampering	5	Physical server vulnerability	4	60	Secured component verification (SCV), Intrusion detection
Staff and student portal	3	Unauthorized open-port attacks	5	Server management vulnerability	4	60	Implement remote attestation control
Computers	3	Buffer overflow	4	Arbitrary code execution	5	60	Regular patch updates
Computers	3	Insider threats	4	Windows Security Center API Remote Code Execution Vulnerability	5	60	Patch update of Windows Security Center API
Computers	3	Spoofing	6	read/write/execute capabilities vulnerability	3	54	Personnel vetting for computer usage
Computers	3	Remote code execution	3	Poor defined access control list	6	54	Regulated usage of VPNs
Firewall	4	Privilege escalation	6	Unknown function of “Licenseinformation.jsp” vulnerability	2	48	Patch the “Licenseinformation.jsp” vulnerability
Computers	3	Buffer overflow	4	Overwriting function pointers	4	48	Regular patch updates
Computers	3	Control list DoS	4	Misconfiguration of access control list	4	48	IP reputation protection
Advantage workstation	3	Hacking	3	Stack-based Buffer Overflow	5	45	upgrade to DICOM Viewer version 2024
Advantage workstation	3	Insider threat	3	Improper Authorization in Handler for Custom URL Scheme	5	45	Minimize network exposure

Staff and student portal	3	Malware Injection	3	Software vulnerability	5	45	Patch as required
Computers	3	preimage attack	3	MD5 thumbprint certificate vulnerability	5	45	Regular patch updates
Nurse call system	2	sabotage	4	Insecure access protocol	5	40	Access granted to authorize personnel
Remote energy monitoring	3	Unauthorized access	3	Remote function calling	3	27	Use virtual private network (VPN), restrict physical access
Switches	2	MAC Address Flooding	3	Inadequate segregation of network traffic	4	24	Coordinated segregation of network
Switches	2	MAC Address Flooding	3	Outdated firmware	4	24	Regular patch update
Switches	2	MAC Address Flooding	3	Unimplemented port security control	4	24	Closed unused ports
Switches	2	Insider threats	4	Incorrect access control list (ACLs)	3	24	Vetted ACLS and proper coordination of access privileges
Switches	2	Insider threats	4	Unchanged default passwords	3	24	Always change default passwords
Switches	2	Insider threats	4	Lack of log review system	3	24	proper coordination of access privileges
Firewall	4	Shell injection	3	Susceptibility to SSL VPN console manipulation	2	24	Patch update
Computers	3	Buffer overflow	4	System bugs leading to infinite loop	2	24	Code vetting
Switches	2	MAC Address Flooding	3	Use of HTTP protocol	3	18	Use HTTPS protocol
Switches	2	VLAN Hopping	3	Not utilizing virtual local area networks (VLANS)	3	18	Proportionately deploy VLANS
Switches	2	MAC Address Flooding	3	Use of Telnet protocol	2	12	Use secure shell protocol (SSH)
Switches	2	VLAN Hopping	3	Network segmentation deficiencies	2	12	Coordinated segregation of network
Intercom server	1	Packet sniffing	3	Unencrypted packets in transit	3	9	End-to-end network encryption
Intercom server	1	SIP (Session Initiation Protocol) spoofing	2	Lack of call session monitoring state	3	6	Implement safeguards such as Session Border controllers
Intercom server	1	Black Storm Attacks	2	Susceptibility to IP manipulation	3	6	Strick session monitoring standards
Intercom server	1	DDoS Attacks	3	Poor network segmentation and multi-tenant service provision	2	6	Implement services such as content delivery networks
Intercom server	1	Jamming attacks	2	Poorly implemented anti jamming algorithm	2	4	Implement signal strength monitoring

							systems, practice good password and encryption hygiene.
Intercom server	1	Vishing attacks	2	Untrained operator staff susceptible to social engineering manipulation	2	4	Good policy practice and staff training

To conclude, the case study of Baze University Hospital demonstrates the critical importance of implementing a comprehensive and structured risk assessment framework that is context specific. From the risk register we developed, it shows that assets such Website servers are high priority due to susceptibility to phishing attacks and introducing technical sophistication and AI/ML can help in mitigating such risks. Electronic Medical Records (EMR) are also key assets of high value as presented by our work. Although we did not implement all mitigation actions for the risk register, we focused on those that require technical actions.

5.3 Case Study 2: AI/ML in anomaly detection

The second relevant case study used in this thesis is the application of AI/ML in healthcare anomaly detection. This application is deployable because of the data streams exchange in a typical modern hospital network as reflected in subsection 5.2. Anomaly detection is understood to be the identification of observations, events or data points that deviate from what is routine, standard or expected, making them inconsistent with the rest of a data set.

The increasing integration of Internet of Things (IoT) technologies in healthcare systems that has given rise to the more specialized domain of Internet of Health Things (IoHT), where physical medical objects are embedded with electronic sensors and networking capabilities for the purpose of real-time medical data exchange. By definition, the Internet of Health Things (IoHT) is a specific adaptation of the IoT concept, uniquely tailored for the healthcare sector where the interconnectivity of the sensory objects are for the sole objective of

patient's health service administration [267], [268], [269]. With this streamlined adaptation, patients can efficiently access services outside the confines of the hospital physical infrastructure by leveraging several forms of sensing devices. The capabilities of the IoHT technology includes electrocardiography [270], heart rate monitoring [271], electroencephalogram [272], and a range of sensors, such as those measuring pulse, body temperature, blood oxygen levels (SpO2), glucose level, blood pressure, galvanic skin response, electromyography, and patient position (accelerometer) [273], [274], [275].

5.3.1 Dataset description

The ECU-IoHT focus on cyberattacks within the healthcare sector, specifically targeting IoHT devices. It includes a variety of attack techniques, methods or scenarios to address the acute inadequacy of publicly available datasets in this domain. Preliminary analysis demonstrates that the dataset has a realistic testbed to ensure that it considerably reflects real-world conditions. Using this dataset can help develop robust cybersecurity measures tailored to the high-risk healthcare industry, encouraging AI research and innovation in IoHT security [11]. Table 5.5: shows the name of dataset, place of production, number of records, number of features, size collection start time and collection end date.

Table 5.5: Dataset Description

SN.	Description	Description
1.	Name of Dataset	ECU-IoHT
2.	Place of production	Edith Cowan University
3.	Number of records	111,000
4.	Number of features	9 (No, Time, Source, Destination, Protocol, Length, Info, Type Type of attack)
5.	Size	5 MB
6.	Collection start time	24-02-2020
7.	Collection end data	12-06-2020

In developing the testbed, the authors leveraged the Libelium MySignals service, an IoHT production company that provides remote surveillance for daily healthcare monitoring

including sleep apnea, asthma respiratory depression, diabetes, high blood pressure, arrhythmias, and tachycardias [276]. As illustrated in Fig. 5.7, The MySignals IoHT kit was connected wirelessly to the internet. With this setup, laptops and phones with the MySignals application can transmit health monitoring data to the cloud via a router with a broadcast address. From this setup we can see cyber criminals compromising computing devices and smart phones to gain access into the network for illicit activities.

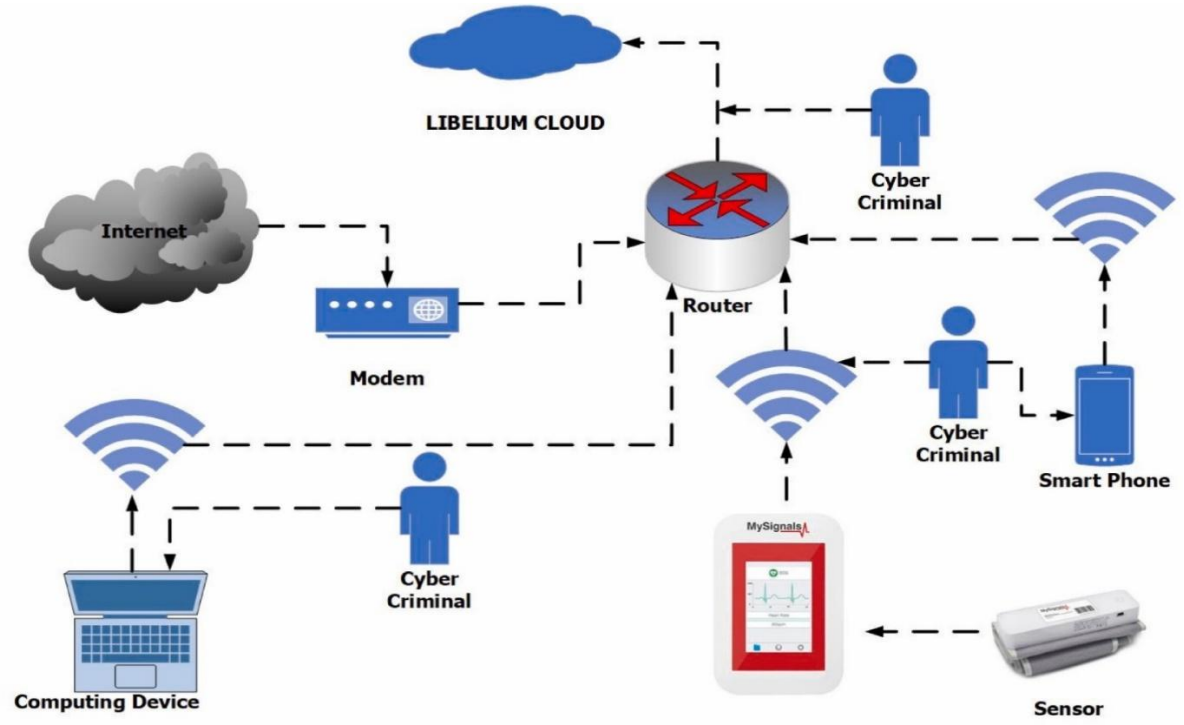


Fig. 5.7: ECU-IoHT Testbed Architecture [11].

5.3.2 Attack scenarios

Cyberattacks are maliciously perpetrated activities directed towards databases, firewalls computers, routers and network equipment to compromise their intended use [277]. These activities usually leads to theft of confidential information, disruptions and overwhelming financial losses [278]. One of the techniques of protecting cyber assets from cyber attackers is the Anomaly based Network Intrusion Detection System (NIDS) [279]

where, profiles are defined to differentiate between the activities of normal and abnormal users [280]. In this work, four types of these cyber attackers are considered that differentiates between the activities of normal use.

1. ***Address Resolution Protocol (ARP) Spoofing*:** The address resolution protocol (ARP) is a protocol used to determine the IP address and the corresponding physical address of a network card [281]. ARP spoofing occurs when an attacker alters the ARP tables of other network hosts, masquerading as a legitimate device [282].
2. ***Nmap (Network Mapper):*** in network mapping, the attacker executes scanning activities to detect underlying vulnerabilities from a target host. The obtained information can further be utilized as a springboard for secondary attacks [283].
3. ***Denial-of-Service (DoS):*** In DoS attack, the cyber-attacker where the perpetrator seeks to make a machine or network resource unavailable to its intended users by temporarily or indefinitely disrupting services of a host connected to the Internet [284].
4. ***Smurf***: Smurf attacks are a specific variant of Distributed Denial-of-Service (DDoS) attack that exploits broadcast networks to amplify malicious traffic. It sends spoofed ICMP echo requests to a broadcast address, causing all hosts on the network to send echo replies to the targeted victim, overwhelming it with traffic [36].

5.3.3 AI/ML algorithms used for anomaly detection

This work used four known ML algorithms that are not originally considered in the research conducted by Ahmed et al. [285]. The selected algorithms are Random Forest (RF), Gradient Boosting (GB), Decision Trees (DT) and Multilayer Perceptron (MLP). The reasons for this selection are to ensure a balance of accuracy, interpretability, and flexibility. Gradient boosting excels in predictive accuracy, while random forest reduces overfitting and handles missing values. Decision trees are simple and fast, making them ideal for interpretation. Multilayer perceptron captures complex relationships and adapt to various problems effectively. Also, tree-based models such as RF, DT and GB are very effective with tabular data [286].

The first is a decision tree which is a classifier expressed as a recursive partition of the instance space [287]. The formula used to create decision trees involves calculating impurity measures that aid in identifying the optimal splits, such as entropy or Gini impurity.

Gini Impurity: Where Pi is the probability of class i in the dataset D [288]

$$Gini\ (D) = 1 - \sum_{i=1}^{n} (Pi)2 \tag{1}$$

Entropy (Information Gain): Where Pi is the probability of class i in the dataset D [288]

$$Entropy\ (D) = \ - \sum_{i=1}^{n} Pi\ \log_2(Pi) \tag{2}$$

Information Gain: Where D is the dataset, A is the attribute, and Dv is the subset of D where attribute A has value v.

$$Information\ Gain\ (D, A) = Entropy\ (D) - \sum_{v \in Values(A)} \frac{|Dv|}{|D|} . Entropy\ (Dv) \tag{3}$$

The second is the random forest algorithm which is a tree-structured classifier consisting of an ensemble of decision trees, where each tree is trained on a random subset of the training data and a random subset of features [289], [290].

Prediction Aggregation for classification using majority voting across all trees is shown in equation 4. Where $\hat{y}$ is the final predicted class label and $\hat{y}_1, \hat{y}_1, \ldots, \hat{y}_k$ are the predictions from the k individual decision trees.

$$\hat{y} = \text{mode}\ (\hat{y}_1, \hat{y}_1, \ldots, \hat{y}_k) \tag{4}$$

The third is gradient boosting machines, where the learning process sequentially fits additional models to enhance the accuracy of predicting the response variable [291]. The three key stages of building gradient boosting machines are given as follows.

Model Initialization: This denotes the start of the model Where $F_0(x)$ is the initial prediction model before any boosting iterations and $\arg\min_c \sum_{i=1}^{n} L\ (y_i, c)$ represents the value c that minimizes the total loss over all n training instances.

$$F_0(x) = \arg\min_c \sum_{i=1}^{n} L\ (y_i, c) \tag{5}$$

Additive Model: Builds additive model by sequentially adding base learners or shallow trees. Where γ_m is a step size or learning rate that scales the contribution of each base learner and at each stage m, a new base learner $h_m(x)$ is fitted to the residual errors or negative gradients of the loss function.

$$F_m(x) = F_{m-1}(x) + \gamma_m h_m(x) \tag{6}$$

Residual Calculation: Residuals are used to guide the learning process in subsequent iterations. The residual r_{im} represents the difference between the true value y_i and the model's current prediction for the $i-\text{th}$ instance at iteration m. The term $\left[\frac{\partial L\left(y_i, F(x_i)\right)}{\partial F(x_i)}\right] F(x) = F_{m-1}(x)$ represents the gradient of the loss function L with respect to the model's prediction $F(x_i)$, evaluated at the current model's prediction $F_{m-1}(x)$

$$r_{im} = -\left[\frac{\partial L\left(y_{i,}F(x_i)\right)}{\partial F(x_i)}\right] F(x) = F_{m-1}(x) \tag{7}$$

Final Model: Describe as the final prediction model after iterations. It is the sum of the initial model and the weighted contributions of all the base learners. Where $F_m(x)$ is the final prediction model after M boosting iterations. $F_0(x)$ is the initial prediction model before any boosting iterations and the summation $\sum_{m=1}^{M} \gamma_m h_m(x)$ represents the contributions of all the base learners like weak decision trees added during each iteration.

$$F_m(x) = F_0(x) + \sum_{m=1}^{M} \gamma_m h_m(x) \tag{8}$$

MLP is an Artificial Neural Network (ANN) that is feedforward. MLP is composed of at least three layers: an input layer with neurons that receive input features; a hidden layer with neurons that apply a weighted sum of their inputs; and an output layer that generates the final predictions after passing it through an activation function [292]. Equation (9) shows a function for creating single layer perceptron unit: Where $f^{(1)}(x)$ represents the activation output of the first layer for a given input vector x. a is the activation function (sigmoid, Relu, tanh) applied to the weighted sum of the inputs. $w_0^{(1)}$ is the bias term for the first layer added to the weighted sum of the inputs before applying the activation function. The expression $\sum_{n=1}^{N} w_n^{(1)} x_n$ represents the weighted sum of the input features [293].

$$f^{(1)}(x) = a\left(w_0^{(1)} + \sum_{n=1}^{N} w_n^{(1)} x_n\right) \tag{9}$$

5.3.4 Methodology and performance metrics

A customized procedural adaptation of the generic machine learning pipeline, as shown in Fig. 5.8, was used in the study. At every stage, a variety of strategies and tactics were employed to achieve the goal of employing four ML models to identify different types of network attacks in an IoHT environment.

The environment for the experiment setup includes (1) a laptop, HP EliteBook Pro, x360 8th Gen, 500GB SSD, Intel Core i7, CPU 2.11GHz, and RAM 16GB. (2) Anaconda Distribution Version 3 (Jupyter Lab—6.4.12 running on localhost 8888, Python 3.9, pandas, SKlearn, and matplotlib).

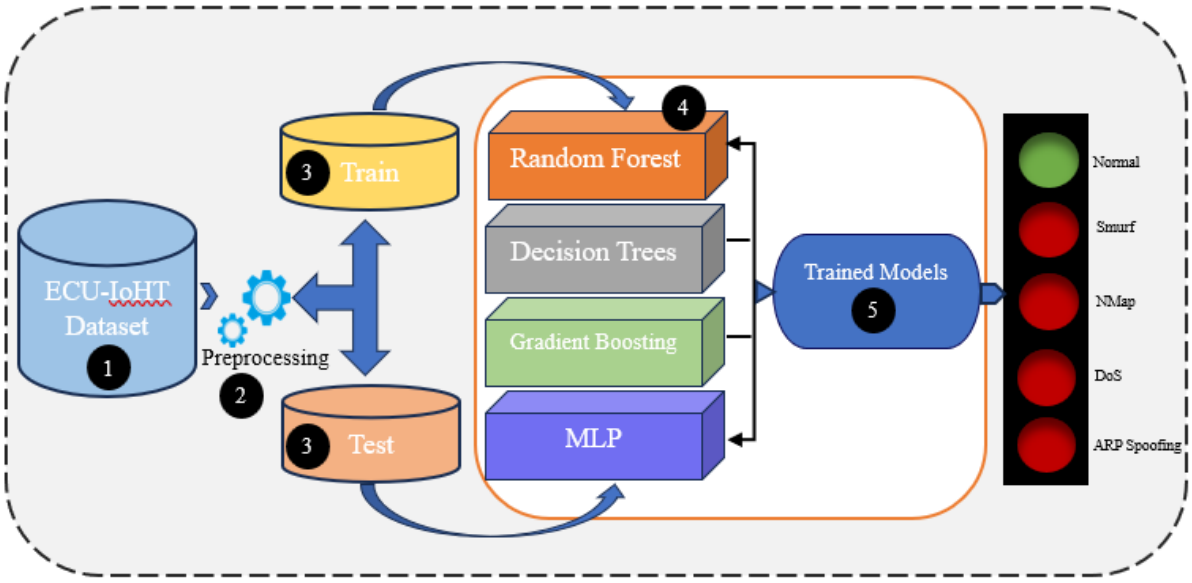


Fig. 5.8: Proposed system

The performance or evaluation metrics denotes the quantitative measures used to assess the performance of ML models [294].

- ***Accuracy:*** Accuracy as captured in equation (10) represents the ratio of correctly predicted instances to the total number of instances [295].

$$Accuracy = (TP + TN)/(TP + TN + FP + FN) \quad (10)$$

- ***Precision:*** Precision as captured in equation (11) is the ratio of correctly predicted abnormal instances to the total number of instances predicted as abnormal [296].

$$Precision = TP/(TP + FP) \quad (11)$$

- ***Recall:*** Recall as captured in equation 12 is the ratio of all the accurately predicted abnormal instances to the entire actual abnormal instance [297].

$$Recall = TP/(TP + FN) \quad (12)$$

- ***F1-Score:*** The F1-score is derived from calculating the harmonic mean of the Precision and Recall as captured in equation (13) [294].

$$F1 - Score = (2 \times Precison \times Recall)/(Precision + Recall) \quad (13)$$

5.3.5 Parameter settings for the models used

For this experiment, we used the following parameter settings for the different models. We ensured that we kept most of the parameters stable for all models so that our comparison can be valid. Finally, we used the 70:30 train test split because they work better with medium sized data.

In selecting the values of our parameters, we followed the recommendations of the sklearn documentation [196]. The sklearn documentation provides a robust guide for choosing parameter values based on the type and size of dataset.

1. **Decision trees:** In sklearn, the decision tree model is initialized by DecisionTreeClassifier(). And we specified the following hyperparameters for the model.

- **Criterion:** This ensures the quality of split. There are two options possible with this parameter (gini or entropy). In this experiment, we used the gini criterion because it is computationally more efficient.
- **Max_depth:** Specifies the maximum depth of the tree. We used the default setting which allows all nodes to be expanded until the leaves are pure.
- **Min_sample_split:** This refers to the minimum samples required to split a node. We used 2 for our min_sample_split to reduce excessive branching.
- **Min_sample_leaf:** This refers to the minimum samples in a leaf node. We used 1 to ensure each node has enough samples and to also ensure consistency with comparison with the other models.
- **Max_features:** The max features refer to the number of features to consider when deciding the best split. We used all the features to decide on the best split.
- **Random_state:** We used 42 for our random state for reproducibility.

2. **Random Forest:** We initialized our random forest by this expression forest_model = RandomForestClassifier(). We then specified the following parameters for the model.

- **N_estimators:** We specified 100 as the number of trees in the forest.
- **Criterion:** We used gini to ensure the quality of the split and ensure consistency with the decision tree model.
- **Max_depth:** We used the default setting which allows all nodes to be expanded until the leaves are pure.
- **Min_sample_split:** We used 2 for our min_sample_split to reduce excessive branching.

- **Min_sample_leaf:** This refers to the minimum samples in a leaf node. We used 1 to ensure each node has enough samples and to also ensure consistency with comparison with the other models.
- **Max_features:** The max features refer to the number of features to consider when deciding the best split. We used all the features to decide on the best split.
- **Random_state:** We used 42 for our random state for reproducibility.
- **Bootstrap:** We used the bootstrap to train each tree on a different subset of the training data. We initialized this value to True.

3. **Gradient boosting:** We initialized our gradient boosting model by this expression gradient_boosting_model = GradientBoostingClassifier(). Then we specified the following parameters for the model.

- **N_estimators:** We specified 100 as the number of boosted trees forest.
- **Learning_rate:** The learning rate shrinks the contribution of each tree to reduce overfitting. We used 0.1 for our learning rate because it ensures better generalization although, it requires more trees and training time.
- **Subsample:** Subsample refers to the fraction of samples to be used for fitting the individual base learners. We use 1.0 in our experiment for each boosting iteration.
- **Max_depth:** We used the default setting which allows all nodes to be expanded until the leaves are pure.
- **Min_sample_split:** We used 2 for our min_sample_split to reduce excessive branching.
- **Min_sample_leaf:** This refers to the minimum samples in a leaf node. We used 1 to ensure each node has enough samples and to also ensure consistency with comparison with the other models.
- **Max_features:** The max features refer to the number of features to consider when deciding the best split. We used all the features to decide on the best split.
- **Random_state:** We used 42 for our random state for reproducibility.

4. **Multilayer perceptron:** We initialized the MLP classifier by the expression mlp_model = MLPClassifier(). Subsequently, we specified the following parameters for the model.

- **Hidden_layer_sizes:** The hidden_layer_sizes parameter in MLPClassifier defines the structure of the hidden layers in an MLP network. We used a single hidden layer with 100 neurons because it is faster to train and less prone to overfitting.
- **Activation:** The activation function determines whether a neuron is activated or not depending on the set of input features. We used the relu activation due to the efficiency in training and reduced vanishing gradient issues.
- **Adam:** We used the Adam optimizer because of its versatility with different data types and it also requires less tuning to function effectively.
- **Alpha:** Alpha is the L2 regularization parameter that is responsible for reducing overfitting by penalizing large weight values. We initialized alpha with the default value = 0.0001.
- **Learning_rate:** In MLP, the learning rate determines how weights are updated. There are two possible options for choosing the learning rate: constant or adaptive. We selected the constant option in this experiment.
- **Max_iter:** This parameter defines the maximum number of iterations. We set the max_iter at 400 to balance efficiency and the time needed for training.

5.3.6 Visualizations, results and discussions

In this sub-section, the proposed system's results are presented. Fig. 5.9 shows 2D and 3D t-SNE (t-Distributed Stochastic Neighbor Embedding) used in EDA (Exploratory Data Analysis) plot. Fig. 5.10 shows the evaluation metrics for the compared ML models and Fig. 5.11 shows the confusion matrices for the misclassification of the classes after predictions have been made.

Dimensions 1, 2, and 3 in Fig 5.9A (3D) and Dimensions 1 and 2 in Fig. 5.9B (2D), are the same representation of the reduced dimensions created by the t-SNE algorithm. It is worth mentioning that the t-SNE method is optimized for visualization rather than precise classification. To understand the figures, each point represents a datapoint in the ECU-IoHT dataset, and different colors correspond to the different cyberattack classes generated by the network traffic. ARP spoofing is shown in dark blue, DoS attacks in orange, Nmap port scans in light gray, normal traffic in dark gray, and Smurf attacks in light blue. The plots help to visually separate different categories of network traffic, and the datapoints are separated based on similarities. The distribution of the colors indicates how distinguishable different traffic types are. We can deduce that points with the same color tend to form clusters, indicating that similar attacks may have similar features. We can also see that the orange clusters (DoS) are well separated, which suggests that they have similar characteristics. We can also see that the light grey clusters (Normal) and dark grey clusters (Nmap port scan) have overlapping regions, implying that this can potentially be challenging for ML models to classify.

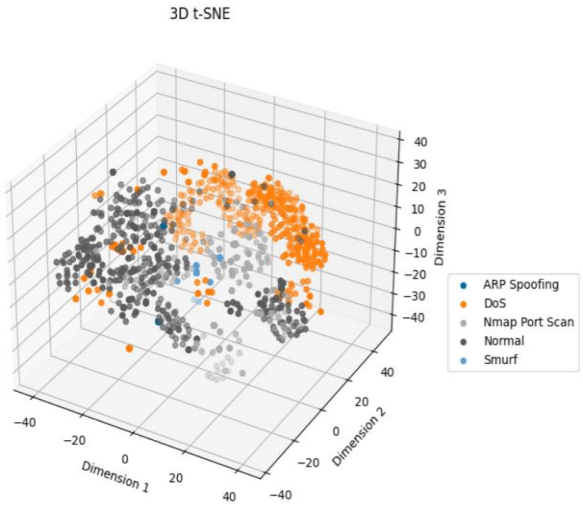


Fig 5.9A: 3D t-SNE EDA

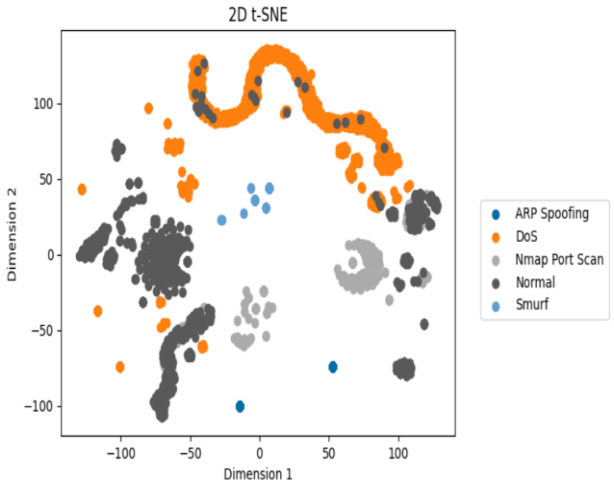


Fig 5.9B: 2D t-SNE EDA

Fig. 5.9: 3D and 2D t-SNE EDA

The bar graph shown in Fig. 5.10 Accuracy, Precision, Recall, and F1-Score are the four-evaluation metrics that we used to compare the performance of the four applied machine learning algorithms.

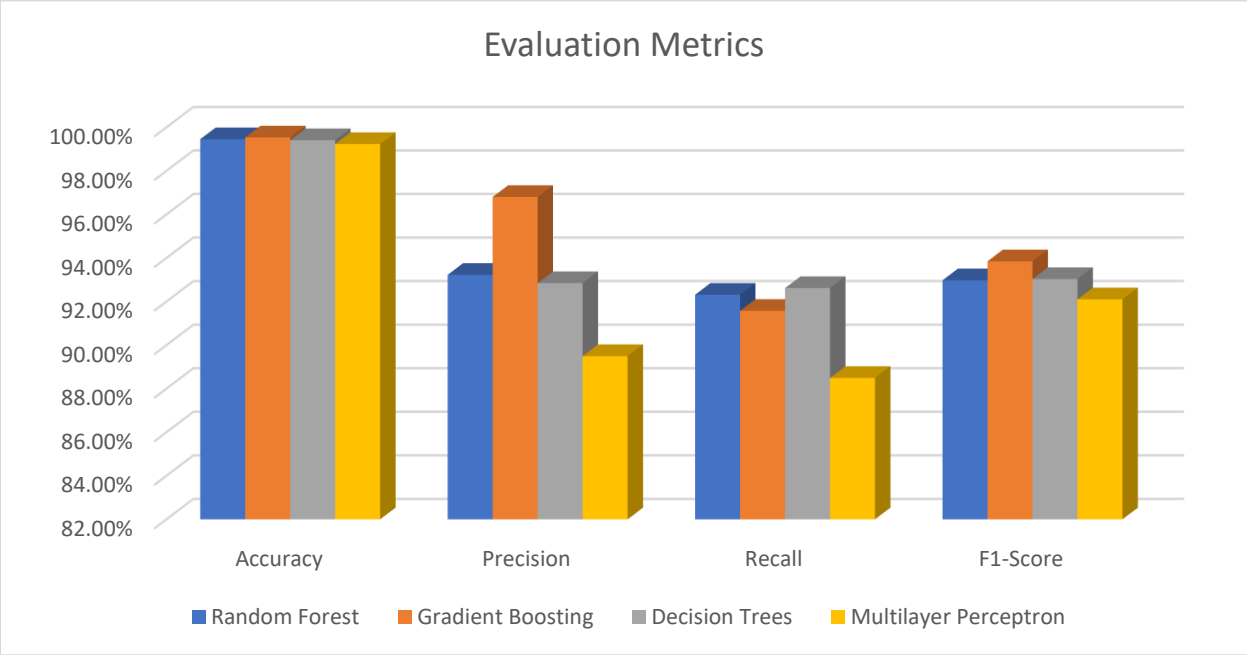


Fig. 5.10: Comparison of Evaluation Metrics

The experiment shows high accuracy with Random Forest model which reached up to 99.45%. This demonstrates that it is quite robust in handling ECU-IoHT dataset. It is apparent that this model benefits from strengths of ensemble approach that mitigates overfitting by averaging multiple decision trees [298]. Furthermore, RF models strike a balance between true positive (TP) identification and minimizing false positives (FP) as indicated by precision and recall values of 93.22% and 92.31%, respectively. With F1-Score at 92.97%, the overall performance of the model reveals its ability to maintain a high level of classification suggesting that Random Forest can be used reliably for detecting normal traffic behavior and attack classes.

The GB model outperforms the other models in terms of accuracy and F1-Score achieving 99.53% and 93.84%, respectively. This model's sequential learning approach of boosting the accuracy of weak classification from previous instances into more accurate

predictions in subsequent models [299]. This is achieved by correcting errors from previous iterations that lead to enhanced precision and the gradient boosting model was able to significantly attain that with a precision of 96.80%. Although, a recall of 91.57% is marginally lower than the Random Forest and Decision Trees algorithms, GB is still exceptionally good at reducing false positives. The high F1-Score is indicative of the model's ability to provide balance and robustness in detection making it reliable for the ECU-IoHT dataset that has nuanced and varied intrusion patterns.

The Decision Trees (DT) model similarly shows high accuracy of 99.41%. A recall of 92.61% also suggests that it is effective at classifying true positive instances. However, a precision score of 92.84% is marginally lower than that of Gradient Boosting and Random Forest due to susceptibility to overfitting [300]. The F1-Score of 93.03% also indicates a strong balance between precision and recall.

The MLP model shows the lowest performance among the evaluated models with an accuracy of 99.23%. It exhibits a precision of 89.49% and a recall of 88.49% suggesting difficulty in capturing all classes within the ECU dataset. Despite these challenges, the MLP achieves a decent F1-Score of 92.10%, indicating it can still be leveraged as a valuable model in the detection of network intrusions.

Overall, the Random Forest and Gradient Boosting models exhibit superior performance, with Gradient Boosting slightly leading in precision and overall F1-Score. Decision Trees also perform robustly, especially in recall, while MLP requires further tuning to reach its full potential. These findings underscore the importance of selecting appropriate machine learning techniques tailored to the specific characteristics of the dataset and the targeted application in network intrusion detection.

Furthermore, Fig 5.11 provides more insight into the misclassification of the four ML models experimented. The confusion matrix provides a reliable means for encoding the

specification of misclassifications problems [301]. Since this study used the supervised ML approach to the classification problem, the confusion matrices as present in Fig. 5.11 are well suited to evaluate errors in classification problems [302]. Brighter colors indicate a higher number of correctly or incorrectly classified instances. The Decision Trees model (Fig. 5.11A) accurately identified all instances of Smurf and Normal traffic. It misclassified some Nmap Port Scan traffic as normal. Figure 5.11B shows similar results, achieving slightly better precise classification of 9,291 instances of Normal traffic, yet still misclassifying some Nmap Port Scan traffic instances. Gradient Increasing (Fig. 5.11C) improves normal traffic detection, accurately identifying 9,349 instances, along with experiencing some difficulty with Nmap Port Scan traffic, resulting in 75 misclassifications. The Multilayer Perceptron (Fig. 5.11D) exhibited the highest error rate, particularly with Nmap Port Scan traffic, misclassifying 203 instances. Overall, all models showed varying degrees of misclassification with Nmap port scan attacks but showed high accuracy with smurf and ARP spoofing attacks indicating that these attack traffic have more distinguishable characteristics.

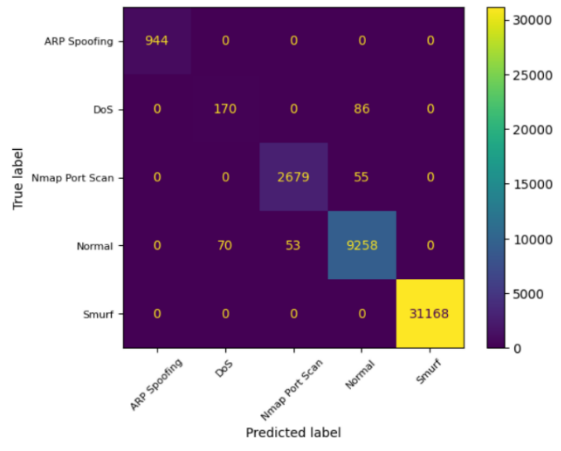


Fig. 5.11A: Decision Trees

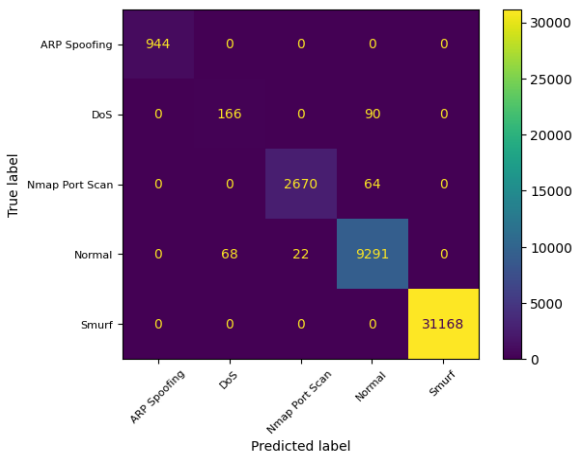


Fig. 5.11B: Random Forest

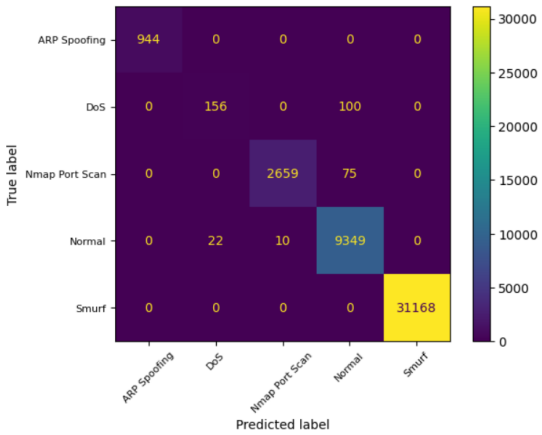


Fig. 5.11C: Gradient Boosting

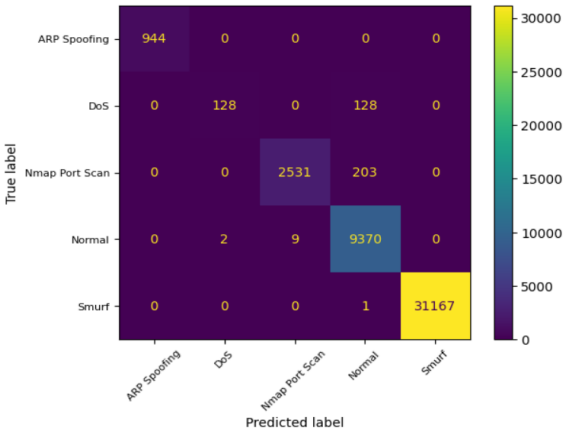


Fig. 5.11D: Multilayer Perceptron

Fig. 5.11: Confusion matrices

DT and MLP models as shown in Fig. 5.11A and 5.11D exhibited higher misclassification rates in distinguishing between Nmap Port Scan and Normal instances or classes. Random Forest has better classification accuracy by reducing variance, while Gradient Boosting achieves the best performance by minimizing misclassifications across all the classes. Overall, GB and RF outperform DT and MLP, which suggest a superiority of the ensemble methods in handling multi-label classification problems.

5.4 Case study 3: AI/ML in website phishing detection

The third relevant case study used in this thesis is the application of AI/ML in the detection of phishing websites. As shown in Table. 5.4, some of the widely known cyberattack incidents in the healthcare sector are phishing attacks. Apart from the outcomes of our risk register that places phishing attacks with a ranking of 405, several studies in the literature also corroborate this fact [303], [133], [304], [305], [306].

Phishing is a form of social engineering which involves psychologically manipulating an individual through persuasion to reveal sensitive or confidential information [307], [308]. Amongst the popular types of social engineering attacks such as trojan horse, shoulder

surfing and dumpster diving [309], phishing stands out as the most frequently employed technique [213], [310].

5.4.1 Dataset and description

This case study leverages Abdelhamid et al.'s work for the dataset used [311]. where a dataset of 1353 websites from various sources was gathered, and several characteristics associated with legitimate and phishing websites were identified. Phishing websites were sourced from the Phishtank data archive (www.phishtank.com), a community website that enables users to submit, verify, track, and share phishing data. Legitimate websites were sourced from Yahoo and starting point directories. By integrating the PHP script with a browser, the authors were able to gather 548 valid websites from a total of 1353. There were 702 phishing URLs and 103 suspicious URLs in the dataset. A website is considered suspicious if it exhibits characteristics of both legitimate and phishing websites, suggesting that it possesses both authentic and fraudulent qualities [35].

The dataset's application of the Chi-Square measure [312], a statistical technique for assessing the degree of independence or association between two categorical variables, is yet another noteworthy feature. In order to compare these differences to their expected values under the assumption of independence, the difference between observed and expected frequencies of occurrence is calculated in a contingency table. This allowed for the creation of precise, succinct features and the elimination of data noise that might not be crucial for producing reliable predictions. Out of the 16 features that were first identified, nine had the highest correlation with the class attribute after the Chi-square test was applied. Table 5.6: shows the features and description for the dataset which include Request URL, Domain Age, HTTPS and SSL, Website Traffic, Long URL, SFH, Pop-Up window, Anchor URL, Redirect URL, and IP Address usage are some of these features.

Table 5.6: Features and Definition

SN.	Feature	Description
1.	Server Form Handler (SFH)	When user submits information, it is transferred to a server for processing, typically on the same domain, but phishers may use an empty server form handler or transfer the information to a different domain.
2.	Pop-up Windows	High prevalence of pop-up windows pre-empting users to enter their personal identifiable information is more likely to be associated with phishing sites.
3.	SSL Certificate	HTTPS protocol presence indicates a legitimate website, but phishers may use fake HTTPS, hence verify HTTPS by trusted issuers (e.g., GeoTrust, GoDaddy, VeriSign).
4.	Web Redirect	Phishers often use link redirection to deceive users into submitting their information to a fraudulent site, making it difficult for users to detect the real link they are being directed to.
5.	@ Symbol	The ‘‘@’’ symbol leads the browser to ignore everything prior it and redirects the user to the link typed after it.
6.	Web Traffic	Phishing websites have low web traffic and short life, whereas legitimate websites have high traffic and lower rank, typically less than or equal to 150,000 according to Alexadatabase.
7.	Long URL	Phishers may hide parts of the URL to redirect user information or upload pages to suspicious domains, with no reliable length to distinguish phishing from legitimate URLs, but a length greater than 54 characters may indicate a phishing URL [313]
8.	Age of domain	Websites with a duration of less than 1 year of online presence may be deemed risky.
9.	IP address	Using an IP address in the domain name of the URL is an indicator someone is trying to access the personal information. This trick involves links that may begin with an IP address that most companies do not commonly use any more.

5.4.2 Proposed system

As illustrated in Fig. 5.12. The first step in this method is to identify the analysis of the phishing dataset and preprocessing to determine whether it is suitable for ML tasks. We then determined the class label (an output variable) and the independent features (an input variable). These components must be precisely defined to train an ML model. Next, based on the dataset's features, we determined the kind of phishing issue we wished to address. To find the relationships between the dataset's variable pairs, we used a par plot. We identified this as a classification problem based on the pair plot's findings. After that, we constructed and trained the model using Support Vector Machines (SVM). Furthermore, we applied the RBF and polynomial functions optimization kernels of SVM. The model must then be integrated into a web framework in order to be deployed. Integrating database technology with Flask is the last essential step, particularly to facilitate database growth.

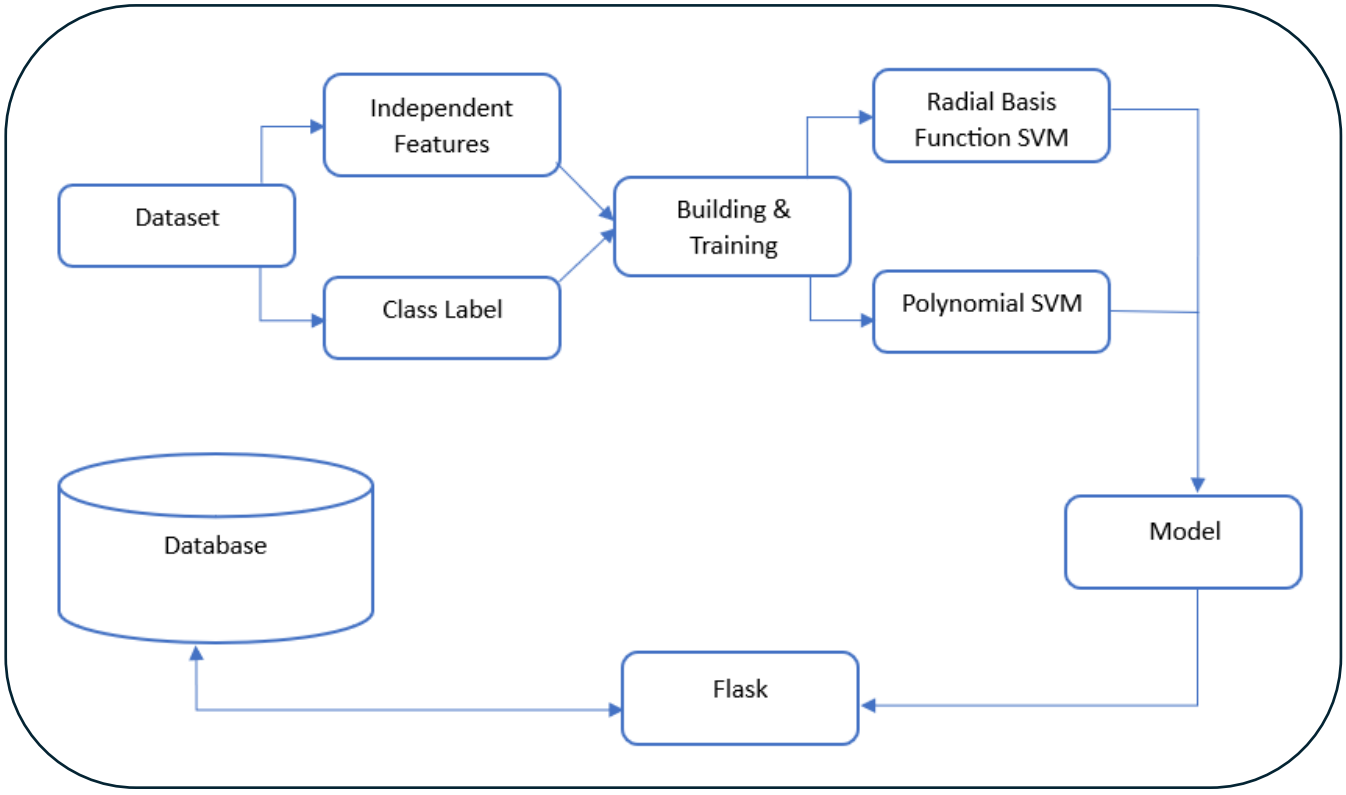


Fig. 5.12: Proposed System

5.4.3 Justification for algorithm used

Various data types and problems are better suited for different algorithms. The problem we are trying to solve, the particulars of the data, and the trade-offs between computational efficiency, interpretability, and model complexity all influence the algorithm that is chosen for this work [314].

For instance, logistic regression is used in classification problems where the datapoints are linearly separable however, that will not be applicable in this case given that our data is a three-label multiclass classification problem [315]. There are also theoretical dimensions that are abundantly articulated in the literature on the use of SVMs in solving multiclass classification problems [316], [317], [318] and [319]. Some of these inherent advantages include.

Since the data is a three-label multiclass classification problem, logistic regression, for example, is not applicable in this situation. Instead, it is used in classification problems where the datapoints are linearly separable [315]. Also, the literature on the application of SVMs to multiclass classification problems also extensively articulates certain theoretical aspects [316], [317], [318] and [319]. Among these benefits are the following.

1. **Handling Nonlinear Relationships:** SVMs are effective in capturing complex nonlinear relationships in data.
2. **Performance in Imbalanced Datasets:** SVMs can handle imbalanced datasets well and are not heavily influenced by class distribution.
3. **Memory Efficiency:** SVMs are memory-efficient, particularly when dealing with high-dimensional datasets.

5.4.4 Support Vector Machines (SVM)

In supervised learning, the Support Vector Machines (SVMs) family of generalized linear classification techniques is employed for both classification and regression tasks [320]. They are known as Maximum Margin Classifiers because of their unique ability to maximize the geometric margin and minimize the empirical classification error at the same time. SVMs operate by mapping the input vectors into a higher dimensional space where a maximal separating hyperplane is built. This process is based on the Structural Risk Minimization (SRM) principle. To separate the data, two parallel hyperplanes are built on either side of the separating hyperplane. The separating hyperplanes is the one that optimizes the separation between these parallel hyperplanes. A greater margin or separation between the parallel hyperplanes is thought to improve the classifier's generalization error [320], [321].

To efficiently differentiate between various data elements and categorize them as phishing, legitimate, or suspicious depending on the dataset, the SVM algorithm in this work makes use of a hyperplane. The hyperplane guarantees the optimal data separation by dividing the features. The category is then predicted by SVM based on which side of the gap the input or point falls on after it has been mapped into the same space. Additionally, two kernels, the polynomial and the radial basis function—were tested on the dataset to identify which one performed best with it when SVM was implemented.

Equation (14) shows the kernel used in the radial basis function (RBF) of SVM. The mathematical equation for the RBF kernel function is presented as thus [321]:

Where x_i, x_j are input vectors, $||x_i - x_j||$^2 denotes the squared Euclidean distance between x_i, x_j, and γ is a hyperparameter that controls the width of the kernel.

This kernel function maps the input data into a higher-dimensional feature space where it becomes separable by a linear decision boundary.

The RBF kernel is widely used in support vector machines due to its ability to handle complex, nonlinear decision boundaries in the data [321].

$$K(x_i, x_j) = \backslash exp(-\Upsilon\ ||x_i - x_j||\hat{}2) \quad (14)$$

The second kernel used in this work is the polynomial kernel which is mathematically represented by Equation (15) [321]:

Where x_i, x_j denote two datapoints in the feature space, $x_i\hat{}T\ x_j$ compute the dot product of their feature vectors and raise it to the power of d, and also adding a constant r. The width of the kernel and the smoothness of the decision boundary are determined by the value of γ. To improve the separation of data points, the input space is transformed into a higher dimensional space using the polynomial kernel [321].

$$K(x_i, x_j) = (\gamma\ x_i\hat{}T\ x_j + r)\hat{}d \quad (15)$$

5.4.5 Parameter settings for the experiment

There are five major parameters used with the SVM model. Specification for the kernel type, the decision function shape, the regularization parameter, the train-test split and random state.

1. **Kernel type:** When using the SVM model, we had to specify the kernel type. The radial basis function is activated and called by kernel = 'rbf' while the polynomial kernel is activated by kernel = 'poly'. The RBF kernel is better when there are nonlinear

relationships between features while the polynomial kernel is better at capturing polynomial relationships.

2. **Decision function shape:** The decision function shape determines or specifies how multiclass classification problems are handled. There are two major options we can use for our experiment. One-vs-one (OVO) and one-vs-rest (OVR). We used the OVO decision function shape because there are few classes to predict (e.g., < 10). In our case, our dataset has only three classes (legitimate, suspicious and phishing). OVR are more useful when the size of the data is large and the number of classes to predict are much (e.g., >10). When computational cost is a priority then OVR is ideal but when accuracy is a priority then OVO is the most ideal.

3. **Regularization parameter:** The regularization parameter is a hyperparameter used to prevent overfitting by adding a penalty to the model's complexity. It controls how much the model should generalize rather than just memorize the training data. The regularization parameter in SVM is denoted by C. We use c =1 as the regularization parameter in our experiment because a lower value maintains lower misclassification because a higher value can lead to overfitting.

4. **Train- test split:** We used 80% for training and 20% for testing as it produced the best results for this experiment.

5. **Random state:** The random state ensures the reproducibility of the split and we used 42 for this experiment. Thie random state can be denoted by any fixed number. If this parameter is not specified, then the split will be different every time.

5.4.6 Visualizations, results and discussions

As shown in Fig. 5.13, the Exploratory Data Analysis (EDA) provides an overview of the features within the dataset, illustrating how these features are distributed across the three

class labels: -1, 0, and 1. From the specifics of the dataset of -1 represents phishing instances, 0 denotes suspicious cases, and 1 indicates legitimate instances. From the output of the EDA, we were able to understand patterns and potential correlations that is present in the dataset. This is very important in understanding underlying structures as a preparatory for applying ML models.

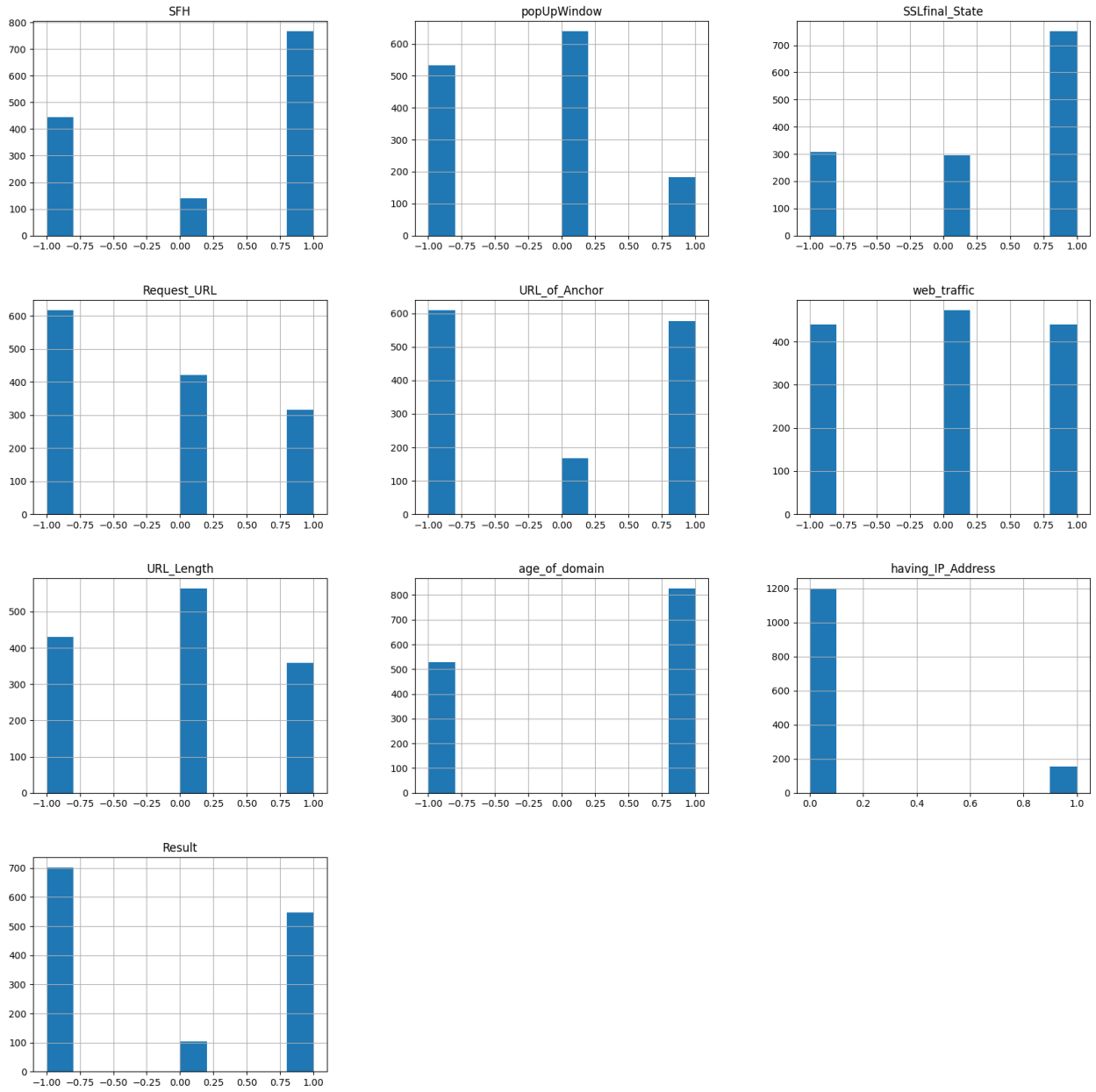


Fig. 5.13: EDA result

Fig 5.14 compares the performance of two different Support Vector Machine (SVM) kernel functions, the Radial Basis Function (RBF) and the Polynomial function, in predicting phishing classification outputs. Each plot visualizes actual values in blue versus predicted values in red, providing insight into how well each kernel function aligns with the real data. On the Y-axis, the three distinct classes are [-1 = phishing, 0 = suspicious, 1 = legitimate]. The X-axis represents the datapoints as indexed in the dataset which corresponds to the sample size of the phishing dataset. In Fig 5.14A (Radial Basis Function kernel), most of the predicted values (red) closely align with the actual values (blue), indicating that the RBF kernel performs well in classifying data points correctly with an accuracy of 82.6%. However, there are a few misclassifications, as seen in the scattered red points that deviate from the blue values. In Fig 5.14B (Polynomial Function kernel), the model also follows the actual values closely with an accuracy of 84.5%. There are less misclassified points, especially around the lower range of the plot in comparison to the RBF, where red points appear in areas without corresponding blue points. From these plots, we can deduce that both kernel functions provide reasonable classification performance, but the polynomial kernel appears to be slightly more accurate, with fewer misclassified points.

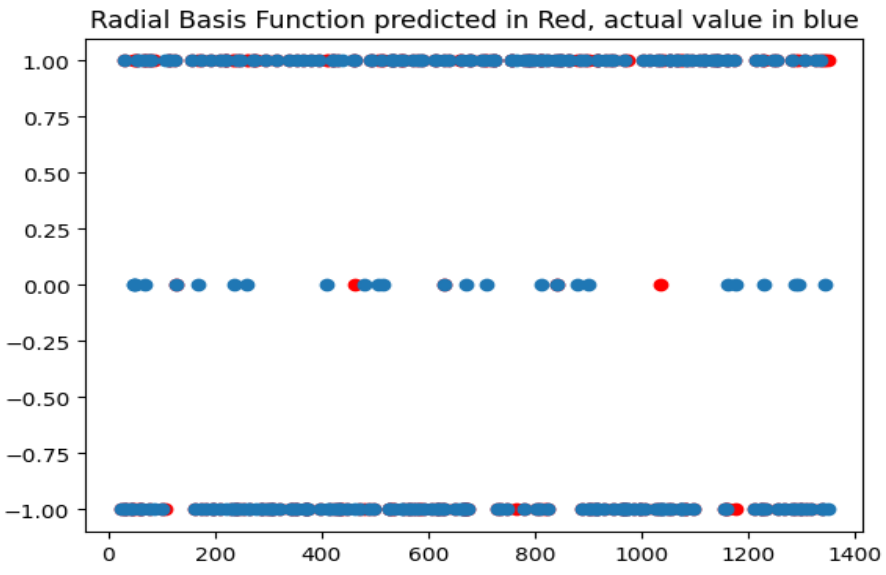


Fig. 5.14A: Radial basis function

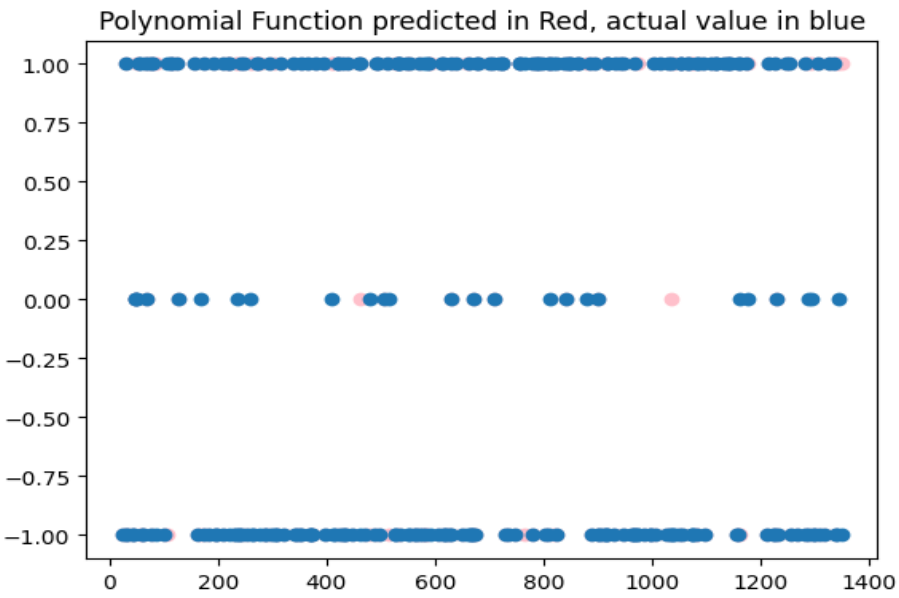


Fig. 5.14B: Polynomial function

Fig. 5.14: Output of phishing plot using SVM

Although not by a huge margin, the polynomial SVM performed better than radial basis function on the phishing dataset (84.5% to 82.6% respectively). However, we can fairly conclude that our models performed reasonably well and can both be used in solving website phishing problems or similar forms of classification problems where the data set has complex patterns or multiple class distinctions.

In comparison to other works, our model performed reasonable well as the work of [322] emerged with an accuracy score of 77%. Other related work [323], [324], [325], [326] approach the phishing problem as a binary classification problem with either phishing or normal instances and that reduced the classification complexity of the ML models used by the authors.

Whilst putting future work into perspective, it is important to note that phishing attacks are highly dynamic in nature, and that once cybercriminals are caught in their schemes, they use other strategies to get around the issue and obtain an advantage. This work heavily relies on well-known intrinsic characteristics of websites and chooses the most salient features that are common and significantly correlate with the legitimacy or illegitimacy of unidentified websites that are subjected to the ML algorithm. With the proliferation of several web technologies like Yii, Motion UI, Django, and Meteor JS, these nine features may not be enough to support the more recent trends in cybercriminals' websites. Thus, investigating more recent developments in website phishing will be the focus of this work in the future.

5.5 Case Study 4: AI/ML in intrusion detection

The fourth relevant case study used in this thesis is the application of AI/ML in intrusion detection systems. An intrusion detection system in general terms is perceived as

a hardware or software program that checks malicious activity or policy infractions on a network or systems [327].

Network security is very important to most organizations, particularly the healthcare sector. As IoT and medical devices communicate over the network increases the risk of cyberattacks. In order to improve Machine Learning (ML)-based Network Intrusion Detection Systems (NIDS), this study explores Cross-Machine Learning (Cross-ML), which at the core involves utilizing data from various sources to generalize on network traffic. Therefore, this case study presents a clear framework for Cross-ML and shows how to use it on three publicly available datasets: UNSWNB15, X-IIoTID, and ToN-IoT.

The concept of Cross-ML as articulated in this work refers to a clear and repeatable method for training and testing phases of ML models. The idea behind this concept is because ML-models have a tendency of concept drift when deployed in production [328]. So, Cross-training, involves engaging the entire ML pipeline on various datasets as opposed to standard training which focuses on one, with the likelihood that it improves the model's capacity for generalization. Assessing a model's generalization ability and performance on novel data which are more representative of the data examined in real-world scenarios involves testing and analyzing the model's performance on different datasets. It is crucial to use datasets with identical or similar features when using cross-ML. The Cross-ML pipeline's training and testing stages will become complicated otherwise. Given that there is paucity of data in conducting NID research, this case study is highly relevant because

- Explain the steps involved in implementing ToN-IoT, X-IIoTID, and UNSWNB15 ML techniques.
- Provide a framework for Cross-ML and the significance of this approach in the field of intelligent IoT NIDS
- Stress how important it is for feature names and attack type definitions to be standardized globally.

5.5.1 Used datasets and description

All the datasets analyzed in this study are linked within the NIDS field and are publicly accessible. However, because each dataset offered distinct features, not all of them were used. As mentioned, common features across the various datasets are the necessary precondition for the application of Cross-ML. Other examples of datasets that are publicly available for NIDS research but were not used are the ECU-IoHT and Edge-IIoT datasets as they do not satisfy the criteria in terms of feature names consistency.

The first dataset is ToN-IoT dataset which is used to evaluate the precision and efficacy of different cybersecurity applications that rely on AI/ML [329]. Among other uses, they are designed for adversarial ML, intrusion detection, threat intelligence, and privacy-preserving models. A range of data sources were included in the dataset, such as telemetry datasets from IoT and IIoT sensors, network traffic datasets, and operating system datasets from Windows 7 and 10 and Ubuntu 14 and 18 TLS. It was developed in the UNSW Canberra Cyber's IoT Lab where a realistic and large testbed network was used to collect the data sets in order to replicate the complexity and scalability of IIoT and Industry 4.0 networks. Numerous cloud and fog platforms, IoT sensors, physical systems, hacking platforms, and virtual machines were all connected by this network. To effectively use the Ton-IoT data we leveraged the work conducted in [330] and [329] for a thorough description of this testbed [331]. As shown in Table 5.7, The data is provided in CSV format, The sub dataset used includes 461,043 flows, 161,043 malicious flows and 300,000 benign flows.

Table 5.7: Samples from ToN-IoT dataset

Type of attack	**Number of samples**
Normal	300,000
Scanning	20,000
Dos	20,000

Injection	20,000
DDoS	20,000
Password	20,000
Xss	20,000
Ransomware	20,000
Backdoor	20,000
Mitm	1,043

The second dataset used in this study is the X-IIoTID dataset, which include 421,417 normal records, 399,417 malicious records, and 59 features, were collected from device resources, network traffic, and device/alert logs. SMTP, TCP, MQTT, Modbus, and CoAP are the protocols that are used. The malicious records were produced using three different frameworks: ATT&CK, MALC, and CKC. The testbed attack process consists of several attack classes as seen Table 5.8., including reconnaissance, weaponization, exploitation, lateral movement, command and control, exfiltration, tampering, crypto-ransomware, and ransom DoS [332].

Table 5.8: Samples from X-IIoTID dataset

Type of attack	**Number of samples**
Normal	421,417
RDOS	141,261
Reconnaissance	127,590
Weaponization	67,260
Lateral movement	31,596
Exfiltration	22,134
Tampering	5,122
C&C	2,863
Exploitation	1,133

Crypto-ransomware	458

The third dataset used in this case study is the UNSW-NB 15 dataset which includes recent network traffic from both common and uncommon events, such as low-footprint attacks. For reliable evaluation of Network Intrusion Detection systems, the data is more appropriate because it is available in a clean format and lacks redundancy. It contains about 2,540,044 data instances in total collected in a CSV format. From these records, 82,332 data instances are used as the test set and 175,341 instances as the training set [333].

The UNSW-NB 15 dataset's raw network packets were produced by the UNSW Canberra Cyber Range Lab's IXIA PerfectStorm tool, which combined real-world modern everyday activities with simulated modern attack behaviors. Using the tcp dump tool, 100 gigabytes of raw traffic were recorded. Attacks on this dataset include Worms, shellcode, reconnaissance, backdoors, analysis, fuzzers, exploits, and DoS. Using the Argus and Bro-IDS tools, 49 features with the class label was generated [334]. Table 5.9: shows the various classes of the sub dataset used.

Table 5.9: Samples from UNSW-NB15 dataset

Type of attack	**Number of samples**
Normal	93,000
Generic	58,871
Exploits	44,525
Fuzzers	24,246
Reconnaissance	13,987
DoS	16,353
Analyse	2,677
Backdoor	2,329
Shellcode	1,511

Worms	174

5.5.2 AI/ML algorithms used for intrusion detection

To implement Cross-ML, we emphasize the interpretability of the model. Therefore, we chose the decision tree model. In addition to the easy interpretability of the model, it also has a low computational cost and offers access to more complex ensemble models. The mathematical expression for the decision tree model is shown in equations (1), (2) and (3).

5.5.3 The proposed cross-ML framework

In this sub-section, we present our proposed framework for the application of Cross-ML.

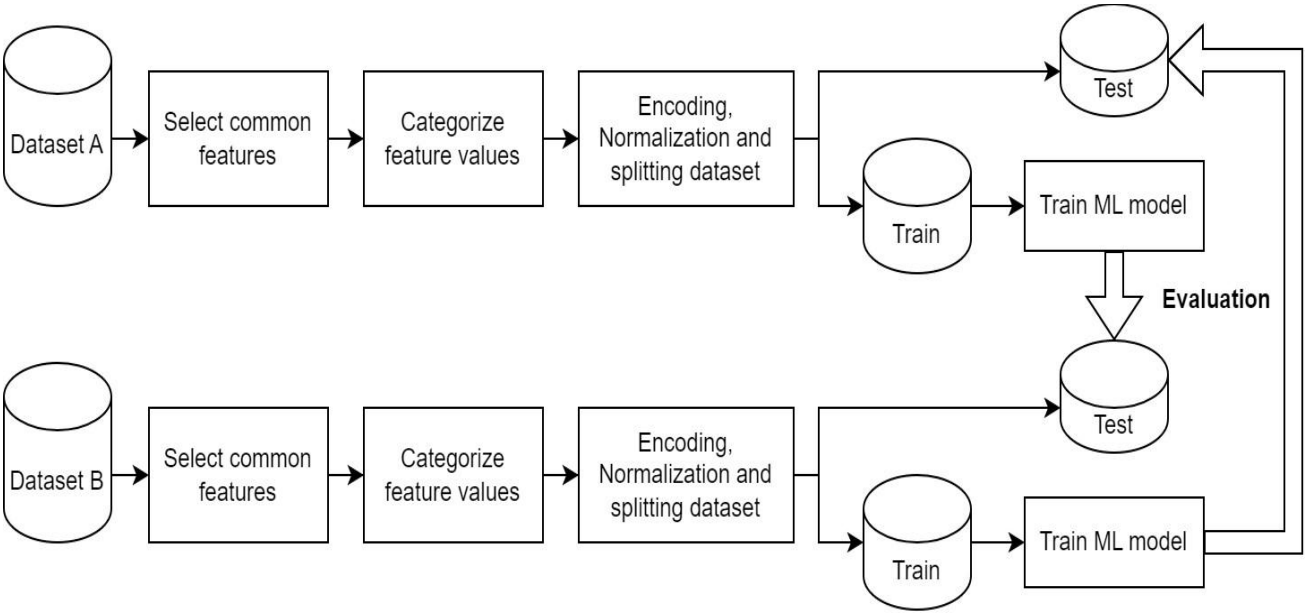


Fig. 5.15: Proposed Cross-ML Framework

As shown in Fig. 5.15, the key steps involved in implementing the Cross-ML framework are described below.

- We begin the process by considering at least two datasets (Dataset A, Dataset B.... Dataset *N*). All datasets contain similar or the same features.

- The next step is the selection of common features across all the datasets since it is anticipated that they may have structural inconsistencies or varying feature sets. It is crucial to identify the common features between the datasets and ensure they are compatible, renaming feature names is necessary.
- The next step is the categorization of feature values which comes after the selection of common features. The key step of this phase is to ensure that feature values are symmetrically aligned to ensure that they have a common meaning across the datasets.
- After categorizing the feature values, the categorical values are encoded into numerical values to make them usable for ML tasks. Normalization is then performed to scale the numerical features to a common range to stabilize the performance of the model. The final step in this category is splitting the data into training and testing subsets.
- After both datasets are preprocessed and split, the next step is to select the model to use. In this case, a decision tree model was adopted. Dataset A, is trained using the training portion of A, while dataset B is trained using the training portion of B. This is important so that the model can learn based on the characteristics of the individual dataset.
- The next step is to test each model based on their test set and after this process, each model is trained on the test set of the other dataset. Intra-dataset evaluation [Dataset A tested on Dataset A, while Dataset B is tested on Dataset B]. This at least provides a baseline score for each dataset. Cross-Dataset evaluation [Where Dataset A is tested on dataset B, while dataset B is tested on dataset A]. This establishes the generalizability of the model.
- Lastly, the results of both intra-dataset and cross-dataset evaluation is compared, using standard performance metrics as F1-score, accuracy and recall. If both results are high, it implies that the model can generalize well.

5.5.4 Common features categorization

As a precondition for cross-ML as stated, the features in the dataset must be similar. To achieve this, this work compared common features across the datasets and presented it in a Venn diagram as shown in Fig. 5.16. Where ToN-IoT features are shown in blue, UNSW-NB15 is shown in red and X-IIoTID is shown in yellow.

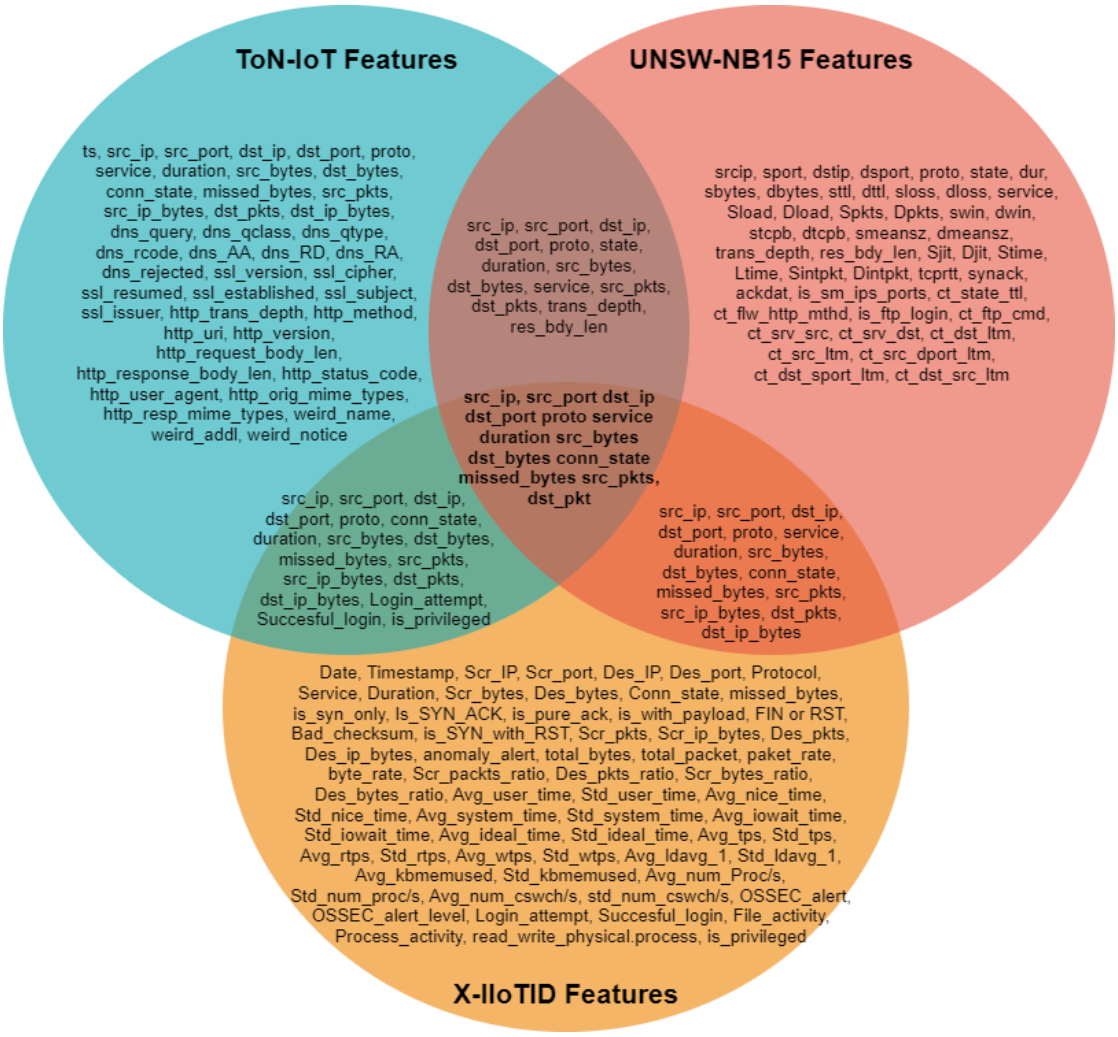


Fig. 5.16: Common features Ton-IoT, X-IIoTID and UNSW-NB15

Furthermore, Table 5.10 shows the common features and description. The categorical features include “proto”, “service”, and “type”. Since our study addresses a

binary classification problem, we set the “type” feature to either “Normal” or “Attack”. For “proto” feature, we retained the values “udp”, “tcp”, and “icmp”, introducing an “other” category for miscellaneous values.

Table 5.10: Common features and description

Feature name	Description
“proto”	Protocol used in the communication
“service”	Type of service for the communication
“duration”	Duration of the communication
“src bytes”	Number of bytes sent by the source
“dst bytes”	Number of bytes received by the destination
“missed bytes”	Number of bytes missed during communication
“src pkts”	Number of packets sent by the source
“dst pkts”	Number of packets received by the destination
“type”	Type of communication

5.5.5 Encoding, normalization and split

To ensure consistency across datasets which at this stage have common features, we encoded and normalized the values from the three datasets. For the categorical features, one hot encoding is applied. One hot encoding is a commonly used technique used for the conversion of categorical features into numerical features. It works by creating binary columns for each category and represents them with a 1 indication a presence of a category and a 0 representing the absence of a category [297]. Afterwards, the numerical features are normalized which involves scaling numerical features to a standard range typically between -1 and 1 as used in this study. Failure to normalize could stop models from converging and manipulating models to assert more importance to values that do not fall in the same range. Normalizing also ensures that models are not biased towards certain

dominant features as given that they do not fall in the same range [329]. Afterwards we used a 60/40 split ratio for training and testing of the dataset.

5.5.6 Training and testing procedure

We used the decision tree model majorly because of computational efficiency and fast training time. The key objective of this case study is to cross the training and testing processes on three datasets. Therefore, we needed a model that is fast and requires low computational overhead. Therefore, we trained a decision tree model for the respective dataset as follows.

1. Trained ToN-IoT tested ToN-IoT, trained X-IIoTID tested on X-IIoTID and trained UNSW-NB15 tested on UNSW-NB15
2. Trained Combined dataset tested on the combined test dataset, Trained Combined dataset tested on ToN-IoT, trained Combined dataset tested on X-IIoTID and trained Combined dataset tested on UNSW-NB15
3. Trained ToN-IoT tested X-IIoTID, trained ToN-IoT tested UNSW-NB15, trained X-IIoTID tested on ToN-IoT, trained X-IIoTID tested on UNSW-NB15 and trained UNSW-NB15 tested on ToN-IoT, trained UNSW-NB15 tested on X-IIoTID

5.5.7 Parameter settings for the experiment

For this experiment, we used the 60:40 train test split because we needed sufficient testing data when crossing.

- **Criterion:** We used gini criterion for quality split.
- **Max_depth:** We used the default setting which allows all nodes to be expanded until the leaves are pure.
- **Min_sample_split:** We used 2 for our min_sample_split to reduce excessive branching.

- **Min_sample_leaf:** We used 1 to ensure each node has enough samples and to also ensure consistency with comparison with the other models.
- **Max_features:** We used all the features to decide on the best split.
- **Random_state:** We used 7 for our random state for reproducibility.

5.5.8 Results and discussions

From the experiments, results are largely satisfactory. As shown in Fig. 5.17, the binary classification results recorded very high scores ranging from 92.0% to 98.1% in terms of F1-score. This is despite using just eight features as the uniformity of feature set is the precondition for the Cross-ML approach used in this studies. The results do not vary much in comparison with models trained on the full feature set of each dataset as provided in the work of Tareq et al. [335]. Their work also provides a comprehensive summary of prior studies on ToN-IoT and UNSW-NB15 and results typically ranged from 96.10% to 99.93% for ToN-IoT and 84.24% to 99.59% on UNSW-NB15 dataset. Whilst using all the features for the X-IIoTID dataset, the work of Verma et al. [336], shows that the accuracy score was between 91.48% and 99.79%. Although the results shown in Fig. 5.17, are trained on the corresponding dataset, the high results helped assert and baseline that the eight feature set are good enough for the individual models trained and tested.

Moreover, we suspect that our results may have been better with other tree-based models such as random forest and gradient boosting because they provide more parameters for optimizing the binary decision tree model. However, it is likely not going to be by a large margin as observed from the experimental results of AI/ML in anomaly detection (subsection 5.3.3). For models such as SVM and MLP, we did not conduct experiments to establish whether they are suited for this kind of problem therefore, we cannot ascertain their applicability. Furthermore, the goal of this case study is to propose a solution for generalizability challenges across multiple datasets of similar features

therefore, we prioritized establishing the applicability of the cross-ML framework than the actual performance of the model.

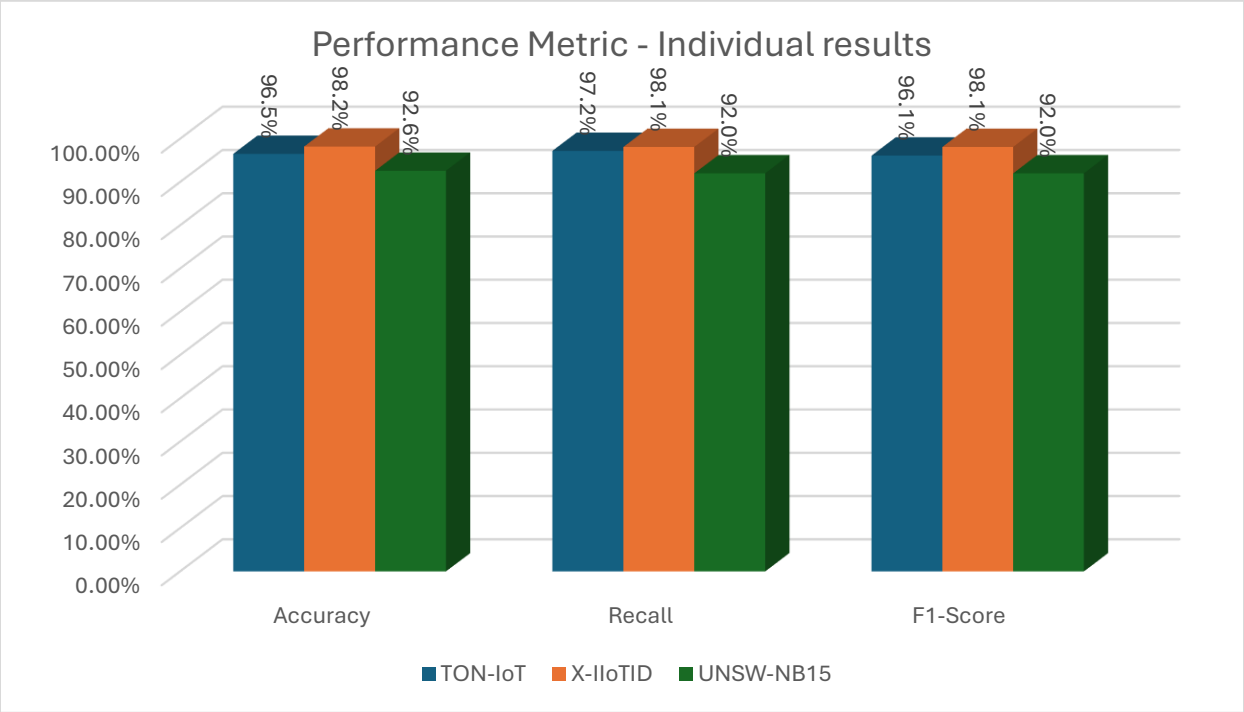


Fig. 5.17: Performance metric Ton-IoT, X-IIoTID and UNSW-NB15

The next step is having a combined model on all three datasets as a whole and evaluating the results which shows an accuracy, recall and F1-score of 90.9%, 84.0% and 84.9% respectively as shown in Fig. 5.18. Afterwards, the trained combined model is tested on the individual testing datasets. Expectedly, there is a significant drop in the model performance in comparison to individual model presented in Fig. 5.17. however, the accuracy scores still showed promise ranging from 82.4% to 97.8%. Limiting the number of features could have been the reason for the drop in performance or fundamentally just using datasets that are captured under different conditions. A study in the literature titled “ToN-IoT: The Role of Heterogeneity and the Need for Standardization of Features and Attack Types in IoT Network Intrusion Data Sets” [331] reinforces our assertion.

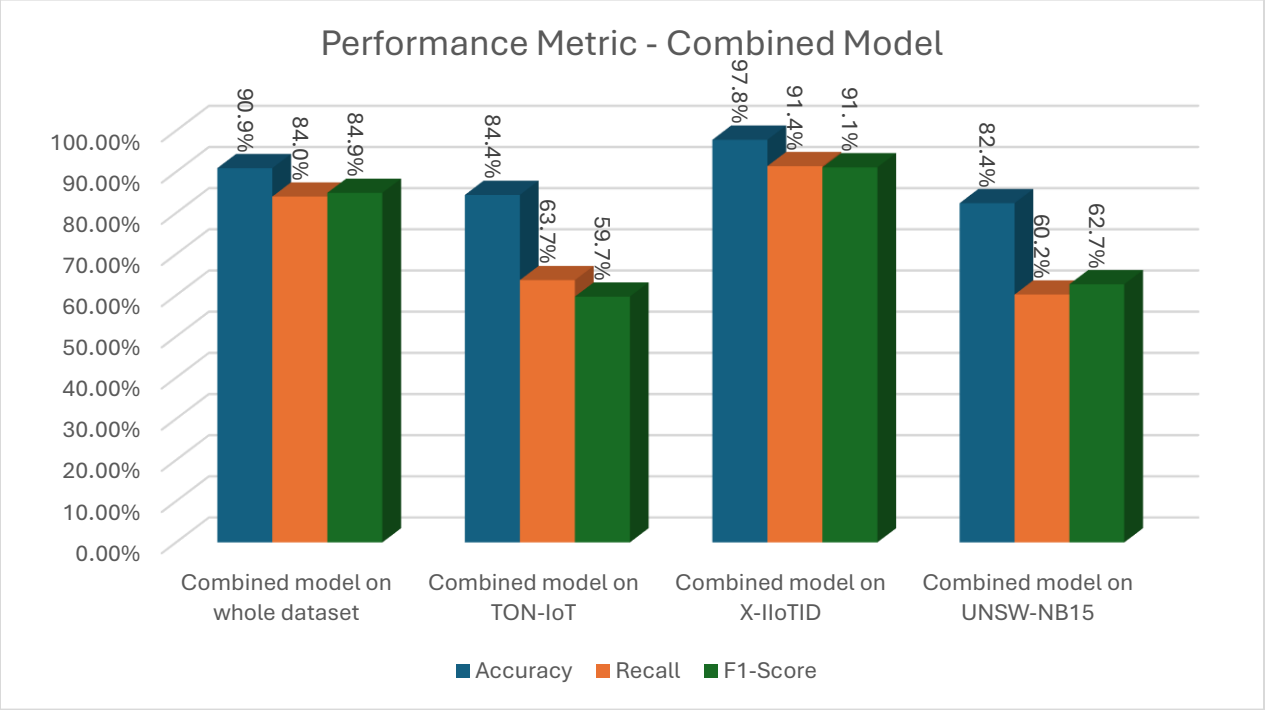


Fig. 5.18: Combined model results

Finally, to test the generalizability of models across datasets, Table 5.19. show the performance of Cross-ML classifications. This was done following the method described in sub-section 5.5.6 (3). [Trained ToN-IoT tested X-IIoTID, trained ToN-IoT tested UNSW-NB15, trained X-IIoTID tested on ToN-IoT, trained X-IIoTID tested on UNSW-NB15 and trained UNSW-NB15 tested on ToN-IoT, trained UNSW-NB15 tested on X-IIoTID]. The results show marginally better performance on the X-IIoTID dataset and ToN-IoT dataset.

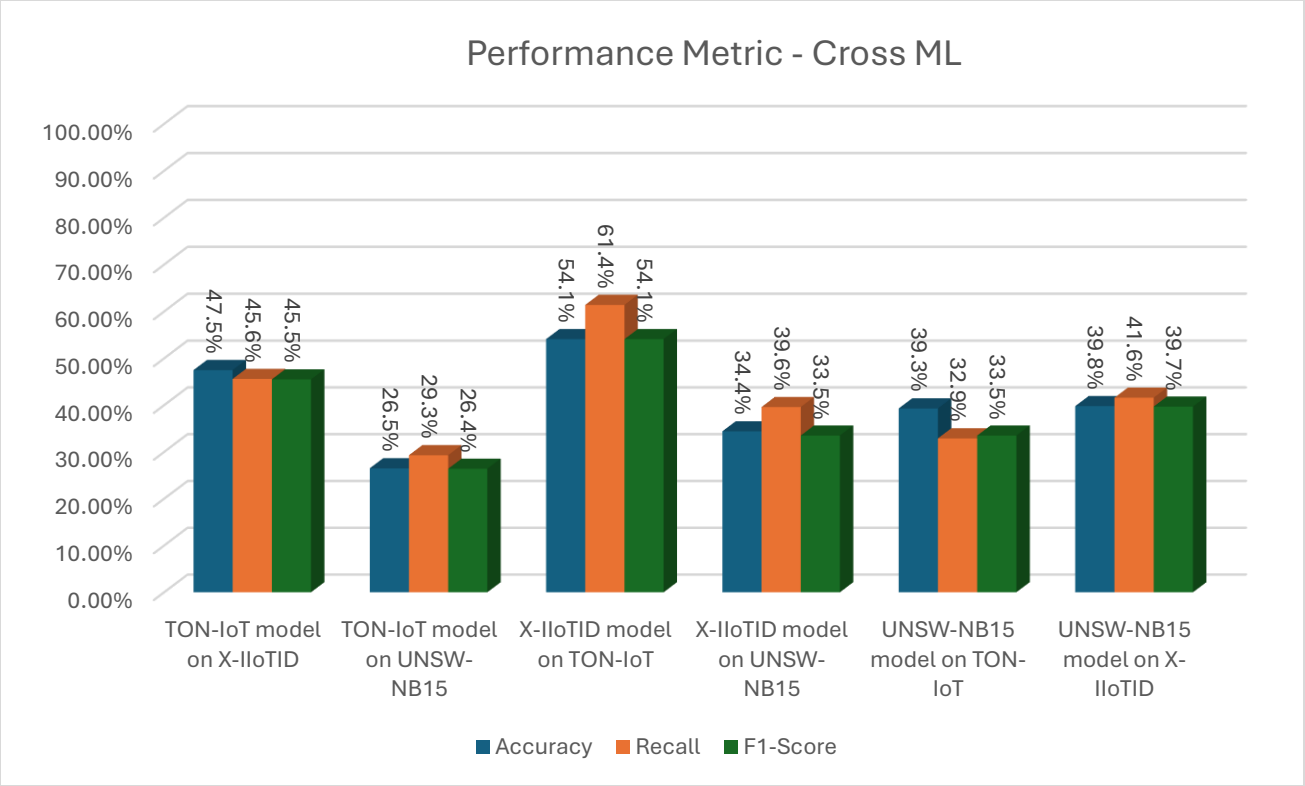


Fig. 5.19: Cross ML performance

In this sub-section, the proposed system's results are presented. Fig. 5.20 shows 2D and 3D t-SNE (t-Distributed Stochastic Neighbor Embedding) used in EDA (Exploratory Data Analysis).

Fig. 5.20 compares the visualization of the network traffic data in reduced dimensional space for the ToN-IoT, X-IIoTID and UNSW-NB15 dataset. Each dot in the images represents a data point, which is a record of network activity. The blue dots correspond to normal network traffic, while the orange dots represent attack traffic. Both the 2D and 3D view provide different perspectives of the same data. Fig. 5.20A, which is ToN-IoT dataset suggests a clear separation between attack and normal traffic, meaning ML models might easily distinguish between normal and attack traffic. In contrast, Fig. 5.20B (X-IIoTID dataset) and Fig. 5.20C (UNSW-NB15 dataset) shows more overlap between

normal and attack traffic, indicating that some attacks may share similar characteristics with normal traffic, making them harder to detect.

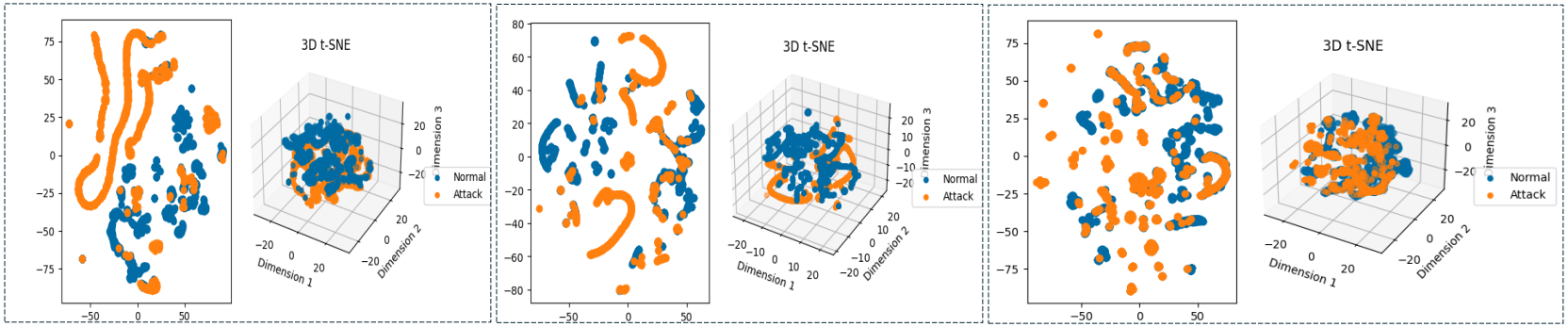


Fig. 5.20A: t-SNE on ToN-IoT

Fig. 5.20B: t-SNE on X-IIoTID

Fig. 5.20C: t-SNE on UNSW-NB15

Fig. 5.20: t-SNE EDA results on all three datasets

Now that we have evaluated the performance of the ML model, it is important to gain insight into the actual reason for the misclassification in some specific instances. Fig. 5.21 shows how well the decision tree model detects attacks in the network traffic. Each confusion matrix is a 2x2 grid for the ToN-IoT, X-IIoTID, and UNSW-NB15 datasets. Top-left is the (True Positives—TP): Correctly classified attacks, Top-right (False Negatives—FN): Attacks misclassified as normal traffic, Bottom-left (False Positives—FP): Normal traffic misclassified as an attack; bottom-right (True Negatives—TN): Correctly classified normal traffic. The color scale on the right shows the frequency of each value, with brighter colors indicating higher counts. We can deduce that Fig. 5.21A (ToN-IoT) correctly identified most attacks (55,894) but made some mistakes by falsely labeling normal traffic as an attack (523). However, normal traffic (114,345) was properly classified with only 5,656 misclassified as attack traffic. For the X-IIoTID in Fig. 5.21B, 119,659 attacks were properly classified as attacks, and 3022 were misclassified as normal traffic, while 116,308 were properly classified as normal traffic, and 2,259 were misclassified as attacks. Lastly, Fig.

5.21C (UNSW-NB15) classified 62,129 attack traffic accurately and misclassified attack traffic as normal traffic 3,741 times. On the other hand, normal traffic was classified accurately in 33,358 cases and misclassified as attacks in 3,842 cases. To conclude, Fig. 5.21B (X-IIoTID) showed better overall performance with fewer negatives and false positives, while Fig. 5.21C (UNSW-NB15) showed the worst performance as it has higher false negatives and false positives relative to the instances of the network traffic.

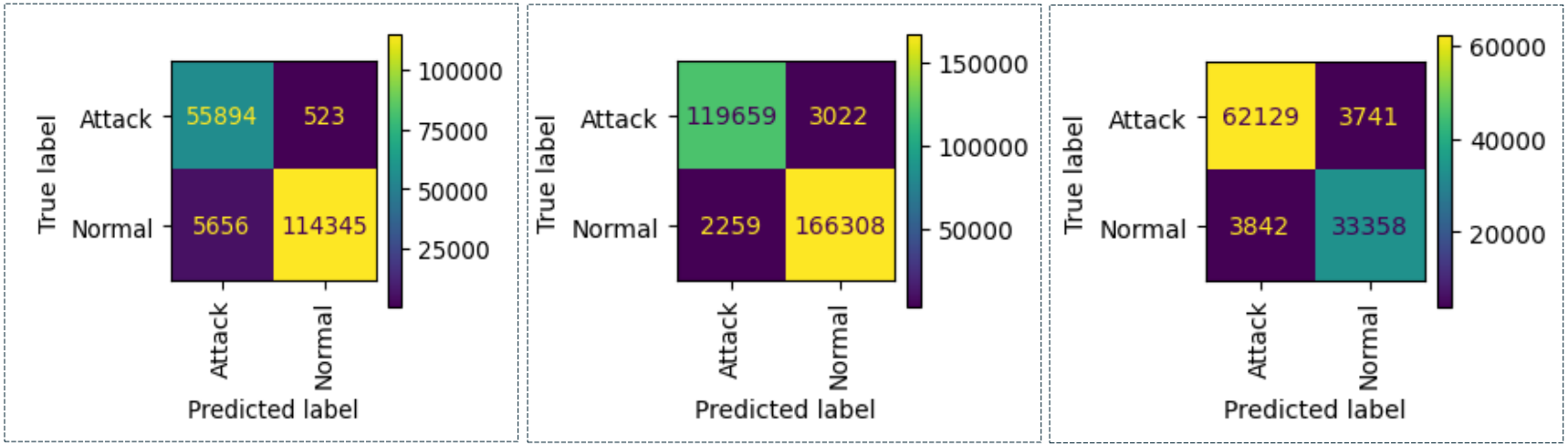


Fig. 5.21A: Confusion matrix of ToN-IoT Fig. 5.21B: Confusion matrix of X-IIoTID Fig. 5.21C: Confusion matrix of UNSW-NB15

Fig. 5.21: Confusion matrices of all three datasets

The Receiver Operating Characteristic (ROC) curve is a visual representation of a classification model's performance at various thresholds. We used the receiver operating characteristic (ROC) curve, shown in Fig. 5.22, because it is simple but provides a

complete empirical description of the decision threshold effect, indicating all possible combinations of the relative frequencies of the various types of correct and incorrect decisions for the normal and attack classes in all three datasets [337]. Our findings are 0.97, 0.98, and 0.92 for ToN-IoT, X-IIoTID, and UNSW-NB15, respectively. In most cases, a ROC curve above 0.5 indicates very strong model performance. The closer the value is to one, the greater the performance, which applies to all of our experiments. This is also consistent with the results of the confusion matrix in Fig. 5.21. Furthermore, the X-IIoTID and ToN-IoT datasets provide the best performance, indicating that these datasets allow the model to discriminate classes most effectively, whilst the UNSW-NB15 reduced score of 0.92 suggests a lower but still strong performance.

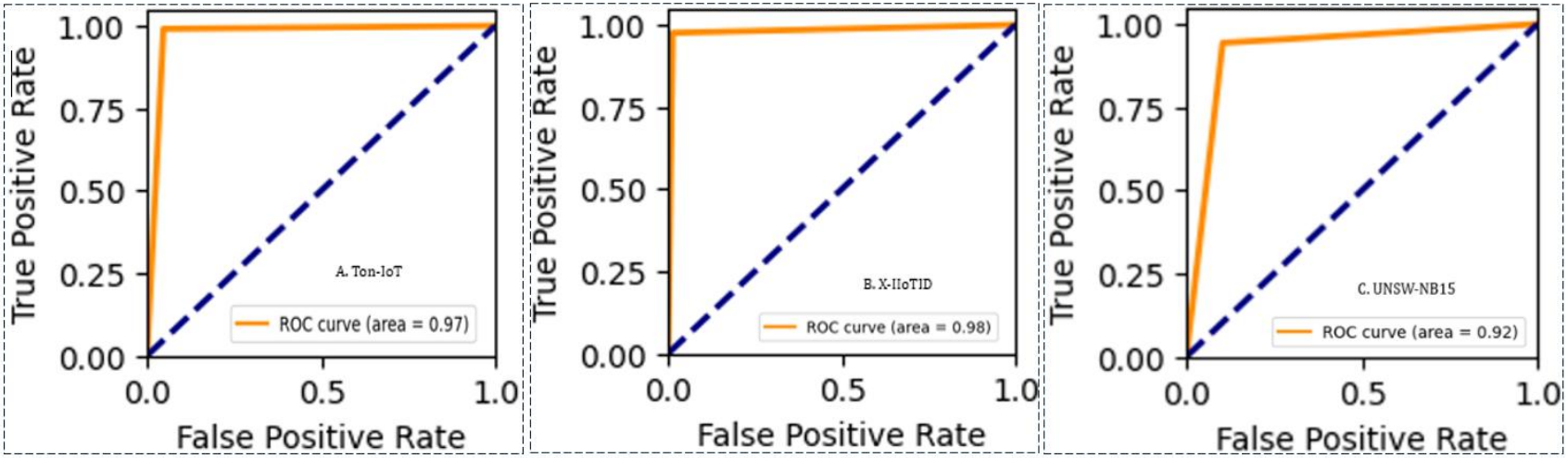


Fig. 5.22A: ROC curve ToN-IoT

Fig. 5.22B: ROC curve of X-IIoTID

Fig. 5.22C: ROC curve of UNSW-NB15

Fig. 5.22: ROC curve of all three datasets

5.6 Conclusion

In conclusion, this chapter discusses case studies and applications, results, discussions, and analysis. This chapter provided a risk assessment framework named the BUHCRAF model strictly for the cyber infrastructure of a typical healthcare facility. Based on the outcomes, targeted technical controls involving AI methods are proffered. These methods include phishing detection using SVMs, anomaly detection in IoHT devices using decision trees, random forests, gradient boosting, and multilayer perceptron. Lastly, we also proposed a cross-ML framework that is based on three datasets (ToN-IoT, X-IIoTID and UNSW-NB15). This proposed framework provides a general, adaptable pipeline that can be used to address the inherent challenges of generalization and concept drift in AI/ML cybersecurity.

CHAPTER 6 – CONCLUSION

6.1 Introduction

This thesis investigated a multilayered cybersecurity approach that is rooted in database usage, the application of AI/ML techniques for cybersecurity, and a streamlined risk assessment framework that is tailored to a specific use case (Baze University Hospital). As evolving cyberattacks continue to pose significant risk to critical infrastructure such as healthcare, this research reiterates the need to move cybersecurity from a reactive to a more proactive approach in order to address challenges relating to the huge attack surface that is largely caused by the integration of electronic medical devices. Therefore, we iterate two main aspects of the work in this chapter. First, the key contributions of this work and secondly, the future direction for research.

6.2 Contributions

Some of the contributions of this work that finely correspond with the objectives put forward in Chapter 1 are.

A proposal for a cyber event reporting mechanism that functions as a database to enable public reporting of cyberattack incidents. Our research shows that such cyber incident databases are scarce, especially for the healthcare sector, and are also closely coupled as per the specific requirements of the users. While our paradigm, however, lays more emphasis on public accessibility and unrestricted reporting to the system. What is paramount about this specific contribution is the possibility of capturing the evolving threat landscape in terms of attack patterns and newer trends.

This study also established the applicability of AI/ML techniques in analyzing healthcare-specific cyberattack datasets. This contribution works in synergy with the

tailored risk assessment framework that establishes technical controls linked to cyber assets that are susceptible to cyberattack compromises. The first of this AI/ML application is in anomaly detection in the ECU-IoHT dataset that captures ARP spoofing, smurf, Nmap and DoS attacks. Our work perceived this to be a classification and predictive modeling problem that is particular to the healthcare sector, and we utilized decision trees, random forest, gradient boosting, and MLP to classify and predict these threats. From our experiments, the ensemble techniques of random forest and gradient boosting performed well, with the gradient boosting technique achieving an accuracy of 99.53%. Apart from what the significance of this result implies to the body of knowledge, it also highlights the potential of AI/ML in strengthening cybersecurity measures.

Another contribution of this research is the application of Support Vector Machines (SVMs) with radial basis function and polynomial kernel types in predicting or classifying web page phishing attack types. The prevalence of phishing attacks plaguing the healthcare sector is abundantly clear in the literature. Our results show accuracy exceeding 84.5%, which also demonstrates the practical application of AI/ML in reducing social engineering attacks, which in most cases are often used as a springboard for other devasting attack types. This contribution highlights that tailored AI/ML models can effectively detect phishing activities, reducing vulnerabilities to such attacks in healthcare environments.

Another key contribution in the domain of Network Intrusion Detection Systems (NIDS) that is aimed at tackling the problem of ML model generalizability is the proposal of a Cross-ML framework. The experiments were conducted by leveraging three datasets (ToN-IoT, X-IIoTID, and UNSW-NB15). This approach addresses the inherent challenge posed by heterogeneous data because of differences in context. Our method seeks to improve model generalization, and our results show F1-scores ranging from 92% to 98.1% in binary classification tasks using the decision tree model. This innovation demonstrates how the Cross-ML technique can effectively adapt NIDS to diverse real-world scenarios.

Another notable contribution of this research relates to the introduction of risk assessment schemes within the framework of the multilevel model in perspective. Although other contributions may have diverse areas of application, this contribution applies specifically to Baze University Hospital (BUH) as a typical healthcare facility. Therefore, the proposed BUHCRAF model we proposed that streamlined key components of the NIST and ISO 27000 series allowed us to make specific recommendations for volatile cyber assets that can be adapted to other contexts with minimal modifications to the framework.

6.3 Future work

While this thesis has made significant progress in articulating computing paradigms that address security challenges in the healthcare sector, we acknowledge the necessity for further research just like in any other research work.

At the conception of our research work, we explored the application of AI/ML methods directly to cyberattack databases. But as the work progressed, we identified that our approach, although intuitive, could not effectively work in synergy with the data that is in the public domain or already existing cyber event databases. Since cyber incidents are haphazardly reported in journals, news outlets, and cybersecurity reports, future work can consider a framework that can automate the reporting of these events. Furthermore, for the known incidents documented, they do not contain usable numerical features or align with the independent and identically distributed random variables (IID), which form a large prerequisite for the conditions to be met in order to use AI/ML techniques on most data.

Although we achieved good results with the application of AI/ML techniques for cyberattack detection, we acknowledge the need for further research with other AI techniques such as deep learning, reinforcement learning, and hybrid ensemble methods. There are way too few datasets for research in the domain of cybersecurity in general and even far too little in cybersecurity of the healthcare sector, as representative data always

triggers the issue of privacy and ethical concerns as stipulated in regulations such as GDPR and HIPAA. Therefore, future research into this specific issue can be explored.

Finally, while our proposed BUHCRAF framework provides a structured approach to risk assessment that is context specific, our methodology may not be flexible enough to accommodate real-time risk monitoring to capture and analyze real-time risk profiles. Also, since our framework was adapted for a specific use case, future work can also consider a mechanism that will aggregate input data that comes from different health care facilities. Progressing the research in this direction can potentially improve data granularity and accuracy, improve dynamic and adaptive risk models, and ensure cross-sector application.

6.4 Conclusion

In summary, this thesis has laid very strong groundwork for the pressing issue of cybersecurity in the healthcare sector. The research leveraged historical data that feeds into a tailored risk assessment framework while using AI/ML methods to address common threats such as phishing, anomaly detection, and network intrusion detection. Since the nature of cyber threats is dynamic, continuous innovation and collaboration will be required. It is hoped that the present work, besides satisfying a curious academic interest in the outstanding investigative methodologies, may also become an incentive to start new research and efforts to protect the digital infrastructure of the health sector for the safety and well-being not only of patients but also of healthcare professionals.

REFERENCES

[1] Y. Li and Q. Liu, "A comprehensive review study of cyber-attacks and cyber security; Emerging trends and recent developments," *Energy Reports*, vol. 7, pp. 8176–8186, Nov. 2021, doi: 10.1016/j.egyr.2021.08.126.

[2] D. Strickland, "The Impact of Cyberattacks on Healthcare," CurrentWare. Accessed: Nov. 11, 2022. [Online]. Available: https://www.currentware.com/blog/the-impact-of-cyberattacks-on-healthcare/

[3] A. Banafa, "Challenges Facing Using AI in Cybersecurity," OpenMind. Accessed: Jan. 08, 2023. [Online]. Available: https://www.bbvaopenmind.com/en/technology/artificial-intelligence/challenges-facing-using-ai-in-cybersecurity/

[4] H. Saleous *et al.*, "COVID-19 pandemic and the cyberthreat landscape: Research challenges and opportunities," *Digital Communications and Networks*, Jun. 2022, doi: 10.1016/j.dcan.2022.06.005.

[5] D. Milmo, "NHS ransomware attack: what happened and how bad is it?," *The Guardian*, Aug. 11, 2022. Accessed: Nov. 14, 2022. [Online]. Available: https://www.theguardian.com/technology/2022/aug/11/nhs-ransomware-attack-what-happened-and-how-bad-is-it

[6] "WannaCry ransomware attacks cost the NHS £92m," *Computer Fraud & Security*, vol. 2018, no. 11, pp. 1–3, Nov. 2018, doi: 10.1016/S1361-3723(18)30102-7.

[7] "Oxford Health: Cyber attack continues to hit NHS trust's services," *BBC News*, Sep. 27, 2022. Accessed: Nov. 16, 2022. [Online]. Available: https://www.bbc.com/news/uk-england-oxfordshire-63046401

[8] H. N. Security, "Healthcare organizations at risk: The attack surface is expanding," Help Net Security. Accessed: Nov. 16, 2022. [Online]. Available: https://www.helpnetsecurity.com/2021/11/15/healthcare-organizations-attack-surface/

[9] C. Gregersen, "Connected medical devices brought security loopholes mainstream," Help Net Security. Accessed: Nov. 16, 2022. [Online]. Available: https://www.helpnetsecurity.com/2021/04/26/connected-medical-devices-security/

[10] T. A. Mattei, "Privacy, Confidentiality, and Security of Health Care Information: Lessons from the Recent WannaCry Cyberattack," *World Neurosurgery*, vol. 104, pp. 972–974, Aug. 2017, doi: 10.1016/j.wneu.2017.06.104.

[11] 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, Nov. 2021, doi: 10.1016/j.adhoc.2021.102621.

[12] A. B. Arrieta *et al.*, "Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges toward Responsible AI," Dec. 26, 2019, *arXiv*: arXiv:1910.10045. Accessed: Jan. 09, 2023. [Online]. Available: http://arxiv.org/abs/1910.10045

[13] B. Thuraisingham, "Cyber Security and Artificial Intelligence for Cloud-based Internet of Transportation Systems," in *2020 7th IEEE International Conference on Cyber Security and Cloud Computing (CSCloud)/2020 6th IEEE International Conference on Edge Computing and Scalable Cloud (EdgeCom)*, Aug. 2020, pp. 8–10. doi: 10.1109/CSCloud-EdgeCom49738.2020.00011.

[14] L. Salazar, W. D. Parreira, A. Fernandes, and V. Leithardt, "No-Show in Medical Appointments with Machine Learning Techniques: A Systematic Literature Review," *Information*, vol. 13, p. 507, Oct. 2022, doi: 10.3390/info13110507.
[15] A. Moharram, S. Altamimi, and R. Alshammari, "Data Analytics and Predictive Modeling for Appointments No-show at a Tertiary Care Hospital," in *2021 1st International Conference on Artificial Intelligence and Data Analytics (CAIDA)*, Apr. 2021, pp. 275–277. doi: 10.1109/CAIDA51941.2021.9425258.
[16] I. H. Sarker, "Machine Learning: Algorithms, Real-World Applications and Research Directions," *SN COMPUT. SCI.*, vol. 2, no. 3, p. 160, Mar. 2021, doi: 10.1007/s42979-021-00592-x.
[17] B. Ślusarczyk, "Industry 4.0 – Are We Ready?," *Polish Journal of Management Studies*, vol. 17, no. 1, pp. 232–248, 2018, doi: 10.17512/pjms.2018.17.1.19.
[18] R. Jafari-Marandi, "Supervised or unsupervised learning? Investigating the role of pattern recognition assumptions in the success of binary predictive prescriptions," *Neurocomputing*, vol. 434, pp. 165–193, Apr. 2021, doi: 10.1016/j.neucom.2020.12.063.
[19] A. L. Buczak and E. Guven, "A Survey of Data Mining and Machine Learning Methods for Cyber Security Intrusion Detection," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 2, pp. 1153–1176, 2016, doi: 10.1109/COMST.2015.2494502.
[20] I. H. Sarker, A. S. M. Kayes, S. Badsha, H. Alqahtani, P. Watters, and A. Ng, "Cybersecurity data science: an overview from machine learning perspective," *J Big Data*, vol. 7, no. 1, p. 41, Jul. 2020, doi: 10.1186/s40537-020-00318-5.
[21] I. H. Sarker, Y. B. Abushark, F. Alsolami, and A. I. Khan, "IntruDTree: A Machine Learning Based Cyber Security Intrusion Detection Model," *Symmetry*, vol. 12, no. 5, Art. no. 5, May 2020, doi: 10.3390/sym12050754.
[22] J. Han, J. Pei, and H. Tong, *Data Mining: Concepts and Techniques*. Morgan Kaufmann, 2022.
[23] J. Chigada and R. Madzinga, "Cyberattacks and threats during COVID-19: A systematic literature review," *SA Journal of Information Management*, vol. 23, no. 1, Art. no. 1, Feb. 2021.
[24] L. Coventry and D. Branley, "Cybersecurity in healthcare: A narrative review of trends, threats and ways forward," *Maturitas*, vol. 113, pp. 48–52, Jul. 2018, doi: 10.1016/j.maturitas.2018.04.008.
[25] M. Bansal, V. Sirpal, T. Jain, Ujjwal, and U. Nangia, "Cybersecurity challenges in the healthcare domain and theroad ahead," presented at the AIP Conference Proceedings, 2023. doi: 10.1063/5.0149167.
[26] M. Lehto, P. Neittaanmäki, J. Pöyhönen, and A. Hummelholm, "Cyber Security in Healthcare Systems," in *Cyber Security: Critical Infrastructure Protection*, M. Lehto and P. Neittaanmäki, Eds., Cham: Springer International Publishing, 2022, pp. 183–215. doi: 10.1007/978-3-030-91293-2_8.
[27] "Why Hospital Cyber Protection Is a Hard Nut to Crack | GREYCORTEX." Accessed: Nov. 22, 2022. [Online]. Available: https://www.greycortex.com/blog/why-hospital-cyber-protection-hard-nut-crack
[28] S. Sharon, "What is Cybersecurity? Everything You Need to Know | TechTarget," SearchSecurity. Accessed: Nov. 23, 2022. [Online]. Available: https://www.techtarget.com/searchsecurity/definition/cybersecurity

[29] C. Hughes, “How to Mitigate Primary and Secondary Impacts in Incident Response,” Acceleration Economy. Accessed: Aug. 22, 2024. [Online]. Available: https://accelerationeconomy.com/cybersecurity/how-to-mitigate-primary-and-secondary-impacts-in-incident-response/
[30] “The Effects of Cyber-Attacks | Cybersecurity Threats.” Accessed: Aug. 22, 2024. [Online]. Available: https://www.securiwiser.com/blog/the-effects-of-cyber-attacks/
[31] Ö. A. Aslan and R. Samet, “A Comprehensive Review on Malware Detection Approaches,” *IEEE Access*, vol. 8, pp. 6249–6271, 2020, doi: 10.1109/ACCESS.2019.2963724.
[32] O. Or-Meir, N. Nissim, Y. Elovici, and L. Rokach, “Dynamic Malware Analysis in the Modern Era—A State of the Art Survey,” *ACM Comput. Surv.*, vol. 52, no. 5, p. 88:1-88:48, Sep. 2019, doi: 10.1145/3329786.
[33] Y. Ye, T. Li, D. Adjeroh, and S. S. Iyengar, “A Survey on Malware Detection Using Data Mining Techniques,” *ACM Comput. Surv.*, vol. 50, no. 3, p. 41:1-41:40, Jun. 2017, doi: 10.1145/3073559.
[34] M. Alawida, A. E. Omolara, O. I. Abiodun, and M. Al-Rajab, “A deeper look into cybersecurity issues in the wake of Covid-19: A survey,” *Journal of King Saud University - Computer and Information Sciences*, vol. 34, no. 10, Part A, pp. 8176–8206, Nov. 2022, doi: 10.1016/j.jksuci.2022.08.003.
[35] E. S. Shombot, G. Dusserre, R. Bestak, and N. B. Ahmed, “An application for predicting phishing attacks: A case of implementing a support vector machine learning model,” *Cyber Security and Applications*, vol. 2, p. 100036, Jan. 2024, doi: 10.1016/j.csa.2024.100036.
[36] S. Chattopadhyay, A. K. Sahoo, and S. Jasola, “An Enhanced DDoS Attack Detection in Software-Defined-Networks using Ensemble Learning,” *SN COMPUT. SCI.*, vol. 5, no. 5, p. 580, May 2024, doi: 10.1007/s42979-024-02938-7.
[37] L. F. Eliyan and R. Di Pietro, “DoS and DDoS attacks in Software Defined Networks: A survey of existing solutions and research challenges,” *Future Generation Computer Systems*, vol. 122, pp. 149–171, Sep. 2021, doi: 10.1016/j.future.2021.03.011.
[38] P. O’Kane, S. Sezer, and D. Carlin, “Evolution of ransomware,” *IET Networks*, vol. 7, no. 5, pp. 321–327, 2018, doi: 10.1049/iet-net.2017.0207.
[39] “Know the types of cyber threats | Mass.gov.” Accessed: Aug. 22, 2024. [Online]. Available: https://www.mass.gov/info-details/know-the-types-of-cyber-threats
[40] “Detection and prediction of insider threats to cyber security: a systematic literature review and meta-analysis | Big Data Analytics.” Accessed: Jan. 20, 2025. [Online]. Available: https://link.springer.com/article/10.1186/s41044-016-0006-0
[41] L. Liu, O. De Vel, Q.-L. Han, J. Zhang, and Y. Xiang, “Detecting and Preventing Cyber Insider Threats: A Survey,” *IEEE Communications Surveys & Tutorials*, vol. 20, no. 2, pp. 1397–1417, 2018, doi: 10.1109/COMST.2018.2800740.
[42] “Impact and Key Challenges of Insider Threats on Organizations and Critical Businesses.” Accessed: Jan. 20, 2025. [Online]. Available: https://www.mdpi.com/2079-9292/9/9/1460
[43] “Healthcare Cybersecurity: The Biggest Stats and Trends in 2024,” SafetyDetectives. Accessed: Aug. 22, 2024. [Online]. Available: https://www.safetydetectives.com/blog/healthcare-cybersecurity-statistics/
[44] C. Thyagarajan, S. Suresh, N. Sathish, and D. S. Suthir, “A Typical Analysis And Survey On Healthcare Cyber Security,” vol. 9, no. 03, 2020.

[45] S. Kakarla, D. N. Rao, G. Kakarla, and S. Gorla, “Statistical Trend in Cyber Attacks and Security Measures,” in *Computational Intelligent Security in Wireless Communications*, CRC Press, 2023.
[46] S. A. E. Hoffman, “Cybersecurity Threats in Healthcare Organizations: : Exposing Vulnerabilities in the Healthcare Information Infrastructure,” *World Libraries*, vol. 24, no. 1, Art. no. 1, Jul. 2020, Accessed: Aug. 22, 2024. [Online]. Available: https://worldlibraries.dom.edu/index.php/worldlib/article/view/588
[47] J. Pool, S. Akhlaghpour, F. Fatehi, and A. Burton-Jones, “A systematic analysis of failures in protecting personal health data: A scoping review,” *International Journal of Information Management*, vol. 74, p. 102719, Feb. 2024, doi: 10.1016/j.ijinfomgt.2023.102719.
[48] S. Ksibi, F. Jaidi, and A. Bouhoula, “A Comprehensive Study of Security and Cyber-Security Risk Management within e-Health Systems: Synthesis, Analysis and a Novel Quantified Approach,” *Mobile Netw Appl*, vol. 28, no. 1, pp. 107–127, Feb. 2023, doi: 10.1007/s11036-022-02042-1.
[49] L. Wasserman and Y. Wasserman, “Hospital cybersecurity risks and gaps: Review (for the non-cyber professional),” *Frontiers in Digital Health*, vol. 4, 2022, Accessed: Jan. 05, 2023. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fdgth.2022.862221
[50] A. H. Seh *et al.*, “Healthcare Data Breaches: Insights and Implications,” *Healthcare (Basel)*, vol. 8, no. 2, p. 133, May 2020, doi: 10.3390/healthcare8020133.
[51] S. T. Argaw, N.-E. Bempong, B. Eshaya-Chauvin, and A. Flahault, “The state of research on cyberattacks against hospitals and available best practice recommendations: a scoping review,” *BMC Medical Informatics and Decision Making*, vol. 19, no. 1, p. 10, Jan. 2019, doi: 10.1186/s12911-018-0724-5.
[52] M. Simon and V. Looten, “Description of Data Breaches Notifications in France and Lessons Learned for the Healthcare Stakeholders,” in *Studies in health technology and informatics*, vol. 275, 2020. doi: 10.3233/SHTI200721.
[53] C. S. Kruse, B. Frederick, T. Jacobson, and D. K. Monticone, “Cybersecurity in healthcare: A systematic review of modern threats and trends,” *THC*, vol. 25, no. 1, pp. 1–10, Feb. 2017, doi: 10.3233/THC-161263.
[54] M. Muthuppalaniappan and K. Stevenson, “Healthcare cyber-attacks and the COVID-19 pandemic: an urgent threat to global health,” *International Journal for Quality in Health Care*, vol. 33, no. 1, p. mzaa117, Feb. 2021, doi: 10.1093/intqhc/mzaa117.
[55] E. G. Spanakis *et al.*, “Cyber-attacks and threats for healthcare – a multi-layer thread analysis,” in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, Jul. 2020, pp. 5705–5708. doi: 10.1109/EMBC44109.2020.9176698.
[56] A. Bendovschi, “Cyber-Attacks – Trends, Patterns and Security Countermeasures,” *Procedia Economics and Finance*, vol. 28, pp. 24–31, Jan. 2015, doi: 10.1016/S2212-5671(15)01077-1.
[57] M. Barad, “Linking Cyber Security Improvement Actions in Healthcare Systems to Their Strategic Improvement Needs,” *Procedia Manufacturing*, vol. 39, pp. 279–286, Jan. 2019, doi: 10.1016/j.promfg.2020.01.335.
[58] R. Collier, “NHS ransomware attack spreads worldwide,” *CMAJ*, vol. 189, no. 22, pp. E786–E787, Jun. 2017, doi: 10.1503/cmaj.1095434.

[59] S. Askarifar, N. A. A. Rahman, and H. Osman, “A REVIEW OF LATEST WANNACRY RANSOMWARE: ACTIONS AND PREVENTIONS,” no. 7.
[60] “Brno University Hospital in Czech Republic Suffers Cyberattack During COVID-19 Outbreak | 2020-03-17 | Security Magazine.” Accessed: Aug. 23, 2024. [Online]. Available: https://www.securitymagazine.com/articles/91921-brno-university-hospital-in-czech-republic-suffers-cyberattack-during-covid-19-outbreak
[61] “Best collaboration app 2022: Top team tools | ZDNET.” Accessed: Aug. 23, 2024. [Online]. Available: https://www.zdnet.com/article/best-collaboration-app/
[62] “Ransomware Attacks on Hospitals Have Changed | Cybersecurity | Center | AHA.” Accessed: Aug. 23, 2024. [Online]. Available: https://www.aha.org/center/cybersecurity-and-risk-advisory-services/ransomware-attacks-hospitals-have-changed
[63] “Brno University Hospital ransomware attack (2020),” International cyber law: interactive toolkit. Accessed: Aug. 23, 2024. [Online]. Available: https://cyberlaw.ccdcoe.org/wiki/Brno_University_Hospital_ransomware_attack_(2020)
[64] “Kaiser Permanente Website Tracker Breach Affects 13.4 Million Individuals.” Accessed: Aug. 23, 2024. [Online]. Available: https://www.hipaajournal.com/kaiser-permanente-website-tracker-breach-affects-13-4-million-individuals/
[65] S. Alder, “Healthcare Data Breach Statistics,” The HIPAA Journal. Accessed: Aug. 23, 2024. [Online]. Available: https://www.hipaajournal.com/healthcare-data-breach-statistics/
[66] S. Alder, “11.27 Million HCA Healthcare Patients Affected by Recent Cyberattack,” The HIPAA Journal. Accessed: Aug. 23, 2024. [Online]. Available: https://www.hipaajournal.com/hca-healthcare-cyberattack-data-breach-2023/
[67] M. Mocydlarz-Adamcewicz, B. Bajsztok, S. Filip, J. Petera, M. Mestan, and J. Malicki, “Management of Onsite and Remote Communication in Oncology Hospitals: Data Protection in an Era of Rapid Technological Advances,” *J Pers Med*, vol. 13, no. 5, p. 761, Apr. 2023, doi: 10.3390/jpm13050761.
[68] “Health | European Data Protection Supervisor.” Accessed: Aug. 26, 2024. [Online]. Available: https://www.edps.europa.eu/data-protection/our-work/subjects/health_en
[69] E. Guardelli, *Cybersecurity in Healthcare : Protecting Critical Data and Patients*. MedTechBiz, 2024.
[70] A. M. Ibrahim *et al.*, “Balancing confidentiality and care coordination: challenges in patient privacy,” *BMC Nurs*, vol. 23, no. 1, p. 564, Aug. 2024, doi: 10.1186/s12912-024-02231-1.
[71] D. McGraw and K. D. Mandl, “Privacy protections to encourage use of health-relevant digital data in a learning health system,” *NPJ Digit Med*, vol. 4, p. 2, Jan. 2021, doi: 10.1038/s41746-020-00362-8.
[72] H. Mumtaz *et al.*, “Current challenges and potential solutions to the use of digital health technologies in evidence generation: a narrative review,” *Front. Digit. Health*, vol. 5, Sep. 2023, doi: 10.3389/fdgth.2023.1203945.
[73] “Nursing Care Coordination for Patients with Complex Needs in Primary Healthcare: A Scoping Review | International Journal of Integrated Care.” Accessed: Aug. 26, 2024. [Online]. Available: https://ijic.org/articles/10.5334/ijic.5518
[74] T. Muhammad, M. T. Munir, M. Z. Munir, and M. Zafar, “Integrative Cybersecurity: Merging Zero Trust, Layered Defense, and Global Standards for a Resilient Digital Future,” vol. 1, p. 35, Jan. 2017.

[75] "A multi-layered security architecture for modelling complex systems | Proceedings of the 4th annual workshop on Cyber security and information intelligence research: developing strategies to meet the cyber security and information intelligence challenges ahead." Accessed: Aug. 26, 2024. [Online]. Available: https://dl.acm.org/doi/abs/10.1145/1413140.1413180
[76] N. Hewitt, "The Strategic Advantage of a Multi-Layered Cybersecurity Approach • TrueFort," TrueFort. Accessed: Aug. 26, 2024. [Online]. Available: https://truefort.com/multi-layered-cybersecurity/
[77] "Multi-Layered Security Approach: What Is Defense-in-Depth?" Accessed: Aug. 26, 2024. [Online]. Available: https://www.druva.com/glossary/multi-layered-security
[78] "What Is Defense in Depth? An Intro to Multi-Layered Security." Accessed: Aug. 26, 2024. [Online]. Available: https://prolion.com/blog/defense-in-depth-the-importance-of-multi-layered-security/
[79] E. Wai and C. K. M. Lee, "Seamless Industry 4.0 Integration: A Multilayered Cyber-Security Framework for Resilient SCADA Deployments in CPPS," *Applied Sciences*, vol. 13, no. 21, Art. no. 21, Jan. 2023, doi: 10.3390/app132112008.
[80] J. Chen *et al.*, "A Multi-Layer Security Scheme for Mitigating Smart Grid Vulnerability against Faults and Cyber-Attacks," *Applied Sciences*, vol. 11, no. 21, Art. no. 21, Jan. 2021, doi: 10.3390/app11219972.
[81] F. Zhang, H. A. D. E. Kodituwakku, J. W. Hines, and J. Coble, "Multilayer Data-Driven Cyber-Attack Detection System for Industrial Control Systems Based on Network, System, and Process Data," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 7, pp. 4362–4369, Jul. 2019, doi: 10.1109/TII.2019.2891261.
[82] S. Sahoo, T. Dragičević, and F. Blaabjerg, "Multilayer Resilience Paradigm Against Cyber Attacks in DC Microgrids," *IEEE Transactions on Power Electronics*, vol. 36, no. 3, pp. 2522–2532, Mar. 2021, doi: 10.1109/TPEL.2020.3014258.
[83] A. Hahn, R. K. Thomas, I. Lozano, and A. Cardenas, "A multi-layered and kill-chain based security analysis framework for cyber-physical systems," *International Journal of Critical Infrastructure Protection*, vol. 11, pp. 39–50, Dec. 2015, doi: 10.1016/j.ijcip.2015.08.003.
[84] "Defense in Depth - an overview | ScienceDirect Topics." Accessed: Sep. 02, 2024. [Online]. Available: https://www.sciencedirect.com/topics/computer-science/defense-in-depth
[85] K. I. Gubbi *et al.*, "Optimized and Automated Secure IC Design Flow: A Defense-in-Depth Approach," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 71, no. 5, pp. 2031–2044, May 2024, doi: 10.1109/TCSI.2024.3364160.
[86] S. Rose, O. Borchert, S. Mitchell, and S. Connelly, "Zero Trust Architecture," National Institute of Standards and Technology, Aug. 2020. doi: 10.6028/NIST.SP.800-207.
[87] "A zero trust architecture for health information systems | Health and Technology." Accessed: Sep. 02, 2024. [Online]. Available: https://link.springer.com/article/10.1007/s12553-023-00809-4
[88] "Zero Trust Architecture (ZTA): A Comprehensive Survey | IEEE Journals & Magazine | IEEE Xplore." Accessed: Sep. 02, 2024. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9773102

[89] “Flexible zero trust architecture for the cybersecurity of industrial IoT infrastructures - ScienceDirect.” Accessed: Sep. 02, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1570870524000258
[90] “A critical analysis of Zero Trust Architecture (ZTA) - ScienceDirect.” Accessed: Sep. 02, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0920548924000011
[91] “A Survey on Zero Trust Architecture: Challenges and Future Trends - He - 2022 - Wireless Communications and Mobile Computing - Wiley Online Library.” Accessed: Sep. 02, 2024. [Online]. Available: https://onlinelibrary.wiley.com/doi/full/10.1155/2022/6476274
[92] “Migrating to Zero Trust Architecture: Reviews and Challenges - Teerakanok - 2021 - Security and Communication Networks - Wiley Online Library.” Accessed: Sep. 02, 2024. [Online]. Available: https://onlinelibrary.wiley.com/doi/full/10.1155/2021/9947347
[93] “Zero Trust Architecture: A Brief Introduction - SSL.com.” Accessed: Sep. 02, 2024. [Online]. Available: https://www.ssl.com/blogs/zero-trust-architecture-a-brief-introduction/
[94] E. John, *Database Technologies: Concepts, Methodologies, Tools, and Applications: Concepts, Methodologies, Tools, and Applications*. IGI Global, 2009.
[95] B. A, “Evaluating Database Security and Cyber Attacks: A Relational Approach,” *J Internet Bank Commer*, vol. 20, no. 2, 2015, doi: 10.4172/1204-5357.1000115.
[96] S. M. Toapanta, O. A. Escalante Quimis, L. E. M. Gallegos, and M. R. Maciel Arellano, “Analysis for the Evaluation and Security Management of a Database in a Public Organization to Mitigate Cyber Attacks,” *IEEE Access*, vol. 8, pp. 169367–169384, 2020, doi: 10.1109/ACCESS.2020.3022746.
[97] M. Ishaq *et al.*, “Advances in database systems education: Methods, tools, curricula, and way forward,” *Educ Inf Technol*, vol. 28, no. 3, pp. 2681–2725, Mar. 2023, doi: 10.1007/s10639-022-11293-0.
[98] S. McGregor, “Preventing Repeated Real World AI Failures by Cataloging Incidents: The AI Incident Database,” Nov. 17, 2020, *arXiv*: arXiv:2011.08512. doi: 10.48550/arXiv.2011.08512.
[99] “Federal Aviation Administration.” Accessed: Jun. 23, 2023. [Online]. Available: https://www.faa.gov/
[100] “ASRS Report Sets - Aviation Safety Reporting System.” Accessed: Jun. 22, 2023. [Online]. Available: https://asrs.arc.nasa.gov/search/reportsets.html
[101] “AIN Summary Page.” Accessed: Jun. 23, 2023. [Online]. Available: https://www.asias.faa.gov/apex/f?p=100:93:::NO:::
[102] “NASA,” NASA. Accessed: Jun. 23, 2023. [Online]. Available: http://www.nasa.gov/index.html
[103] C. Jr, J. Strong, and C. Zorn, “Analyzing aviation safety: Problems, challenges, opportunities,” *Research in Transportation Economics*, vol. 43, pp. 148–164, Jul. 2013, doi: 10.1016/j.retrec.2012.12.001.
[104] “IATA Releases 2022 Airline Safety Performance.” Accessed: Jun. 23, 2023. [Online]. Available: https://www.iata.org/en/pressroom/2023-releases/2023-03-07-01/
[105] K. Oosthoek and C. Doerr, “SoK: ATT&CK Techniques and Trends in Windows Malware,” in *Security and Privacy in Communication Networks*, vol. 304, S. Chen, K.-K. R. Choo, X. Fu, W. Lou, and A. Mohaisen, Eds., in Lecture Notes of the Institute for

Computer Sciences, Social Informatics and Telecommunications Engineering, vol. 304. , Cham: Springer International Publishing, 2019, pp. 406–425. doi: 10.1007/978-3-030-37228-6_20.
[106] "ATTACK_Design_and_Philosophy_March_2020.pdf." Accessed: Jun. 23, 2023. [Online]. Available: https://attack.mitre.org/docs/ATTACK_Design_and_Philosophy_March_2020.pdf
[107] R. Al-Shaer, J. M. Spring, and E. Christou, "Learning the Associations of MITRE ATT&CK Adversarial Techniques," May 12, 2020, *arXiv*: arXiv:2005.01654. Accessed: Jun. 23, 2023. [Online]. Available: http://arxiv.org/abs/2005.01654
[108] S. S. Bagui *et al.*, "Introducing UWF-ZeekData22: A Comprehensive Network Traffic Dataset Based on the MITRE ATT&CK Framework," *Data*, vol. 8, no. 1, Art. no. 1, Jan. 2023, doi: 10.3390/data8010018.
[109] "MITRE ATT&CK®." Accessed: Jun. 23, 2023. [Online]. Available: https://attack.mitre.org/
[110] S. Roy, E. Panaousis, C. Noakes, A. Laszka, S. Panda, and G. Loukas, "SoK: The MITRE ATT&CK Framework in Research and Practice," Apr. 14, 2023, *arXiv*: arXiv:2304.07411. Accessed: Jun. 23, 2023. [Online]. Available: http://arxiv.org/abs/2304.07411
[111] O. of the Commissioner, "U.S. Food and Drug Administration," FDA. Accessed: Jun. 23, 2023. [Online]. Available: https://www.fda.gov/home
[112] "FDA Adverse Events Reporting System (FAERS) Public Dashboard - FDA Adverse Events Reporting System (FAERS) Public Dashboard | Sheet - Qlik Sense." Accessed: Jun. 23, 2023. [Online]. Available: https://fis.fda.gov/sense/app/95239e26-e0be-42d9-a960-9a5f7f1c25ee/sheet/7a47a261-d58b-4203-a8aa-6d3021737452/state/analysis
[113] H. Fang *et al.*, "Exploring the FDA Adverse Event Reporting System (FAERS) to Generate Hypotheses for Disease Monitoring," *Clin Pharmacol Ther*, vol. 95, no. 5, pp. 496–498, May 2014, doi: 10.1038/clpt.2014.17.
[114] A. Alsuhibani, M. Alrasheed, M. Gari, A. L. Hincapie, and J. J. Guo, "Descriptive analysis of reported adverse events associated with anti-obesity medications using FDA Adverse Event Reporting System (FAERS) databases 2013–2020," *Int J Clin Pharm*, vol. 44, no. 1, pp. 172–179, Feb. 2022, doi: 10.1007/s11096-021-01330-2.
[115] Information and Operations Management Department Texas A&M University, R. Sen, G. Heim, Information and Operations Management Department Texas A&M University, Q. Zhu, and Information and Operations Management Department Texas A&M University, "Artificial Intelligence and Machine Learning in Cybersecurity: Applications, Challenges, and Opportunities for MIS Academics," *CAIS*, vol. 51, no. 1, pp. 179–209, 2022, doi: 10.17705/1CAIS.05109.
[116] E. Bout, V. Loscri, and A. Gallais, "How Machine Learning Changes the Nature of Cyberattacks on IoT Networks: A Survey," *IEEE Communications Surveys & Tutorials*, vol. 24, no. 1, pp. 248–279, 2022, doi: 10.1109/COMST.2021.3127267.
[117] S. O. Uwagbole, W. J. Buchanan, and L. Fan, "Applied Machine Learning predictive analytics to SQL Injection Attack detection and prevention," in *2017 IFIP/IEEE Symposium on Integrated Network and Service Management (IM)*, May 2017, pp. 1087–1090. doi: 10.23919/INM.2017.7987433.
[118] J. Gardiner and S. Nagaraja, "On the Security of Machine Learning in Malware C&C Detection: A Survey," *ACM Comput. Surv.*, vol. 49, no. 3, p. 59:1-59:39, Dec. 2016, doi: 10.1145/3003816.

[119] N. Alzahrani and D. Alghazzawi, “A Review on Android Ransomware Detection Using Deep Learning Techniques,” in *Proceedings of the 11th International Conference on Management of Digital EcoSystems*, in MEDES ’19. New York, NY, USA: Association for Computing Machinery, Jan. 2020, pp. 330–335. doi: 10.1145/3297662.3365785.
[120] T. Gangavarapu, C. D. Jaidhar, and B. Chanduka, “Applicability of machine learning in spam and phishing email filtering: review and approaches,” *Artif Intell Rev*, vol. 53, no. 7, pp. 5019–5081, Oct. 2020, doi: 10.1007/s10462-020-09814-9.
[121] “A Comprehensive Survey for Intelligent Spam Email Detection | IEEE Journals & Magazine | IEEE Xplore.” Accessed: Sep. 08, 2024. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/8907831
[122] Y. Xin *et al.*, “Machine Learning and Deep Learning Methods for Cybersecurity,” *IEEE Access*, vol. 6, pp. 35365–35381, 2018, doi: 10.1109/ACCESS.2018.2836950.
[123] G. Apruzzese *et al.*, “The Role of Machine Learning in Cybersecurity,” *Digital Threats*, vol. 4, no. 1, p. 8:1-8:38, Mar. 2023, doi: 10.1145/3545574.
[124] R. Jayaraj, A. Pushpalatha, K. Sangeetha, T. Kamaleshwar, S. Udhaya Shree, and D. Damodaran, “Intrusion detection based on phishing detection with machine learning,” *Measurement: Sensors*, vol. 31, p. 101003, Feb. 2024, doi: 10.1016/j.measen.2023.101003.
[125] “The 5 Biggest Phishing Scams of All Time - IT Governance Blog En.” Accessed: Sep. 08, 2024. [Online]. Available: https://www.itgovernance.eu/blog/en/the-5-biggest-phishing-scams-of-all-time
[126] F. Feng, Q. Zhou, Z. Shen, X. Yang, L. Han, and J. Wang, “The application of a novel neural network in the detection of phishing websites,” *J Ambient Intell Human Comput*, vol. 15, no. 3, pp. 1865–1879, Mar. 2024, doi: 10.1007/s12652-018-0786-3.
[127] “[2408.01667] Automated Phishing Detection Using URLs and Webpages.” Accessed: Sep. 08, 2024. [Online]. Available: https://arxiv.org/abs/2408.01667
[128] R. J. van Geest, G. Cascavilla, J. Hulstijn, and N. Zannone, “The applicability of a hybrid framework for automated phishing detection,” *Computers & Security*, vol. 139, p. 103736, Apr. 2024, doi: 10.1016/j.cose.2024.103736.
[129] N. Altwaijry, I. Al-Turaiki, R. Alotaibi, and F. Alakeel, “Advancing Phishing Email Detection: A Comparative Study of Deep Learning Models,” *Sensors*, vol. 24, no. 7, Art. no. 7, Jan. 2024, doi: 10.3390/s24072077.
[130] “Curated Datasets and Feature Analysis for Phishing Email Detection with Machine Learning | IEEE Conference Publication | IEEE Xplore.” Accessed: Sep. 08, 2024. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/10585821
[131] M. Somesha and A. R. Pais, “DeepEPhishNet: a deep learning framework for email phishing detection using word embedding algorithms,” *Sādhanā*, vol. 49, no. 3, p. 212, Jul. 2024, doi: 10.1007/s12046-024-02538-4.
[132] Y. Han and Y. Shen, “Accurate spear phishing campaign attribution and early detection,” in *Proceedings of the 31st Annual ACM Symposium on Applied Computing*, Pisa Italy: ACM, Apr. 2016, pp. 2079–2086. doi: 10.1145/2851613.2851801.
[133] J. James, S. L., and C. Thomas, “Detection of phishing URLs using machine learning techniques,” Dec. 2013, pp. 304–309. doi: 10.1109/ICCC.2013.6731669.
[134] C. Zhang, D. Jia, L. Wang, W. Wang, F. Liu, and A. Yang, “Comparative research on network intrusion detection methods based on machine learning,” *Computers & Security*, vol. 121, p. 102861, Oct. 2022, doi: 10.1016/j.cose.2022.102861.

[135] “Network intrusion detection system: A systematic study of machine learning and deep learning approaches - Ahmad - 2021 - Transactions on Emerging Telecommunications Technologies - Wiley Online Library.” Accessed: Sep. 10, 2024. [Online]. Available: https://onlinelibrary.wiley.com/doi/full/10.1002/ett.4150
[136] C. Sinclair, L. Pierce, and S. Matzner, “An application of machine learning to network intrusion detection,” in *Proceedings 15th Annual Computer Security Applications Conference (ACSAC’99)*, Dec. 1999, pp. 371–377. doi: 10.1109/CSAC.1999.816048.
[137] R. Sommer and V. Paxson, “Outside the Closed World: On Using Machine Learning for Network Intrusion Detection,” in *2010 IEEE Symposium on Security and Privacy*, May 2010, pp. 305–316. doi: 10.1109/SP.2010.25.
[138] M. Zaman and C.-H. Lung, “Evaluation of machine learning techniques for network intrusion detection,” in *NOMS 2018 - 2018 IEEE/IFIP Network Operations and Management Symposium*, Apr. 2018, pp. 1–5. doi: 10.1109/NOMS.2018.8406212.
[139] S. K. Rajwar, Dr. I. Mukherjee, and Dr. P. K. Manjhi, “Machine Learning Methods for Network Intrusion Detection,” *SSRN Journal*, 2020, doi: 10.2139/ssrn.3610618.
[140] H. Liu and B. Lang, “Machine Learning and Deep Learning Methods for Intrusion Detection Systems: A Survey,” *Applied Sciences*, vol. 9, no. 20, Art. no. 20, Jan. 2019, doi: 10.3390/app9204396.
[141] Y. K. Ever, B. Sekeroglu, and K. Dimililer, “Classification Analysis of Intrusion Detection on NSL-KDD Using Machine Learning Algorithms,” in *Mobile Web and Intelligent Information Systems*, I. Awan, M. Younas, P. Ünal, and M. Aleksy, Eds., Cham: Springer International Publishing, 2019, pp. 111–122. doi: 10.1007/978-3-030-27192-3_9.
[142] J. Li, Z. Zhao, R. Li, and H. Zhang, “AI-based Two-Stage Intrusion Detection for Software Defined IoT Networks,” Jun. 07, 2018, *arXiv*: arXiv:1806.02566. doi: 10.48550/arXiv.1806.02566.
[143] “Zero-Day Malware Classification and Detection Using Machine Learning | SN Computer Science.” Accessed: Sep. 10, 2024. [Online]. Available: https://link.springer.com/article/10.1007/s42979-023-02404-w#Sec11
[144] “EvadeDroid: A practical evasion attack on machine learning for black-box Android malware detection - ScienceDirect.” Accessed: Sep. 10, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167404823005862
[145] “Automated machine learning for deep learning based malware detection - ScienceDirect.” Accessed: Sep. 10, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167404823004923
[146] “Android malware detection using time-aware machine learning approach | Cluster Computing.” Accessed: Sep. 10, 2024. [Online]. Available: https://link.springer.com/article/10.1007/s10586-024-04484-6#Sec1
[147] “PermDroid a framework developed using proposed feature selection approach and machine learning techniques for Android malware detection | Scientific Reports.” Accessed: Sep. 10, 2024. [Online]. Available: https://link.springer.com/article/10.1038/s41598-024-60982-y
[148] “AndyWar: an intelligent android malware detection using machine learning | Innovations in Systems and Software Engineering.” Accessed: Sep. 10, 2024. [Online]. Available: https://link.springer.com/article/10.1007/s11334-023-00530-5

[149] “Machine learning aided malware detection for secure and smart manufacturing: a comprehensive analysis of the state of the art | International Journal on Interactive Design and Manufacturing (IJIDeM).” Accessed: Sep. 10, 2024. [Online]. Available: https://link.springer.com/article/10.1007/s12008-023-01578-0
[150] “Enhanced detection of obfuscated malware in memory dumps: a machine learning approach for advanced cybersecurity | Cybersecurity.” Accessed: Sep. 10, 2024. [Online]. Available: https://link.springer.com/article/10.1186/s42400-024-00205-z
[151] X. Liu, Y. Lin, H. Li, and J. Zhang, “A novel method for malware detection on ML-based visualization technique,” *Computers & Security*, vol. 89, p. 101682, Feb. 2020, doi: 10.1016/j.cose.2019.101682.
[152] V. Bolbot, G. Theotokatos, E. Boulougouris, and D. Vassalos, “A novel cyber-risk assessment method for ship systems,” *Safety Science*, vol. 131, p. 104908, Nov. 2020, doi: 10.1016/j.ssci.2020.104908.
[153] I. Meriah and L. B. Arfa Rabai, “Comparative Study of Ontologies Based ISO 27000 Series Security Standards,” *Procedia Computer Science*, vol. 160, pp. 85–92, Jan. 2019, doi: 10.1016/j.procs.2019.09.447.
[154] M. Toussaint, S. Krima, and H. Panetto, “Industry 4.0 data security: A cybersecurity frameworks review,” *Journal of Industrial Information Integration*, vol. 39, p. 100604, May 2024, doi: 10.1016/j.jii.2024.100604.
[155] “Compendium of Risk Management Frameworks with Potential Interoperability,” ENISA. Accessed: Jul. 11, 2024. [Online]. Available: https://www.enisa.europa.eu/publications/compendium-of-risk-management-frameworks
[156] M. Ekstedt, Z. Afzal, P. Mukherjee, S. Hacks, and R. Lagerström, “Yet another cybersecurity risk assessment framework,” *Int. J. Inf. Secur.*, vol. 22, no. 6, pp. 1713–1729, Dec. 2023, doi: 10.1007/s10207-023-00713-y.
[157] S. Lopes, P. Leite, S. Carvalho, and P. Teixeira, “Using ITIL as part of the NIST Cybersecurity Framework,” in *2024 12th International Symposium on Digital Forensics and Security (ISDFS)*, Apr. 2024, pp. 1–6. doi: 10.1109/ISDFS60797.2024.10527256.
[158] M. Alshar’e, “CYBER SECURITY FRAMEWORK SELECTION: COMPARISION OF NIST AND ISO27001,” *Applied computing Journal*, pp. 245–255, Feb. 2023, doi: 10.52098/acj.202364.
[159] “ISO - ISO/IEC 27000 family — Information security management,” ISO. Accessed: Jul. 11, 2024. [Online]. Available: https://www.iso.org/standard/iso-iec-27000-family
[160] F. Kitsios, E. Chatzidimitriou, and M. Kamariotou, “The ISO/IEC 27001 Information Security Management Standard: How to Extract Value from Data in the IT Sector,” *Sustainability*, vol. 15, no. 7, Art. no. 7, Jan. 2023, doi: 10.3390/su15075828.
[161] F. Alazzawi, “Assessment of Information Security Risk Management System based on ISOIEC27005 in the Independent High Electoral Commission,” *Review of International Geographical Education Online*, vol. 11, pp. 4633–4656, Aug. 2021, doi: 10.48047/rigeo.11.05.339.
[162] “ISO 27005 | IT Governance UK.” Accessed: Jul. 12, 2024. [Online]. Available: https://itgovernance.co.uk/iso27005
[163] M. E. Whitman and H. J. Mattord, *Principles of information security*, Sixth Edition. Australia: Cengage Learning, 2018.

[164] “An introduction to buildings cybersecurity framework | IEEE Conference Publication | IEEE Xplore.” Accessed: Jul. 12, 2024. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/8285228
[165] W. Wang, S. M. Sadjadi, and N. Rishe, “A Survey of Major Cybersecurity Compliance Frameworks,” in *2024 IEEE 10th Conference on Big Data Security on Cloud (BigDataSecurity)*, May 2024, pp. 23–34. doi: 10.1109/BigDataSecurity62737.2024.00013.
[166] “Integrating cost–benefit analysis into the NIST Cybersecurity Framework via the Gordon–Loeb Model | Journal of Cybersecurity | Oxford Academic.” Accessed: Jul. 12, 2024. [Online]. Available: https://academic.oup.com/cybersecurity/article/6/1/tyaa005/5813544
[167] National Institute of Standards and Technology, “Framework for Improving Critical Infrastructure Cybersecurity, Version 1.1,” National Institute of Standards and Technology, Gaithersburg, MD, NIST CSWP 04162018, Apr. 2018. doi: 10.6028/NIST.CSWP.04162018.
[168] M. Parmar and D. A. Miles, “Cyber Security Frameworks (CSFs): An Assessment Between the NIST CSF v2.0 and EU Standards”.
[169] E. Hollnagel, “How Resilient Is Your Organisation? An Introduction to the Resilience Analysis Grid (RAG),” 2010.
[170] S. Duchek, “Organizational resilience: a capability-based conceptualization,” *Bus Res*, vol. 13, no. 1, pp. 215–246, Apr. 2020, doi: 10.1007/s40685-019-0085-7.
[171] M. M. Weber, B. Pedell, and P. G. Rötzel, “Resilience-oriented management control systems: a systematic review of the relationships between organizational resilience and management control systems,” *J Manag Control*, vol. 35, no. 4, pp. 563–620, Dec. 2024, doi: 10.1007/s00187-024-00385-2.
[172] B. Rathnayaka, C. Siriwardana, D. Robert, D. Amaratunga, and S. Setunge, “Improving the resilience of critical infrastructures: Evidence-based insights from a systematic literature review,” *International Journal of Disaster Risk Reduction*, vol. 78, p. 103123, Aug. 2022, doi: 10.1016/j.ijdrr.2022.103123.
[173] J. Carthey, “Institutional resilience in healthcare systems,” *Quality in Health Care*, vol. 10, no. 1, pp. 29–32, Mar. 2001, doi: 10.1136/qhc.10.1.29.
[174] H. B. Lyng *et al.*, “Capacities for resilience in healthcare; a qualitative study across different healthcare contexts,” *BMC Health Serv Res*, vol. 22, no. 1, p. 474, Apr. 2022, doi: 10.1186/s12913-022-07887-6.
[175] K. Hausken, “Cyber resilience in firms, organizations and societies,” *Internet of Things*, vol. 11, p. 100204, Sep. 2020, doi: 10.1016/j.iot.2020.100204.
[176] I. Linkov and A. Kott, “Fundamental Concepts of Cyber Resilience: Introduction and Overview,” Jun. 07, 2018, *arXiv*: arXiv:1806.02852. doi: 10.48550/arXiv.1806.02852.
[177] P. E. Roege *et al.*, “Bridging the Gap from Cyber Security to Resilience,” in *Resilience and Risk*, I. Linkov and J. M. Palma-Oliveira, Eds., Dordrecht: Springer Netherlands, 2017, pp. 383–414. doi: 10.1007/978-94-024-1123-2_14.
[178] “Healthcare cybersecurity challenges: a look at current and future trends - ScienceDirect.” Accessed: Sep. 12, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/B9780443132681000030

[179] M. Javaid, A. Haleem, R. P. Singh, and R. Suman, “Towards insighting cybersecurity for healthcare domains: A comprehensive review of recent practices and trends,” *Cyber Security and Applications*, vol. 1, p. 100016, Dec. 2023, doi: 10.1016/j.csa.2023.100016.
[180] D. N. Burrell, A. S. Sabie-Aridi, A. Shufutinsky, J. B. Wright, C. Nobles, and M. Dawson, “Exploring Holistic Managerial Thinking to Better Manage Healthcare Cybersecurity,” *IJHSTM*, vol. 2, no. 1, pp. 1–13, Jan. 2022, doi: 10.4018/IJHSTM.300337.
[181] C. Abraham, D. Chatterjee, and R. R. Sims, “Muddling through cybersecurity: Insights from the U.S. healthcare industry,” *Business Horizons*, vol. 62, no. 4, pp. 539–548, Jul. 2019, doi: 10.1016/j.bushor.2019.03.010.
[182] A. Parisi, *Hands-on artificial intelligence for cybersecurity: implement smart AI systems for preventing cyber attacks and detecting threats and network anomalies*. Birmingham, UK: Packt Publishing, 2019.
[183] A. H. Salem, S. M. Azzam, O. E. Emam, and A. A. Abohany, “Advancing cybersecurity: a comprehensive review of AI-driven detection techniques,” *J Big Data*, vol. 11, no. 1, p. 105, Aug. 2024, doi: 10.1186/s40537-024-00957-y.
[184] Z. O’Leary, *The Essential Guide to Doing Research*. SAGE, 2004.
[185] A. A. Glatthorn and R. L. Joyner, *Writing the Winning Thesis Or Dissertation: A Step-by-Step Guide*. Corwin Press, 2005.
[186] P. Pedamkar, “Types of Research Methodology,” EDUCBA. Accessed: May 28, 2023. [Online]. Available: https://www.educba.com/types-of-research-methodology/
[187] B. Johnson, “Introducing Visual Studio Code,” 2019, pp. 1–12. doi: 10.1002/9781119588238.ch1.
[188] D. Rolon-Merette, M. Ross, T. Rolon-Merette, and K. Church, “Introduction to Anaconda and Python: Installation and setup,” *The Quantitative Methods for Psychology*, vol. 16, pp. S3–S11, May 2020, doi: 10.20982/tqmp.16.5.S003.
[189] “pandas - Python Data Analysis Library.” Accessed: Jan. 22, 2025. [Online]. Available: https://pandas.pydata.org/
[190] C. A. Győrödi, D. V. Dumşe-Burescu, R. Ş. Győrödi, D. R. Zmaranda, L. Bandici, and D. E. Popescu, “Performance Impact of Optimization Methods on MySQL Document-Based and Relational Databases,” *Applied Sciences*, vol. 11, no. 15, Art. no. 15, Jan. 2021, doi: 10.3390/app11156794.
[191] J. Letkowski, “Doing database design with MySQL,” *Journal of Technology Research*, vol. Volume 6, Jan. 2015.
[192] D. A. Harris, “Journal of Information Systems Education,” vol. 14, 2019.
[193] “Flask Vs Django: Which Python Framework to Choose?,” InterviewBit. Accessed: May 28, 2023. [Online]. Available: https://www.interviewbit.com/blog/flask-vs-django/
[194] “1.4. Support Vector Machines — scikit-learn 1.6.1 documentation.” Accessed: Jan. 22, 2025. [Online]. Available: https://scikit-learn.org/stable/modules/svm.html
[195] “1.10. Decision Trees — scikit-learn 1.6.1 documentation.” Accessed: Jan. 24, 2025. [Online]. Available: https://scikit-learn.org/stable/modules/tree.html
[196] “1.11. Ensembles: Gradient boosting, random forests, bagging, voting, stacking — scikit-learn 1.6.1 documentation.” Accessed: Jan. 24, 2025. [Online]. Available: https://scikit-learn.org/stable/modules/ensemble.html
[197] “1.17. Neural network models (supervised) — scikit-learn 1.6.1 documentation.” Accessed: Jan. 24, 2025. [Online]. Available: https://scikit-learn.org/stable/modules/neural_networks_supervised.html

[198] “System Analysis | System Design,” GeeksforGeeks. Accessed: Sep. 26, 2024. [Online]. Available: https://www.geeksforgeeks.org/system-analysis-system-design/
[199] “Wayback Machine.” Accessed: Sep. 26, 2024. [Online]. Available: https://web.archive.org/web/20190828154643/https://www.cms.gov/Research-Statistics-Data-and-Systems/CMS-Information-Technology/XLC/Downloads/SelectingDevelopmentApproach.pdf
[200] A. Dennis, B. H. Wixom, and R. M. Roth, *Systems analysis and design*, 6. ed. Hoboken, NJ: Wiley, 2015.
[201] D. Falessi, G. Cantone, and C. Grande, *A Comparison of Structured Analysis and Object Oriented Analysis - An Experimental Study.* 2007, p. 221.
[202] R. Malan, H.-P. Company, D. Bredemeyer, and B. Consulting, “Functional Requirements and Use Cases”.
[203] K. Bittner and I. Spence, *Use Case Modeling*. Addison-Wesley Professional, 2003.
[204] F. Cremer *et al.*, “Cyber risk and cybersecurity: a systematic review of data availability,” *Geneva Pap Risk Insur Issues Pract*, vol. 47, no. 3, pp. 698–736, 2022, doi: 10.1057/s41288-022-00266-6.
[205] W. Raghupathi, V. Raghupathi, and A. Saharia, “Analyzing Health Data Breaches: A Visual Analytics Approach,” *AppliedMath*, vol. 3, no. 1, Art. no. 1, Mar. 2023, doi: 10.3390/appliedmath3010011.
[206] “Home,” Center for International and Security Studies at Maryland. Accessed: Jun. 24, 2023. [Online]. Available: https://cissm.umd.edu/
[207] C. Harry, “A Proposed Hierarchical Taxonomy for Assessing the Primary Effects of Cyber Events: A Sector Analysis 2014-2016”.
[208] S. I. S. Center, “InfoSec Handlers Diary Blog,” SANS Internet Storm Center. Accessed: Jun. 24, 2023. [Online]. Available: https://isc.sans.edu/diary.html?storyid=0
[209] C. Harry and N. Gallagher, “Categorizing and Assessing the Severity of Disruptive Cyber Incidents,” Apr. 2017, doi: 10.13016/M2QZ22H5J.
[210] “The European Repository of Cyber Incidents,” EuRepoC: European Repository of Cyber Incidents. Accessed: Jan. 08, 2025. [Online]. Available: https://eurepoc.eu/about-us/
[211] H. T. Neprash *et al.*, “Trends in Ransomware Attacks on US Hospitals, Clinics, and Other Health Care Delivery Organizations, 2016-2021,” *JAMA Health Forum*, vol. 3, no. 12, p. e224873, Dec. 2022, doi: 10.1001/jamahealthforum.2022.4873.
[212] N. James, “ 1 - 80+ Healthcare Data Breach Statistics 2023.” Accessed: Jun. 06, 2023. [Online]. Available: https://www.getastra.com/blog/security-audit/healthcare-data-breach-statistics/
[213] A. Yasin, R. Fatima, L. Liu, J. Wanga, R. Ali, and Z. Wei, “Counteracting social engineering attacks,” *Computer Fraud & Security*, vol. 2021, no. 10, pp. 15–19, Oct. 2021, doi: 10.1016/S1361-3723(21)00108-1.
[214] K. Chetioui, B. Bah, A. O. Alami, and A. Bahnasse, “Overview of Social Engineering Attacks on Social Networks,” *Procedia Computer Science*, vol. 198, pp. 656–661, Jan. 2022, doi: 10.1016/j.procs.2021.12.302.
[215] L. Coppolino, S. D’Antonio, L. Romano, L. Sgaglione, M. Magliulo, and R. Pacelli, “Protecting Critical Business Processes of Smart Hospitals from Cyber Attacks,” in *2019 15th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS)*, Nov. 2019, pp. 363–367. doi: 10.1109/SITIS.2019.00065.

[216] “BAZE FOCUS MAGAZINE (2022 CONVOCATION EDITION) by Baze University - Issuu.” Accessed: Jul. 09, 2024. [Online]. Available: https://issuu.com/bazeuniversity/docs/baze_focus_magazine_2022_-merged/44
[217] M. Jofre *et al.*, “Cybersecurity and Privacy Risk Assessment of Point-of-Care Systems in Healthcare—A Use Case Approach,” *Applied Sciences*, vol. 11, no. 15, Art. no. 15, Jan. 2021, doi: 10.3390/app11156699.
[218] M. E. Whitman and H. J. Mattord, *Management of information security*, Fourth edition. Stamford, CT, USA: Cengage Learning, 2014.
[219] I. Lee and C. Choi, “Camp2Vec: Embedding cyber campaign with ATT&CK framework for attack group analysis,” *ICT Express*, vol. 9, no. 6, pp. 1065–1070, Dec. 2023, doi: 10.1016/j.icte.2023.05.008.
[220] H. Sun, H. Shu, F. Kang, Y. Zhao, and Y. Huang, “Malware2ATT&CK: A sophisticated model for mapping malware to ATT&CK techniques,” *Computers & Security*, vol. 140, p. 103772, May 2024, doi: 10.1016/j.cose.2024.103772.
[221] “CVE Website.” Accessed: Jul. 26, 2024. [Online]. Available: https://www.cve.org/
[222] M. A, K. Kota, A. S. Babu, and S. V. S, “CVE Severity Prediction From Vulnerability Description - A Deep Learning Approach,” *Procedia Computer Science*, vol. 235, pp. 3105–3117, Jan. 2024, doi: 10.1016/j.procs.2024.04.294.
[223] “Predicting Vulnerability Type in Common Vulnerabilities and Exposures (CVE) Database with Machine Learning Classifiers | IEEE Conference Publication | IEEE Xplore.” Accessed: Jul. 26, 2024. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9513723
[224] R. J. Mejias, M. A. Shepherd, M. Fronmueller, and R. A. Huff, “Using Threat Vulnerability Asset (TVA) Methodology to Identify Cyber Threats and System Vulnerabilities: A Student Field Project Case Study”.
[225] “Network Attacks Unveiled: Understanding the Threats to Your Switches | FS Community.” Accessed: Jul. 26, 2024. [Online]. Available: https://community.fs.com/article/network-attacks-unveiled-understanding-the-threats-to-your-switches.html
[226] “What Is a Network Switch?” Accessed: Jul. 26, 2024. [Online]. Available: https://www.sophos.com/en-us/cybersecurity-explained/network-switch
[227] DigitalFortresslk, “Common Attack Types on Switches,” Digital Fortress. Accessed: Jul. 28, 2024. [Online]. Available: https://digitalfortresslk.wordpress.com/2018/03/22/common-attack-types-on-switches/
[228] “Routing & Switching: Security Issues & Solutions,” study.com. Accessed: Jul. 28, 2024. [Online]. Available: https://study.com/academy/lesson/routing-switching-security-issues-solutions.html
[229] “Switch Attacks - Tutorial.” Accessed: Jul. 26, 2024. [Online]. Available: https://www.vskills.in/certification/tutorial/switch-attacks/
[230] “CVE - CVE-2020-6977.” Accessed: Jul. 28, 2024. [Online]. Available: https://cve.mitre.org/cgi-bin/cvename.cgi?name=CVE-2020-6977
[231] “CVE Record | CVE.” Accessed: Jul. 28, 2024. [Online]. Available: https://www.cve.org/CVERecord?id=CVE-2020-6977
[232] “GE Healthcare Ultrasound Products (Update A) | CISA.” Accessed: Jul. 26, 2024. [Online]. Available: https://www.cisa.gov/news-events/ics-medical-advisories/icsma-20-049-02

[233] “NVD - CVE-2020-6977.” Accessed: Jul. 28, 2024. [Online]. Available: https://nvd.nist.gov/vuln/detail/CVE-2020-6977
[234] “GE Healthcare Ultrasound Kiosk Mode Vulnerability,” NHS England Digital. Accessed: Jul. 28, 2024. [Online]. Available: https://digital.nhs.uk/cyber-alerts/2020/cc-3372
[235] “GE HealthCare Guidance on Cyber - Support.” Accessed: Jul. 28, 2024. [Online]. Available: https://www.gehealthcare.com/security/cvd
[236] “GE Healthcare MDHexRay Vulnerabilities,” NHS England Digital. Accessed: Jul. 28, 2024. [Online]. Available: https://digital.nhs.uk/cyber-alerts/2020/cc-3699
[237] “Severe MDHexRay bug affects 100+ GE Healthcare imaging systems,” BleepingComputer. Accessed: Jul. 28, 2024. [Online]. Available: https://www.bleepingcomputer.com/news/security/severe-mdhexray-bug-affects-100-plus-ge-healthcare-imaging-systems/
[238] “GE Healthcare Imaging and Ultrasound Products | CISA.” Accessed: Jul. 28, 2024. [Online]. Available: https://www.cisa.gov/news-events/ics-medical-advisories/icsma-20-343-01
[239] R. Crisp, “Medical Device Cybersecurity Vulnerability Detected in GE Equipment,” CE-Tech. Accessed: Jul. 28, 2024. [Online]. Available: https://ce-tech.com/ge-medical-device-cybersecurity-vulnerability/
[240] “NVD - CVE-2020-25175.” Accessed: Jul. 28, 2024. [Online]. Available: https://nvd.nist.gov/vuln/detail/CVE-2020-25175
[241] “CWE - CWE-523: Unprotected Transport of Credentials (4.15).” Accessed: Jul. 28, 2024. [Online]. Available: https://cwe.mitre.org/data/definitions/523.html
[242] “CWE - CWE-522: Insufficiently Protected Credentials (4.15).” Accessed: Jul. 28, 2024. [Online]. Available: https://cwe.mitre.org/data/definitions/522.html
[243] “CISA Discovers Vulnerabilities in GE Healthcare Radiological Devices - News.” Accessed: Aug. 08, 2024. [Online]. Available: https://control.com/news/cisa-discovers-vulnerabilities-in-ge-healthcare-radiological-devices/
[244] “MicroDicom DICOM Viewer | CISA.” Accessed: Jul. 28, 2024. [Online]. Available: https://www.cisa.gov/news-events/ics-medical-advisories/icsma-24-163-01
[245] “Mitsubishi Electric FA Engineering Software Products | CISA.” Accessed: Jul. 29, 2024. [Online]. Available: https://www.cisa.gov/news-events/ics-advisories/icsa-24-030-02
[246] A. Maatallaoui, H. Touil, and L. Setti, “The Impact of Radio Frequency (RF) Attacks on Security and Privacy: A Comprehensive Review,” in *Proceedings of the 6th International Conference on Networking, Intelligent Systems & Security*, in NISS ’23. New York, NY, USA: Association for Computing Machinery, Nov. 2023, pp. 1–23. doi: 10.1145/3607720.3607771.
[247] “CVE - Search Results.” Accessed: Jul. 29, 2024. [Online]. Available: https://cve.mitre.org/cgi-bin/cvekey.cgi?keyword=sip+spoofing
[248] K. Seshadri, “Using your SBC to Defend Against VoIP Spoofing Attacks.” Accessed: Jul. 29, 2024. [Online]. Available: https://assertion.cloud/blog/securing-your-sbcs-how-they-defend-against-voip-spoofing-attacks/
[249] S. Gauci, “Full Disclosure: [ES2021-07] FreeSWITCH does not authenticate SIP MESSAGE requests, leading to spam and message spoofing.” Accessed: Jul. 29, 2024. [Online]. Available: https://seclists.org/fulldisclosure/2021/Oct/44
[250] “Sophos fixes SQL injection vulnerability in their Cyberoam OS,” BleepingComputer. Accessed: Jul. 29, 2024. [Online]. Available:

https://www.bleepingcomputer.com/news/security/sophos-fixes-sql-injection-vulnerability-in-their-cyberoam-os/
[251] G. R. via Fulldisclosure, “Full Disclosure: Cyberoam NetGenie (C0101B1-20141120-NG11VO) - Cross Site Scripting (XSS).” Accessed: Jul. 29, 2024. [Online]. Available: https://seclists.org/fulldisclosure/2021/Aug/20
[252] “CWE - 2022 CWE Top 25 Most Dangerous Software Weaknesses.” Accessed: Jul. 29, 2024. [Online]. Available: https://cwe.mitre.org/top25/archive/2022/2022_cwe_top25_supplemental.html#problematicMappingDetails
[253] bhdresh, “Bugtraq: Sophos Cyberoam Cross-site scripting (XSS) vulnerability.” Accessed: Jul. 29, 2024. [Online]. Available: https://seclists.org/bugtraq/2017/Jun/4
[254] “Cyber Resilient Security in Dell EMC PowerEdge Servers”.
[255] “10 Common Web Security Vulnerabilities,” Gcore. Accessed: Jul. 30, 2024. [Online]. Available: https://gcore.com/learning/10-common-web-security-vulnerabilities/
[256] “Dell Technologies Supply Chain Security: Secured Component Verification for PowerEdge | Dell Technologies Info Hub.” Accessed: Jul. 30, 2024. [Online]. Available: https://infohub.delltechnologies.com/en-us/p/dell-technologies-supply-chain-security-secured-component-verification-for-poweredge/
[257] “ARP Poisoning.” Accessed: Jul. 30, 2024. [Online]. Available: https://www.radware.com/security/ddos-knowledge-center/ddospedia/arp-poisoning/
[258] “INTEL-SA-00582,” Intel. Accessed: Jul. 30, 2024. [Online]. Available: https://www.intel.com/content/www/us/en/security-center/advisory/intel-sa-00582.html
[259] “CVE - Search Results.” Accessed: Jul. 30, 2024. [Online]. Available: https://cve.mitre.org/cgi-bin/cvekey.cgi?keyword=windows+11
[260] “NVD - CVE-2022-21836.” Accessed: Jul. 30, 2024. [Online]. Available: https://nvd.nist.gov/vuln/detail/CVE-2022-21836
[261] “CVE-2022-21836 - Security Update Guide - Microsoft - Windows Certificate Spoofing Vulnerability.” Accessed: Jul. 30, 2024. [Online]. Available: https://msrc.microsoft.com/update-guide/en-US/advisory/CVE-2022-21836
[262] “CWE - CWE-122: Heap-based Buffer Overflow (4.15).” Accessed: Jul. 30, 2024. [Online]. Available: https://cwe.mitre.org/data/definitions/122.html
[263] F. D. Patterson and K. Neailey, “A Risk Register Database System to aid the management of project risk,” *International Journal of Project Management*, vol. 20, no. 5, pp. 365–374, Jul. 2002, doi: 10.1016/S0263-7863(01)00040-0.
[264] T. M. Willams, “Using a risk register to integrate risk management in project definition,” *International Journal of Project Management*, vol. 12, no. 1, pp. 17–22, Feb. 1994, doi: 10.1016/0263-7863(94)90005-1.
[265] T. Kendrick and I. ebrary, *Identifying and managing project risk [electronic resource] : essential tools for failure-proofing your project*. New York : AMACOM, 2003. Accessed: Aug. 12, 2024. [Online]. Available: http://archive.org/details/identifyingmanag00tomk
[266] M. C. Leva, N. Balfe, B. McAleer, and M. Rocke, “Risk registers: Structuring data collection to develop risk intelligence,” *Safety Science*, vol. 100, pp. 143–156, Dec. 2017, doi: 10.1016/j.ssci.2017.05.009.
[267] A. S. R, P. M, Selvapandian. A, J. V, and A. S, “Design of an EEG-based transceiver with data decomposition for IoHT applications,” in *2021 Fifth International Conference on I-*

SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), Nov. 2021, pp. 258–262. doi: 10.1109/I-SMAC52330.2021.9641040.
[268] M. Ianculescu, A. Alexandru, N.-D. Nicolau, G. Neagu, and O. Bica, "IoHT and Edge Computing, Warrants of Optimal Responsiveness of Monitoring Applications for Seniors. A Case Study," in *2019 22nd International Conference on Control Systems and Computer Science (CSCS)*, May 2019, pp. 655–661. doi: 10.1109/CSCS.2019.00118.
[269] M. Bansal and Priya, "Application Layer Protocols for Internet of Healthcare Things (IoHT)," in *2020 Fourth International Conference on Inventive Systems and Control (ICISC)*, Jan. 2020, pp. 369–376. doi: 10.1109/ICISC47916.2020.9171092.
[270] S. Deb, S. Md. R. Islam, J. RobaiatMou, and Md. T. Islam, "Design and implementation of low cost ECG monitoring system for the patient using smart device," in *2017 International Conference on Electrical, Computer and Communication Engineering (ECCE)*, Feb. 2017, pp. 774–778. doi: 10.1109/ECACE.2017.7913007.
[271] C. Li, X. Hu, and L. Zhang, "The IoT-based heart disease monitoring system for pervasive healthcare service," *Procedia Computer Science*, vol. 112, pp. 2328–2334, Jan. 2017, doi: 10.1016/j.procs.2017.08.265.
[272] P. M. Vergara, E. de la Cal, J. R. Villar, V. M. González, and J. Sedano, "An IoT Platform for Epilepsy Monitoring and Supervising," *Journal of Sensors*, vol. 2017, no. 1, p. 6043069, 2017, doi: 10.1155/2017/6043069.
[273] L. Catarinucci *et al.*, "An IoT-Aware Architecture for Smart Healthcare Systems," *IEEE Internet of Things Journal*, vol. 2, no. 6, pp. 515–526, Dec. 2015, doi: 10.1109/JIOT.2015.2417684.
[274] Y. Yin, Y. Zeng, X. Chen, and Y. Fan, "The internet of things in healthcare: An overview," *Journal of Industrial Information Integration*, vol. 1, pp. 3–13, Mar. 2016, doi: 10.1016/j.jii.2016.03.004.
[275] A. Azizan, W. Ahmed, and A. H. A. Razak, "Sensing health: a bibliometric analysis of wearable sensors in healthcare," *Health Technol.*, vol. 14, no. 1, pp. 15–34, Jan. 2024, doi: 10.1007/s12553-023-00801-y.
[276] "Libelium promotes health monitoring in the workplace with MySignals - Libelium." Accessed: Jun. 17, 2024. [Online]. Available: https://www.libelium.com/libeliumworld/success-stories/libelium-promotes-health-monitoring-in-the-workplace-with-mysignals/
[277] H. Ünözkan, M. Ertem, and S. Bendak, "Using attack graphs to defend healthcare systems from cyberattacks: a longitudinal empirical study," *Netw Model Anal Health Inform Bioinforma*, vol. 11, no. 1, p. 52, Nov. 2022, doi: 10.1007/s13721-022-00391-1.
[278] I. H. Sarker, M. H. Furhad, and R. Nowrozy, "AI-Driven Cybersecurity: An Overview, Security Intelligence Modeling and Research Directions," *SN COMPUT. SCI.*, vol. 2, no. 3, p. 173, Mar. 2021, doi: 10.1007/s42979-021-00557-0.
[279] J. Zhang and M. Zulkernine, "Anomaly Based Network Intrusion Detection with Unsupervised Outlier Detection," in *2006 IEEE International Conference on Communications*, Jun. 2006, pp. 2388–2393. doi: 10.1109/ICC.2006.255127.
[280] G. Kim, S. Lee, and S. Kim, "A novel hybrid intrusion detection method integrating anomaly detection with misuse detection," *Expert Systems with Applications*, vol. 41, no. 4, Part 2, pp. 1690–1700, Mar. 2014, doi: 10.1016/j.eswa.2013.08.066.
[281] M. S. Song, J. D. Lee, Y.-S. Jeong, H.-Y. Jeong, and J. H. Park, "DS-ARP: a new detection scheme for ARP spoofing attacks based on routing trace for ubiquitous

environments," *ScientificWorldJournal*, vol. 2014, p. 264654, 2014, doi: 10.1155/2014/264654.
[282] V. Hnamte and J. Hussain, "Enhancing security in Software-Defined Networks: An approach to efficient ARP spoofing attacks detection and mitigation," *Telematics and Informatics Reports*, vol. 14, p. 100129, Jun. 2024, doi: 10.1016/j.teler.2024.100129.
[283] S. Liao *et al.*, "A Comprehensive Detection Approach of Nmap: Principles, Rules and Experiments," in *2020 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC)*, Oct. 2020, pp. 64–71. doi: 10.1109/CyberC49757.2020.00020.
[284] K. P. Vijayakumar, K. Pradeep Mohan Kumar, K. Kottilingam, T. Karthick, P. Vijayakumar, and P. Ganeshkumar, "An adaptive neuro-fuzzy logic based jamming detection system in WSN," *Soft Comput*, vol. 23, no. 8, pp. 2655–2667, Apr. 2019, doi: 10.1007/s00500-018-3636-5.
[285] M. Ahmed, S. Byreddy, A. Nutakki, L. Sikos, and P. Haskell-Dowland, "ECU-IoHT," *Research Datasets*, Jul. 2020, doi: 10.25958/5f1f97b837aca.
[286] L. Grinsztajn, E. Oyallon, and G. Varoquaux, "Why do tree-based models still outperform deep learning on typical tabular data?".
[287] L. Rokach and O. Maimon, "Decision Trees," in *Data Mining and Knowledge Discovery Handbook*, O. Maimon and L. Rokach, Eds., Boston, MA: Springer US, 2005, pp. 165–192. doi: 10.1007/0-387-25465-X_9.
[288] "Gini Index: Decision Tree, Formula, Calculator, Gini Coefficient in Machine Learning," Quantitative Finance & Algo Trading Blog by QuantInsti. Accessed: Jun. 18, 2024. [Online]. Available: https://blog.quantinsti.com/gini-index/
[289] "An improved random forest classifier for multi-class classification - ScienceDirect." Accessed: Jun. 18, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2214317316300099
[290] "Random forest in remote sensing: A review of applications and future directions - ScienceDirect." Accessed: Jun. 18, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0924271616000265
[291] "Frontiers | Gradient boosting machines, a tutorial." Accessed: Jun. 18, 2024. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fnbot.2013.00021/full
[292] "Deep neural networks in the cloud: Review, applications, challenges and research directions - ScienceDirect." Accessed: Jun. 19, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0925231223004502
[293] "13.1 Multi-layer perceptrons (MLPs)." Accessed: Jun. 19, 2024. [Online]. Available: https://kenndanielso.github.io/mlrefined/blog_posts/13_Multilayer_perceptrons/13_1_Multi_layer_perceptrons.html
[294] G. Zachos, I. Essop, G. Mantas, K. Porfyrakis, J. C. Ribeiro, and J. Rodriguez, "An Anomaly-Based Intrusion Detection System for Internet of Medical Things Networks," *Electronics*, vol. 10, no. 21, Art. no. 21, Jan. 2021, doi: 10.3390/electronics10212562.
[295] D. Chicco and G. Jurman, "The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation," *BMC Genomics*, vol. 21, no. 1, p. 6, Jan. 2020, doi: 10.1186/s12864-019-6413-7.
[296] M. Sokolova, N. Japkowicz, and S. Szpakowicz, "Beyond Accuracy, F-Score and ROC: A Family of Discriminant Measures for Performance Evaluation," in *AI 2006: Advances in*

Artificial Intelligence, A. Sattar and B. Kang, Eds., Berlin, Heidelberg: Springer, 2006, pp. 1015–1021. doi: 10.1007/11941439_114.
[297] K. P. Vijayakumar, K. Pradeep, A. Balasundaram, and M. R. Prusty, "Enhanced Cyber Attack Detection Process for Internet of Health Things (IoHT) Devices Using Deep Neural Network," *Processes*, vol. 11, no. 4, Art. no. 4, Apr. 2023, doi: 10.3390/pr11041072.
[298] G. Biau and E. Scornet, "A random forest guided tour," *TEST*, vol. 25, no. 2, pp. 197–227, Jun. 2016, doi: 10.1007/s11749-016-0481-7.
[299] A. Mayr, H. Binder, O. Gefeller, and M. Schmid, "The Evolution of Boosting Algorithms," *Methods Inf Med*, vol. 53, no. 6, pp. 419–427, 2014, doi: 10.3414/ME13-01-0122.
[300] S. B. Kotsiantis, "Decision trees: a recent overview," *Artif Intell Rev*, vol. 39, no. 4, pp. 261–283, Apr. 2013, doi: 10.1007/s10462-011-9272-4.
[301] E. Beauxis-Aussalet and L. Hardman, "Simplifying the Visualization of Confusion Matrix".
[302] E. Beauxis-Aussalet and L. Hardman, "Visualization of Confusion Matrix for Non-Expert Users".
[303] A. Basit, M. Zafar, X. Liu, A. R. Javed, Z. Jalil, and K. Kifayat, "A comprehensive survey of AI-enabled phishing attacks detection techniques," *Telecommun Syst*, vol. 76, no. 1, pp. 139–154, Jan. 2021, doi: 10.1007/s11235-020-00733-2.
[304] S. W. Liew, N. F. M. Sani, Mohd. T. Abdullah, R. Yaakob, and M. Y. Sharum, "An effective security alert mechanism for real-time phishing tweet detection on Twitter," *Computers & Security*, vol. 83, pp. 201–207, Jun. 2019, doi: 10.1016/j.cose.2019.02.004.
[305] J. Abawajy and A. Kelarev, "A Multi-tier Ensemble Construction of Classifiers for Phishing Email Detection and Filtering," in *Cyberspace Safety and Security*, Y. Xiang, J. Lopez, C.-C. J. Kuo, and W. Zhou, Eds., in Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, 2012, pp. 48–56. doi: 10.1007/978-3-642-35362-8_5.
[306] P. Bountakas and C. Xenakis, "HELPHED: Hybrid Ensemble Learning PHishing Email Detection," *Journal of Network and Computer Applications*, vol. 210, p. 103545, Jan. 2023, doi: 10.1016/j.jnca.2022.103545.
[307] P. Suresh *et al.*, "Chapter 10 - Contemporary survey on effectiveness of machine and deep learning techniques for cyber security," in *Machine Learning for Biometrics*, P. P. Sarangi, M. Panda, S. Mishra, B. S. P. Mishra, and B. Majhi, Eds., in Cognitive Data Science in Sustainable Computing. , Academic Press, 2022, pp. 177–200. doi: 10.1016/B978-0-323-85209-8.00007-9.
[308] K. C. Bourne, "Chapter 15 - Security," in *Application Administrators Handbook*, K. C. Bourne, Ed., Boston: Morgan Kaufmann, 2014, pp. 242–267. doi: 10.1016/B978-0-12-398545-3.00015-7.
[309] S. D. Applegate, "Social Engineering: Hacking the Wetware!," *Information Security Journal: A Global Perspective*, vol. 18, no. 1, pp. 40–46, Feb. 2009, doi: 10.1080/19393550802623214.
[310] A. Mughaid, S. AlZu'bi, A. Hnaif, S. Taamneh, A. Alnajjar, and E. A. Elsoud, "An intelligent cyber security phishing detection system using deep learning techniques," *Cluster Comput*, vol. 25, no. 6, pp. 3819–3828, Dec. 2022, doi: 10.1007/s10586-022-03604-4.

[311] N. Abdelhamid, A. Ayesh, and F. Thabtah, “Phishing detection based Associative Classification data mining,” *Expert Systems with Applications*, vol. 41, no. 13, pp. 5948–5959, Oct. 2014, doi: 10.1016/j.eswa.2014.03.019.
[312] I. H. Witten and E. Frank, “Data mining: practical machine learning tools and techniques with Java implementations,” *SIGMOD Rec.*, vol. 31, no. 1, pp. 76–77, Mar. 2002, doi: 10.1145/507338.507355.
[313] R. M. Mohammad, F. Thabtah, and L. McCluskey, “An assessment of features related to phishing websites using an automated technique,” in *2012 International Conference for Internet Technology and Secured Transactions*, Dec. 2012, pp. 492–497.
[314] “Choosing the right estimator,” scikit-learn. Accessed: Nov. 14, 2023. [Online]. Available: https://scikit-learn/stable/tutorial/machine_learning_map/index.html
[315] “1.1. Linear Models,” scikit-learn. Accessed: Nov. 14, 2023. [Online]. Available: https://scikit-learn/stable/modules/linear_model.html
[316] D. Anguita, A. Ghio, N. Greco, L. Oneto, and S. Ridella, “Model selection for support vector machines: Advantages and disadvantages of the Machine Learning Theory,” in *The 2010 International Joint Conference on Neural Networks (IJCNN)*, Jul. 2010, pp. 1–8. doi: 10.1109/IJCNN.2010.5596450.
[317] “4. Supervised Learning: Models and Concepts - Machine Learning and Data Science Blueprints for Finance [Book].” Accessed: Nov. 14, 2023. [Online]. Available: https://www.oreilly.com/library/view/machine-learning-and/9781492073048/ch04.html
[318] E. A. Zanaty, “Support Vector Machines (SVMs) versus Multilayer Perception (MLP) in data classification,” *Egyptian Informatics Journal*, vol. 13, no. 3, pp. 177–183, Nov. 2012, doi: 10.1016/j.eij.2012.08.002.
[319] C.-W. Hsu and C.-J. Lin, “A comparison of methods for multiclass support vector machines,” *IEEE Trans Neural Netw*, vol. 13, no. 2, pp. 415–425, 2002, doi: 10.1109/72.991427.
[320] S. N. Wan Ahmad, “Comparative Performance of Machine Learning Methods for Classification on Phishing Attack Detection,” *IJATCSE*, vol. 9, no. 1.5, pp. 349–354, Sep. 2020, doi: 10.30534/ijatcse/2020/4991.52020.
[321] D. K. Srivastava and L. Bhambhu, “DATA CLASSIFICATION USING SUPPORT VECTOR MACHINE,” 2005.
[322] D. Wahyudi, M. Niswar, and A. A. P. Alimuddin, “WEBSITE PHISING DETECTION APPLICATION USING SUPPORT VECTOR MACHINE (SVM),” *JITU*, vol. 5, no. 1, pp. 18–24, Jun. 2022, doi: 10.56873/jitu.5.1.4836.
[323] M. Nabet and L. George, “Phishing Attacks Detection by Using Support Vector Machine,” *Journal of Al-Qadisiyah for Computer Science and Mathematics*, vol. 15, Sep. 2023, doi: 10.29304/jqcm.2023.15.2.1242.
[324] D. Aksu, A. Abdulwakil, and M. A. Aydin, “Detecting phishing websites using support vector machine algorithm,” presented at the Pressacademia, Jun. 2017, pp. 139–142. doi: 10.17261/Pressacademia.2017.582.
[325] A. Altaher, “Phishing Websites Classification using Hybrid SVM and KNN Approach,” *ijacsa*, vol. 8, no. 6, 2017, doi: 10.14569/IJACSA.2017.080611.
[326] R. Karnik and D. G. M. Bhandari, “Support Vector Machine Based Malware and Phishing Website Detection,” 2016. Accessed: Nov. 14, 2023. [Online]. Available: https://www.semanticscholar.org/paper/Support-Vector-Machine-Based-Malware-and-Phishing-Karnik-Bhandari/ffea603ec9f33931c9de630ba1a6ac71924f1539

[327] “Advancing Network Security in Industrial IoT: A Deep Dive into AI-Enabled Intrusion Detection Systems - ScienceDirect.” Accessed: Nov. 12, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1474034624003331
[328] G. Apruzzese *et al.*, “The Role of Machine Learning in Cybersecurity,” *Digital Threats*, vol. 4, no. 1, p. 8:1-8:38, Mar. 2023, doi: 10.1145/3545574.
[329] “TON_IoT Telemetry Dataset: A New Generation Dataset of IoT and IIoT for Data-Driven Intrusion Detection Systems | IEEE Journals & Magazine | IEEE Xplore.” Accessed: Aug. 05, 2024. [Online]. Available: https://ieeexplore.ieee.org/document/9189760
[330] N. Moustafa, M. Keshky, E. Debiez, and H. Janicke, “Federated TON_IoT Windows Datasets for Evaluating AI-Based Security Applications,” in *2020 IEEE 19th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom)*, Dec. 2020, pp. 848–855. doi: 10.1109/TrustCom50675.2020.00114.
[331] T. M. Booij, I. Chiscop, E. Meeuwissen, N. Moustafa, and F. T. H. D. Hartog, “ToN_IoT: The Role of Heterogeneity and the Need for Standardization of Features and Attack Types in IoT Network Intrusion Data Sets,” *IEEE Internet Things J.*, vol. 9, no. 1, pp. 485–496, Jan. 2022, doi: 10.1109/JIOT.2021.3085194.
[332] “Edge-IIoTset: A New Comprehensive Realistic Cyber Security Dataset of IoT and IIoT Applications for Centralized and Federated Learning | IEEE Journals & Magazine | IEEE Xplore.” Accessed: Oct. 22, 2024. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9751703
[333] R. A. Disha and S. Waheed, “Performance analysis of machine learning models for intrusion detection system using Gini Impurity-based Weighted Random Forest (GIWRF) feature selection technique,” *Cybersecurity*, vol. 5, no. 1, p. 1, Jan. 2022, doi: 10.1186/s42400-021-00103-8.
[334] N. Moustafa and J. Slay, “UNSW-NB15: a comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set),” Nov. 2015. doi: 10.1109/MilCIS.2015.7348942.
[335] I. Tareq, B. M. Elbagoury, S. El-Regaily, and E.-S. M. El-Horbaty, “Analysis of ToN-IoT, UNW-NB15, and Edge-IIoT Datasets Using DL in Cybersecurity for IoT,” *Applied Sciences*, vol. 12, no. 19, Art. no. 19, Jan. 2022, doi: 10.3390/app12199572.
[336] P. Verma, J. G. Breslin, and D. O’Shea, “FLDID: Federated Learning Enabled Deep Intrusion Detection in Smart Manufacturing Industries,” *Sensors*, vol. 22, no. 22, Art. no. 22, Jan. 2022, doi: 10.3390/s22228974.
[337] “Basic principles of ROC analysis - ScienceDirect.” Accessed: Jan. 02, 2025. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0001299878800142

APPENDICES

Appednix A: Visit to BUH

This section shows our visit to Baze University Hospital on 29-31 January 2024. The goal of our visit is to assess the cyber infrastructure of the healthcare facility to understand the critical assets that are vulnerable to cyberattacks. This section shows some of the medical equipment that we assessed during our visit. Our second goal for this visit is to be able to practically understand how the cyber architecture of a typical healthcare facility is set up to correlate why the healthcare sector is always vulnerable.

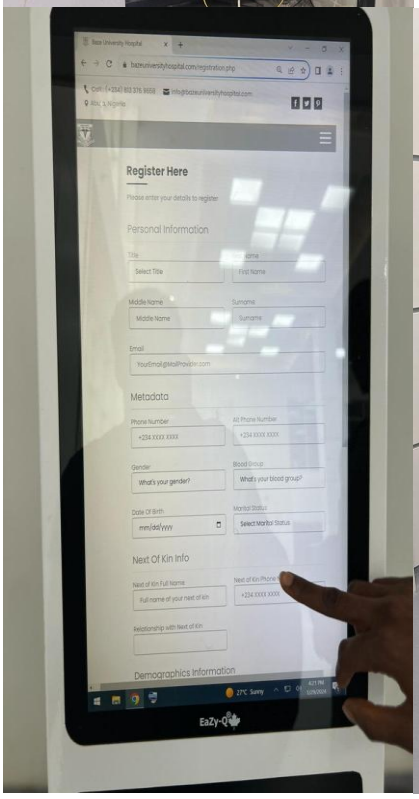




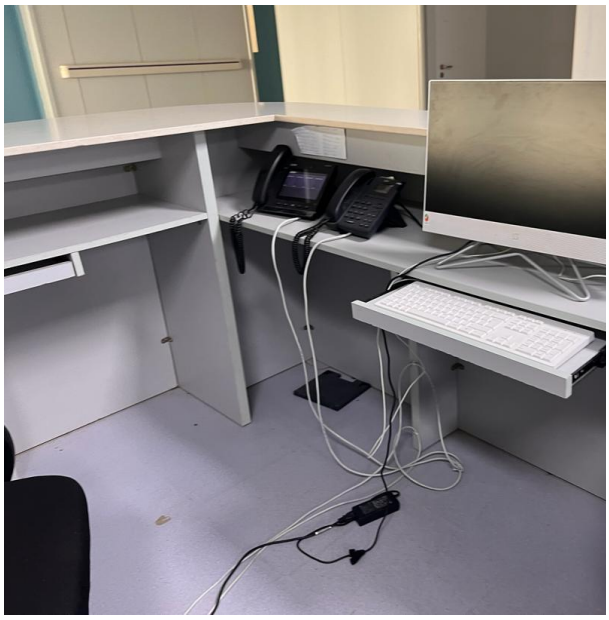





BAZE UNIVERSITY
BUTH


Register Here
Personal Information
Metadata
Next Of Kin Info
Demographics Information
EaZy-Q







Appednix B: Analysis of OCR datasource

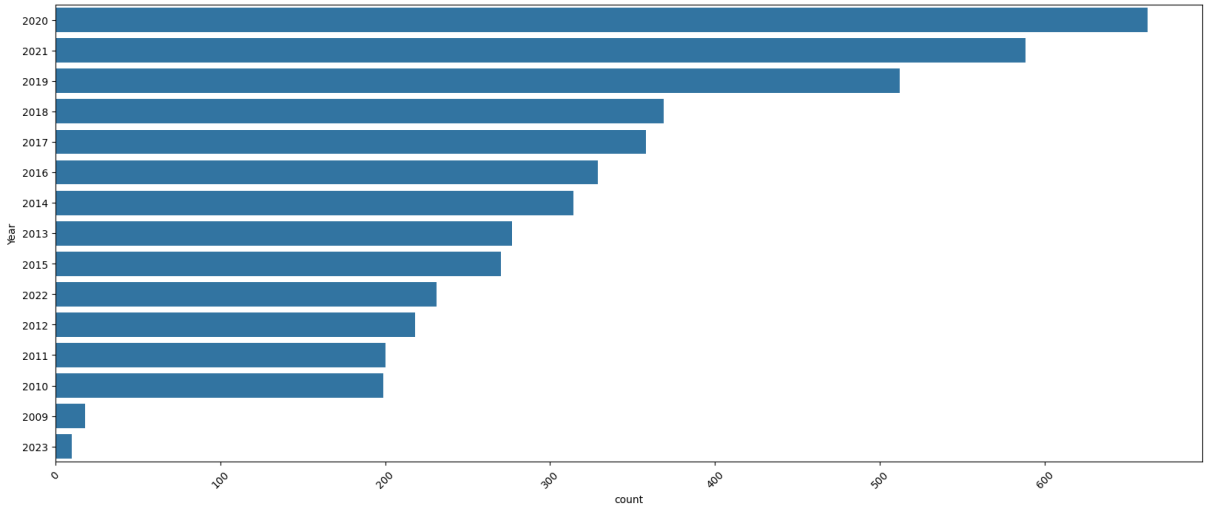


Fig. B1: Number of Incidents by year (OCR)

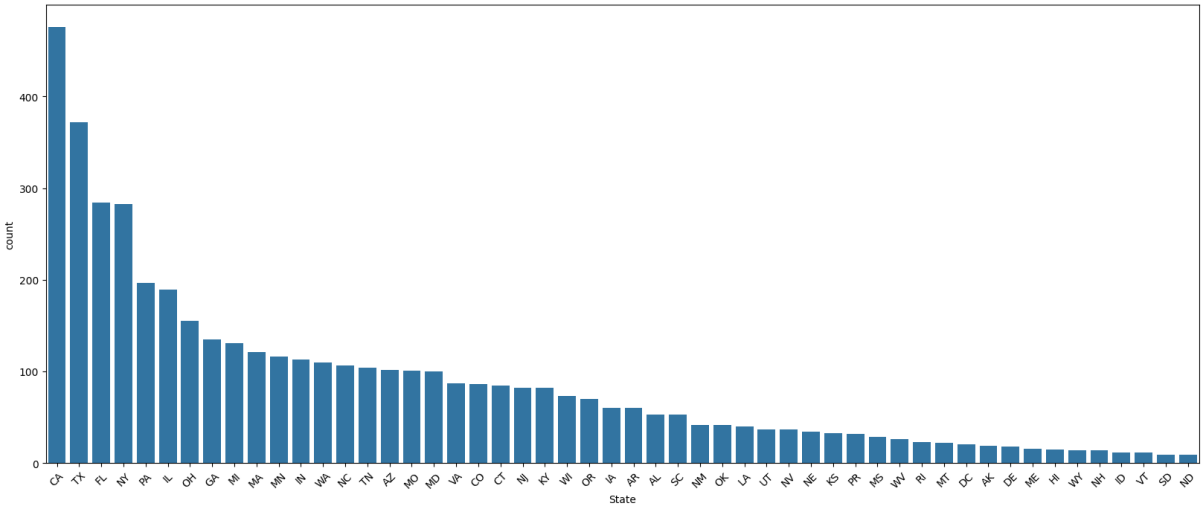


Fig.B2: Number of Incidents by State (OCR)

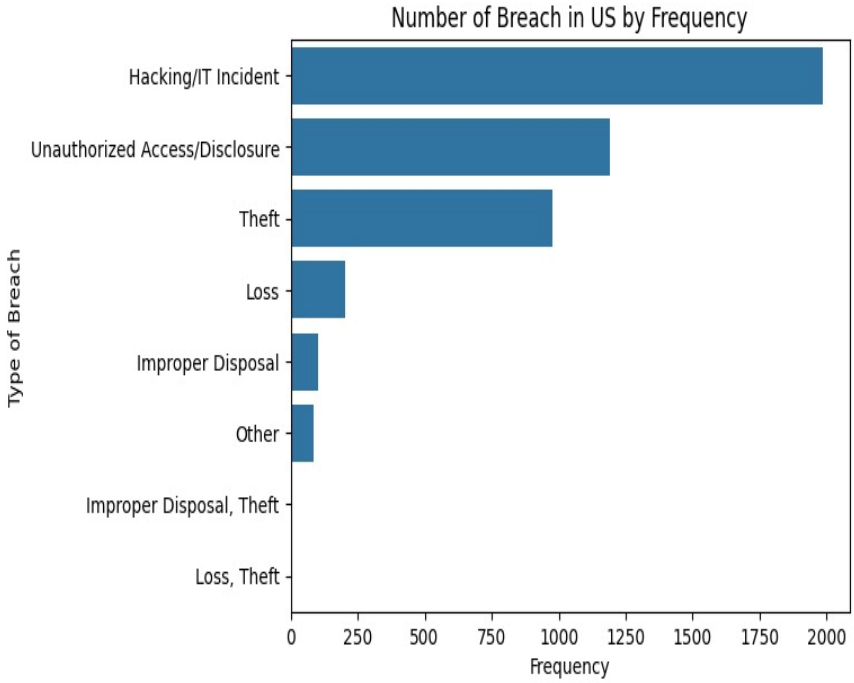


Fig. B3: Number of attacks by Type (OCR)

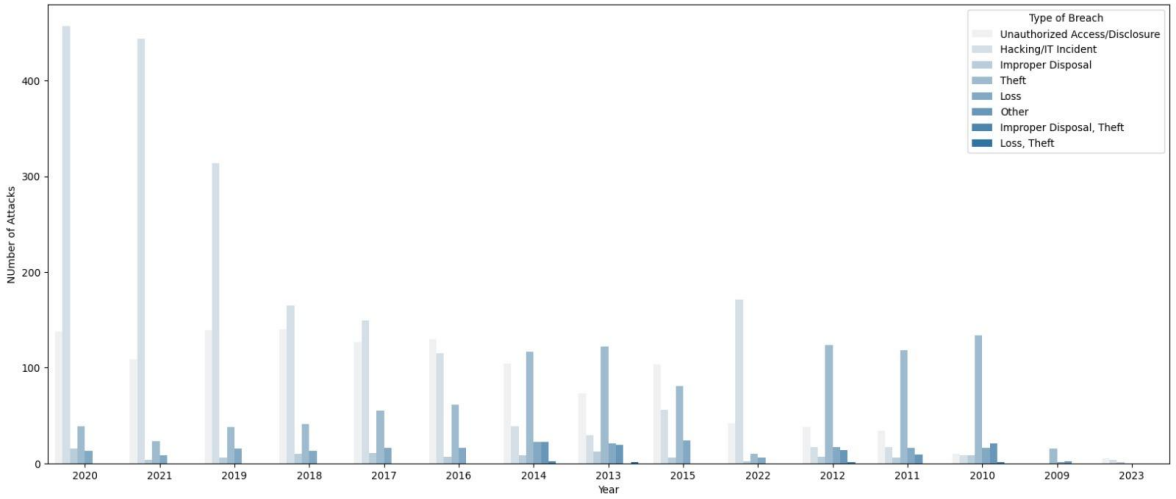


Fig. B4: Number of attacks by breach type by year (OCR)

Appednix C: Analysis of CISSM Datasource

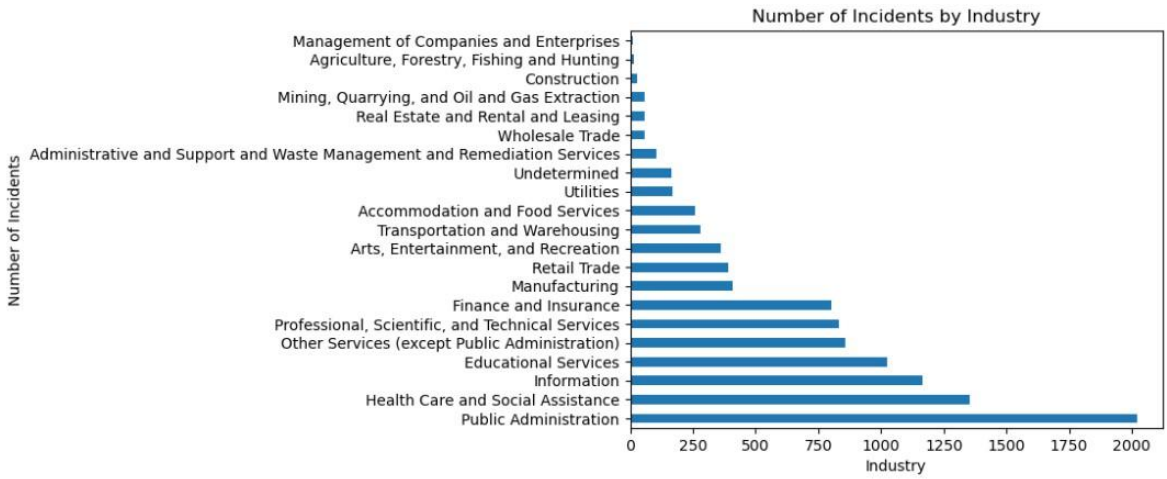


Fig. C1: Number of incidents by industry (CISSM)

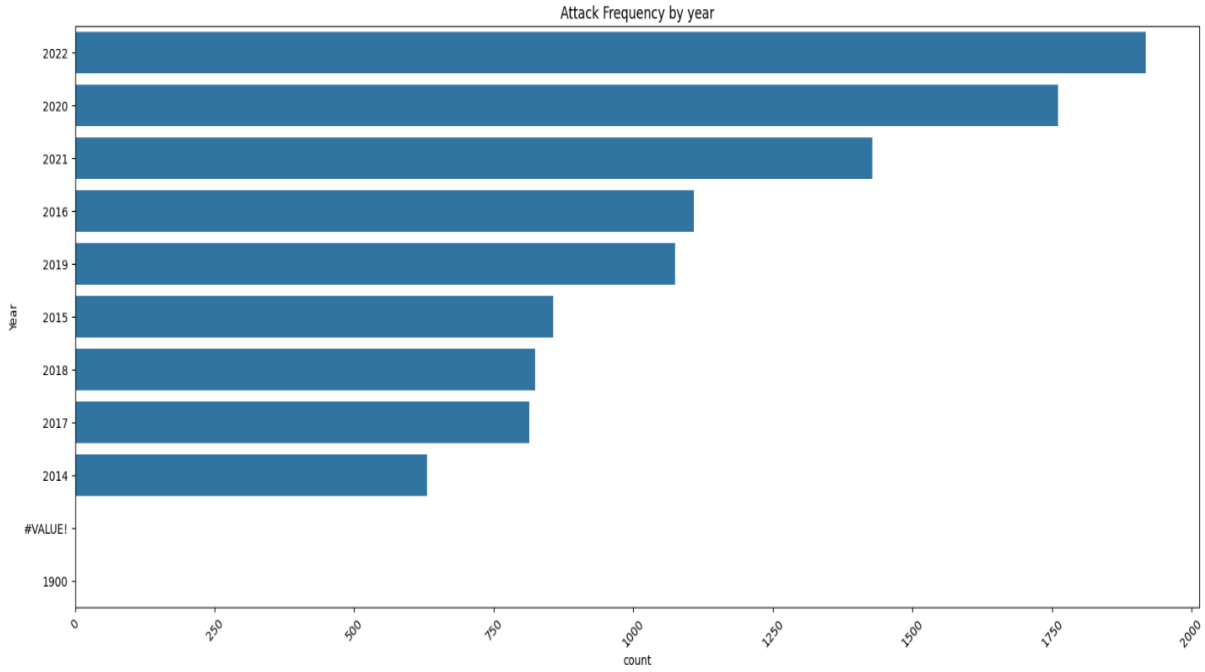


Fig. C2: Number of incidents by year (CISSM)

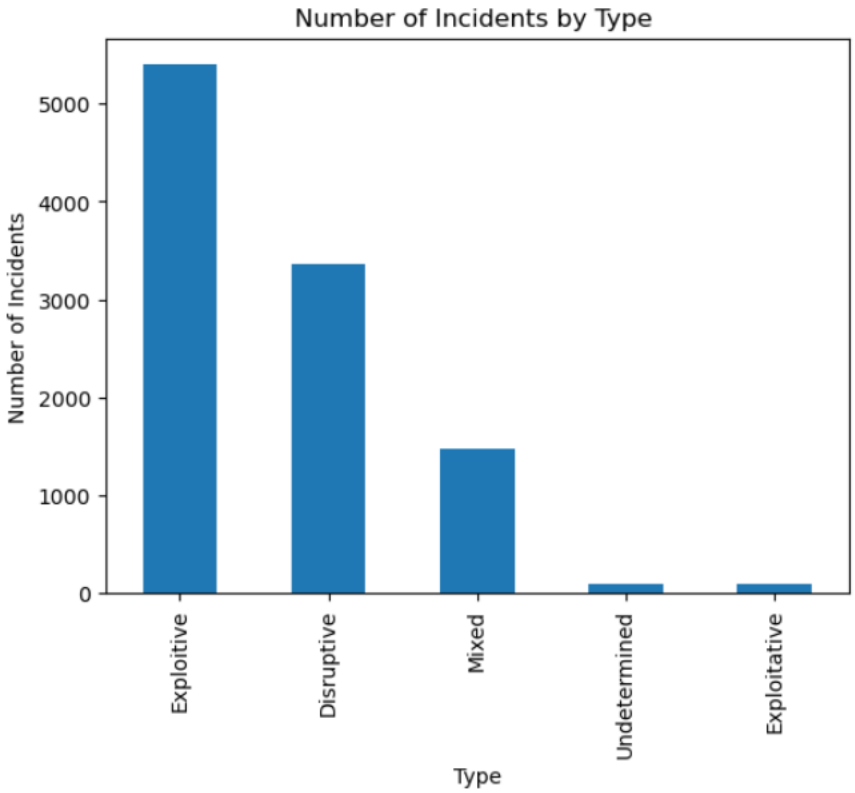


Fig. C3: Number of incidents by type (CISSM)

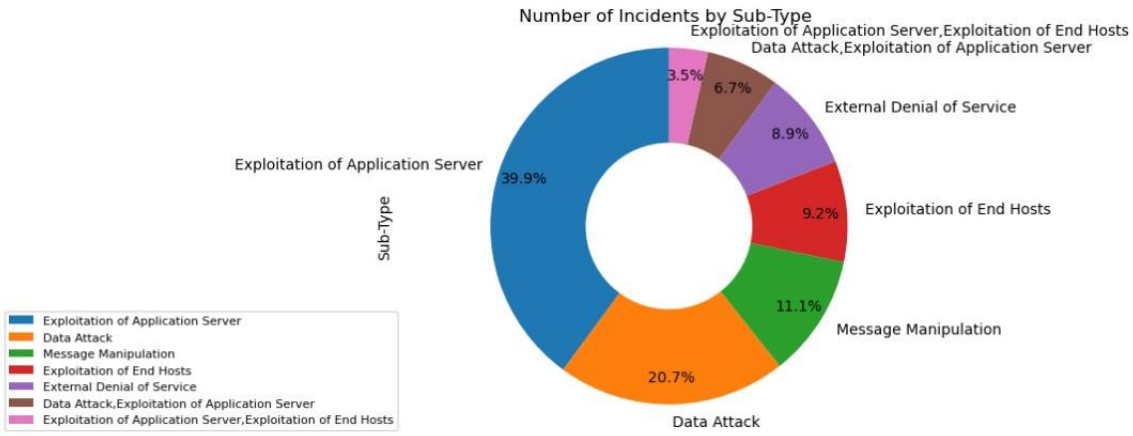


Fig. C4: Incidents by sub-type (CISSM)

Appednix D: Healthcare related cyber incidents (CISSM)

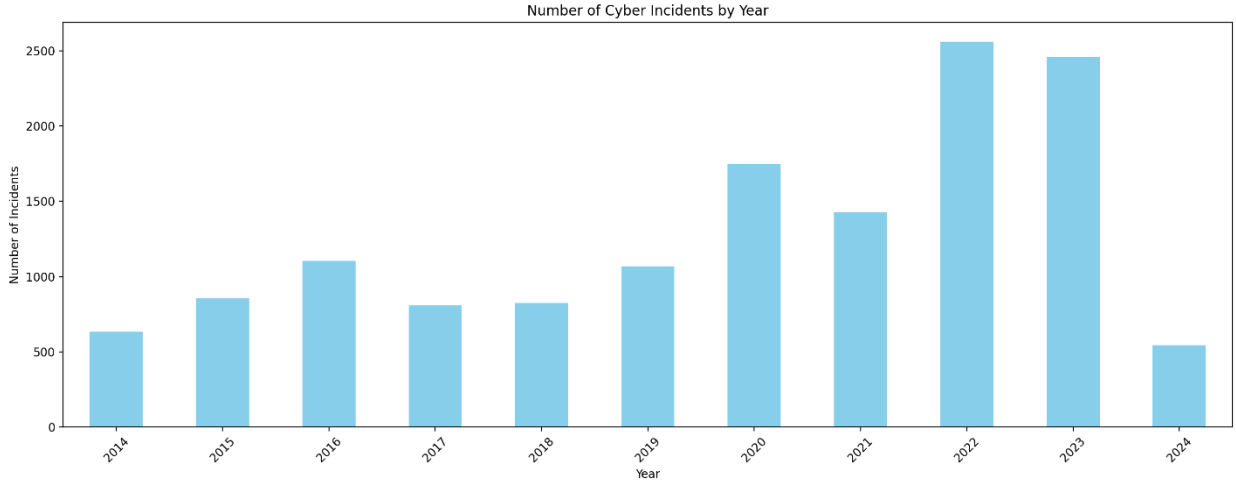


Fig. D1: Number of incidents by year (CISSM-Healthcare related)

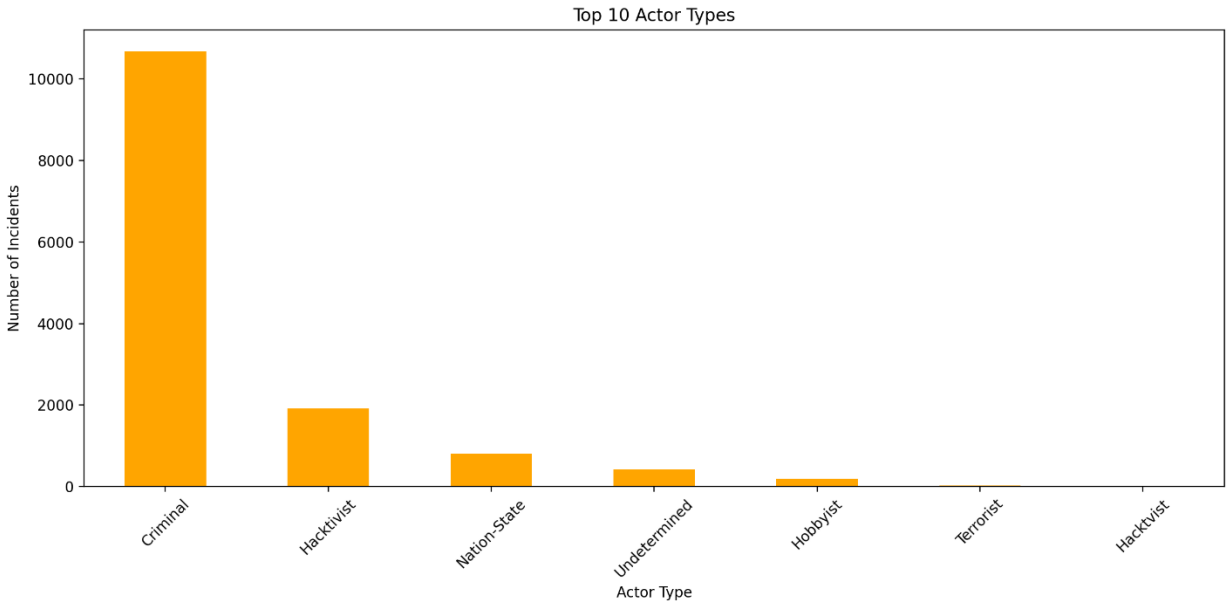


Fig. D2: Top ten actors (CISSM-Healthcare related)

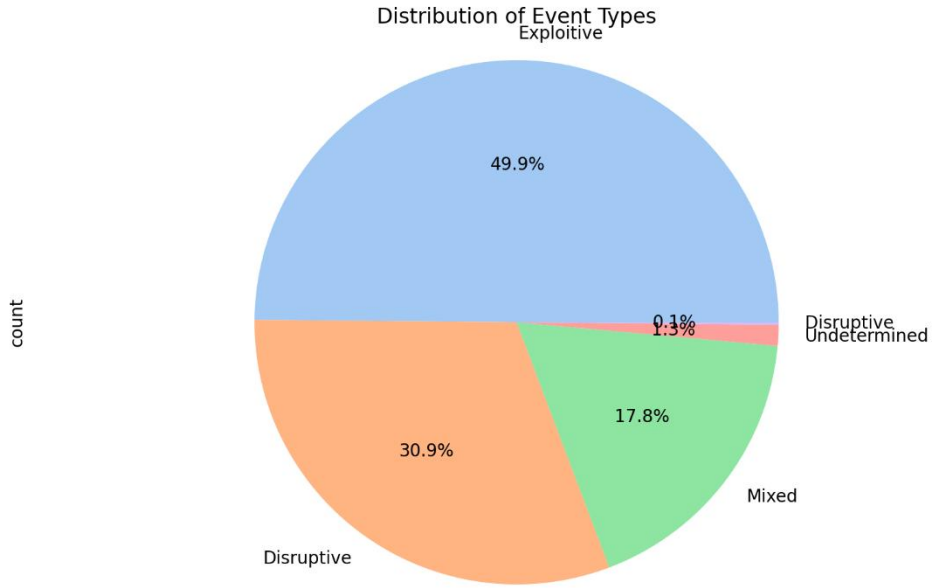


Fig. D3: Distribution of Events by type (CISSM-Healthcare related)

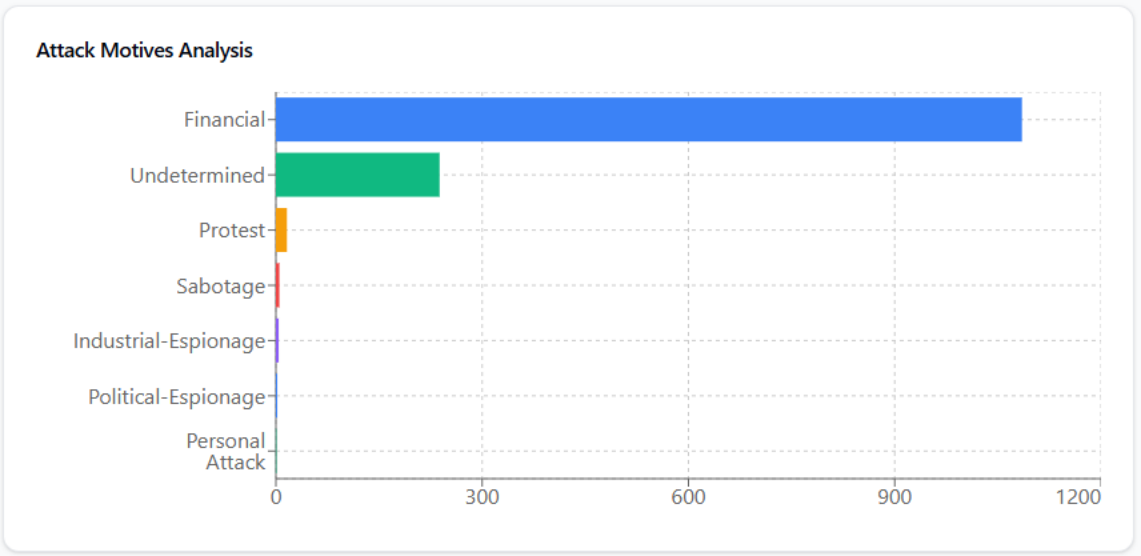


Fig. D4: Attack motives (CISSM-Healthcare related)

ABSTRACT

In recent years, there has been a growing frequency and sophistication in cyberattacks, which have made cybersecurity in the healthcare sector a pressing global concern. Some of the factors responsible for this fragility include overdependence of healthcare facilities on linked digital infrastructures, sensitive patient information, and the use of antiquated security systems. This thesis explores a multilayered cybersecurity approach in healthcare using database technology, artificial intelligence (AI), machine learning (ML), and risk assessment methodologies. We begin by assessing the current threat landscape through a web application that leverages publicly available cyber incident records mainly from the Center for International and Security Studies Maryland (CISSM), which contains over 1,500 healthcare-related cases, and the Office for Civil Rights (OCR), which has documented over 800 investigated incidents. The developed platform enables users to add, delete, search, view, and download cyber-related incidents. To address cybersecurity vulnerabilities, we introduce the Baze University Hospital Cyber Risk Assessment Framework (BUHCRAF), applied at Baze University Hospital as a case study. This framework identifies over 50 security risks, with some linked to known threats documented in the Common Vulnerabilities and Exposures (CVE) database. Key risks include unpatched software, unsecured medical IoT devices, and weak access controls. Based on these insights, AI/ML-driven cybersecurity solutions are proposed. The study includes several case studies demonstrating AI/ML applications in cybersecurity. The first case focuses on anomaly detection in the Internet of Health Things (IoHT), using the ECU-IoHT dataset containing 111,000 records and five attack classes (Normal, Nmap, DoS, ARP, Smurf). We compare four ML models—Random Forest (RF), Gradient Boosting (GB), Decision Trees (DT), and Multilayer Perceptron (MLP)—where GB achieves the highest accuracy of 99.53%. The second case study applies AI/ML to phishing detection, analyzing features such as request URL, domain age, HTTPS/SSL status, website traffic, and pop-ups. Support Vector Machines (SVM) were used with two kernel functions—Radial Basis Function (RBF) and Polynomial. The Polynomial kernel outperformed RBF, achieving 84.5% accuracy compared to 82.6%. The third case study explores AI/ML in intrusion detection by proposing a Cross-ML framework using three public datasets: UNSW-NB15, X-IIoTID, and ToN-IoT. Individual models implies models trained on their respective datasets achieved F1-scores ranging from 92.0% to 98.1%, while the combined dataset approach recorded an accuracy of 90.9%, recall of 84.0%, and an F1-score of 84.9%. Cross-dataset evaluations showed that the X-IIoTID model tested on ToN-IoT achieved the best performance, reaching 75% accuracy and a 74% F1-score. This research highlights the importance of a multilayered cybersecurity approach that integrates database-driven approach, AI/ML-enhanced security applications, and structured risk assessment frameworks to strengthen cybersecurity resilience in healthcare organizations.

Keywords: *Cybersecurity, Artificial Intelligence, Machine Learning, Databases, Risk Assessment*