Advanced Privacy-Preserving Decentralized Federated Learning for Insider Threat Detection in Collaborative Healthcare Institutions

by King David D. Newman

B.Eng. in Computer Engineering, December 2016, University of Rwanda
M.S. in Cybersecurity, December 2021, Walden University

A Praxis submitted to

The Faculty of
The School of Engineering and Applied Science
of The George Washington University
in partial fulfillment of the requirements
for the degree of Doctor of Engineering

January 10, 2025

Praxis directed by

Domingo Rivera
Professorial Lecturer in Engineering and Applied Science

The School of Engineering and Applied Science of The George Washington University certifies that King David Don Newman has passed the Final Examination for the degree of Doctor of Engineering as of December 20, 2024 This is the final and approved form of the Praxis.

Advanced Privacy-Preserving Decentralized Federated Learning for Insider Threat Detection in Collaborative Healthcare Institutions

King David D. Newman

Praxis Research Committee:

Amir Etemadi, Associate Professor of Engineering and Applied Science, Praxis Chair

Domingo Rivera, Professorial Lecturer in Engineering and Applied Science, Praxis Director

Linda Rodriguez, Professorial Lecturer in Engineering and Applied Science, Committee Member

© Copyright 2024 by King David D. Newman
All rights reserved

Dedication

The author wishes to dedicate this Praxis to God for the guidance that has been the source of strength and perseverance. The author also dedicates this Praxis to his wife, Clementine; your love and belief in me have been the cornerstone of my achievements. The author extends this dedication to his children, Davis, Avis, and Ava, who inspire him to aim higher and strive for a better future. Thank you for being the inspiration and motivation during this journey.

Acknowledgements

The author wishes to acknowledge God for providing the strength, wisdom, and guidance throughout this academic journey. To God be the glory.

The author also wishes to acknowledge his Praxis advisor, Dr. Domingo Rivera, for his invaluable time and expertise, as well as for inspiring him to strive for the highest quality of work by providing guidance whenever possible. Your high standards and encouragement have greatly advanced my personal and professional growth. I feel privileged to have had the opportunity to learn from you and am deeply grateful for your mentorship.

The author is deeply grateful to Dr. Amir Etemadi, the esteemed Chair of the Praxis Committee, and Dr. Linda Rodriguez, Committee Member, for their leadership and guidance during our committee meeting.

Lastly, the author wishes to acknowledge everyone who significantly contributed to this journey. Your support made this academic endeavor possible.

Abstract of Praxis

Advanced Privacy-Preserving Decentralized Federated Learning for Insider Threat Detection in Collaborative Healthcare Institutions

The healthcare sector is embracing a digital transformation driven by technological advancements and the quest for more effective healthcare delivery. This transformation has led to an exponential increase in the volume and diversity of data generated by healthcare institutions. Such large datasets offer the potential to reveal significant insights, including patterns that can predict, detect, and prevent insider threats in collaborative healthcare settings (Brauneck et al., 2023). However, this data growth also intensifies risks related to insider threats, which often originate from within an organization and pose significant challenges due to digital health adoption and privacy concerns with centralized methods, leading to data breaches and increasing regulatory challenges (Alder, 2023; Vasko and Gardner, 2019; Fisher and Phillips, 2024; Teo et al., 2024; U.S. Department of Health and Human Services, 2022).

Additionally, the rising adoption of machine learning (ML) for big data analysis and research involves centralized machine learning methods widely applied to tackle these threats. However, these methods face challenges such as privacy violations, ownership disputes, and regulatory non-compliance (Goyal & Malviya, 2023; Tabassum et al., 2024; Paracha et al., 2024; Naresh, Thamarai & Allavarpu, 2023). To mitigate these risks, the centralized federated learning (CFL) method offers improvements through a collaborative approach. However, CFL relies on a main server to aggregate updates from distributed models, which could introduce a single point of failure.

This study proposes an Advanced Privacy-Preserving Decentralized Federated Learning (APPDFL) system as a solution for insider threat detection that adopts collaborative healthcare institution participation for intelligence sharing. Leveraging a decentralized federated learning communication architecture, this approach ensures privacy, scalability with large datasets, and flexibility compared to traditional methods (Nguyen et al., 2022). We aim to apply security by design by implementing server-less peer-to-peer (P2P) communication in a horizontal decentralized federated learning with differential privacy mechanism. This mechanism has been widely implemented in numerous domains, such as deep learning research, to prevent indirect data leakage (Abadi et al., 2016; McMahan, 2017). APPDFL allows multiple healthcare institutions to collaboratively train shared threats intelligence without exchanging confidential data, eliminating the need for a central server (Roy et al., 2019; He et al., 2020; LeCun, Bengio, and Hinton, 2015; Shayan et al., 2019).

Additionally, federated learning complies with laws and regulations, ensuring data privacy preservation and data security (Saha et al., 2024; Truong et al., 202; Chalamala et al., 2022; Sharma & Kumar, 2023), making it ideal for sensitive sectors such as healthcare, financial institutions, and government agencies. For this Praxis implementation, we utilized the most widely used dataset for insider threat from Carnegie Mellon University-CERT, specifically versions r4.2, r5.2, and r6.2. We applied various balancing techniques to balance the dataset and simulate real-world cases of insider threat detection across multiple healthcare institutions. It is worth noting that to address the challenge presented by literature works that most commonly use only one or two versions of CERT's dataset and to ensure a fair comparison for validating RH1, traditional, modern, and centralized

federated learning approaches were tested on the same dataset versions used for our Advanced Privacy-Preserving Decentralized Federated Learning (APPDFL) system as a proposal solution. The performance of the proposed APPDFL system as a proposed solution is then assessed and compared to both previous works review and centralized methods tested. The trade-off between accuracy and privacy enhancement is evaluated to determine the noise level tolerance that maintains acceptable model accuracy with privacy-preservation among all participants (Ahamed et al., 2024).

The proposed APPDFL system as a solution for insider threat detection in collaborative healthcare institutions significantly improves accuracy, enhances privacy, and outperforms both the centralized methods tested and existing work in the field. The findings clearly indicate that collaboration among different participating healthcare institutions for insider threat intelligence sharing is possible. This collaboration maintains superior effectiveness while ensuring the accuracy of the learning process and acceptable noise tolerance. Furthermore, this approach adheres to the core principles of information security: ensuring confidentiality, maintaining integrity, and guaranteeing availability of data—commonly known as the "CIA Triad." For the APPDFL System graphical abstract summary, refer to Figure A-1.

Table of Contents

Dedication ..... **iv**

Acknowledgements ..... **v**

Abstract of Praxis ..... **vi**

List of Tables ..... **xv**

List of Equations ..... **xvii**

List of Symbols ..... **xviii**

List of Acronyms ..... **xix**

Chapter 1—Introduction **1**

1.1 Background 1

1.2 Research Motivation 3

1.3 Problem Statement 7

1.4 Thesis Statement 10

1.5 Research Objectives 11

1.6 Research Questions and Hypotheses 11

1.7 Scope of Research 12

1.8 Research Limitations 13

1.9 Organization of Praxis 14

Chapter 2—Literature Review **15**

2.1 Introduction 15

2.2 Understanding Insider Threats and their Consequences 16

2.2.1 Insider Threat Types and Methods 16

2.2.2 Common Types of Insider Threats 18

2.2.3 Most Common Methods Used by Insiders 20
2.2.4 Real-life cases of insider threats to healthcare 22
2.2.5 Insider Threats in Healthcare Cybercrime 23
2.3 Machine Learning-Based Approaches for Insider Threat Detection 29
2.4 Federated Learning as a Potential Solution in Healthcare 38
2.4.1 Centralized Federated Learning Approach 45
2.4.2 Decentralized Federated Learning Approach 50
2.5. Data Imbalance and Balancing Techniques in Insider Threat Detection 51
2.6 Privacy-preserving techniques for data sharing 52
2.7 Differential Privacy-Preservation in Federated Learning 55
2.8 Summary and conclusion 58
Chapter 3—Methodology 59
3.1 Introduction 59
3.2 Implementing the Environment 62
3.2.1 Hardware and Software Setup 62
3.2.2 Data Sources Description 66
3.3 Data Preparation and Balancing Techniques 68
3.3.1 Data Preprocessing and Feature Engineering 68
3.3.2 Data Balancing and Data Splitting 74
3.4 Model Architecture Selection 79
3.4.1 Optimized Deep Neural Networks 81
3.4.2 Optimized Long Short-Term Memory 87
3.4.3 Hyperparameter Tuning 89

3.4.4 Comparative Analysis of Optimized DNN and Optimized LSTM 90
3.4.5 APPDFL Model Results and Performance using optimized DNN 92
Model Architecture 92
3.5 Balancing Privacy and Accuracy in Decentralized Federated Learning 93
Implementation Approach 93
3.5.1 Privacy Preservation Strategies in Decentralized Federated Learning .. 93
3.5.2 Horizontal Peer-to-Peer as Serverless Communication Method 94
3.5.3 Local Model Training and Model Aggregation Synchronous Advantage
..... 95
3.6. Performance Metrics of Applied Models 95
Chapter 4—Results 101
4.1 Introduction 101
4.2 Centralized Machine Learning Methods 103
4.2.1 Traditional and Modern Machine Learning Implementation Process . 103
4.2.1.1 Traditional and Modern Machine Learning Data Preparation 103
4.2.1.2 Traditional and Modern Machine Learning Summary 104
4.2.2. Modern Machine Learning Model Results and Performance 106
4.3 Centralized Federated Learning Method 107
4.3.1 Centralized Federated Learning Implementation Process 108
4.3.2 Centralized Federated Learning Model Results and Performance 109
4.3.3 Balancing Techniques in Centralized Federated Learning 111
4.4 Advanced Privacy-Preserving Decentralized Federated Learning 116
4.4.1 Effect of Noise Levels Budget on Accuracy of Advanced Privacy 121

Preserving Decentralized Federated Learning (APPDFL) System 121
4.4.2 The Optimal Privacy Budget for Balancing Accuracy and F1- 121
Score in APPDFL System 121
4.5 Comparative Analysis of TML, CFL and APPDFL testing evaluations 126
4.5.1 Traditional Machine Learning (TML) 126
4.5.2 Centralized Federated Learning evaluation 127
4.5.3 Proposed Advanced Privacy Preserving Decentralized Federated 127
Learning approach evaluation 127
4.6 Analysis of APPDFL Using Optimized DNN and Optimized LSTM 129
4.7 Hypothesis Evaluation and Validation 132
4.7.1 Hypothesis 1 (RH1) Validation 134
4.7.2 Hypothesis 2 (RH2) Validation 138
4.7.3 Hypothesis 3 (RH3) Validation 139
Chapter 5—Discussion and Conclusions 143
5.1 Discussion 143
5.2 Conclusions 144
5.3 Contributions to Body of Knowledge 146
5.4 Recommendations for Future Research 149
References 153
Appendix A 181

List of Figures

Figure 1. Characteristics of common data loss or exfiltration and their impact 5

Figure 2. Healthcare victim organization relationship to insiders. 8

Figure 3. Healthcare data breaches. 9

Figure 4. HIPAA penalty structure for violation. 9

Figure 5. Potential consequences of insider threat in healthcare institutions. 26

Figure 6. Federated Learning Structure Overview. 39

Figure 7. Common Federated Learning Topologies. 41

Figure 8. Centralized Federated Learning Architecture. 46

Figure 9. Decentralized Federated Learning Architecture. 50

Figure 10. Top Level Methodology Overview. 61

Figure 11. Interest over time - PyTorch (blue) vs TensorFlow (red). 65

Figure 12. Comparison of Insider and Non Insider in CERT's dataset. 75

Figure 13. Standard Base DNN Model Architecture Design. 84

Figure 14. Optimized DNN Model Architecture Design. 84

Figure 15. Performance Report of APPDFL using an optimized LSTM Model. 89

Figure 16. APPDFL results - Optimized DNN Model Architecture. 92

Figure 17. Insider Threat Confusion Matrix. 97

Figure 18. Modern Machine Learning Model Results. 107

Figure 19. Accuracy Results suing Balancing Techniques (CFL). 113

Figure 20. F1-Score Comparison with Balancing Techniques in (CFL). 115

Figure 21. Performance Report of CFL. 116

Figure 22. APPDFL Accuracy Comparison with Balancing Techniques..... 118

Figure 23. F1-Score Comparison Balancing Techniques in APPDFL. 120

Figure 24. Optimal Privacy-Accuracy Trade-Off in APPDFL. 125

Figure 25. APPDFL-DNN vs. APPDFL-LSTM Performance. 130

Figure A- 1. APPDFL System Graphical Abstract Summary. 181

Figure A- 2. High level workflow diagram for the implementation process. 181

Figure A- 3. Interest for PyTorch vs TensorFlow by regions. 182

Figure A- 4. Interest for PyTorch vs TensorFlow by sub-region. 182

Figure A- 5. Comparative Analysis of Metrics and Strategies used in this Praxis. 186

List of Tables

Table 1. Summary of the Commonly used CERT's Insider Threat Dataset Versions 67

Table 2. Hyperparameters of APPDFL 90

Table 3. Comparison of optimized DNN and optimized LSTM Model Architectures 91

Table 4. Pseudo-code for proposed APPDFL System 93

Table 5. Binary Confusion Matrix for Classification (Fahmy, 2022) 97

Table 6. The Performance Metrics 100

Table 7. Summary of Hyperparameters for TML 105

Table 8. Pseudo-code for TML 106

Table 9. Pseudo-code for Centralized Federated Learning 110

Table 10. Accuracy Results suing Balancing Techniques (CFL) 112

Table 11. F1-Score Comparison with Balancing Techniques in (CFL) 114

Table 12. APPDFL Accuracy Comparison with Balancing Techniques 117

Table 13. F1 Scores for APPDFL Models with Different Balancing Techniques 119

Table 14. Accuracy with Different Privacy Budget Levels 122

Table 15. F1 Score with Different Privacy Budget Level 124

Table 16. APPDFL vs. CFL: DNN and LSTM Results 130

Table 17. APPDFL-DNN vs. APPDFL-LSTM Clients Performance 131

Table 18. APPDFL vs Tested Methods 135

Table 19. APPDFL vs. Centralized Federated Learning Methods include tested CFL .. 136

Table 20. Results and Strategies: APPDFL vs. Tested Methods and Previous Works .. 136

Table 21. Hypothesis 1 (RH1) Evaluation Summary 137

Table 22. Hypothesis 2 (H2) Evaluation Summary 139

Table 23. Mean Performance comparison- Optimized-DNN and optimized- LSTM 141

Table 24. Hypothesis 3 (H3) Evaluation Summary 142

Table A- 1. Potential Consequences of Insider Threat in Healthcare Institutions. 183

List of Equations

Equation 1. Mathematical foundation of differential privacy condition..... 56

Equation 2. Adapted of differential privacy condition of two datasets 56

Equation 3. An optimized DNN (LeCun et al., 2015) 82

Equation 4. Optimized LSTM- Forget Gate Activation 87

Equation 5. Optimized LSTM- Input Gate Activation 87

Equation 6. Optimized LSTM- Output Gate Activation..... 87

Equation 7. Optimized LSTM- Candidate Cell State 88

Equation 8. Optimized LST- Cell State Update..... 88

Equation 9. Optimized LST- Hidden State Update 88

Equation 10. Accuracy (Bonnet, 2023)..... 98

Equation 11. F1-Score (Bobbitt, 2021*)*..... 98

Equation 12. Precision (Bonnet, 2023). 98

Equation 13. Recall (Bonnet, 2023)..... 99

Equation 14. Epsilon (Dwork, 2006). 99

Equation 15. The local model weights were updated using gradient descent in CFL 108

Equation 16. Aggregation at Central Server in CFL (McMahan et al., 2017)..... 109

List of Symbols

δ	Delta often represents a small quantity or error term
ε	Epsilon is a privacy loss parameter in differential privacy
σ	Sigma Standard deviation of a population
Pr	Probability
M(x)	Mechanism applied to dataset x
M(y)	Mechanism applied to dataset y
S	Subset of possible outputs
e	The base of the natural logarithm (approximately equal to 2.71828)
$\in$	Privacy loss parameter in differential privacy (epsilon)

List of Acronyms

1 NN	Nearest Neighbor
AE	Autoencoder
AI	Artificial Intelligence
AMD	Advanced Micro Devices
ANN	Artificial Neural Network
APPDFL	Advanced Privacy-Preserving Decentralized Federated Learning
CERT	Computer Emergency Response Team
CFL	Centralized Federated Learning
CIA	Central Intelligence Agency
CISA	Cybersecurity and Infrastructure Security Agency
CMU	Carnegie Mellon University
CNN	Convolutional Neural Network
DFL	Decentralized Federated Learning
DLP	Data Loss Prevention
DNN	Deep Neural Network
DP	Differential Privacy
EDR	Endpoint Detection and Response
EHR	Electronic Health Records
FBI	Federal Bureau of Investigation
FL	Federated Learning
FNN	Feedforward Neural Network
FN	False Negative
FP	False Positive
GANs	Generative Adversarial Networks
GDPR	General Data Protection Regulation
HFL	Horizontal Federated Learning
HIPAA	Health Insurance Portability and Accountability Act
HPH	Healthcare and Public Health

IAM	Identity and Access Management
IBM	International Business Machines Corporation
IDS	Intrusion Detection Systems
IF	Isolation Forest
IoHT	Internet of Health Things
IoMT	Internet of Medical Things
IoT	Internet of Things
LSTM	Long Short-Term Memory
MHz	Megahertz
ML	Machine Learning
MLP	Multi-Layer Perceptron
NAS	Neural Architecture Search
NATO	North Atlantic Treaty Organization
NIST	National Institute of Standards and Technology
NSA	National Security Agency
OCSVM	One-Class Support Vector Machine
P2P	Peer-to-Peer
PAM	Privileged Access Management
ReLU	Rectified Linear Unit
RF	Random Forest
SIEM	Security Information and Event Management
SNN-MLP	Self-normalizing Neural Network-based Multi-Layer Perceptron
SVM	Support Vector Machine
TML	Traditional Machine Learning
TP	True Positive
TN	True Negative
UAM	User Activity Monitoring
UEBA	User and Entity Behavior Analytics
VAE	Variation Autoencoder
VFL	Vertical Federated Learning

Chapter 1—Introduction

1.1 Background

Technological advancements and the need for more efficient and effective healthcare delivery have driven a significant transformation in the healthcare industry in recent years, which is crucial because our well-being and quality of life heavily depend on this sector (Amao, 2023). However, this digital transformation has resulted in a 35% increase in data breaches, and alarmingly, 82% of healthcare organizations struggle to assess the true extent of insider attacks' damage (Alder, 2023; U.S. Department of Health and Human Services, 2022). As defined by the Cybersecurity and Infrastructure Security Agency (CISA), insider threats are described as “the threat that an insider will use their authorized access to harm an organization” (CISA, n.d.). Additionally, while North Atlantic Treaty Organization [NATO] Cooperative Cyber Defense Centre of Excellence, (n.d.) highlighted the significant risks of ”good guys gone bad”. These threats often originate from employees or other individuals who have inside knowledge about an organization’s security protocols, sensitive information, and information technology systems, making them difficult to detect and prevent, often leading to significant damage.

Healthcare institutions typically collect and store substantial amounts of network data using various tools and solutions for insider threat detection, prevention, and protection, such as platforms that aggregate and analyze security data, such as those for Security information and event management (SIEM), and those that automate and coordinate security responses, such as security orchestration, automation, and response (SOAR) platforms, syslog servers, platforms that monitor and assess users and entity activities, refer to as “user behavior analytics (UEBA)”(Grafton, 2023), data loss

prevention (DLP) tools, systems for managing user identities and their access, known as "identity and access management (IAM) systems", and solutions for detecting and responding to threats at endpoints, known as "endpoint detection and response (EDR) solutions" (Vaideeswaran, 2023; Microsoft Security, n.d.), and privileged access management (PAM) tools. These platforms not only aid in network monitoring and security but also offer valuable data for training machine learning models, thereby enhancing threat detection and response capabilities.

However, traditional and modern machine learning methods that centralize data are used to detect insider threats and present significant challenges. These methods involve collecting information from multiple sources organizing it in a central location for processing and analysis. This centralization poses risks such as potential privacy violations, ownership disputes, and issues of regulatory non-compliance (Soares dos Santos, 2022). Sharing confidential information between multiple organizations to improve detection effort, without ensuring proper measures, can lead legal and logistical complications. This includes potential Health Insurance Portability and Accountability Act (HIPAA) violations, which could result in severe fines and penalties, as shown in Figure 4, detailing how the HIPAA penalty are structured and different categories and tiers of violation.

These challenges are influenced by multiple factors. These include the lack of uniform data collection methods, concerns about possible harm to an organization's reputation, challenges in building trust among participant, potential breaches of privacy or antitrust regulations, strict government classification requirements, and varying legal systems across different regions (Nepal, 2024). As Zibak and Simpson (2019) highlighted,

“these concerns have hampered the progress in collaborative threat detection and intelligence sharing”.

This Praxis proposes an innovative approach to address these challenges by implementing advanced privacy-preserving decentralized federated learning for insider threat detection in collaborative healthcare institutions. This approach facilitates large-scale, cross-institutional collaboration to detect insider threats, focusing specifically on healthcare institutions. The approach guarantees the confidentiality of sensitive information, including protected health information, and prevents its exposure during the insider threat detection process. The proposed Advanced Privacy-Preserving Decentralized Federated Learning for Insider Threat Detection in Collaborative Healthcare Institutions presents a solution that not only detects insider threats but also facilitates the adoption of intelligence sharing among collaborative healthcare institution participants. This practice provides a method for insider threat detection that leverages shared intelligence while maintaining data privacy, resulting in more secure and privacy-preserving practices in healthcare institutions.

1.2 Research Motivation

Nothing is more important than our health (Batra et al., 2019). Healthcare is an essential aspect of human life, with every individual relying on healthcare systems at various stages. In 2017, healthcare expenditures in the United States exceeded $3.5 trillion, making up 17.9% of the nation's gross domestic product, which amounted to approximately $10,739 per person (Centers for Medicare & Medicaid Services, 2018). This level of expenditure underscores the critical and essential role healthcare institutions play in providing care and treatment (Batra et al., 2019). Batra et al. (2019) also emphasized that

information regarding people, communities, organizations, and the environment will be central to the future of health. However, it is crucial to secure this data to prevent its alteration or exfiltration. Also, according to World Economic Forum (2019), hospitals generate around 50 petabytes of data annually, which emphasizes the importance of data in improving healthcare and highlights the need for innovative solutions, as the future of health is expected to be shaped by digital transformation, driven by highly interoperable data and secure, open platforms.

Digitalization has led to an exponential growth in healthcare data, enabling innovative approaches in patient care but also increasing the likelihood of insider threats. These threats pose a significant challenge, given their internal origin, which make them harder to detect and prevent. Insider threats are a crucial security threat for businesses, institutions, and government agencies, with healthcare institutions being no exception. As the sophistication and frequency of these threats increase, healthcare institutions face an enormous challenge in protecting themselves from these internal network threats. Proofpoint's report on the healthcare industry, a Ponemon Institute (2023) survey, highlighted that Malicious insiders and accidental data loss are the top two causes of data loss and exfiltration, which can disrupt patient care and potentially increase mortality rates. This report added that 46% of institutions stated that the data loss or exfiltration incident impacted patient care and increase mortality rate, as shown in the Figure 1, underscoring the critical need for enhanced security measures. Since individuals engage with the healthcare system in different capacities throughout their lives (Batra et al., 2019), these findings serve as key motivations for this research. The main objective is to advance

effective, scalable, and collaborative ways of insider threat intelligence sharing as a healthcare detection solution,

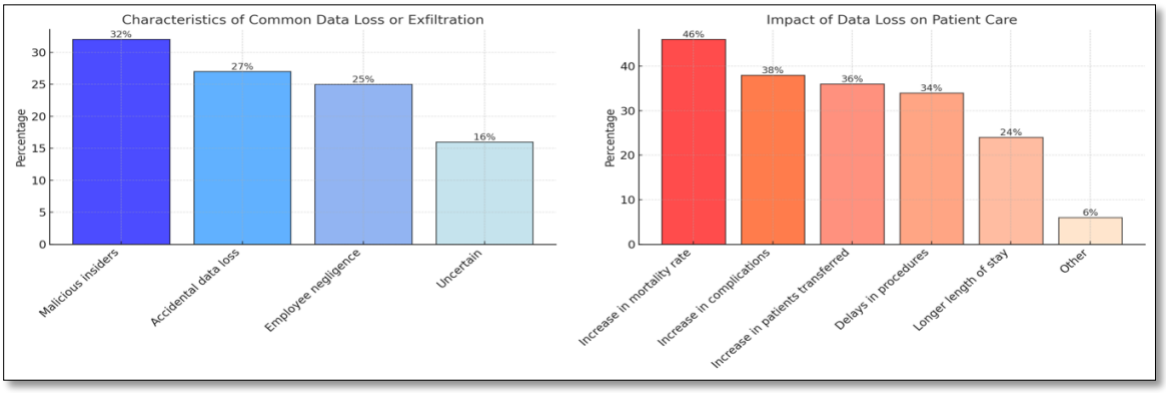


Figure 1. Characteristics of common data loss or exfiltration and their impact

Adapted from Ponemon Institute Report (2023).

recognizing the universal importance of healthcare improvements that can have widespread and lasting impacts on a large portion of the population.

Insider threats not only jeopardize the confidentiality, integrity, and availability of healthcare information but also have a direct impact on patient care (Compliancy Group, n.d.). Additionally, the designation of insiders cannot be defined simply by their employment status (Wilder, 2022). Protecting healthcare data is particularly challenging in complex organizations that involve internal employees, external stakeholders, and third-party contractors. A report from the Ponemon Institute reveals that insider threat incidents increased by 44% between 2020 and 2022, with costs per incident rising to as much as USD 15.38 million (Ponemon Institute, 2022). This evidence emphasizes the need for healthcare organizations to innovate new approaches to managing insider threats, particularly by examining insider applications and data access entitlements.

Existing traditional and modern machine learning methods for detecting insider threats offer both advantages and disadvantages. Machine learning algorithms are capable

of processing vast quantities of datasets and detecting patterns that human analysts might overlook, potentially leading to more accurate detection of insider threats and preventing costly data breaches. However, using machine learning as a solution presents various challenges, particularly when it comes to privacy. The Ponemon Institute (2022) reported that data breaches resulted in an estimated annual cost of $6.2 billion for the US healthcare industry. Moreover, the use of machine learning techniques frequently necessitates access to sensitive information, which can put patient privacy at risk if not properly managed. The American Medical Association (AMA)'s survey found that "more than 92% of patients believe privacy is a right and that their health data should not be available for purchase" (AMA, 2022; Lagasse, 2022), which demonstrates the significance of privacy concerns in healthcare. Furthermore, an earlier AMA report from 2021 found that "nearly 75% of patients expressed concern about protecting the privacy of personal health data, with only 20% saying they knew the scope of companies and individuals with access to their data" (AMA, 2021). These privacy concerns highlight the importance of finding solutions that balance the need for effective insider threat detection with the requirement to protect patient data.

Given the growing complexity of insider threats and the critical need for enhanced security, this research is motivated by the pressing necessity to protect healthcare data while enabling collaborative for insider threat detection. "One way organizations can stay ahead of attacks is through collaborative threat detection" (Nepal, 2024). This can be achieved by adopting and integrating advanced threat detection approaches that not only identify potential threats but also facilitate intelligence sharing among healthcare institutions. This research proposes an Advanced Decentralized Federated Learning

approach that detects insider threats collaboratively while allowing multiple healthcare institutions to participate privately. To the best of our knowledge, this is the first study to implement a decentralized federated learning-based approach for insider threat detection in healthcare, offering a novel way to leverage shared insider threat intelligence in healthcare institutions. The pressing need to protect healthcare data while enabling collaborative efforts in threat detection motivated this research.

1.3 Problem Statement

Existing centralized machine learning approaches for insider threat detection in healthcare institutions face significant challenges related to scalability, privacy, and the effective adoption of collaboration that facilitates intelligence sharing among participants.

The current landscape of insider threat detection largely relies on traditional machine learning techniques, and healthcare is no exception. While existing traditional centralized machine learning approaches have proven to be effective in addressing these existing threats in certain scenarios, they exhibit limitations, particularly when applied to sensitive data such as those found in healthcare settings. The following points, based on the HIPAA Journal and Ponemon Institute's 2023 reports, provide more specific insights into the problem of insider threats in healthcare:

- The HIPAA Journal report on insider threats in healthcare revealed that 53% of all healthcare data breaches were caused by insiders (Alder, 2022).
- The HIPAA Journal report highlighted that over 50% of organizations have experienced an insider threat within the past year (Alder, 2023).

- The Ponemon Institute report (2023) indicated that “insider attacks have become more frequent, with 74% of organizations stating that they are at least moderately vulnerable or worse to insider threats”.

These threats have seen a significant rise in recent years so is the compliance violation penalty. The Figures 2, illustrates the growing severity of breaches caused by direct and indirect insider threats, which can include employee errors, negligence, unauthorized access to medical records, and intentional data theft by malicious insiders. To put it in perspective, from 2009 to 2023, Alder (2024) from the HIPAA Journal stated that “there were 5,887 reported instances of healthcare data breaches involving 500 or more records.” The author added that “these breaches led to the exposure or unauthorized disclosure of 519,935,970 healthcare records that resulted in HIPAA violation fines” (Alder, 2024). This trend emphasizes the pressing need for innovative approaches to detect and prevent insider breaches in healthcare institutions. For more details, refer to Figure 3 and Figure 4, adapted from HIPAA Journal by Alder (2024).

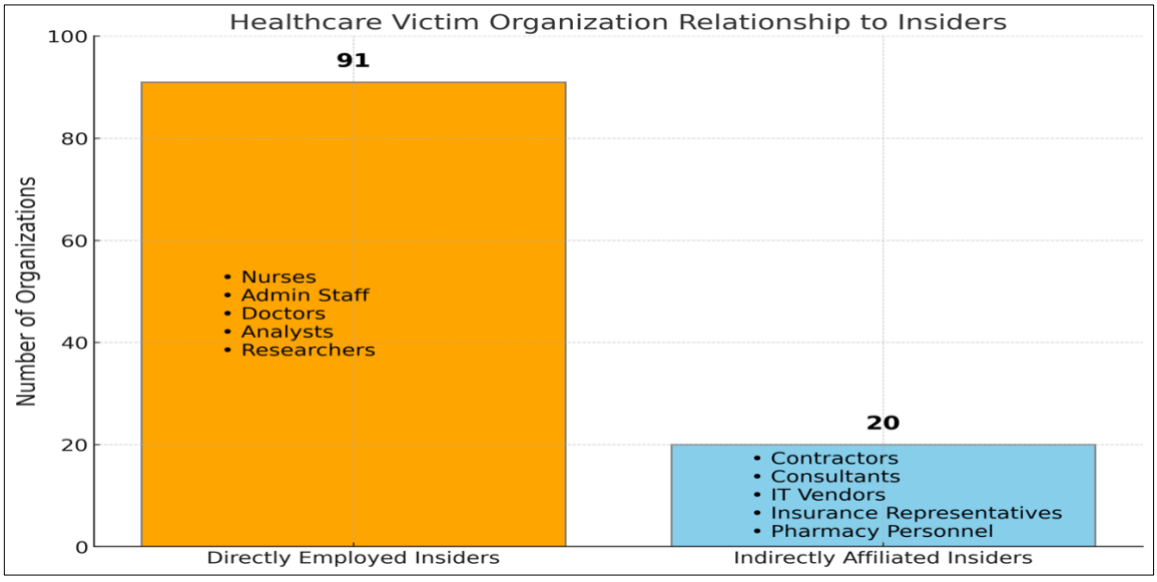


Figure 2. Healthcare victim organization relationship to insiders.

Adapted from CERT (2019).

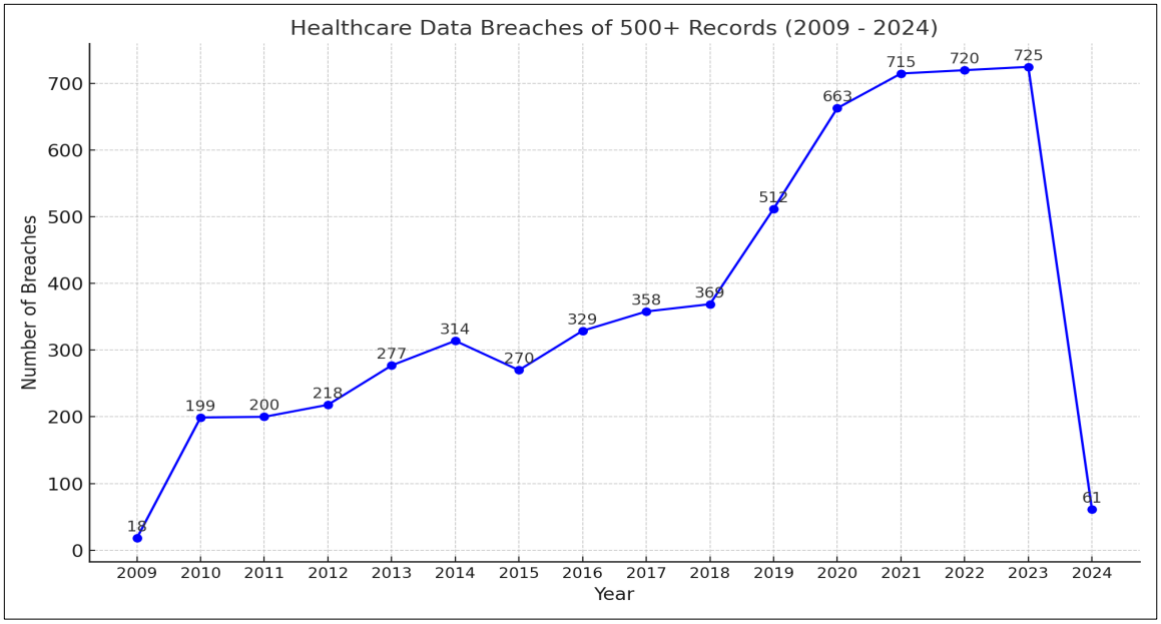


Figure 3. Healthcare data breaches.

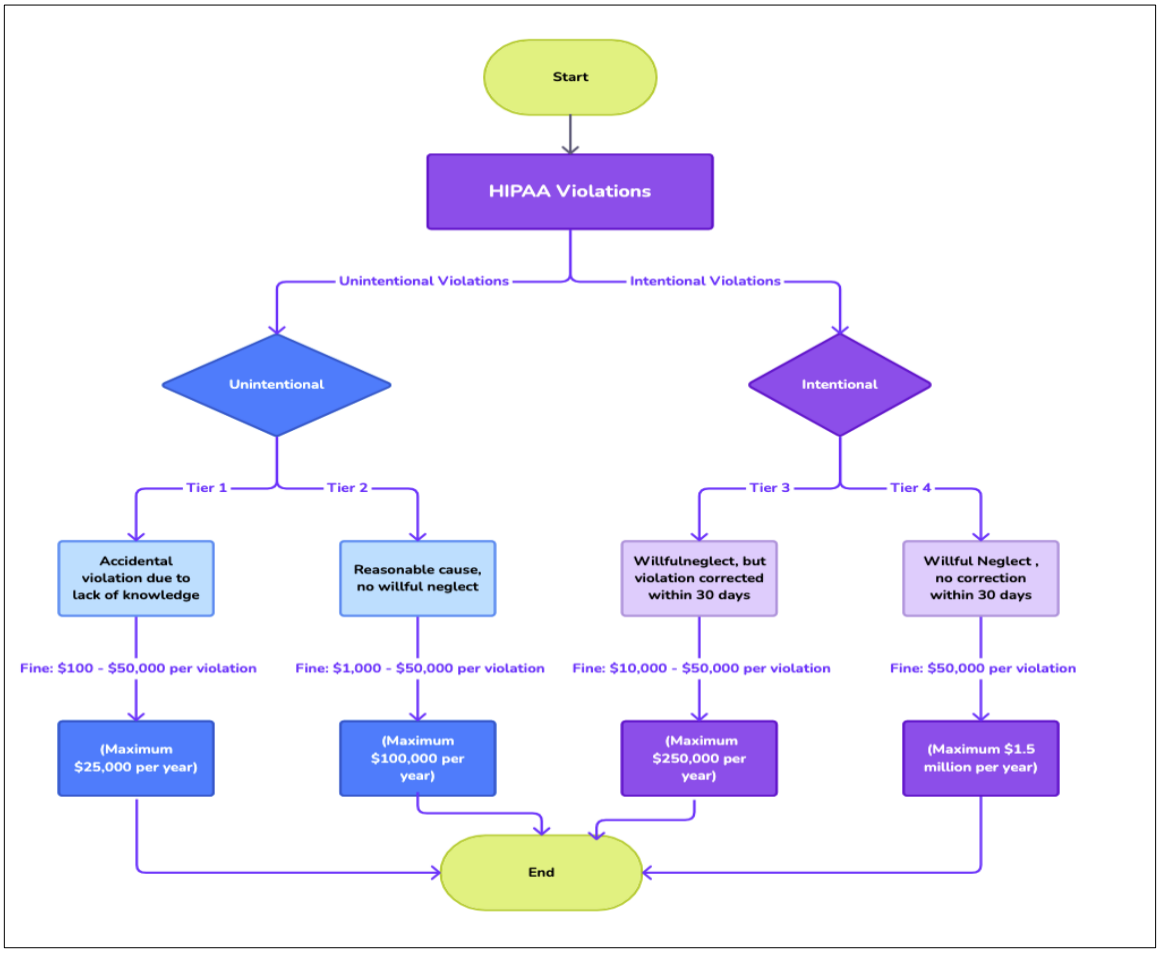


Figure 4. HIPAA penalty structure for violation.

The adoption and integration of healthcare artificial intelligence (AI) technologies are emerging rapidly, prompting an increasing discourse on how to regulate their development. A primary area of concern involves the access to, use of, and control over patient data in healthcare analysis and research, which could involve insider threats if not properly protected, as insiders with full access may misuse their privileges, either unintentionally or with malicious intent. Murdoch (2021) demonstrates that the deployment of AI in healthcare raises major concerns about data privacy and security, as "AI algorithms often require access to large amounts of data, posing privacy risks if not properly managed" (Sindhwani et al., 2024).

Furthermore, these methods of detecting insider threats struggle with effective threat intelligence sharing among organizations, including healthcare institutions (Burrell, et al., 2022; Youssef, et al., 2023). This is due to obstacles such as "the need for standardization, reputational risks, trust establishment challenges, and varying legal frameworks across jurisdictions" (Zibak & Simpson, 2019). Consequently, the healthcare sector has not been able to fully leverage the benefits of collaborative insider threat detection because healthcare institutions, such as hospitals, are particularly impacted by data privacy concerns.

1.4 Thesis Statement

The proposed Advanced Privacy-Preserving Decentralized Federated Learning System for insider threat detection in collaborative healthcare institutions maintains confidentiality without the need for a central coordinator and enables secure and effective insider threat detection in collaborative healthcare institutions without compromising data privacy.

An advanced privacy-preserving decentralized federated learning approach offers a novel way of effectively detecting insider threats and sharing threat intelligence data among collaborative healthcare institutions. This approach addresses the security and privacy issues associated with existing traditional centralized machine learning methods along with centralized federated learning approaches, enabling healthcare institutions to collaborate on insider threat detection and intelligence-sharing initiatives.

1.5 Research Objectives

The objective of this Praxis research is to implement an advanced privacy-preserving decentralized federated learning system that aims to enhance security, accuracy, and efficiency for insider threat detection among participants in collaborative healthcare institutions.

This approach also facilitates intelligence sharing and collaborative adoption, addressing existing limitations and outperform current insider threat detection approaches in healthcare institutions.

1.6 Research Questions and Hypotheses

Given the above Praxis research objectives, this research addresses the following research questions (RQs) and research hypotheses (RHs):

RQ1: How does the proposed Advanced Privacy-Preserving Decentralized Federated Learning System perform in terms of detecting insider threats, compared to centralized federated learning?

RQ2: In the proposed Advanced Privacy-Preserving Decentralized Federated Learning System what is the optimal trade-off point value to ensure sufficient protection and acceptable accuracy of insider threat

detection across collaborative healthcare institutions?

RQ3: Which Model Architecture, Deep Neural Network (DNN) or Long Short-Term Memory (LSTM), yields higher classification performance for insider threat detection when integrated with selected balancing techniques (SMOTE, ADASYN, Near Miss, SMOTE-ENN, Tomek Links, or SMOTE-Tomek Links) within the Advanced Privacy-Preserving Decentralized Federated Learning System?

H1: The proposed Advanced Privacy-Preserving Decentralized Federated Learning System will detect insider threats with greater accuracy, than centralized federated learning.

H2: The analysis of the proposed Advanced Privacy-Preserving Decentralized Federated Learning System will identify an optimal trade-off point value that ensures sufficient privacy protection while maintaining acceptable accuracy in insider threat detection across collaborative healthcare institutions.

H3: The optimized Deep Neural Network (DNN) Model Architecture, integrated with the Tomek Links balancing technique in the proposed Advanced Privacy-Preserving Decentralized Federated Learning System, yields higher classification performance for insider threat detection than the Long Short-Term Memory (LSTM) architecture.

1.7 Scope of Research

This research focuses on the implementation of an Advanced Privacy-Preserving Decentralized Federated Learning (APPDFL) System aimed at detecting insider threats in

collaborative healthcare institutions. The Praxis covers how decentralized federated learning, integrated with differential privacy mechanisms and optimized model architectures such as deep neural networks (DNN) or long short-term memory (LSTM), integrates with various data balancing techniques to yield higher classification performance for insider threat detection while ensuring privacy preservation and collaborative adoption in healthcare institutions.

In this research, we used the three most commonly utilized CERT insider threat dataset versions r4.2, r5.2, and r6.2, each with their diverse insider threat simulation scenarios. Due to the absence of utilization of three or more CERT insider threat datasets versions, and the lack of decentralized federated learning cases for insider threat detection in previous work, this research also tested centralized methods on the same CERT datasets used in our proposal. This research does not cover datasets from real-world healthcare institutions due to the common challenge of finding such real-world datasets for insider threat detection. For the overall implementation process, refer to Appendix A-2.

1.8 Research Limitations

Our approach is designed to demonstrate that the advanced privacy-preserving decentralized federated learning system for insider threat detection in collaborative healthcare institutions outperforms existing machine learning methods for insider threats. Although our approach outperforms existing insider threat detection methods by facilitating the collaborative adoption of insider threat detection in a privacy-preserving way, certain limitations that may impact the application of our solution in the real world must be acknowledged.

This Praxis did not assess how well this solution would accurately detect insider threats and scale in non-healthcare institution environments. This research additionally evaluated the utilization of the CERT insider threat dataset, focusing only on known attack scenarios to simulate real-world case scenarios (Le, 2021; Kim et al., 2019). For the details recommendations for future research, Refer to Chapter 5.

The detection approach is evaluated only on metrics such as insider threat detection accuracy, F1-score, privacy-preserving capability, and the facilitation of insider threat detection intelligence sharing that allows healthcare institution participants to collaborate, compared to centralized insider threat detection methods.

1.9 Organization of Praxis

This Praxis is organized into five chapters. Chapter 1 presents the research topic. Chapter 2 delivers a comprehensive literature review of previous research on insider threat detection generally and particularly in healthcare institutions, highlighting the methods and mechanisms used in this research domain. Chapter 3 explains our methodology, while Chapter 4 presents our results and analysis, comparing its performance to other works as discussed in the literature review. Chapter 5 summarizes our findings, discusses limitations, and provides recommendations for future research within the interdisciplinary field of decentralized federated learning, differential privacy as privacy-preservation, to advance collaborative approaches for healthcare institutions in insider threat detection.

Chapter 2—Literature Review

2.1 Introduction

Insider threats pose a significant risk to organizations across all industries, and healthcare is no exception. Organizations across all industries face substantial risks due to insider threats (Mazzarolo & Jurcut, 2019), and healthcare is no exception. For example, the second and third annual reports by Ponemon Institute [Proofpoint](2023; 2024), and McKeon (2024) revealed that in healthcare, insider threats remained a top risk to healthcare institutions in 2023 and 2024.

This chapter is organized into seven clearly defined main sections to provide the context and background of the problem addressed in this Praxis, along with the previous works to be discussed in this literature review. Section 2.1 presents this chapter and its structure. Section 2.2 delves into understanding insider threats, defining their types and methods, discussing their complexity within healthcare, and highlighting the consequences of such threats, including compliance issues like HIPAA and GDPR. Section 2.3 dives into the most common machine learning approaches for insider threat detection, explores how they each work, summarizing previous works, along with their strengths and limitations. Section 2.4 introduces federated learning (FL) as a potential privacy-preserving solution in healthcare, presenting FL approaches, their architectures, and categories, explaining how they each work, reviewing FL applications in general, and highlighting each FL approaches advantages and disadvantages, particularly highlighting the potential benefits of adopting FL for collaborative insider threat detection in healthcare institutions. Section 2.5 explores the challenges of data imbalance, examines various techniques used to handle imbalanced datasets, reviews previous works addressing this

issue, and highlights their contributions and impact on model accuracy and reliability. Section 2.6 discusses privacy-preserving strategies in federated learning, explores their applicability in healthcare, highlights the importance of these mechanisms in maintaining data confidentiality, integrity, and availability, and emphasizes the need for collaborative insider threat detection intelligence sharing by leveraging decentralized federated learning (DFL) incorporating differential privacy mechanism for collaborative healthcare institutions. Section 2.7 discuss about differential privacy in federated learning, finally, Section 2.8 summarizes the literature review, identifying gaps and limitations in existing works. It also, clearly highlights our decision to implement traditional and modern machine learning, as well as centralized federated learning methods, motivated by the need to conduct a fair comparison using the same datasets for our proposed approach. Our proposed approach will be evaluated and compared to both existing work and these tested methods, with detailed results analysis presented in Chapter 4 (For further details, refer to "Chapter 4—Results"), establishing the necessity for this Praxis to develop an advanced privacy-preserving decentralized federated learning approach tailored for insider threat detection in collaborative healthcare institutions.

2.2 Understanding Insider Threats and their Consequences

2.2.1 Insider Threat Types and Methods

Insider threats pose a significant security risk in many sectors, especially in healthcare, given the highly critical nature of the data, which must be kept private, including electronic health records (EHR) and patient information (Appari & Johnson, 2010). The complexity of these threats stems from the diversity of potential threat actors and their methods. Within the Healthcare and Public Health field, insider threats can

originate from individuals within healthcare organizations or contractors who access sensitive information (CISA, n.d.; Gayathri et al., 2024; Alder, 2022). Such misuse can severely compromise patient privacy, institutional integrity, and operational capacity. Insider threats often result in severe consequences, including data breaches and disruptions in patient care, making detection and prevention particularly challenging because these individuals have legitimate access to systems.

Insider threats can be broadly classified based on motivation and methods, a crucial step for developing effective detection and prevention strategies. According to Chinchani and Iyer (2005), the three main types include malicious insiders, negligent insiders, and third-party insiders. Those with malicious intent intentionally misuse their access for harmful purposes, often driven by personal grievances or the desire for financial gain (McKeon, 2024). In contrast, negligent insiders pose risks through unintentional errors due to carelessness or lack of awareness, potentially leading to significant legal and reputational consequences (Kim et al., 2019). Additionally, third-party insiders, such as contractors, may inadvertently or deliberately misuse their access, increasing the risk of data breaches, particularly when security protocols do not adequately cover these external relationships (U.S. Department of Health and Human Services, 2022).

Understanding these categories and the methods employed by insiders is essential for developing robust detection and prevention strategies in healthcare environments. Common methods include data theft, sabotage, and unauthorized access. Data theft involves insiders stealing sensitive information, which may cause considerable financial losses and harm to reputation (Chinchani & Iyer, 2005). Sabotage can disrupt critical operations, posing direct threats to patient safety (McKeon, 2024). Unauthorized

access occurs when insiders misuse their privileges, potentially leading to incorrect treatments or diagnoses (U.S. Department of Health & Human Services, 2022). Real-life incidents, such as the Stradis Healthcare hack and the Texas Hospital breach, emphasize the seriousness of internal threats, highlighting the urgent need for strong protective measures to secure essential healthcare data and operations.

2.2.2 Common Types of Insider Threats

Healthcare organizations need to recognize the various types of insider threats that can compromise their systems. Each type whether driven by malicious intent, accidental negligence, or external influences has the potential to cause significant harm. The following analysis explains the characteristics of these insider threats and their impact on healthcare security:

- **Malicious Insiders:** These individuals, either direct employees or associated with the organization, misuse their authorized access to critical systems or data with harmful intentions. As highlighted by Chinchani and Iyer (2005) and Fortinet (n.d.), these insiders with malicious intent may steal patient information, compromise data, commit fraud, or sabotage healthcare operations. In healthcare institutions, the risks are notably significant, as malicious actions can result in financial losses as well as jeopardize patient safety. Additionally, McKeon (2024) underscores that personal grievances, financial gain, or external coercion frequently motivate malicious insiders, rendering them a particularly dangerous and unpredictable threat. Malicious insiders pose a significant challenge because they often have deep knowledge of the organization's systems, making it difficult to detect their activities early.

- **Unintentional Insiders:** Unintentional insiders pose significant risks to healthcare organizations through accidental security breaches. These breaches often occur due to carelessness or a lack of cybersecurity awareness. Kim et al. (2020) note that unintentional insiders can overlap with negligent insiders, as both arise from a lack of vigilance or awareness. Negligent insiders may ignore protocols, misplace devices or documents containing sensitive data, or use weak security practices such as opting for simple passwords or writing them down in unsecured places. Additionally, there are other types of unintentional insiders, including ignorant insiders who lack proper cybersecurity training and may unknowingly compromise security, phished insiders who are tricked by phishing schemes and accidentally reveal sensitive information, and misled insiders who can be manipulated through social engineering tactics into disclosing confidential information. Essentially, negligent insiders act as if security isn't their problem-until it is (Alder, 2023).

  In healthcare, where compliance with laws like HIPAA and GDPR is mandatory, such errors can lead to significant legal and reputational consequences. Although these insiders do not intend harm, their actions can result in severe outcomes. und erscores the importance of thorough employee training and robust security awareness initiatives (Kim et al., 2019; Alder, 2022; Vasko and Gardner, 2019).
- **Third-Party Insiders:** Third-party insiders refer to external contractors, vendors, or business associates who have legitimate access to an organization’s systems but may misuse this access, either deliberately or through negligence. The U.S. Department of Health & Human Services (2022) points out that third-party insiders are an increasing concern in healthcare, as institutions frequently collaborate with

external service providers, ranging from IT vendors to medical suppliers. The management of insider threats becomes more complex due to these relationships, as the organization's security policies may not fully integrate third-party partners' access to critical infrastructure. This access raises the likelihood of unintentional data breaches, particularly if third parties do not maintain the same level of security as the healthcare institution itself. Thus, healthcare organizations must ensure stringent third-party security protocols are in place.

- **Collusive Threats:** Collusive threats are types of employees working with external attackers to compromise the organization (Fortinet, n.d.). When employees team up with external attackers, the threat goes beyond everyday breach. This mix of internal access and external attack methods makes detection tough, as it can exploit both human and system vulnerabilities. In healthcare particularly, this level of collaboration can lead to severe consequence include patient data theft, fraudulent insurance claim like patient data manipulation to file false insurance claims, Medical theft by using stolen identity to temper with medical supplies or pharmaceuticals, and supply chain compromise by collaborating to tamper with medical supplies or pharmaceuticals.

2.2.3 Most Common Methods Used by Insiders

Insiders, whether malicious or negligent, can use a variety of methods to compromise security in healthcare organizations. The most common methods include data theft, sabotage, and unauthorized access, each of which has the potential to cause severe operational, financial, and legal consequences. The following analysis outlines the primary methods used by insiders to undermine the security of healthcare institutions:

- **Data Theft:** Data theft remains one of the most common and damaging methods used by insiders, particularly in healthcare. Insiders, especially malicious ones, may steal confidential information such as patient records for personal gain, to sell on the black market, or to facilitate identity theft or medical fraud. Chinchani and Iyer (2005) highlight that the high value of medical records makes healthcare institutions especially attractive targets for such theft. Data theft can result in the disclosure of sensitive patient information, which may damage the reputation of the institution and lead to costly legal settlements and regulatory fines.
- **Sabotage:** These are intentional acts designed to harm or interrupt the organization's activities, also known as insider sabotage. McKeon (2024) illustrates that sabotage in healthcare can involve corrupting patient records, disabling critical medical devices, or causing system outages. Given the reliance on accurate data and functioning systems for patient care, sabotage can have catastrophic consequences, including delayed or incorrect treatments that may endanger patient lives. This attack method underscores the importance of advancing intelligence sharing for insider threat detection in a collaborative way that allows robust monitoring systems and swift incident response protocols to prevent operational damage from insiders.
- **Unauthorized Access**: Unauthorized access occurs when insiders misuse their privileges to retrieve, alter, or delete sensitive data without proper authorization. The U.S. Department of Health and Human Services (2022) explains that in healthcare, unauthorized access may involve the manipulation of patient records, leading to incorrect diagnoses or treatments. Personal reasons, such as curiosity or

malicious intent, may motivate such actions. The potential for unauthorized access underscores the importance of strict access control policies, role-based access management, and continuous auditing of system activity to detect abnormal insider behavior.

2.2.4 Real-life cases of insider threats to healthcare

The following are some notable recent real-world incidents caused by insider threats in the healthcare industry, emphasizing the urgent need for advancing insider threat detection solutions:

⇒ **Stradis healthcare hack:** In 2020, following his furlough, Christopher Dobbins, a disgruntled former employee, used a covert account to access and delete critical shipping data at Stradis Healthcare. As highlighted by both the U.S. Attorney's Public Affairs Office (2020) and the Federal Bureau of Investigation (FBI, 2021), his actions delayed the delivery of essential personal protective equipment (PPE), significantly impacting the organization's operations during a critical period. This case illustrates how insider threats can directly disrupt supply chains and essential services in healthcare.

⇒ **Texas hospital breach:** The Center for Internet Security (n.d.) details the case of Jesse McGraw, a hospital employee in Texas, who created a botnet within the hospital's network, compromising medical devices and the HVAC system. McGraw's actions posed a direct threat to patient safety by disrupting critical hospital operations. This incident underscores the dangers posed by insiders with technical expertise and access to sensitive infrastructure.

⇒ **Pharmaceutical espionage:** A case of industrial espionage occurred when an insider at a U.S. pharmaceutical company downloaded over 12,000 confidential records before leaving to work for a competitor. The U.S. Department of Health and Human Services (2022) highlights this case as an example of how insider threats can extend beyond hospitals to include pharmaceutical companies, where the theft of proprietary information can cause significant financial and reputational damage.

These cases, along with others not stated here, highlight the severity of internal threats within the healthcare sector. Losing or misuse of sensitive data regarding financial and operational disruptions, patient safety, and institutional reputation may result in costly legal settlements and regulatory fines. These incidents underscore the critical importance of developing advanced detection models to proactively identify and mitigate risks, thereby protecting sensitive patient information and ensuring the integrity and trust of healthcare systems.

2.2.5 Insider Threats in Healthcare Cybercrime

Healthcare organizations face unique and complex challenges with insider threats, particularly in the context of cybercrime. The sensitive nature of healthcare data, combined with the extensive collaboration required across multiple entities to deliver patient care, makes the sector an attractive target for cybercriminals. Traditional cybersecurity measures often focus on defending against external threats, but insider threats present a distinct and significant risk. Insiders, who have legitimate access to internal systems, complicate detection and prevention efforts (Probst et al., 2010). This challenge is exacerbated by the

collaborative nature of healthcare institutions, where multiple entities-including laboratories, pharmacies, and external vendors-require access to sensitive data.

Insider threats can be driven by various motivations, vulnerabilities, and opportunities. According to Al-Mhiqani et al. (2020), insider threats are often categorized into three primary categories: traitors, masqueraders, and unintentional insiders. Traitors refer to individuals who intentionally misuse their access in order to cause harm or steal information for their own benefit. For example, a researcher at Nationwide Children's Hospital, located in Columbus, Ohio, was convicted of selling research data to China, compromised valuable research and patient safety (Office of Public Affairs-U.S. Department of Justice, 2021; Alder, 2023). Similarly, "a clinic worker stole the protected health information of mentally ill patients and sold it to identity thieves, resulting in a $232,612 tax loss for the IRS" (Alder, 2017). These examples demonstrate how traitors can exploit their legitimate access for financial or political gain, severely compromising healthcare institutions.

Masqueraders are individuals who illegitimately gain access to an insider’s credentials, posing as the insider to perform malicious activities. This can include external attackers or even other insiders. In one case, a medical biller in Tampa Bay, Florida, faced conviction for healthcare fraud and identity theft, having used stolen credentials to perpetrate fraud (Alder, 2018). Masqueraders present a significant challenge because they often go undetected, appearing as legitimate insiders.

Unintentional insiders unknowingly expose sensitive data or compromise systems due to negligence or lack of awareness. This group often includes employees who fall victim to phishing attacks or inadvertently violate security protocols (Greitzer, Moore, &

Cappelli, 2008). A notable example involved a former employee who retained access to sensitive systems after leaving the organization, resulting in HIPAA fines (Alder, 2020). Although these insiders may not have malicious intent, their actions can still lead to severe consequences for healthcare institutions.

In addition to these types of threats, personal grievances often lead to insider attacks. Employees may abuse their access to healthcare records out of spite or revenge. A typical example involves a former coordinator at the University of Pittsburgh Medical Center who accessed patients' medical records without authorization, then used the information to cause harm. This individual was sentenced to one year in prison (Alder, 2019). These types of breaches violate patient privacy and can cause long-term reputational damage to healthcare institutions.

Healthcare institutions also face threats from snooping, where insiders access patient records for unauthorized purposes. In one case, security guards at Washington Hospital accessed medical records without proper authorization, claiming they did so for entertainment purposes. Such incidents highlight the importance of access control systems to prevent unauthorized access (Alder, 2023). Compromised insiders—employees coerced or blackmailed by external attackers pose a unique challenge. External entities can use stolen data to blackmail employees into providing additional access, creating a situation where compromised insiders enable malicious activities (Alder, 2020). These insiders complicate detection effort, as their actions often seems legitimate but are stem from manipulation or extortion. These case examples illustrate that insider threats in healthcare are diverse and complex, with each case presenting its own set of challenges for detection and prevention.

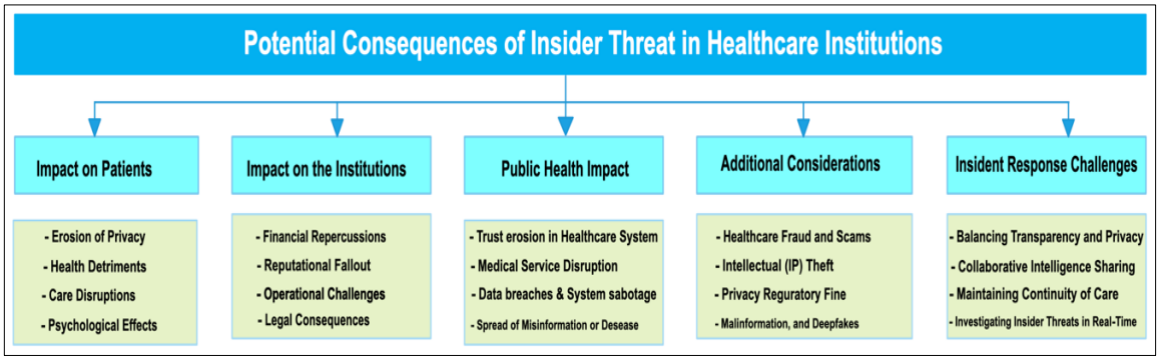


Figure 5. Potential consequences of insider threat in healthcare institutions.

The consequences of insider threats can be significant, leading to financial loss, reputational damage, and even compromised patient care, as illustrates in Figure 5. Gaining unauthorized access or disclosure of patient data can results in violations of patient privacy, with potential legal and financial implications for the institution. As highlighted by Appari and Johnson (2010), the sensitivity of healthcare data, which includes personal health information, financial data, and other confidential information, amplifies the severity of these threats.

Also, insider threats can have devastating consequences, causing financial loss, reputational damage, and even harm to patients in the healthcare sector as detailed in Table A-1, in the Appendix. Moreover, the alteration or destruction of patient data can directly impact patient care, potentially leading to incorrect diagnoses or treatments, which could result in severe harm or even loss of life. The financial impact of these threats is also substantial, as healthcare institutions must invest in remediation, legal fees, and potential fines while also dealing with the long-term damage to their reputation (Greitzer et al., 2010).

Detecting insider threats in healthcare cybercrime presents unique challenges due to the overlap of legitimate internal access with malicious intent. Insiders with existing access to the organization's systems are particularly dangerous, as traditional security tools often fail to identify deviations in behavior when compared to external attacks (Kim et al., 2019). Unlike industries where system access is more compartmentalized, healthcare institutions must allow widespread access to different departments, providers, and third-party entities (McKeon, 2024). This level of openness complicates the control and monitoring of data flows, especially when combined with the heavy reliance on external vendors who may not always adhere to the same security standards (Ostendorf, 2023). The interconnected nature of healthcare systems further exacerbates this issue, as a breach in one area can have cascading effects across other critical functions.

Detecting ransomware attacks presents a significant challenge, particularly when insiders unintentionally expose systems through phishing schemes or other malicious activities (U.S. Department of Health and Human Services, 2022). The healthcare sector's reliance on continuous data availability heightens the risk, as ransomware can disrupt essential services and cause life-threatening delays in patient care. Insiders may unknowingly or directly enable these attacks, further complicating detection efforts (McKeon, 2024). Additionally, Internet of Medical Things (IoMT) devices introduce new vulnerabilities. Insiders with technical expertise can tamper with these devices, causing malfunctions or using them as gateways for larger cyberattacks. This not only presents a significant patient safety risk but also destabilizes healthcare operations (Kim et al., 2019).

As highlighted throughout this Praxis, safeguarding sensitive patient data and protection against insider threats in healthcare institutions are paramount. These threats can

arise directly from within the organization or indirectly through third-party vendors (Skahill and West, 2021; UPMC HealthBeat, 2024). For a detailed visual representation of the relationship chart between the healthcare victim organization and insiders, refer to Figure 2.

To detect and mitigate these threats, healthcare institutions increasingly rely on machine learning algorithms, behavioral analytics, and other technologies to identify abnormal behaviors by analyzing patterns and detecting potential threats. They also analyze and monitor user behavior to detect anomalies that may indicate insider threats or compromised accounts (Grafton, 2023).

By identifying unusual patterns, healthcare institutions can proactively address potential security breaches. User Activity Monitoring provides real-time visibility of user actions, aiding in swift detection of suspicious activities. RBAC plays a crucial role by restricting insider access to specific job functions (Kim et al., 2019), thereby ensuring that only authorized individuals have access to sensitive information. This minimizes the risk of unauthorized access and data breaches by insiders (Ataman, 2024). Ataman (2024) highlighted that DLP is essential for protecting sensitive data from loss, misuse, or unauthorized access particularly in the healthcare sector where maintaining patient confidentiality and regulatory compliance are crucial. While IDS typically serves as a tool for detecting external threats, it can also detect insider threats by observing unusual patterns in internal network traffic. It monitors network traffic for signs of malicious activities, helping in early detection and response. However, relying solely on IDS for defending enterprise networks is not advisable (Rivera, 2023).

Aside from their role in healthcare settings, integrating these tools within healthcare institution is critical for enhancing security and protection against both direct and indirect insider threats. This integration is not without challenges, particularly given the strict data privacy laws healthcare institutions must follow, such as healthcare insurance portability and accountability act (HIPAA). Additionally, while these tools can detect suspicious behaviors patters, they are not sufficient on their own for comprehensive security. A multilayered approach, in a combination with machine learning and other advanced analytics to enhance detections capability are necessary (Soleymanzadeh & Kashef, 2023; Ibrahim et al., 2012; Sarhan & Altwaijry, 2023), because many of these tools are not yet fully effective in differentiating between normal and malicious insider behavior, resulting in an increased of false positives and false negatives rates (McKeon, 2024). Continued innovation is necessary to strike a balance between privacy, effective detection, and scalability of insider threat detection systems within healthcare institutions.

2.3 Machine Learning-Based Approaches for Insider Threat Detection

Nowadays, it is debatable that machine learning (ML) has surpassed its parent, artificial intelligence (AI), as it is widely implemented in numerous fields and adopted in various sectors, including data security, image analysis, automotive industry, astronomy, finance, robotics, navigation, military decision-making, ecommerce, education, and healthcare, among others (Karthikeyan & Priyakumar, 2022; Biswal, 2024). Many scholars and organizations assert that a fundamental aspect of intelligent behavior is the ability to learn, thereby making machine learning (ML) one of the fastest-growing areas within the field of AI (Jordan & Mitchell, 2015; McKinsey & Company, 2022; Grand View Research, 2023). Additionally, the George Washington University (2023), highlights that a wide

range of fields can apply machine learning as a powerful tool to solve complex problems. This progression underscores the pivotal role ML plays in advancing AI capabilities, particularly by enabling the analysis of large datasets and identifying patterns that inform AI decision-making processes. This transition demonstrates how ML significantly enhances the effectiveness and precision of AI applications.

According to Rodriguez (2022), machine learning (ML) leverages domain-specific data to detect patterns and generate valuable knowledge. ML's ability to acquire knowledge from data and predict outcomes is the foundation of this capability. These methodologies have been extensively utilized to analyze user behavior, system logs, and network activities. Alpaydin (2014) categorizes ML into four types: Alpaydin (2014) categorizes ML into four types: supervised, unsupervised, semi-supervised, and reinforcement learning. In supervised learning, labeled datasets are employed for training, while unsupervised learning seeks patterns in unlabeled data (Sun et al., 2023). Semi supervised learning combines elements of both, and reinforcement learning involves learning from past decisions to optimize future actions (Alpaydin, 2014). These methodologies have been extensively utilized to analyze network activities, system logs, or user behavior. Modern techniques such as long short-term memory (LSTM) and deep neural networks (DNN) networks enhance traditional approaches. DNNs can identify complex patterns in large datasets, while LSTM networks effectively model sequential data like user activity logs. Together, these advancements lead to a proactive and robust method for detecting insider threats, offering a thorough defense mechanism against potential security breaches.

Several studies have explored both traditional and modern ML techniques for insider threat detection. For example, Sarhan and Altwaijry (2023) emphasize the

importance of these methods in swiftly identifying, analyzing, and mitigating security breaches, including insider threats. In addition, the intersection of machine learning, cybersecurity, and healthcare has seen significant advancements, particularly with the advent of the Internet of Medical-Things (IoMT). The IoMT has revolutionized the operation of medical facilities and the quality of services.

According to Oh et al. (2021); Peng et al. (2024); Atalay et al. (2024), there should be a promising avenue for connecting healthcare data sources while still preserving privacy. This automation allows security analysts to focus on high-priority issues, thereby enhancing the efficiency of threat detection procedures.

Sharma et al. (2020) proposed a long short-term memory (LSTM) approach based on autoencoders for anomaly detection within user behavior analytics, specifically targeting insider threats. Based on their experimental results, the authors reported a best case accuracy of 90.17%. However, the authors did not discuss any further privacy-preserving methods to protect their data, nor did they address the potential for collaborative intelligence sharing among institutions to enhance insider threat detection strategies in their future work.

Manoharan et al. (2023), explored insider threat detection utilizing supervised machine learning with random forest (RF) techniques. The authors stated that they explored the impact of hyper-parameter tuning using CERT r4.2 dataset. The proposed approach yields superior accuracy and an F1-score of 95.9% compared with previous studies like user behavior analysis, LSTM, and deep neural networks (DNN). Although their proposal achieved the highest results, its centralized implementation on a single version of CERT's insider threat dataset potentially introduces data privacy risks and limits the collaborative

nature of intelligence sharing among organizations. In their future work section, authors planned to boost their minority class performance for imbalanced data like CERT's insider threat dataset.

Another significant contribution is the work by Sarhan and Altwaijry (2023). The authors present insider threat detection using the deep feature synthesis (DFS) algorithm to extract behavioral features from past data and achieved an accuracy of 91% for the anomaly detection model. As described in their paper, they used two popular versions of CERT's insider threat detection dataset, r4.2 and r6.2. The authors investigated the effectiveness of their models in mitigating the impact of the imbalanced dataset. Authors experimented with a variety of machine learning (ML) models, testing them on SMOTE oversampling as a balancing technique. Authors found that the support vector machine (SVM) model produced exceptional results, yielding a 100% classification model compared to all other ML models. However, they pointed out that their proposal is not without some drawbacks; they stated the lack of real-world datasets and privacy concerns limits organizations' collaboration.

Kim et al. (2019) proposed a novel approach for detecting insider threats based on user behavior modeling and anomaly detection algorithms. The authors conducted their study using R6.2, the "CERT Insider Threat dataset which is the latest and largest dataset (Glasser, 2013; Sarhan and Altwaijry, 2023). The test results showed that the proposed framework worked well with an unbalanced dataset. By monitoring the top 30% of suspicious instances, they achieved over 90% detection accuracy with daily activity summaries and up to 98.67% accuracy with email contents. The authors highlight the effectiveness of their approach but also acknowledge the limitations of the proposed

framework in terms of the complexity associated with finding a real-word dataset, training each model separately, and labeling diverse forms of anomalous behaviors as specific threat types.

Al-Shehari et al. (2023) proposed an insider threat detection model by using an anomaly-based isolation forest algorithm. The authors stated that they used the CERT r4.2 dataset, which was the most widely used dataset in the literature. They emphasized that their paper concentrated on an anomaly-based detection strategy, utilizing the IF algorithm to counterbalance the highly imbalanced CERT dataset. In their conclusion, they emphasized the importance of exploring various approaches and utilizing real-world datasets for a more comprehensive evaluation, thereby ensuring practical utility in insider threat detection.

Le (2021) proposes a machine learning-based framework for insider threat detection. The author utilized two versions of CERT's insider threat datasets, r4.2 and r5.2, to test and evaluate their autoencoder using percentile representation as a combination for anomaly detection, as described in the conclusion. The proposed detection approach achieved a high accuracy and precision. The overall limitations and recommendations emphasize the need to explore real-world insider threat detection in a secure, transparent, robust, and flexible approach. This is important, especially in an environment with sensitive, personal, and financial information.

Tian et al. (2020) proposed an insider threat detection system (anomaly detection type) using a combination of LSTM and the dempster-shafer theory, utilizing the CERT's insider threat dataset r6.2. The method incorporated a multi-head attention mechanism and applied dempster's conditional rule to update beliefs and enhance

predictions. The author stated that they achieved an accuracy of 95.47%. However, the adoption of privacy and collaboration limits the effectiveness of the system.

This method of data collection raises privacy concerns, especially if applied in sensitive environments that require privacy-preserving mechanisms, as real-world insider threat detection would require monitoring actual employee behavior, which could infringe on privacy (Singh et al., 2022). Additionally, the reliance on a single dataset limits the system's ability to be broadly adopted in collaborative environments like healthcare, where multiple institutions need to share sensitive data without compromising privacy, a challenge noted by Al-Mhiqani et al. (2020*)*, who emphasized the importance of privacy-preserving techniques in collaborative insider threat detection systems. This challenge is especially critical in collaborative environments where institutions need to share threat intelligence while safeguard sensitive, personal, and financial information.

Similarly, Dosh (2021) proposed “a method for detecting insider threats within institutions using a CERT dataset and different ML techniques”. The author emphasized the lack of a real-world public dataset for detecting insider threats, which led them to utilize the CERT’s insider threat dataset in their research. The author also stated that they used the "r4.1 and r4.2" datasets for analysis and investigation to detect insider threats, emphasizing the exploration of three machine learning techniques include random forest, and nearest neighbor, naive bayes. The results were from 89.7% to 94.6%.

Pantelidis et al. (2021) proposed a model using autoencoder (AE) and variational autoencoder (VAE) neural networks for insider threat detection, utilizing the CERT r4.2 dataset. The authors experimented with these deep learning algorithms to identify anomalies in user behavior without human intervention. Their system achieved an accuracy

of 96% with the VAE model, outperforming the AE model, which had a detection accuracy of 95% but with a slightly higher false alarm rate. Although their approach has demonstrated effectiveness in identifying insider threats, if it were proposed to be applied in sensitive sectors such as financial institutions, healthcare, or government agencies, it would face a notable limitation due to the absence of privacy-preserving techniques and collaborative frameworks, which are crucial in multi-institutional environments. Addressing these aspects would enhance the real-world applicability and robustness of the system in such settings.

Ge et al. (2022) proposed a novel insider threat detection model using residual convolutional neural networks, focusing on multi-source data fusion to address the growing challenge of data heterogeneity. Although most experiments widely use the traditional 7:3 splitting ratio for model training and evaluation (Cappelli et al., 2012; Yuan & Wu, 2021), the authors found this approach unreliable for insider threat detection. They adopted the idealistic and realistic settings suggested by Le et al. (2020). For the idealistic setting, they randomly selected 50% of the user-day data from the entire dataset to train their model. Evaluating the model on the CERT r5.2 dataset, they observed that while their approach achieved effective data fusion, it also led to an increase in false alarm rates, highlighting the need for a collaborative approach to improve its real-world applicability.

Lavanya et al. (2024), introduced a novel enhanced bidirectional generative adversarial network (EBiGAN) to address data imbalance in insider threat detection. The model was tested using the CMU CERT r6.1 and r6.2 datasets along with a deep neural network (DNN) that was optimized using the probability of improvement (PI) acquisition function. The developed system achieved a detection accuracy of up to 98.8% with a low

false alarm rate. Despite the system's high accuracy, the authors did not address potential limitations such as privacy concerns or fostering collaboration across institutions.

As highlighted throughout this Praxis, these works and other existing research not mentioned here provide valuable insights into the current state of insider threat detection using traditional and modern centralized machine learning and highlight the ongoing challenges in this field, particularly in the context of healthcare settings. Insider threats constantly adapt their tactics, necessitating continuous updates of ML models to stay ahead of evolving threats, as noted by Sarhan and Altwaijry (2023). This need for continuous updates presents significant logistical and technical challenges. Additionally, Appari and Johnson (2010) emphasize the importance of balancing security needs with user privacy when implementing ML for insider threat detection. In healthcare settings that involve handling sensitive private data, such as patients' personal and financial information, privacy regulations pose a significant challenge. These regulations can restrict the availability of training data for machine learning models. Moreover, different organizations face additional challenges in collaboratively adopting traditional machine learning methods, which train in a centralized manner and require direct access to the dataset.

These factors exacerbate the overall challenge in the healthcare setting, adding complexity to the training and validation process of these models and consequently making it notably challenging to effectively train and validate traditional machine learning methods in healthcare settings. The requirement for continuous updates that directly access trained datasets from centralized methods presents significant logistical and technical challenges. This is because traditional machine learning (ML) does not facilitate collaboration for

intelligence sharing among various institutions or organizations, which is crucial for finding a more effective solution to detect insider threats.

Furthermore, while ML is a powerful tool, it cannot replace human expertise when it comes to final decision. No one can deny its contribution in advancing the decision making process especially its capability of learning and automating tasks, but still the human judgement is also critical and important in decision making. Olaoye and Egon (2024), assert that security analysts are essential for interpreting ML results, conducting investigations, and making informed decisions. The reliance on human expertise underscores the importance of maintaining a skilled cybersecurity workforce, even as organizations increasingly turn to automated solutions. For example, the early detection of threats facilitated by ML allows for quicker response times, thereby minimizing potential damage. This is particularly crucial in the healthcare sector, where timely detection and mitigation of insider threats can restrict unauthorized enrty to sensitive patient data and ensure the continuity of critical healthcare services.

Human factors, including user behavior and psychology, play a crucial role in insider threats. Insiders often manipulate trust and exploit access privileges to carry out malicious activities. In healthcare settings, specific human factors can contribute to insider threats. For instance, insiders may find access to sensitive healthcare data attractive, and they can use social engineering tactics to manipulate healthcare professionals. As Wilder (2022) states, the designation of insiders cannot be defined simply by their employment status, highlighting the complexity of this issue. The sensitivity of healthcare data, encompassing personal health information, financial data, and other confidential information, intensifies the severity of these threats, particularly due to the increasing use

of digital technologies in healthcare, which brings forth concerns about the security and privacy of digital healthcare systems (National Institute of Standards and Technology [NIST], 2022; Xiang and Cai, 2021; Jawad, 2024). Therefore, it is essential to address these concerns while developing and implementing insider threat detection methods. In the next section, we will delve into federated learning as a promising solution for insider threat detection.

2.4 Federated Learning as a Potential Solution in Healthcare

The growing concern over data privacy and the stringent regulations, along with strict regulations like the Health Insurance Portability and Accountability Act (HIPAA) in the United States and the General Data Protection Regulation (GDPR) in Europe, have made the sharing of sensitive data, particularly in healthcare, increasingly difficult. These regulations, while necessary for protecting individual privacy, create significant challenges for institutions attempting to build machine learning (ML) models that require large datasets from various sources (Li et al., 2020). As highlighted throughout this Praxis, traditional machine Learning (TML), also known as centralized machine learning or classical machine learning, involves collecting raw data from various sources and transferring it to a centralized location for model development (Asad et al., 2022).

This centralization process requires institutions to gather sensitive information, such as patient or financial data, into a single server where the machine learning model is trained. Although this approach can generate accurate and effective models, it poses significant privacy concerns. Centralizing large amounts of raw data increases the risk of security breaches or unauthorized access, as all sensitive information is stored and

Figure 6. Federated Learning Structure Overview.

processed in one place (Asad et al., 2022). Furthermore, industries that operate under strict privacy regulations, like healthcare, may struggle with compliance when using this approach due to the heightened exposure of sensitive data.

To address these challenges, Federated Learning (FL) has emerged as a promising innovative approach as a solution that enables institutions to collaborate on ML model training while preserving data privacy. Introduced by Google in 2016, FL allows

decentralized model training, where participants, such as hospitals, train machine learning models locally on their own data without transferring sensitive information to a centralized server (Konečný et al., 2016). Instead of sharing raw data, participants exchange only model updates, which reduces the risk of data breaches and alleviates concerns about compliance with privacy regulations (Intel, 2024). This decentralized approach is particularly valuable in industries like healthcare, where safeguarding patient data is a top priority. By facilitating collaborative learning while keeping data secure and localized, FL enables healthcare institutions to enhance their machine learning capabilities without compromising patient privacy (GeeksforGeeks, 2024).

While Google first introduced federated learning (FL) in 2016 as a distributed machine learning technique (Konečný et al., 2016), federated networks or systems for sharing decentralized data had existed prior. For example, Jacobson et al. (2015) discussed a federated network used for translational cancer research, where clinical data and biospecimens were shared across institutions to facilitate research while keeping data localized. Similarly, Constable et al. (2015) introduced privacy-preserving genome-wide association studies (GWAS) on federated genomic datasets. However, it was Google's introduction of FL in 2016 that formalized the concept of training machine learning models within decentralized data without requiring data to be shared directly, marking a significant advancement in privacy-preserving machine learning.

Since its formalization by Google in 2016, federated learning (Konečný et al., 2016; McMahan et al., 2017) has drawn significant interest from both scholars and organizations (Yang & Dong, 2022), as it allows for collaborative learning across different institutions

while preserving privacy. This capability is particularly beneficial in fields handling

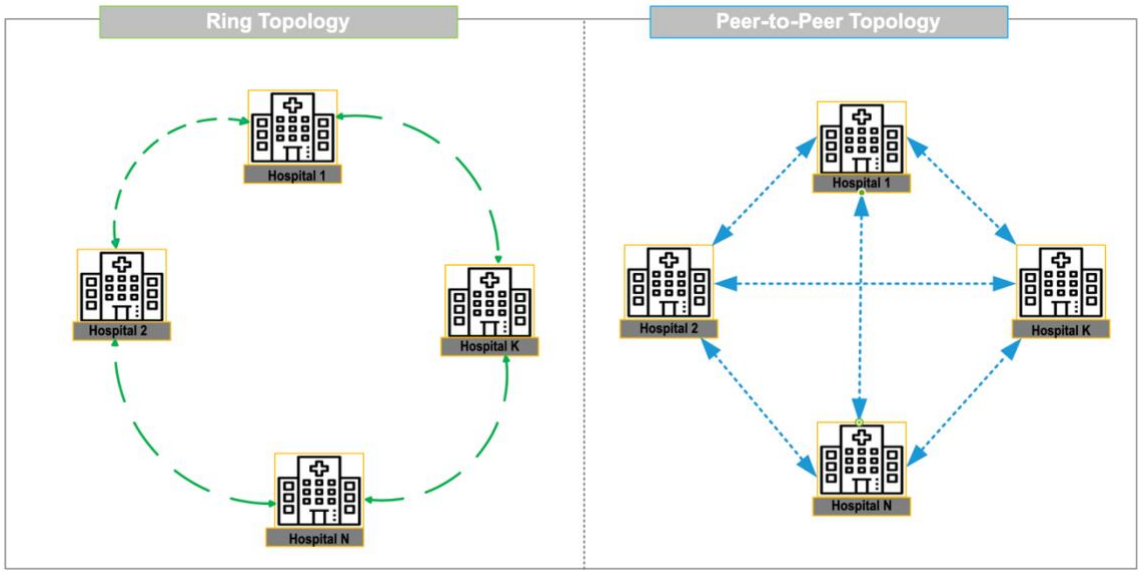


Figure 7. Common Federated Learning Topologies.

sensitive, personal, or financial information, such as financial institutions, government agencies, or healthcare institutions, where patient information is highly sensitive and subject to strict regulatory controls. Konečný et al. (2016) highlight that federated learning (FL) addresses privacy concerns by separating the training process from direct access to raw data. In healthcare, FL allows healthcare institutions include hospitals, clinics, or research institutions to collaboratively train machine learning models, improving diagnostic tools, patient care, and predictive analytics without compromising patient privacy (Rieke et al., 2020). Furthermore, FL overcomes the problem of data fragmentation in healthcare by distributing the model training process across multiple institutions, ensuring that no raw data is shared externally (Asad, 2022).

The architecture of communication methods in federated learning (FL) includes centralized federated learning (CFL) as (Client-server communication, which utilizes a central server to coordinate learning, and decentralized federated learning (DFL) includes Peer to peer (P2P) communication, which operates without a central server, enhancing scalability and fault tolerance (Joshi et al., 2022). FL can also be categorized based on how data is distributed among participants, enhancing its applicability to insider threat detection. Horizontal federated learning (HFL), also referred to as sample-based FL, is used when participants share the same feature space but have different sample spaces (GeeksforGeeks, 2024). For example, in healthcare, multiple hospitals may handle similar types of patient data but for different patients. Horizontal Federated Learning (HFL) enables these institutions to collaborate on model training without exchanging raw data, thus preserving privacy across datasets.

In contrast, vertical federated learning (VFL), also known as feature-based FL, is utilized when "participants share the same sample space but have different feature spaces"(Intel, 2024; GeeksforGeeks, 2024). This method is particularly effective when different departments or organizations hold complementary data about the same individuals, allowing for enhanced insider threat detection while safeguarding privacy (Joshi et al., 2022).

Similarly, hybrid federated learning is another approach that integrates aspects of both horizontal and vertical FL. This type of FL is used when participants differ in both sample and feature spaces, allowing institutions to combine distinct datasets and apply transfer learning techniques to enhance model accuracy while maintaining privacy (Zhang et al., 2021). Hybrid FL can be particularly useful in complex real-world scenarios where

data fragmentation is widespread, such as insider threat detection, by enabling collaboration across institutions with diverse datasets and features.

Additionally, federated learning Topologies are important factors in architecture design process especially in decentralized federated learning approach, they help manages communication and model updates efficiently without a central server. Examples include Mesh topology also known as peer-to-peer (P2P) where nodes communicate directly with each other, enhancing fault tolerance and privacy, while ring topology is the design where nodes form a closed loop, passing updates sequentially, with can streamline communication and reduce overhead (Martínez et al., 2022; Lian et al., 2022; Qi et al., 2021), as it is illustrated in Figure 7.

Recently, automated machine learning (AutoML) has gained significant traction, particularly in the form of neural architecture search (NAS), which automates the architecture design of neural network. The NAS allows for the optimization of deep neural networks (DNNs) in federated settings, ensuring that models are not only privacy-preserving but also highly efficient and accurate (Zhang et al., 2021). This automation has proven especially beneficial in healthcare and other sensitive domains, where insider threat detection models can be trained more effectively without requiring manual intervention in the architecture design process.

FL architectures are further categorized into cross-silo and cross-device settings (Joshi et al., 2022), which determine how participants are structured. cross-silo FL typically involves a small number of organizations, such as hospitals or data centers, which collaborate to train a global model while ensuring sensitive data remains localized. This setup is especially suitable for insider threat detection in healthcare, where privacy

regulations limit direct data sharing (Zhang et al., 2021). On the other hand, Cross-Device Federated Learning (FL) involves a larger number of participants, such as mobile devices or Internet of Things (IoT) devices, where each contributes smaller datasets. This architecture is essential for decentralized environments, where privacy and collaboration are critical (Constable et al., 2015).

In the context of FL model architecture, DNNs have become one of the most popular architectures for model training due to their effectiveness in managing complex, high-dimensional data. Long Short-Term Memory networks (LSTMs), on the other hand, are more suited to tasks involving sequential or time-series data, such as detecting temporal patterns in insider threat behavior. Both architectures can be utilized within FL, but DNNs are frequently preferred for handling large-scale, decentralized datasets that require deep learning across multiple layers. Notably, according to the above facts and the complexity nature of CERT's insider threat dataset versions utilized in this praxis, we implemented and tested both DNNs and LSTMs as part of our model architectures for insider threat detection, and based on the results, selected the most suitable architecture model for our proposed Advanced Privacy-Preserving Decentralized Federated Learning (APPDFL), as highlighted in "Chapter 3: Methodology".

By choosing to implement either centralized federated learning (CFL) or decentralized federated learning (DFL) and selecting the appropriate approach—whether horizontal federated learning (HFL), hybrid federated learning, or vertical federated learning (VFL), also based on the field and data collection methods—organizations can significantly enhance insider threat detection capabilities without compromising data privacy or security. These architectural model choices allow institutions, particularly in

privacy-sensitive sectors such as healthcare, to collaborate effectively while ensuring robust data protection. Therefore, federated learning represents a powerful and promising solution for healthcare institutions (Xu et al., 2021; Islam et al., 2022), enabling the application of collaborative machine learning while maintaining stringent privacy standards. This section explores two primary communication methods of FL: Centralized Federated Learning (CFL) and decentralized federated learning (DFL). For the Federated Learning organization structure overview, refer to Figure 6.

2.4.1 Centralized Federated Learning Approach

Centralized Federated Learning (CFL) emerges as powerful tool for collaborative learning in general, and healthcare particularly, addressing the critical challenge of scalability, which refers to the capacity to handle an increasing amount of work or the ability to expend to accommodate growth. CFL facilitates the adoption of collaborative intelligence sharing via a central server. This approach enables multiple entities, referred to as "Clients" or Participants" to train a shared machine learning model across multiple clients without needing to directly access their raw data, keeping data decentralized on their own devices or local servers.

According to Konečný et al. (2016), in CFL a central Server orchestrates the training process by aggregating and uploading the model based on input from distributed clients. "Every participant independently trains a local model on their private data and then communicates the updated model parameters to the central server" (Li et al., 2020). The server collects these updates, refines the global model, and redistributes the improved model to each client. This centralized aggregation

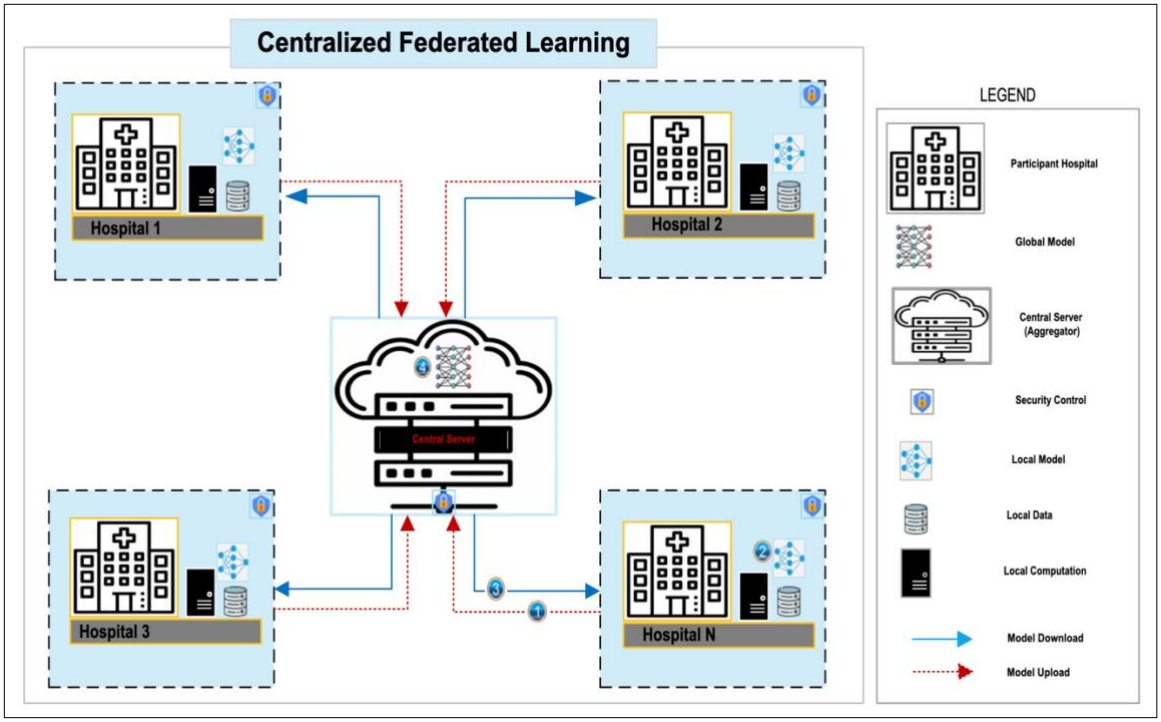


Figure 8. Centralized Federated Learning Architecture.

allows CFL to leverage insights from various data sources while preserving individual privacy.

Figure 8, visually represents the communication architecture of centralized federated learning (CFL), showing how the central server aggregates and redistributes model updates. It also presents a conceptual overview of how a central server enables collaborative learning among clients, ensuring the exchange of no raw data. It also shows how each institution trains the model on its own patient records and transmits only the trained parameters (e.g., gradients or weights) to the central server. Each institution receives the enhanced global model once the server aggregates these parameters.

CFL architecture is one of the most commonly used approaches due to its collaborative nature and the ability to secure both local and global models, offering a degree of privacy. However, it is essential to acknowledge that, although CFL is widely

recognized, it is not the sole architecture employed in federated learning. According to Bonawitz et al. (2016), the model parameters solely encapsulate each client's contribution, thereby preventing the disclosure of sensitive information and mitigating the risks of data breaches.

Previous studies have highlighted Centralized Federated Learning's efficacy in various domains, including insider threat detection. For example, Amiri-Zarandi et al. (2023) made a significant contribution by proposing and implementing a federated and explainable approach for insider threat detection in Internet of Things (IoT) environments. This approach leverages collaboration among different data owners by simulating various collaborative departments using CERT r6.2, enhancing the volume of data available for model training and subsequently improving the performance of the machine learning model. The authors highlight that the approach not only protects data privacy but also addresses a common problem in deep learning-based threat systems, which often act as black boxes. This model provides explainable decision, clarifying why specific behaviors are flagged as potential threats, thereby resolving a significant drawback in traditional threat methods.

The authors stated that their experimental performance results showed that the proposed solution is almost comparable to the performance of central solutions while addressing privacy concerns related to data sharing. Authors then recommend future research to consider using different datasets for validation, as the commonly used CERT's insider threat dataset is extremely imbalanced, which may lead to biased and inaccurate detection results. From the observation view point, the central server in their proposed

threat detection system, which is critical component commonly used for centralized federated learning plays a main central role, introduces a potential single point of failure.

Kumar and Kim (2024) proposed novel approach for detecting multiple cyberattacks on the Internet of Health Things (IoHT) using a Federated Learning (FL) framework integrated with embedded Long Short-Term Memory networks (EFL-STM). This EFL-LSTM framework for detecting cyberattacks on the Internet of Health Things (IoHT) enables IoHT devices to train local models on device-specific data and share model updates with a central server for aggregation, which enhances security and privacy by preventing raw data transmission. Authors utilized LSTM networks, "known for the effectiveness in handling time-series data, to capture and analyze temporal patterns indicative of cyberthreats" (Kumar and Kim, 2024). The experimental results demonstrate that the EFL-LSTM model maintains high detection accuracy of 97.16%. However, from the observation viewpoint, the structure of the proposed EFL-LSTM model includes a central server (aggregator) that consolidates updates from individual devices. This reliance on a central server creates a potential single point of failure, introducing a vulnerability that could compromise the entire system if the server is attacked (Li et al., 2020).

Gayathri et al. (2024) introduced FedAT. Stand for "Federated Adversarial Training for distributed insider threat detection in IoT environments" Gayathri et al. (2024). This approach combines federated learning (FL) with adversarial training (AT) to enhance detection accuracy and robustness against adversarial attacks while preserving data privacy. The authors utilized the CERT r4.2 and r5.2 datasets to evaluate their model and employed FedAvg and FedProx as federated optimization methods. The proposed FedAT model relies on a central server for aggregating model updates from participating

devices, which introduces a potential single point of failure. Experimental results demonstrated that FedAT outperforms other federated learning and adversarial training baselines in detecting insider threats. For future work, the authors plan to enhance the AT model by incorporating mechanisms to defend against backdoor and model inversion attacks in Centralized Federated Learning (FL).

CFL' s centralized structure introduces certain vulnerabilities. Internal or external attackers could compromise the serves, jeopardizing the entire collaborative learning process. Bagdasaryan et al. (2020), highlighted that an attacker could exploit CFL, causing the global model to reach 100% accuracy on a backdoor task. This reliance on a single central server raises concerns about both reliability and security, presenting unique challenges as clients usually communicate only with the server, which can act as a system bottleneck. Dwork and Roth (2014), discussed techniques for mitigating those risks, such as differential privacy. From the above work reviewed, single point of failure is a notable drawback. As Joshi et al. (2022) discussed, while CFL offers a promising solution for privacy preserving and collaborative learning compared to traditional centralized machine learning approaches, the reliance on a centralized server remains a key vulnerability.

That is where decentralized federated learning (DFL) comes into play. Unlike CFL, Decentralized Federated Learning (DFL) offers a different approach to address some of the challenges associated with CFL. In the following section, 2.4.2, we will delve into the concept of DFL, its architecture, and how it can potentially enhance the privacy and efficiency of federated learning in the healthcare domain.

2.4.2 Decentralized Federated Learning Approach

Decentralized Federated Learning (DFL) is an innovative approach that addresses some of the challenges associated with Centralized Federated Learning (CFL) and traditional machine learning methods. Unlike CFL, DFL does not rely on a central server

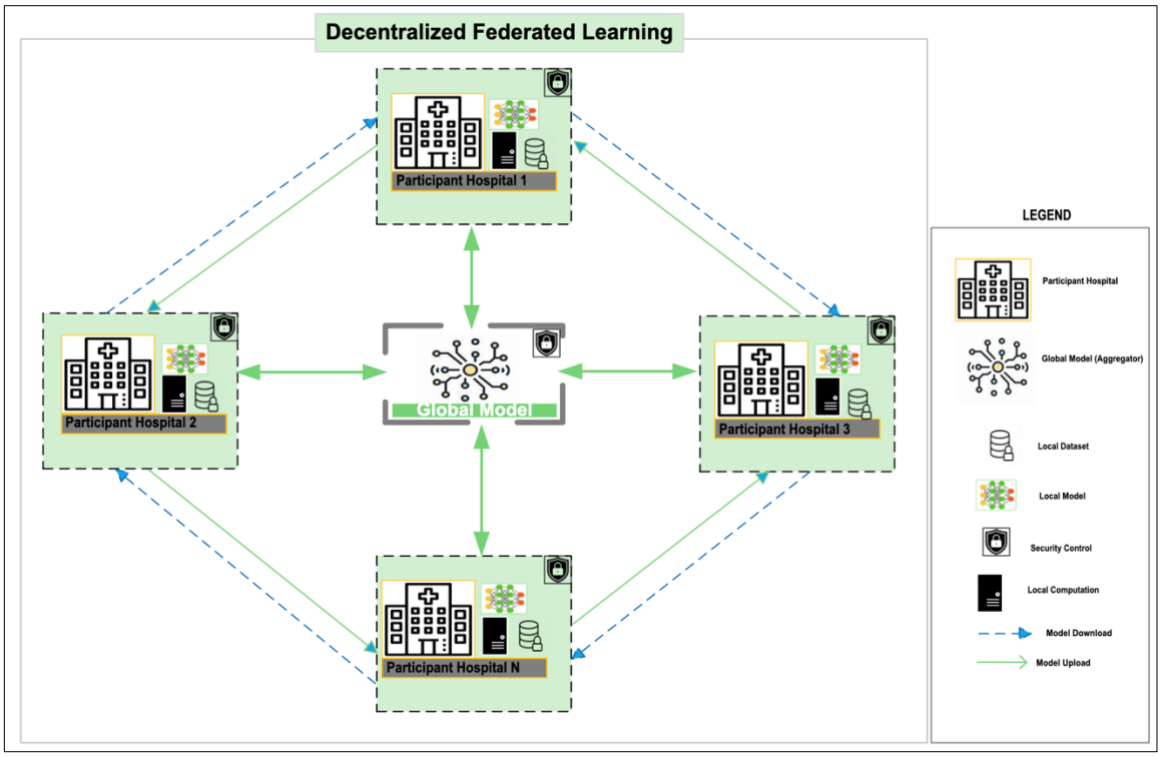


Figure 9. Decentralized Federated Learning Architecture.

to aggregate model updates. Instead, it facilitates direct interaction among clients, resulting in considerable savings of communication resources (Li et al., 2020). For example, Decentralized Federated Learning (DFL) provides enhanced privacy protection compared to CFL and traditional machine learning methods. In DFL, "clients communicate and share model parameters with each other without any server, which eliminates the risk of a single point of failure and strengthens data privacy"(Lu et al., 2022). This is particularly beneficial in healthcare settings where large volumes of sensitive data are involved (Li et al., 2020).

In addition, DFL provides enhanced privacy protection compared to CFL and traditional machine learning methods. In DFL, clients communicate and share model parameters with each other without any server, which eliminates the risk of a single point of failure and strengthens data privacy. This is particularly beneficial in healthcare settings where large volumes of sensitive data are involved (Li et al., 2020). As highlighted by McMahan et al. (2017), DFL is more resilient to network failures or disruptions, allowing for continuous training. This resilience is due to the decentralization of training, which allows for continuous operation even in the face of network disruptions. Collaborative learning allows multiple parties to contribute model enhancements without disclosing sensitive information.

DFL communication architecture, such as peer-to peer (P2P) are getting significant attention for collaborative scenarios involving intelligence sharing, particularly in healthcare, Internet of vehicles, wireless networks and financial institutions. Researchers have highlighted decentralized federated learning (DFL) as a viable and innovative alternative, noting its emergence in various fields including healthcare where hospital can collaboratively train models for disease diagnosis without sharing patient data (Yang et al., 2019; Xu et al., 2021; Pfitzner et al., 2021; Yan & Li, 2024). For the decentralized federated learning architecture overview, refer to the Figure 9.

2.5. Data Imbalance and Balancing Techniques in Insider Threat Detection

Insider threat detection models often face imbalanced datasets, where the number of normal user activities significantly exceeds the number of malicious ones. This imbalance makes it difficult for machine learning models to accurately detect insider threats (Gustavo et al., 2004). Common data balancing techniques include:

- **SMOTE :** Stand for "synthetic minority over-sampling technique", which generates synthetic examples for the minority class, which helps improve model performance on imbalanced datasets (Gustavo et al., 2004).
- **ADASYN:** Stand for **"**adaptive synthetic sampling", which is an advanced version of SMOTE that "generates synthetic samples based on the density of the minority class" (Sharma et al., 2023).
- **Near Miss and Tomek Links**: These techniques under-sample the majority class by creating a balanced dataset, which can lead to improved model performance by eliminating redundant or borderline examples from the majority class (Gustavo et al., 2004).

Sharma et al., (2023), highlighted the importance of balancing techniques like SMOTE, authors' paper also shows that a combination method of over-sampling the minority class and under-sampling the majority class can achieve better classifier performance. In healthcare settings, balancing techniques are crucial for improving the accuracy of insider threat detection models. For example, Al-Shehari and Alsowali (2022) used random resampling techniques for handling the imbalance of dataset classes in insider threat detection, their method lacks privacy and scalability capabilities that allow the adoption of collaborative way of sharing insider threat intelligence for all healthcare institutions' participants.

2.6 Privacy-preserving techniques for data sharing

In the era of big data, especial in healthcare institutions privacy concerns have led to the development of various privacy-preserving techniques. These techniques aim to protect

sensitive information while allowing data to be shared for beneficial purposes such as research and development. For example, as Danger (2022) highlighted :

- Differential Privacy: This approach introduces calibrated noise into the data, providing a mathematical guarantee of privacy. However, these approaches often provide privacy at the cost of reduced model performance or system efficiency" (Aouedi et al., 2023). Despite this trade-off, differential privacy has been widely adopted due to its robust privacy guarantees.
- Homomorphic Encryption: This technique enables computations to be carried out on encrypted data, eliminating the necessity to decrypt data in untrusted environments. This means that sensitive data can remain encrypted while still being useful for computations, providing a secure method for preserving privacy in data sharing.
- Secure Multiparty Computation : SMC enables multiple parties to perform computations on their inputs while maintaining their privacy. This is especially beneficial when sensitive datasets are spread across various data owners. This is particularly useful when sensitive datasets are distributed across different data owners, as it allows for privacy-preserving learning without the need to share the raw data, without requiring the sharing of raw data.

These techniques provide robust mechanisms for preserving privacy in data sharing. However, it is worth noting that each technique comes with its own set of trade-offs between privacy, utility, and efficiency. In our proposed Advanced Decentralized Federated Learning with Differential Privacy approach, differential privacy provides a robust framework to protect patient data. First suggested by Dwork et al. in 2006, this

privacy-preserving technique prevents information leakage by adding noise to data, making it particularly valuable in healthcare where patient confidentiality is paramount (Dwork et al., 2006). Differential privacy is essential for ensuring data confidentiality, as it guarantees that individual records remain indistinguishable even when model parameters are shared across institutions (Abadi et al., 2016). Furthermore, differential privacy supports data availability, enabling healthcare institutions to work together on model training while safeguarding sensitive information.

This strategic focus aligns with our Research Question 2 (RQ2), which explores the optimal trade-off between privacy protection and acceptable accuracy in insider threat detection. In the collaborative healthcare environment, differential privacy offers a balanced approach that mitigates risks without hindering the detection capabilities essential for effective insider threat management. While federated learning (FL) enables distributed model training without centralized data collection, it does not inherently guarantee robust privacy protections. As Boenisch et al. (2023) and Carlini et al. (2023) indicate, FL provides limited privacy assurances on its own and therefore requires the integration of differential privacy and secure aggregation techniques (Bonawitz et al., 2016; Pelikan et al., 2023). By adding Gaussian noise to client updates ("delta" values), the differential privacy mechanism guarantees the privacy of each data point even in a distributed setting, which is crucial for our insider threat detection objectives (Pelikan et al., 2023).

Integrating Applying differential privacy requires careful management of privacy budgets and noise multiplier, as excessive noise can diminish model performance while insufficient noise may fail to adequately protect privacy. This balancing act is essential to maintaining confidentiality, availability, and integrity within federated learning. Adaptive

noise strategies, such as those discussed by Zhang et al. (2022), can further improve accuracy while adhering to differential privacy principles. Ensuring that data integrity is maintained across the federated learning framework is critical for accurate model training and reliable insider threat detection outcomes. “Selecting an optimal trade-off between privacy protection and model accuracy is vital” (Zhao et al., 2020; McMahan et al., 2023). This primarily influenced by the differential privacy parameter, epsilon (ε), which controls the amount of noise added. Smaller ε values increase privacy by introducing more noise, potentially reducing model accuracy, while larger ε values can enhance accuracy at a possible cost to privacy.

Subramanian (2023) study suggest that ε values between 0.1 and 1.0 may provide an effective balance between privacy and utility. However, the ideal ε value is often determined empirically, depending on data sensitivity and the acceptable trade-off between privacy and acceptable accuracy. This approach enables a flexible balance, safeguarding sensitive healthcare data without significantly compromising the accuracy of insider threat detection. By carefully calibrating ε, our model achieves a degree of privacy suitable for the healthcare setting while maintaining effective detection capabilities. For more about RH2 validation, refer to Chapter 4.

2.7 Differential Privacy-Preservation in Federated Learning

According to Owkin, (2022) “Not sharing data is not enough to ensure privacy, we need formal guarantee”. Owkin (2022) also, highlighted that machine learning is susceptible to various attacks on data, differential privacy induces a privacy-utility trade-off which depends on the trust model. For the case of decentralized approach, “full decentralization can amplify privacy guarantee, providing a new incentive for using such approaches

beyond the usual motivation od scalability"(Owkin, 2022). Differential privacy is a technique that has attracted considerable interest in the privacy-preserving data domain. It provides a mathematical guarantee of privacy by introducing a carefully calibrated amount of random noise into the data. This ensures that the privacy of individual data points is preserved, even when the data is used for analysis or machine learning. According to Dwork & Roth, (2014), "a randomized algorithm M satisfy satisfies (ϵ, δ)-differential privacy if for any two adjacent input x and (y) where x and y differ by at most one entry) and for all subsets of output S ⊆ Range (M)", based on Dwork and Roth (2014), the following condition is satisfied:

$$Pr[M(x) \in S] \leq e^{\epsilon} . Pr[M(y) \in S] + \delta$$

Equation 1. Mathematical foundation of differential privacy condition

Source: (Dwork & Roth, 2014).

In the equation, ε (epsilon) is a non- negative parameter that measures the privacy loss, while δ (Delta) is a small value representing the probability of possible breach. See Chapter 3, for more formulas. To put this formula into easy way is to imagine if for all dataset Versions V1 and V2 different on at most one element, and for all S that are contained within the Range (M), also based on Dwork and Roth (2014), the following condition is satisfied:

$$Pr[M(V1) \in S] \leq e^{\epsilon} . Pr[M(V2) \in S] + \delta$$

Equation 2. Adapted of differential privacy condition of two datasets

Adapted from the above original equation (Dwork & Roth, 2014).

This ensures that the presence or absence of individual's data does not affect the output, providing a strong privacy guarantee (Dwork & Roth, 2014). This equation essentially states that the probability of any output (set of outputs S) when the input is x

should be almost the same as when the input is y. The degree of similarity is controlled by the privacy parameter ε, where a smaller ε provides stronger privacy guarantees. However, "these methods often provide privacy at the cost of reduced model performance or system efficiency" (Dwork & Roth, 2014). This is because the noise added to preserve privacy can also add uncertainty to the learning process, potentially affecting the accuracy of the model. It is also worth noting that in the context of federated learning, a form of machine learning where the training data remains on the original device, differential privacy plays a crucial role. It can be used to prevent the leakage of information from the client's data during the learning process. This is achieved by addition of small noise to the model parameters at the client's side before they are aggregated at the server. This approach has been shown to satisfy differential privacy under various protection levels.

A study by Banse et al. (2024) demonstrates how the number of clients and the integration of differential privacy mechanisms impact the model's performance across various data type. The findings indicate that non-ideal and small datasets experience the greatest decline in performance within a distributed and differentially private environment. Another paper by Danger, (2022) deeply summarized the most concepts about differentia privacy includes the relationship of differential privacy with synthetic data generation. Therefore, striking a balance between preserving privacy and optimizing model performance is essential in federated learning. This involves carefully choosing the amount of noise to add in order to satisfy differential privacy, while also ensuring that the model's performance is not significantly affected.

2.8 Summary and conclusion

As already indicated, most prior works focused on centralized methods for insider threat detection, which could create a potential single point of failure. Generally, we noted that previous researchers have not considered the decentralized approach or used multiple datasets from three dataset versions or more to simulate collaborative participation to enhance the effectiveness of insider threat detection methods, particularly in healthcare institutions. The integration of decentralized machine learning methods with differential privacy can create a robust system for insider threat intelligence sharing.

This research will not only improve the detection of insider threats, but also safeguard privacy, enabling the adoption of collaborative approaches that facilitate data sharing for artificial intelligence in healthcare (Youssef et al., 2023). This research contributes to the interdisciplinary advancement of federated learning, insider threat detection, and collaborative healthcare, fostering future innovation and enhanced cybersecurity.

Chapter 3—Methodology

3.1 Introduction

This chapter presents the implementation of the Advanced Privacy-Preserving Decentralized Federated Learning (APPDFL) Approach, which serves as the core methodology of this praxis. The primary goal of this praxis, as highlighted in Chapter 1, Section 1.5, is to develop a solution that enhances insider threat detection in collaborative healthcare institutions by improving security, accuracy, and efficiency. APPDFL specifically addresses critical issues in existing centralized methods, including privacy protection challenges, collaborative adoption inefficiencies, and regulatory compliance risks.

The APPDFL methodology leverages decentralized learning to enable institutions to securely train and share model updates through peer-to peer communication without relying on a central server. This decentralized approach mitigates privacy concerns by keeping sensitive data localized within each institution, thereby lowering the risk of data breaches. Integration of differential privacy (DP) mechanism as a technique, ensure the privacy of these updates by introducing statistical noise to data updates, thereby further protecting the privacy of individual healthcare institutions and enabling meaningful intelligence sharing among all participating healthcare institutions.

Although APPDFL is the primary methodology, a review identified gaps in the exploration of decentralized federated learning for insider threat in collaborative healthcare, particularly when leveraging three or more datasets, such as the popular versions of CERT's insider threat dataset. This motivates our choice to developing the APPDFL approach and evaluate its effectiveness against existing centralized methods.

Additionally, as already indicate, we've focused mainly on APPDFL as our proposed solution. Previous researchers have not considered simulating collaboration using multiple diverse and complex dataset versions closer to real-world scenarios (Amiri- Zarandi et al., 2023). The APPDFL approach also aims to bridge the exisiting gap in insider threat detection research, where traditional and modern centralized machine learning approaches dominate the literature, with few exploring a collaborative approach to improve insider threat detection in general.

To our knowledge, all methods in the literature are centralized solutions, which present a significant risk due to reliance on a main server, especially in the environments with sensitive , personal, or financial information, include healthcare institutions. Existing works have utilized only one version or a combination of two versions of CERT's insider threat dataset, whereas we aim to apply a collaborative approach that leverages a wide array of data consisting of three versions of the CERT dataset to simulate different local data from various collaborating institutions.

This gap underscores the significance of our decision to expand and contemplate the implementation of both traditional and modern machine learning models (TML) and centralized federated learning (CFL) models, drawing from popular researchers' choices as presented in existing works. We utilized the same versions of the CERT insider threat dataset for training, testing, and validation purposes. The goal of this comparison study is to prove our research hypothesis 1 (RH1). This option enables a more comprehensive evaluation of our primary methodology, the APPDFL approach, supported by unbiased intelligence decision-making. Chapter 4, provides in-depth examination of the results, comparing APPDFL's performance with the existing work summarized in Chapter 2.

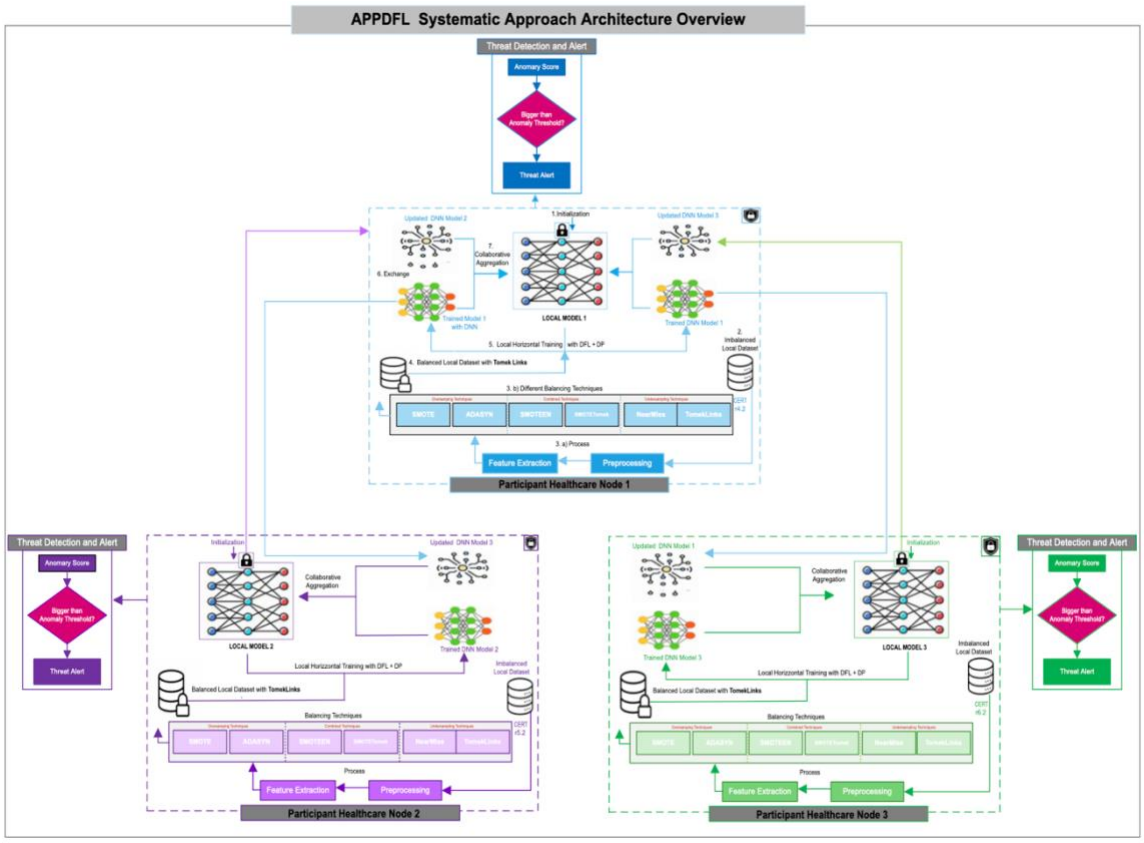


Figure 10. Top Level Methodology Overview.

It highlights the differences in accuracy, privacy preservation mechanisms, and collaborative adoption strategies between APPDFL and the centralized methods. By focusing on the unique advantages of APPDFL, including its ability to enhance insider threat detection while addressing privacy concerns and inefficiencies, this methodology demonstrates a scalable and secure alternative to exisiting methods, particularly in healthcare institutions.

The Figure 10, provides a high-level overview of the methodology, including the identification of data sources, data processing, balancing techniques, model architecture optimization selection, testing and evaluation process. We delve into these stages in detail in the following sections. For a more comprehensive view of the flowchart, refer to the appendix, Figure A-2.

3.2 Implementing the Environment

3.2.1 Hardware and Software Setup

To conduct our experiments, we began by creating three distinct virtual environments to simulate real-world-scenarios using the Anaconda distribution of Python, Visual Studio Code, DataSpell, Sublime text editor and Jupyter Notebook. Anaconda manages the environment dependencies and libraries required for the project, ensuring consistency and compatibility across various experimental scenarios. DataSpell used for data analysis with some exploratory process as its supports notebook. We used Visual Studio Code to set up extension and set up our working environments a code editor for debugging tasks while Jupyter Notebook as a web based interactive computing environment, offers an interactive development environment for writing, executing, and visualizing results in real time, facilitating efficient experimentation and iterative model refinement. Sublime text was an optional preference for code editing, especially for writing scripts to automate tasks before merging the project for comparison purpose of our proposed system and centralized method tested in our praxis implementation

We implement and test each experiment using Python as a main programming language for implementing the models, utilizing popular libraries include PyTorch and scikit-learn which are among the heavyweights in the machine learning world (Pedregosa et al., 2011). These are some of the most widely used machine learning libraries in Python, providing tools and algorithms for building and training models for various cybersecurity applications, such as anomaly detection, intrusion detection, and malware classification (George Washington University, 2024.).

When deciding between two popular deep learning libraries-PyTorch, created by Facebook's AI Research lab or TensorFlow, a Google's brainchild. For the effectiveness of theriacal methodology solution and the facts that our solution will at the end be compared to centralized method tested and previous research works to validate our decision making, from the starting of this praxis implementation, we were inspired by the theory triangulation that encourages researchers of "not relying solely on a single theory that may limit the comprehensiveness and robustness of research findings"(QDacity, n.d.), it's actually fairly common in research to leverage the strengths of different frameworks for various tasks as long as they are compatible with our environment and also benefit our research outcomes. we

The benefits of integrating different tools is to enhance the deepen the analysis and achieve a more robust and comprehensive validation of our research. This comparative analysis theorical resonate with our Praxis implementation as we highlighted in review works that most papers use one or two CERT's dataset in their work and that most of them are centralized methods that implemented mostly using TensorFlow.

Based on the requirements and the facts that it is actually common fairly in research to leverage the strengths of different deep learning framework or libraries as tools to develop, train different deep learning models as long as they are compatible with our environment and also benefit in terms of flexibility, performance and adaptability towards a big picture which is developing a collaborative solution for insider threat detection that allowing participant healthcare institutions to collaborate for intelligence sharing. It is like having different experts working together on a project towards the end solution. For example, PyTorch was chosen the proposed solution solely because of its dynamic

computational graph extensive library for machine learning and federated learning tasks involves in the simulation while Scikit-learn was chosen for preprocessing tasks such as data balancing and splitting due to highly imbalance dataset we utilized in this Praxis implementation. For example, Shivanandhan (2024) asserts that researchers often favor PyTorch over TensorFlow because of its dynamic computational graph, which enhances flexibility and simplifies debugging. It is also worth noting that in this praxis implementation in general we leveraged both Torch for the main methodology and TensorFlow for our centralized methods to be tested the same way the reviewed research work adopted in their implementation process. All of this are to get the fair comparison using triangulation theory as a best solution to the problem presented in our literature review and to get the best of both world that help us to validate our hypothesis and showing thoroughness that gives a more comparison towards our final proposal solution.

Google Trends (n.d.) illustrates the preference trends based on interest over time and compared breakdown by region and sub-region's interest for these libraries based on industry adoption and their location worldwide, Figures 11 below, and additional figures (Figure A-3, and Figure A-4) shown in appendix show that the research community highly favors PyTorch for experimentation and prototyping, while TensorFlow is known for production scalability and deployment. Additionally, as already indicate, for this Praxis, PyTorch was selected based on the project's need for flexibility and ease of use, and its ability to support an optimized Deep Neural network (DNN) and long Short-term Memory (LSTM) model architecture selection process. This ensures that the models are fine-tuned for accuracy and efficiency. This combination of hardware and software allowed for an efficient simulation of decentralized federated learning processes across large-scare,

imbalanced datasets such as CERT's insider threat dataset versions we're using in this Praxis. It is worth noting that this does not imply that TensorFlow could not be used in a different capacity. We believe that the choice of tools is not absolute—as Shivanandhan (2024) highlighted, "today's choice might not be our choice tomorrow," and this Praxis implementation process is the proof that based on theory of triangulation of enhancing the deep and validity of research.

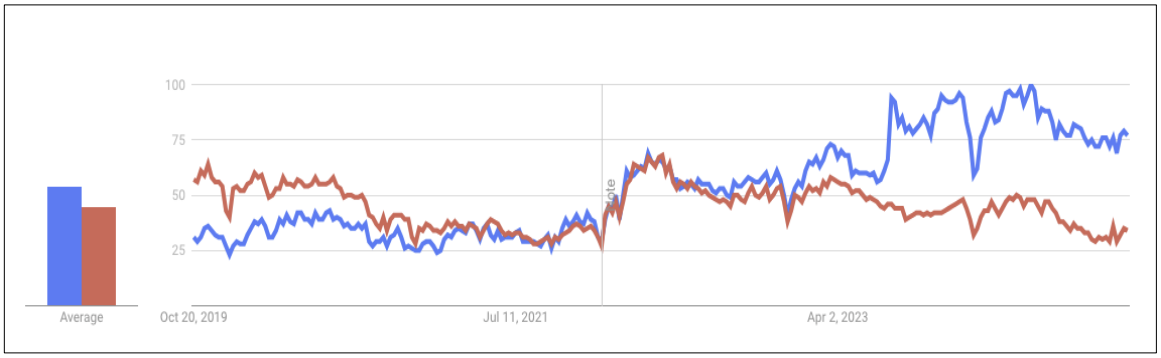


Figure 11. Interest over time - PyTorch (blue) vs TensorFlow (red).

Source: Google Trends (n.d.).

The simulation environment was carried out in a customized Windows-based machine equipped with an AMD Ryzen 7800x3d 8-Core Processor (4201 MHz), 128 GB of RAM, and a 2TB hard disk, chosen for its ability to handle the intensive data processing across multiple large dataset versions like CERT's insider threat dataset we plan to use, and to ensure sufficient computationally intensive power to handle the complex models like Deep Neural Network (DNNs) and Long Short-Term Memory (LSTM) involved in this Praxis implementation process. The setup ensures that each implementation is conducted in a controlled environment, minimizing interference between different implementations and configurations.

3.2.2 Data Sources Description

To validate the proposed Advanced Privacy-Preserving Decentralized Federated Learning (APPDFL) System as a solution for insider threat, we assessed its performance using three publicly available datasets that are widely recognized in academic research for insider threat detection. Specifically, our aim is to use most recent and most commonly utilized versions of the CERT insider threat test datasets, created by the CERT Division at Carnegie Mellon University (CMU). These datasets include both benign and malicious activities, providing a realistic simulation for insider threat detection in collaborative healthcare institutions.

These datasets have been extensively used in previous research, including studies by Glasser et al. (2013); Lindauer (2020); Le et al. (2020); Bartoszewski et al. (2021). These are different versions of the CERT's insider threat dataset ranging from r1 to r6.2, making a total of 10 versions at the time of conducting this experiment. While r1 and r2 are single versions, the newer versions are subdivided, such as r3.1 and r3.2; r4.1 and r4.2; r5.1 and r5.2; and the recent versions r6.1 and r6.2, with r6.2 being the latest and most extensive version to date. All versions provide thorough records of employee activities over long periods making this research a realistic simulated approach for insider threat detection in collaborative healthcare institutions.

Among all versions, r4.2 and r6.2 are the most utilized in academic research (Amir-Zarandi, 2023), with r5.2 being the third most used. These three versions are diverse and complex, strengthening the case for simulating real-world environments where insider threats might emerge. To simulate three different healthcare institutions participating in insider threat detection by sharing intelligence, we selected and used the three most popular

versions in academic research: r4.2, r5.2, and r6.2. This decision enhances the reliability and applicability of our insider threat detection system by simulating the complexity of practical scenarios, as shown in Table 1.

CERT Version	Duration	Users Counts	Feature Counts	Normal Users	Abnormal Users	Scenarios	Data Types	Data Collection
CERT r4.2	72 Weeks: (01/02/2010 - 05/16/2011)	1,000	661	930	70	3 : (30,30, 10)	Logon, Email, Web Access, USB Access, File Access	Simulated environment with realistic scenarios
CERT r5.2	74 Weeks: (01/02/2010 - 06/02/2011)	2,000	1,092	1901	99	4 : (29,30, 30,10)	Logon, Email, device, file, psychom etric	Enhanced simulated environment
CERT r6.2	74 Weeks: (01/02/2010 - 06/01/2011)	4,000	1,176	3995	5	5: (1, 1, 1,1,1)	Logon, Email, device, file, psychometr ic	Comprehensive simulated environment, detailed psychometric profiling

Table 1. Summary of the Commonly used CERT's Insider Threat Dataset Versions

By leveraging these datasets, our research aligns with established academic standards and ensures a comprehensive evaluation of the APPDFL system. This alignment with prior studies not only validates our approach but also enhances the robustness and credibility of our findings. Consequently, the use of these widely recognized datasets supports the generalizability and reliability of our insider threat detection system in collaborative healthcare environments.

3.3 Data Preparation and Balancing Techniques

3.3.1 Data Preprocessing and Feature Engineering

In academic context data balancing are not always necessary if the model performs sufficiently well on balanced dataset. However, due to the CERT's imbalance dataset, fairness, equity, and generalization which are critical to our case to validate our research question 3 (RQ 3). Based on that research question, from the start of this praxis research implementation process, we predicted that the integration of data balancing will play a crucial role in achieving these goals. In this case, data preparation and balancing are essential for developing effective insider threat detection systems, especially within a decentralized federated learning (DFL) approach as a communication protocol using differential privacy (DP) mechanisms. Given that data is distributed across institutions in DFL where each node is a secured, preprocessing must occur locally to maintain data security and privacy, particularly in healthcare environments with stringent data protection regulations.

Due to the nature of the dataset used in this research, extensive preparation was necessary since the CERT Insider Threat Test Dataset contains several imbalanced data that must be resolved before analysis and model training (Zhauniarovich et al., 2018). Ensuring consistency in datasets, we began with standardizing data formats, such as timestamp formats and techniques for encoding categorical variables. For example, timestamps were standardized to a common format, allowing us to aggregate user activity logs accurately.

We standardized data formats, such as timestamps and categorical encoding, ensuring consistency across all datasets. Next, normalization using Scikit-learn's

StandardScaler transformed input features to have “a mean of zero and a standard deviation of one” (Pedregosa et al., 2011). This normalization prevented features with larger ranges from disproportionately influencing the model, ensuring a balanced contribution from each feature. Additionally, Z-score normalization was applied to center and scale the data appropriately, which was crucial for models such as deep neural networks (DNNs) and long short-term memory (LSTM) networks that are sensitive to feature scaling.

Since we’ve different environment for different purpose in this experiment, for our main methodology, we first reflected on previous work to speed up the work since we’ve three (3) different versions that are larger, and to avoid unnecessary effort of building on top of what already exist, we leveraged the exisiting feature extraction for CERT insider threat test dataset which is starts from r4.1 to r6.2 that we utilized in this Praxis (Li et al., 2020). Feature extraction for insider threat detection is done locally at each institution. Features such as unauthorized USB access or abnormal logon times are engineered locally on each participant node, and the insights are fed into the federated model via peer-to peer decentralized pipeline, This was adapted from Lindauer (2020), who provided a foundation for feature extraction from CERT datasets. Each participant healthcare institution (client) performs these preprocessing tasks independently, ensuring that the data used locally is prepared and standardized, contributing to better global model performance.

As it is documented on their GitHub Repository (Le et al., 2020), we generated different data representation to meets our experiments needs for each local node to simulates three participants nodes in a horizontal peer-to-peer (P2P). communication architecture. For data management partition, we followed horizontal federated learning partitioning due to the nature of CERT’s insider threat dataset we’re using for this proposal.

Horizontal data partitioning involves dividing a dataset into multiple smaller tables, where each table contains the same number of columns but a subset of the rows. This is useful for larger datasets like CERT's insider threat dataset as it helps to improve query performance by reducing the amount of data scanned during the peer-to peer process as a serverless communication channel. The choice of peer-to-peer is to avoid the server at all, while adopting horizontal federated learning ensures the same features space across different same spaces. This process enhances load balancing across serverless and facilitates parallel processing. This means that the traffic is distributes across multiple instances (nodes) of our serverless functions to ensure high availability and scalability. Horizontal data partitioning ensures that each partition can be managed, processed, and queried independently , leading to more efficient data handling and scalability on top of privacy preserving mechanism implementation in each participant node.

Each participant healthcare institution (also refer to us, as client) used their data subset—CERT datasets r4.2, r5.2, and r6.2 locally—to address data-specific nuances and imbalances unique to insider threat detection. Horizontal partitioning allowed parallel processing across three healthcare participants, improving scalability and system responsiveness (Thapa et al., 2020). By training models locally, healthcare institutions ensure data remains within their control. For example, the Advanced Privacy-Preserving Decentralized Federated Learning (APPDFL) system implementation with CERT's dataset versions (r4.2, r5.2, r6.2) enables models to adapt to variations in local data distributions while ensuring data remains on-premises. This Advanced Privacy-Preserving Decentralized Federated Learning solution complies with privacy regulations (Smith &

Nguyen, 2021). For the top level architecture design methodology overview, refer to Figure 12.

We create python scripts to automate this process by optimizing the publicly available tools for transforming Cert's data into a format that is easier for our server-less architecture design understand that traditional machine learning. All three datasets are normalized and cleaned in their local nodes, respectively. (Al-Mhiqani et al., 2020) state that normalization is necessary to guarantee that every feature in the dataset contributes equally to the learning process, as certain attributes may vary in scale, with some being binary and others having a greater range of values. Machine learning methods also need the conversion of categorical variables into numerical values. Given the dataset's large dimensionality and processing demands, it comprises hundreds of features. According to Zhauniarovich et al. (2018), dimensionality reduction approaches are consequently required to preserve important information while streamlining the computational process.

To guarantee reliable findings, the dataset was adjusted for anomalies such as outliers or incorrect entries. As incomplete records have a significant impact on model performance, handling them was crucial (Chan et al., 2019). We employed methods to detect and treat missing values, ensuring that the dataset used for training was as complete and accurate as possible. This included imputation techniques for missing values and removal of records that could not be reliably corrected.

Since there are multiple versions different case scenarios and activities include user activity logs, or logs of logons or website visits during the day, weekend, or after hours, the initial step of data preprocessing he data from CERT's to perform data analysis for each

client node. This data analysis is essential to create a cohesive view of user activities across different periods.

Class imbalance is a prevalent concern in CERT's insider threat detection datasets, leading to challenges in accurately identifying minority class threat incidents amidst a majority of normal activities. Models may favor the majority class if appropriate balancing procedures aren't used, producing unreliable findings. According to Salmi et al. (2024), these preprocessing procedures are essential for overall performance comparison to ensure that the model learns from all classes effectively and to validate our research hypotheses. Time-based data in the dataset, such as 'starttime' and 'endtime' columns, which give exact timestamps for the start and end of actions, is one of its most important features. This is especially helpful for time-based user behavior analysis, which helps to pinpoint patterns and deviations across predetermined durations. The collection also records the precise day and week that events transpired, facilitating longitudinal research that distinguishes between typical and abnormal behavior. Additional insights are provided by the binary markers 'isweekday' and 'isweekend,' enabling models to carefully recognize behaviors on weekends vs regular workdays, potentially revealing suspicious activity.

The user column has a unique identification for every user, making monitoring possible over several sessions. This makes it easier to analyze user-specific trends over time, essential for spotting insider threats that are gradual or dispersed over extended periods. The role column gives context for determining if a specific behavior aligns with a user's expected actions based on their work role, containing job titles (e.g., employee, supervisor, IT staff). Additionally, the business and functional units to which users belong are represented by the 'b_unit' and 'f_unit' columns, providing extra information about a

user's position within the organizational structure. Knowing this background helps assess if a given activity complies with assigned tasks and organizational regulations.

The binary ‘insider’ column provides the ground truth label for whether a certain behavior is marked as potentially harmful. The machine learning algorithms designed to identify insider risks require this labeling to be trained and assessed. These labels are used by supervised learning approaches to categorize behaviors as normal or suspicious. Apart from these fundamental variables, we focused on the collection of 508 elements that document many facets of user conduct, such as file accessibility, email correspondence, online exploration, and more network operations. Due to the fascinating scenarios of insiders and based on perspectives of user interactions that these capabilities offer, both malicious insider and accidental insider threat behaviors are thoroughly analyzed. This extensive feature set will be used to assess how well various machine learning models detect insider threats, paying close attention to the trade-offs between accuracy of detection and privacy protection.

Feature engineering was performed to extract meaningful insights from raw data, such as identifying abnormal logon times, system access patterns, and unusual resource access frequencies. This process enabled the model to capture critical user behavior patterns linked to insider threats. To further improve efficiency, Principal Component Analysis (PCA) was employed to reduce high-dimensional datasets, thus addressing the risk of overfitting. By retaining the most relevant components, PCA allowed the model to focus on the essential features, contributing to a more accurate and generalizable threat detection system.

Principal Component Analysis can be adapted to work within a federated learning framework, where data remains distributed across multiple nodes, and only model updates are shared. We did this to ensures that raw data does not leave its original location, maintaining privacy. To further enhance the efficiency of the training process, dimensionality reduction techniques were employed. Given the complexity and high dimensionality of the CERT's dataset versions, with over 500 features recording various user activities, Principal Component Analysis was applied to reduce the number of features while retaining the most important components of variance. This not only reduced the computational load but also helped prevent overfitting, as redundant or irrelevant features were removed. PCA enabled the preservation of essential data while streamlining the dataset, ensuring that the model could focus on the most critical aspects of user behavior.

In summary, our architecture inspired by secured by design principal, as a horizontal peer-to-peer decentralized federated learning system with differential privacy. This approach ensures that data remains confidential and secure throughout the preprocessing and learning process.

3.3.2 Data Balancing and Data Splitting

3.3.2.1 Data Balancing

As we introduced in Chapter 2, balancing techniques are crucial for handling imbalanced datasets, especially in the context of the CERT datasets where insider threat behaviors constitute only a small fraction of the total records. To provide a clear picture of the dataset's initial state before balancing, refer to the Table 1, include the following characteristics of dataset version we're using in this praxis. CERT dataset version r4.2

contains 1,000 employees, with 93% being non-insiders and 7% insiders. The r5.2 version

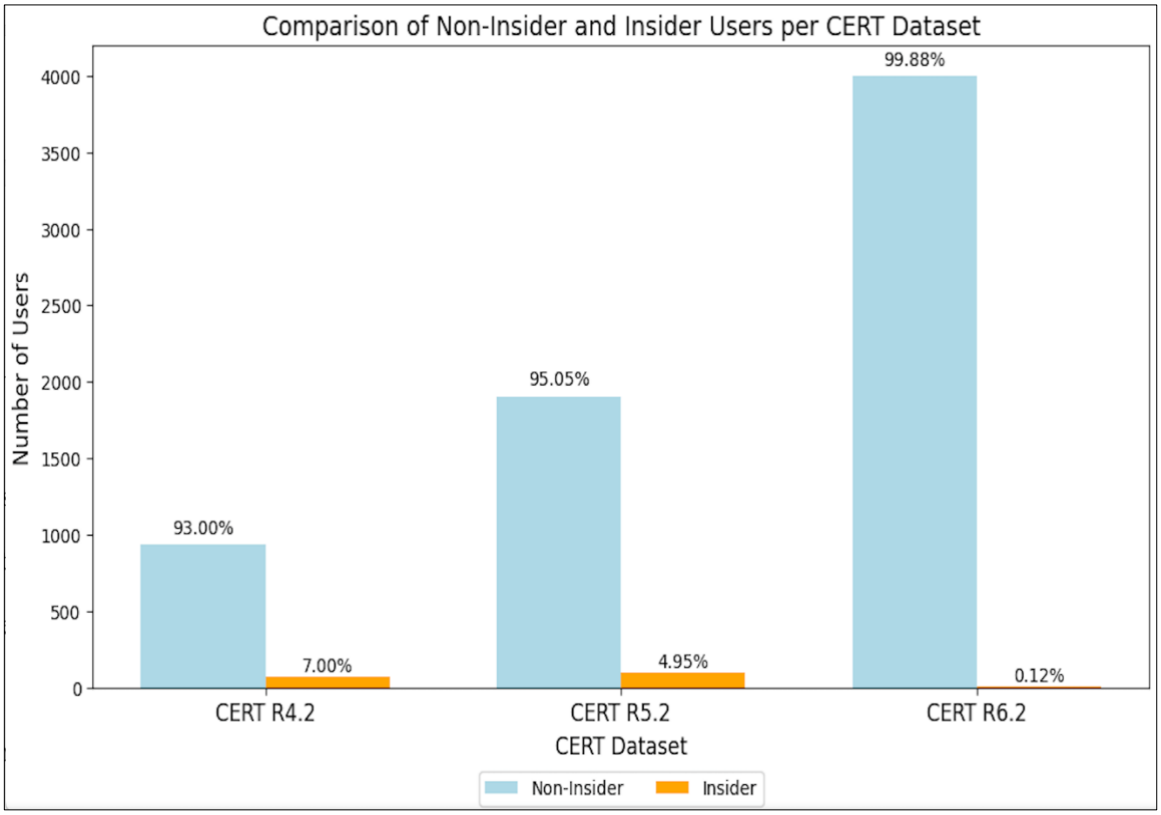


Figure 12. Comparison of Insider and Non Insider in CERT's dataset.

has 2,000 employees, where 95.05% are non-insiders and 4.95% are insiders. The most complex and sophisticated version, r6.2, simulates real-world environments with different case scenarios, including 4,000 employees and only 5 abnormal activities. This means 99.875% represents non-insiders, while only 0.125% are insiders. Given the significant imbalance, the r 6.2 dataset appears almost entirely normal, highlighting the necessity for effective balancing techniques.

Inspired by Cai et al. (2024); Buba et al. (2018), we used the imbalance ratio defined as : *Nmaj/Nmin*, "where *Nmaj* and *Nmin* are the sample size of the majority and minority class"(Cai et al.,2024), for each CERT dataset version utilized in this praxis. Our approach to conducting different balancing techniques was guided by our Research

Question 3 (RQ3) by asking: “Which model architecture, Deep Neural Network (DNN) or Long Short-Term Memory (LSTM), yields higher classification performance for insider threat detection when integrated with selected balancing techniques (SMOTE, ADASYN, Near Miss, SMOTE-ENN, Tomek Links, or SMOTE-Tomek Links) within the Advanced Privacy-Preserving Decentralized Federated Learning Approach?” To address this imbalance and avoid biased predictions toward the majority (non-insider) class, several techniques were implemented and evaluated:

1. *SMOTE:* Given its popularity and effectiveness in other studies (Chawla et al., 2002; Douzas et al., 2018), SMOTE was initially considered a strong candidate for balancing the dataset. However, despite our expectations, SMOTE alone did not completely resolve the imbalance issues, leading us to explore additional popular balancing techniques. SMOTE creates new instances by selecting random examples from the minority class and using them to create "synthetic" data points between existing instances.
2. *ADASYN:* Similar to SMOTE, ADASYN generates more synthetic samples closer to the decision boundary, which tends to improve the model’s performance in difficult-to-classify regions (He et al., 2008). By focusing more on the minority class instances that are harder to classify, ADASYN adapts to the complexity of the decision boundary. This ensures that the learning algorithm pays more attention to those challenging areas, which can significantly enhance the overall accuracy of the model, especially for the minority class.
3. *Near Miss*: This under-sampling technique selects majority class examples that are closest to the minority class, effectively creating a more balanced dataset by

focusing on the most relevant majority instances. Near Miss works by keeping the majority class samples that are most similar to the minority class samples, thereby reducing the impact of the majority class and making the dataset more balanced. This technique helps to remove redundant information from the majority class while maintaining the integrity of the data, which can lead to better performance in classifying minority class instances.

4. *Tomek Links*: By removing majority class data points which seems to be closest to the minority class, Tomek Links enhance class separation and improve the model's ability to distinguish between classes (Tomek, 1976). Tomek Links identify pairs of nearest neighbors from different classes and remove the majority class instances from those pairs. This reduces the overlap between classes and enhances the classifier's ability to differentiate between them. The application of Tomek Links helps to improve the precision of the minority class predictions and reduce false positives.
5. *SMOTE-ENN*: A hybrid method combining SMOTE with Edited Nearest Neighbors, which removes noisy or misclassified data points to create a cleaner and more effective dataset. SMOTE-ENN first applies SMOTE to generate synthetic minority class instances, followed by ENN to clean the dataset by removing examples that are misclassified by their nearest neighbors. This combination helps in reducing noise and enhancing the dataset's quality, ultimately improving the model's performance and robustness.
6. *SMOTE-Tomek Links*: This combination of SMOTE and Tomek Links aims to benefit from both synthetic over-sampling and reduction of class overlap, achieving

a balanced and well-separated dataset. By first generating synthetic minority class instances using SMOTE and then applying Tomek Links to remove the most confusing majority class instances. This dual approach helps to refine the dataset and improve the model's accuracy in detecting minority class instances.

Data balancing is performed on each local node per each healthcare participant (Client), as each one has their own CERT's insider threat dataset version. For example, in our architecture simulation design, healthcare participant 1 uses r4.1, Healthcare Participant 2 uses r5.2, and Healthcare Participant 3 uses r6.2.

3.3.2.2 Data Splitting

To maximize the data used for training in the decentralized federated learning context, the *train_test_split* function using Scikit-learn with an 80-20 split. This decision aligns with findings that suggest better model generalization and performance in federated learning environments with imbalanced datasets compared to the traditional 70-30 split, especially when dealing with imbalanced datasets (Thapa et al., 2020).

The 80-20 split provides a larger training set, which is crucial for federated learning where data is distributed horizontal in a serverless peer-to-peer communication across multiple nodes. This method enables the model to generalize better by allowing more training examples to refine its parameters. Conversely, a 70-30 split might result in insufficient training data, potentially hampering model performance. An 80-15-5 split, allocating 5% to validation, is another approach, but prioritizing training data is particularly important in federated learning to maximize the efficacy of decentralized training.

By using this split, we ensured that 80% of the data is used for training and 20% is reserved for testing. Furthermore, to enhance privacy for each participant client,

differential privacy mechanisms were incorporated. Differential privacy ensures that each healthcare participants' data remains confidential by adding noise to the data during the training process, preventing any single data point from being distinguishable from the aggregated model updates. This approach is particularly significant in decentralized federated learning environments like our architecture design, where model updates from multiple nodes are shares (align with advancing collaborative healthcare for threat intelligence sharing). By ensuring privacy-preserving data analysis, differential privacy supports the robustness of the 80-20 split by allowing more data to be used for training without compromising individual privacy. This strategy also aligns with our aim to build a secure and efficient decentralized federated learning system for insider threat detection.

By incorporating differential privacy, our plan is to secure both local models and exchanged models for each node. Introducing noise to the data can impact accuracy; therefore, using an 80-20 split allows more data for training, mitigating the loss of accuracy due to noise. This approach is essential to maintain high performance and robust privacy in our federated learning system.

3.4 Model Architecture Selection

In this section, we establish comparative baselines to validate one of the three research hypotheses, which guides our Praxis research implementation process and decision making. Each model architecture is examined for its performance results and limitations through a structured comparative analysis, including data preparation, feature scaling, model architecture, training, and evaluation to ensure a fair comparison. This comparative analysis will guide the validation of our research question (RQ3): Which Model Architecture, Deep Neural Network (DNN) or Long Short-Term Memory (LSTM),

yields higher classification performance for insider threat detection when integrated with selected balancing techniques (SMOTE, ADASYN, Near Miss, SMOTE-ENN, Tomek Links, or SMOTE-Tomek Links) within the Advanced Privacy-Preserving Decentralized Federated Learning System?

As highlighted in Chapter 2, Federated Learning (FL) is recognized as a potential solution in the healthcare sector due to its ability to facilitate collaborative learning while preserving data privacy. Both Deep Neural Networks (DNNs) and Long Short-Term Memory networks (LSTMs) have become popular architectures for model training in the literature. DNNs are highly valued for their capacity to manage complex, high-dimensional data, making them application for a wide range of applications. On the other hand, LSTMs are particularly effective for tasks that requires sequential or time-series data, such as detecting temporal patterns in insider threat behavior. Both architectures can be utilized within FL, but Deep Neural Networks (DNN) are often preferred for handling large-scale, decentralized datasets requiring deep learning across multiple layers.

Due to the nature of our datasets, which encompasses over 500 features including time series and behavioral activity scenarios, we initially predicted that an optimized LSTM model architecture would outperform an optimized DNN model architecture. However, during the implementation phase, we began with the standard base DNN and its optimized version due to DNNs' capability to handle complex, high-dimensional data, such as the CERT insider threat dataset. The selection also took into consideration the integration of six balancing techniques at each local node, amounting to a total of 18 balancing processes across the three nodes.

Furthermore, both local and shared model updates were required to comply with general healthcare data privacy regulations. To ensure compliance, we applied the principles of secure by design (Cybersecurity and Infrastructure Security Agency [CISA], 2023; InfoSec Champ, n.d.; Adkins, 2024; Fitzgerald, 2024; National Security Agency, 2023). Our goal was to implement both security best practices and secure by design cybersecurity engineering architecture principles (National Institute of Standards and Technology, 2004; National Institute of Standards and Technology, 2020; National Institute of Standards and Technology, 2022), as the engineering trustworthy secure systems, which involve integrating security measures into the design and development process from the outset. This dual approach ensures that our system adheres to widely accepted security guidelines in healthcare and integrates security measures from the inception of the design process, reinforcing the robustness and reliability of the federated learning approach.

3.4.1 Optimized Deep Neural Networks

We selected a peer-to-peer (P2P) structure for our federated learning (FL) architecture to meet the specific needs of our insider threat detection system for collaborative healthcare institutions simulation context. This setup allows three separate collaborative healthcare institutions to act as decentralized participants (clients) for threat intelligence sharing. Each institution independently securely trains on its local data while also securely sharing model updates without relying on a central server. According to Singh et al. 2022; Wang et al, 2021), P2P-FL is highly effective in scenarios that require decentralized, collaborative learning, as it supports server-less communication and enables direct peer-to-peer interactions between clients. Each client can thus contribute to a shared

goal of enhanced insider threat detection accuracy while maintaining security control over its own data. This architectural choice aligns with healthcare-specific data-sharing restrictions, emphasizing data privacy and preventing the centralization of sensitive information. According to LeCun et al. (2015), the mathematically transformation of an optimized DNN at each layer l(l) can be expressed as:

$$h^{(l)} = f\left(W^{(l)}h^{(l-1)} + b^{(l)}\right)$$

Equation 3. An optimized DNN (LeCun et al., 2015)

Where:

- $h^{(l)}$ = x is the activation of the l-th layer.
- $W^{(l)}$ is the activation from matrix for the l-th layer.
- $b^{(l)}$ is the bias vector for the l-th layer.
- f is the activation function is applied to the linear combination of the input features, such as ReLU in our case.

Note that, the above term "l-th layer" ("l-th layer") refers to a specific layer within a neural network, where "l" ("l") represents the layer's index number. In a neural network with multiple layers, each layer is indexed sequentially starting from the input layer (which can be considered layer 0). For example: The 1st layer (l = 1) is the first hidden layer immediately after the input layer, while the 2nd layer (l = 2) is the second hidden layer, and so on. The last layer, often referred to as the output layer, is typically indexed as the final value of "l". So, "l-th layer" is simply a way of denoting a specific layer in the neural network, using the variable "l" to represent its position in the sequence of layers. This notation is useful for describing the operations or transformations that occur at each layer in a general and abstract manner.

Before creating the optimized DNN, we, first created a standard base Deep Neural Network (DNN) model architecture with an input layer consisting of 502 neurons to represent features. This input layer followed by two hidden layers with 64 and 32 neurons, respectively, and an output layer with 2 neurons. The ReLU (Rectified Linear Unit) activation function was applied after the linear transformations of the first two layers. This model was designed as a Feedforward Neural Network (FNN) or Multi-Layer Perceptron (MLP), featuring three linear layers.

The input layer processed the 502 features, which were then mapped to the hidden layers. The first hidden layer reduced the dimensionality to 64 neurons, and the second hidden layer further reduced it to 32 neurons before reaching the final output layer of 2 neurons for binary classification (insider versus non-insider), as it is illustrating in Figure 15.

Next, for the optimized DNN model, we restructured the architecture with targeted enhancements aimed at improving model performance, generalization, and robustness against overfitting. The input layer is designed to handle 502 features derived from critical insider threat indicators, such as logon activity, file access, and USB device usage patterns. This comprehensive input representation captures the multifaceted nature of insider threats, allowing the model to learn from complex behavioral data. To effectively process this high-dimensional input, the first hidden layer reduces the dimensionality to 128 neurons, which strikes a balance between model complexity and computational efficiency.

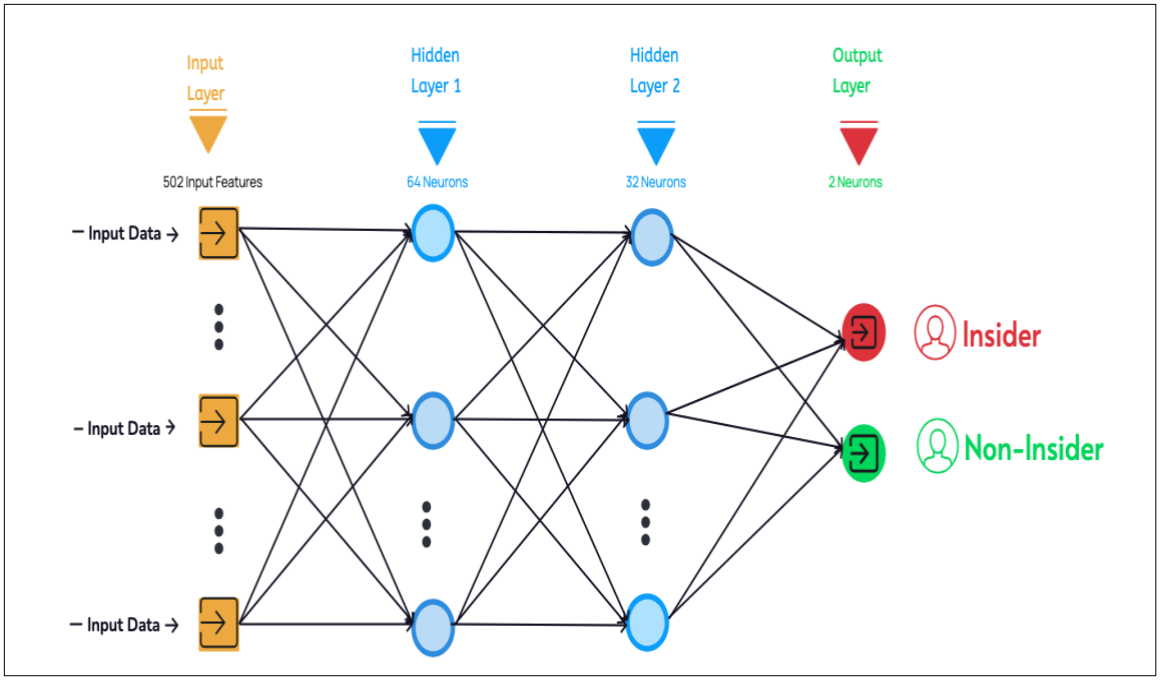


Figure 13. Standard Base DNN Model Architecture Design.

A Multi-Layer Neural Network (Deep Learning Model Architecture Approach).

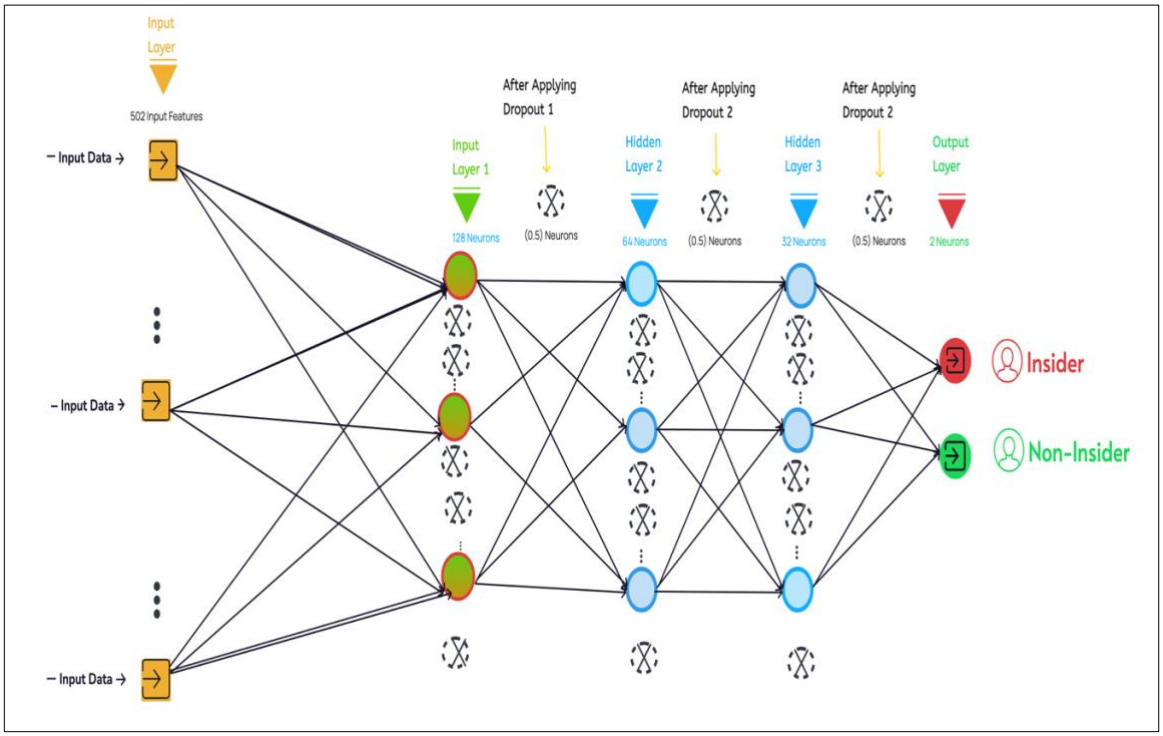


Figure 14. Optimized DNN Model Architecture Design.

A Multi-Layer Neural Network (Deep Learning Model Architecture Approach).

The second hidden layer further reduces this to 64 neurons, followed by a third hidden layer narrowing it down to 32 neurons. This progressive reduction of neurons is a deliberate strategy to create increasingly abstract, discriminative feature spaces at each layer, enabling the model to capture complex patterns in insider threat behaviors. The final output layer produces two neurons, providing binary classification (insider vs. non-insider) capabilities.

To capture the intricate, non-linear relationships present in the data, we incorporated ReLU (Rectified Linear Unit) activation functions in all hidden layers" (Agarap, 2018). ReLU effectively mitigates issues like vanishing gradients, which can be a concern training in deep networks, and thus enables our optimized DNN to learn more complex representations. The output layer, meanwhile, produces logits, which are subsequently converted into probabilities to ensure that the model provides interpretable predictions, making the results usable for downstream applications in threat monitoring and detection.

In terms of optimization, we selected "Adam optimizer for its adaptive learning rate capabilities for its popularity for handling challenging dataset with stability and lower memory requirements" (Vishwakarma, 2024; Kan, 2020; Brownlee, 2021), with a learning rate set to 0.001 which is its often the defaults choice, balancing quick convergence with stability across diverse tasks (Tuychiev, 2024). We used a learning rate of 0.0001 instead of the defaults because we observed oscillation in training loss with high learning rate and that using it causes model to fail converging properly. Additionally, we used a batch size of 64, which maintains a balance between training stability and computational demands, particularly suitable for federated settings where client devices may vary in computational

power. Given the high-dimensional input and complex nature of insider threat data, these choices ensure that the model is not only computationally efficient but also scalable across different healthcare participants (clients) environments.

To further enhance model robustness and prevent overfitting, a critical concern given the limited number of positive cases in insider threat datasets, we applied dropout regularization in each hidden layer. Dropout regularization is a widely-used technique for reducing overfitting in neural networks, particularly in models with limited positive cases, as seen in insider threat detection datasets (Hinton et al., 2012). By randomly "dropping" or deactivating 50% of neurons in each hidden layer during training, dropout forces the model to learn redundant, alternative pathways for data representation rather than relying on specific neurons or features. This encourages the network to generalize better on new, unseen data by reducing its dependency on particular connections. Consequently, dropout regularization helps prevent overfitting, leading to more robust and resilient models in complex tasks like insider threat detection, where data imbalance can otherwise skew model training towards certain features (Srivastava et al., 2014).

Inspired the above widely-used technique, we design our optimized DNN model architecture with a dropout rate set at 0.5, the model randomly drops 50% of neurons in each hidden layer during training, forcing it to learn redundant pathways for data representation. This strategy helps the model generalize better when dealing with the highly imbalanced CERT insider threat datasets, as it reduces reliance on any specific subset of neurons. In addition, batch normalization was included to stabilize the training process and further boost generalization by normalizing the inputs to each layer, thereby reducing sensitivity to shifts in input distributions over the training process.

These architectural and functional adjustments collectively transform the optimized DNN model from a basic structure to one that is more resilient, generalizable, and computationally efficient that improved the performance. For visual illustration of the optimized DNN model architecture design, refer to the Figure 14.

3.4.2 Optimized Long Short-Term Memory

In the implementation of the Long Short-Term Memory (LSTM) model, we designed an architecture specifically tailored for tasks involving sequential or time-series data, such as detecting temporal patterns in insider threat behavior. The LSTM model commenced with an input layer that processed sequences of features derived from logon activity, file access, and USB device usage. To capture these temporal dependencies, the architecture consisted of two LSTM layers, each comprising 64 units. These layers were chosen for their ability to retain information over long sequences, which is crucial for identifying insider threats. According to Hochreiter & Schmidhuber (1997), the operation of an optimized LSTM cell at time step t can be mathematically represented as:

1. Forget Gate Activation:

$$f_t = \sigma\big(W_f \cdot [h_{t-1}, x_t] + b_f\big)$$

Equation 4. Optimized LSTM- Forget Gate Activation

(Hochreiter & Schmidhuber, 1997)

2. Input Gate Activation:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)$$

Equation 5. Optimized LSTM- Input Gate Activation

(Hochreiter & Schmidhuber, 1997)

3. Output Gate Activation:

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o)$$

Equation 6. Optimized LSTM- Output Gate Activation

(Hochreiter & Schmidhuber, 1997)

4. Candidate Cell State:

$$\widetilde{C_t} = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C)$$

Equation 7. Optimized LSTM- Candidate Cell State

(Hochreiter & Schmidhuber, 1997)

5. Cell State Update:

$$C_t = f_t * C_{t-1} + i_t * \widetilde{C_t}$$

Equation 8. Optimized LST- Cell State Update

(Hochreiter & Schmidhuber, 1997)

6. Hidden State Update:

$$h_t = o_t * \tanh(C_t)$$

Equation 9. Optimized LST- Hidden State Update

(Hochreiter & Schmidhuber, 1997)

Where:

- f_t is the forget gate's activation.
- i_t is the input gate's activation.
- o_t is the output gate's activation.
- $\widetilde{C_t}$ is the candidate cell state.
- C_t is the cell state.
- h_t is the hidden state.
- W_f, W_i, W_o, W_C are weights,
- b_f, b_i, b_o, b_C are biases.
- σ represents the sigmoid function
- tanh represents the hyperbolic tangent function.

Recurrent dropout was incorporated within the LSTM layers to prevent overfitting and enhance model generalization. Following the LSTM layers, a fully connected layer was included to reduce dimensionality, culminating in an output layer with 2 neurons for

binary classification of insider versus non-insider threats. We used ReLU activation functions in the intermediate layers “to introduce non-linearity, allowing the model to capture complex patterns effectively” (Hochreiter & Schmidhuber, 1997).

The model optimization process used the Adam optimizer, set with a learning rate of 0.0001 and a batch size of 64. These parameters were meticulously chosen to ensure the model's stability and efficiency during training. Although our initial expected the LSTM model architecture to perform well for its capacity to manage sequential data, empirical evaluations revealed that the optimized DNN model outperformed the LSTM in terms of accuracy and F1-score. Consequently, the DNN model was selected for our APPDFL system's final implementation. For a more detailed comparison between optimized DNN and LSTM, refer to Table 26.

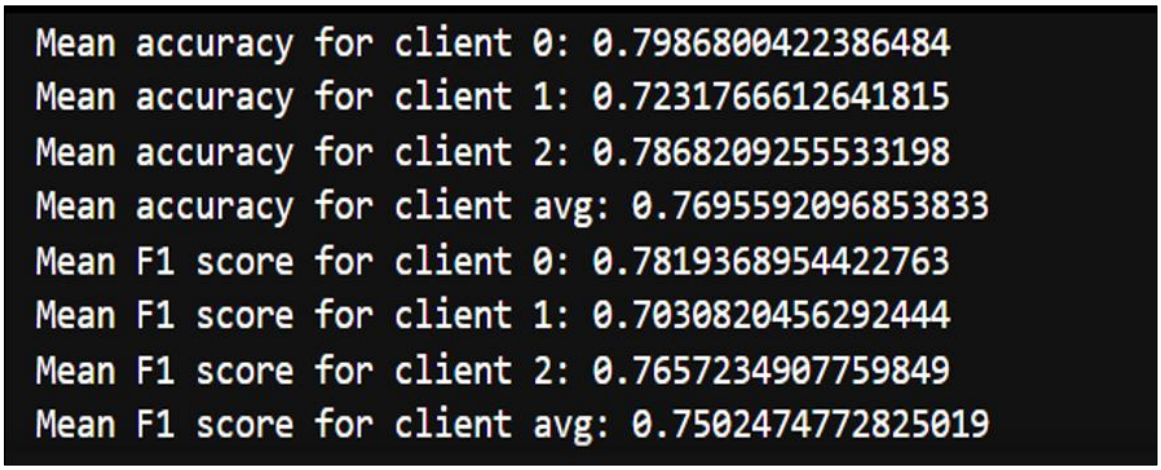


Figure 15. Performance Report of APPDFL using an optimized LSTM Model.

3.4.3 Hyperparameter Tuning

Hyperparameter tuning is crucial for optimizing model performance. For both the DNN and LSTM models, empirical methods were employed to determine the optimal parameters, considering the computational costs and constraints inherent in decentralized learning environments. These parameters were identified through extensive early-stage

Hyperparameter	Value
Learning Rate (η)	0.0001
Local Epochs	10
Global Iterations	10
Batch Size	128
Number of Peers	3
Dropout Rate	0.5

Table 2. Hyperparameters of APPDFL

testing to balance performance and computational efficiency. Similarly, for the LSTM model, the inclusion of recurrent dropout within the LSTM layers and the application of a learning rate and batch size consistent with the DNN model were critical for achieving robust training outcomes. These carefully tuned hyper-parameters facilitated the models' generalizability across diverse datasets while maintaining efficient training processes. The table 2, summarizes the key hyperparameters used in the proposed Advanced Privacy-Preserving Decentralized Federated Learning (APPDFL) system implementation:

3.4.4 Comparative Analysis of Optimized DNN and Optimized LSTM

This section evaluates the use of optimized Deep Neural Networks (DNN) and optimized Long Short-Term Memory networks (LSTM) within Centralized Federated Learning (CFL) and the Advanced Privacy-Preserving Decentralized Federated Learning Approach. This section supports the statement of Hypothesis 3 which proposes that the combination of an optimized DNNs and advanced APPDFL would achieve higher classification accuracy than LSTMs for detecting insider threats. Optimized deep neural

networks (DNNs) are highly efficient because of their deep structures, which enable them to capture complex, nonlinear connections in static data by using hierarchical abstraction.

Conversely, an optimized Long Short-Term Memory (LSTM) networks, a specific kind of Recurrent Neural Network (RNN), are highly proficient in recognizing patterns and preserving long-term connections in sequential data. This makes them valuable for analyzing cumulative user activity logs over a period of time. The assessment in this part aims to determine whether model, either optimized DNN or optimized LSTM, performs better in the advanced APPDFL approach as a solution for detecting insider threats in collaborative healthcare institutions.

Aspect	Optimized DNN	Optimized LSTM
Model Architecture	Fully connected layers	Memory cells with gates
Data Handling	Fixed-size input data	Sequential input data
Training Efficiency	Faster training	Slower training
Model Complexity	Simpler architecture	Complex architecture
Performance	Effective for static data	Effective for sequential data
Privacy Considerations	Differential privacy applicable	Differential privacy applicable

Table 3. Comparison of optimized DNN and optimized LSTM Model Architectures

The optimized DNN is recommended in Federated Learning due to its faster training times, simpler architecture, and effectiveness with static data, making it well-suited for insider threat detection in the Advanced Privacy-Preserving Decentralized Federated

Learning with Differential Privacy (APPDFL) system moreover, the efficiency of optimized DNN validates Hypothesis 3, which suggests that optimized DNN surpasses optimized LSTM in this particular scenario. For more thorough analysis of the results, refer to Chapter 4.

3.4.5 APPDFL Model Results and Performance using optimized DNN Model Architecture

The proposed Advanced Privacy Preserving Decentralized Federated Learning (APPDFL) system was evaluated after 10 epochs and its performance, as it is shown in the Figure 16.

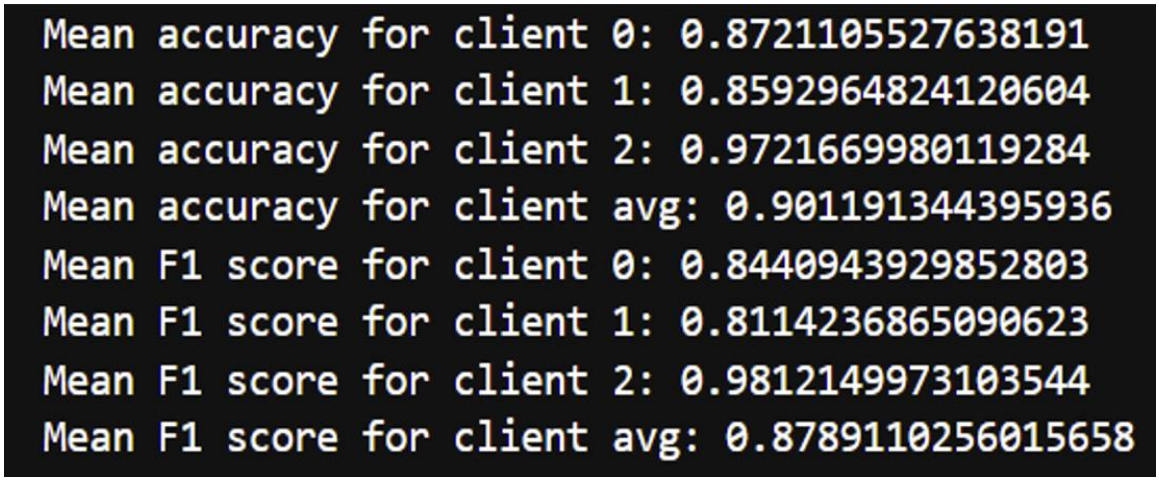


Figure 16. APPDFL results - Optimized DNN Model Architecture.

Pseudo-code for the proposed Advanced Decentralized Federated Learning System
1. Initialization
- Each client initializes its local model independently with same architecture and parameters
2. Local Training at Clients
- For each client:
- Load local dataset
- Train local model for predefined local epochs
- Update local model weights using gradient descent

- Apply differential privacy by small adding noise to the model updates
3. Peer-to-Peer Communication
- Each client randomly selects a few peers to exchange model updates
- Share differentially private model updates with selected peers
4. Model Aggregation at Clients
- Each client aggregates updates received from peers by averaging weights
5. Iteration and Convergence
- Repeat local training, peer-to-peer communication, and aggregation for multiple iterations
6. Evaluation
- Each client evaluates its local model on a validation dataset
Compute : accuracy, precision, recall, F1-Score to monitor performance

Table 4. Pseudo-code for proposed APPDFL System

Adapted from (Chen et al., 2015; Yan & Li, 2024).

3.5 Balancing Privacy and Accuracy in Decentralized Federated Learning Implementation Approach

Our advanced decentralized federated learning system implemented in decentralized federated learning approach with differential privacy mechanism aimed to balance privacy and accuracy through strategic architectural choices and privacy preservation strategies. The primary goal is to develop a detection system compliant with healthcare data privacy regulations, such as the Health Insurance Portability and Accountability Act (HIPAA) and the General Data Protection Regulation (GDPR), while ensuring high performance in insider threat detection.

3.5.1 Privacy Preservation Strategies in Decentralized Federated Learning

To safeguard data privacy, we implemented Differential Privacy (DP) using the Opacus Privacy Engine. The Opacus Privacy Engine was chosen for its ease of integration with PyTorch, efficient batched per-sample gradient computation, and the strong community support that ensures ongoing development and assistance. Opacus injects noise

into gradients before sharing model updates, ensuring that individual data points remain untraceable. This approach aligns with our objective of creating a privacy-preserving federated learning solution for insider threat detection. According to Yousefpour et al. (2022), Opacus provides a practical and user-friendly approach to implementing differential privacy in machine learning models.

The choice adhered to secure by design cybersecurity engineering Architecture Principles, integrating security measures from the inception of the design process. This approach also ensured compliance with security best practices and reinforced the robustness and reliability of our federated learning framework, enhancing data privacy and security from the outset. For more about differential privacy, refer to Chapter 2.

3.5.2 Horizontal Peer-to-Peer as Serverless Communication Method

We adopted a peer-to-peer (P2P) communication topology for our federated learning approach. According to Wang et al. (2021), P2P federated learning is especially effective in scenarios requiring decentralized, collaborative learning, as it supports server-less communication and enables direct peer-to-peer interactions between clients. This topology allowed the three healthcare institutions in our study to function as decentralized clients, each independently training on local data while securely sharing model updates.

The P2P architecture enhances data privacy by keeping sensitive information local to each institution and facilitating collaborative improvement of insider threat detection models without the need for data centralization. This approach aligns with healthcare-specific data-sharing restrictions and supports the secure, decentralized nature of our federated learning system.

3.5.3 Local Model Training and Model Aggregation Synchronous Advantage

Local model training and synchronous updates are fundamental to our federated learning approach. Each client trains its model locally for a fixed number of epochs (for our case it is 10 epochs), followed by synchronous aggregation of model updates. This ensures model consistency across clients at the end of each global training round.

Synchronous updates minimize potential delays or asynchrony due to varied data availability, enhancing the coordination of parallel learning across institutions. This method is effective in maintaining timely and uniform progress in insider threat intelligence sharing. By implementing synchronous updates, we improved the robustness and reliability of the federated learning system, ultimately enhancing classification performance for insider threat detection. Our implementation approach to balancing privacy and accuracy in decentralized federated learning combines advanced privacy preservation strategies, a P2P communication topology, and synchronous model updates. These measures ensure that our federated learning approach is secure, efficient, and capable of delivering high-performance insider threat detection for collaborative healthcare institutions while adhering to stringent data privacy regulations.

3.6. Performance Metrics of Applied Models

As we highlighted throughout this praxis, the effectiveness of the models for detecting insider threats is evaluated using a combination of metrics and strategies. Accuracy and F1-score serve as crucial metrics, providing quantitative insights into the models' performance by comparing predicted outcomes to actual results.

These metrics enable us to assess the precision and recall of the models, ensuring they can accurately identify insider threats without generating excessive false positives or

negatives. Additionally, these metrics help us fine-tune the models to optimize their performance for real-world applications.

Privacy, represented by epsilon, functions as both a metric and a strategy. As a metric, epsilon (ε) quantifies the trade-off between acceptable privacy and model utility, offering a numerical measure such as 0.1. Lower epsilon values indicate stronger privacy but can reduce model accuracy, whereas higher epsilon values enhance performance but weaken privacy guarantees. The privacy budget is balanced with a noise multiplier, which introduces controlled noise to the data or model updates, further guaranteeing the anonymity of individual data points. In our implementation, differential privacy is employed to determine the acceptable noise tolerance level, balancing accuracy and privacy to a degree the system can tolerate. This is crucial for maintaining data confidentiality while achieving effective threat detection. Differential privacy secures both the local data of each participant and the shared model updates. By adding controlled noise through the noise multiplier, it ensures individual data points remain anonymous, protecting sensitive information while enabling collaborative learning across participants.

Collaboration emerges as another key strategy, essential for designing a scalable system that allows multiple healthcare institutions to work together. By sharing intelligence and model updates, institutions can collectively enhance their insider threat detection capabilities without compromising the privacy and security of their data. This collaborative approach not only improves detection accuracy but also fosters a community of practice, where institutions can learn from each other's experiences and continuously improve their security measures. For a comprehensive analysis of evaluated approaches, including detailed metrics and strategies, refer to Figure A-5 in the appendix.

	Predicted Positive	Predicted Negative
Actual Positive	True Positive (TP)	False Negative (FN)
Actual Negative	False Positive (FP)	True Negative (TN)

Table 5. Binary Confusion Matrix for Classification (Fahmy, 2022)

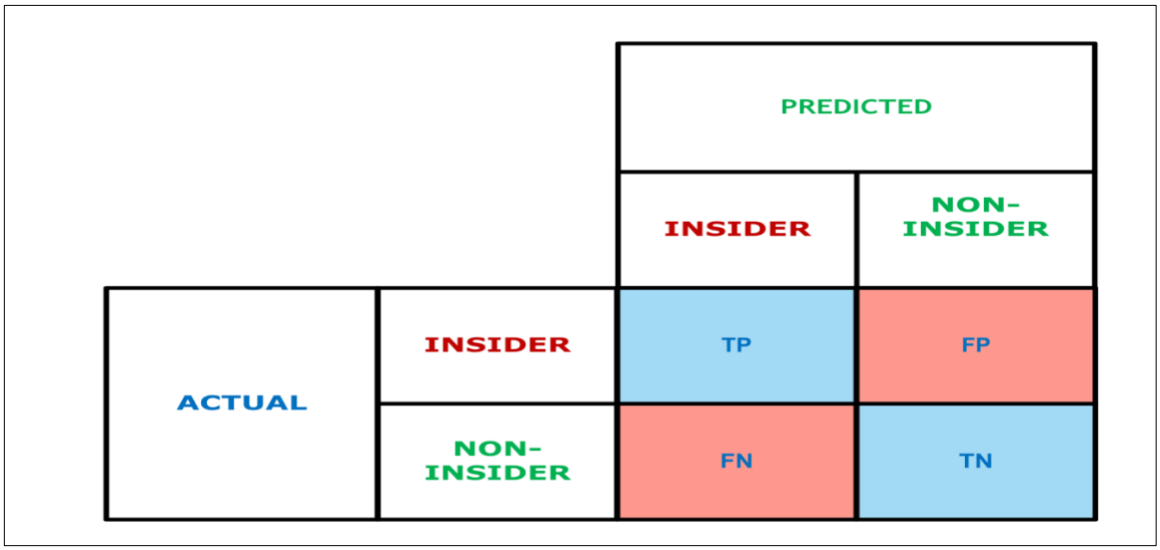


Figure 17. Insider Threat Confusion Matrix.

***Legend**: Blue squares Mean the Data are Correctly Predicted Classification*

According to (Fahmy, 2022), in the above Table 5. and Figure 17, we have:

- *True Positive (TP):* The number of instances where the model correctly predicts an insider threat when it is actually present.
- *False Negative (FN):* The number of instances where the model incorrectly predicts no insider threat when there is actually a threat.
- *False Positive (FP):* The number of instances where the model incorrectly predicts an insider threat when there is no threat.

- *True Negative (TN):* The number of instances where the model correctly predicts no insider threat when there is indeed no threat.

Afterward, using the results determined from the confusion matrix, the performance metrics of the evaluated which are listed below:

- Accuracy: This metric indicates the percentage of correctly classified in- stances out of the total instances which according to Bobbitt (2021), can be calculated as follow:

$$\textbf{Accuracy} = \frac{\text{True Positives} + \text{True Negatives}}{\text{Total Instances}}$$

Equation 10. Accuracy (Bonnet, 2023).

While:

- *F1-Score***:** This metric is "the harmonic mean of precision and recall, providing a single measure of a model's performance in terms of both false positives and false negatives" (Bobbitt, 2021). F1-score can be calculated as:

$$\textbf{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

Equation 11. F1-Score (Bobbitt, 2021*).*

While:

- *Precision:* This metric is "the ratio of correctly predicted positive observations to the total predicted positives" (Bobbitt, 2021). Precision can be calculated as:

$$\textbf{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}}$$

Equation 12. Precision (Bonnet, 2023).

While:

- *Recall:* This metric is "the ratio of correctly predicted positive observations to all observations in the actual class" (Bobbitt, 2021). Recall can be calculated using as follow:

$$\textbf{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$$

Equation 13. Recall (Bonnet, 2023)

While the formula to compute epsilon (**ε**) budget in differential privacy involves different factors, including "the noise multiplier (Sigma), the number of training steps or iterations and the sampling ration"(Dwork, 2006).

$$\textbf{Epsilon } (\varepsilon) = \frac{q \times \sqrt{2} \times T \times \log(1/\delta)}{\sigma}$$

Equation 14. Epsilon (Dwork, 2006).

- (q) is the sampling ratio
- (T) is the given total number of training steps
- (δ) Delta is a small value representing the probability of possible breach.
- (σ) Sigma is the noise multiplier.

Evaluation Measure	Formula	References
Accuracy	$\frac{TP+TN}{TP+TN+FP+FN}$	(Bonnet, 2023)
Precision	$\frac{TP}{TP+FP}$	(Bonnet, 2023)
Recall	$\frac{TP}{TP+FN}$	(Bonnet, 2023)

Evaluation Measure	Formula	References
F1- Score	$2 \times \frac{Precision \times Recall}{Precision + Recall}$	(Bobbitt, 2021)
Epsilon	$\frac{q \times \sqrt{2} \times T \times \log(1/\delta)}{\sigma}$	(Dwork, 2006)

Table 6. The Performance Metric.

Chapter 4—Results

4.1 Introduction

Our research objective is to develop an advanced privacy-preserving decentralized federated learning (APPDFL) system for insider threat in collaborative healthcare institutions. To determine if our proposed APPDFL system satisfy the goal of this Praxis, as highlighted in previous chapters, we also developed and tested traditional and modern machine learning (TML) models alongside a centralized federated learning (CFL), using the same library (TensorFlow Keras) other authors utilized in prior works, and same CERT's insider threat dataset versions (r4.2, r5.2, and r6.2) as we used for our proposed APPDFL system as a solution, ensuring a fair comparison.

This comparison seeks to highlight the benefits of advancing the insider threat detection from centralized methods to a more advanced decentralized federated learning approach with differential privacy mechanism. APPDFL as a server-less with privacy-preserving mechanism built following secure by design architecture principal, provides collaborative healthcare institutions a secure environment for threats intelligence sharing, addressing limitations inherent in centralized methods.

As noted in the prior chapters regarding qualitative research, "understanding complex phenomena necessitates the use of theories that provide a broader viewpoint for interpretation"(QDacity, n.d.). In our study, theory triangulation serves as interpretative framework, that allowed us to integrate insights from centralized and decentralized approaches and enhancing the rigor of our comparative analysis to inform our final decision- making regarding the value added of the proposed advanced privacy-preserving decentralized federated learning (APPDFL) system as the proposed solution.

By examining APPDFL's performance against centralized methods and previous works in this field, we employed theory triangulation to guide our research. This study seeks contributes to the intersection of several key research domains, implement a collaborative approach for insider threat detection specifically tailored for healthcare institutions using an advanced privacy-preserving federated learning solution. Our experience with theorical triangulation in the research process includes underscore the benefit of having a clearly define research objectives aimed at generating an innovative solution. "This clarity ensures that the integration of theories aligns with the research goals and objectives"(QDacity, n.d.). This approach enables strategic advancement from exisiting centralized methods to a server-less decentralized system that foster collaborative participation for insider threat intelligence sharing in a secure environment that protects all participants healthcare institutions. For more details on this this praxis' contributions to the broader body of knowledge, refer to Chapter 5.

Furthermore, we address additional critical research questions designed to validate APPDFL's efficacy. A significant critical aspect of our analysis revolves around these guiding questions: Specifically, explored whether the proposed Advanced Privacy-Preserving Decentralized Federated Learning (APPDFL) solution can identify the optimal trade-off point value to ensures sufficient protection and acceptable accuracy for insider threat detection across collaborative healthcare institutions, addressing Research Hypothesis 2 (RH2).

We investigate which model architecture-Deep Neural Network (DNN) or Long Short-Term Memory (LSTM), yields higher classification performance for insider threat detection when integrated with selected balancing techniques (SMOTE, ADASYN, Near

Miss, SMOTE-ENN, Tomek Links, or SMOTE-Tomek Links) within the Advanced Privacy-Preserving Decentralized Federated Learning System? The goal is to determine which model architecture, integrated with data balancing approaches, yields superior performance. This analysis supports the validation of Research Hypothesis 3 (RH3) presented in Chapter 1.

4.2 Centralized Machine Learning Methods

4.2.1 Traditional and Modern Machine Learning Implementation Process

As discussed in the above, Traditional Machine Learning (TML) aggregates all data from participating institutions into a centralized location for model training, achieving high accuracy due to the large, comprehensive datasets involved. As illustrated in Appendix, Figure A-2, we developed three different traditional centralized machine learning (TML) models: two traditional machine learning models, including logistic regression and random forest, and an artificial neural network (ANN), specifically a deep neural network (DNN), as a modern learning model. Unlike previous works that used only one version or a combination of two versions, all of the traditional centralized methods were developed using CERT's insider threat dataset versions r4.2, r.5.2, and r6.3, the same dataset we used for our proposed solution. Since this is not the main method lather serves as a comparative approach towards a decision-making, the following are the summary of traditional and modern machine learning implementation process**.**

4.2.1.1 Traditional and Modern Machine Learning Data Preparation

The dataset is first loaded and preprocessed. The 'insider' column, which is the target variable indicating whether an activity is an insider threat, is updated for consistency. Features (x) and the target (y) are separated, where x includes all columns except 'insider',

and y is the ’insider’ column. For the feature scaling is utilized to guarantee consistency in the data. This remove the mean and scaling the data to unit variance, StandardScaler improves the uniformity of the features.

4.2.1.2 Traditional and Modern Machine Learning Summary

Since our main methodology is based on Deep Learning, a subset of machine learning which is about using artificial neural network (ANN), let us focus on ANN implementation summary and its findings. These networks consist of layers of interconnected neurons, each layer building upon the last to recognize increasingly complex patterns. Inspired by the book of Goodfellow et al, 2016), the ANN is constructed using TensorFlow’s Keras API which The comprises:

- An input layer matches the number of features we have.
- Two hidden layers with 128 and 64 neurons, respectively, each using the ReLU activation function.
- An output layer with a single neuron and a sigmoid activation function for binary classification.

With a batch size of 128, the model is trained for 10 epochs using the scaled training data. The training process is monitored by the validation set in to avoid model overfit. This ensures that the model generalizes well to unseen data rather than just memorizing the training set. Fine-tuning this balance helps the model maintain high performance without sacrificing accuracy for the unseen data. This approach helps ensure our model won't just perform well on our training data but also remains robust when facing new, real-world scenarios.

Hyperparameter	Value
Learning Rate (η)	0.0001
Batch Size	128
Hidden Layers	2
Neurons in Hidden Layers	128, 64
Activation Function	ReLU (hidden layers), Sigmoid (output layer)
Optimizer	Adam

Table 7. Summary of Hyperparameters for TML

The learning rate (η) is set at 0.0001, utilizing the Adam optimizer, enabling efficient gradient updates that facilitate faster and more stable convergence. The architecture included two hidden layers with 128 and 64 neurons respectively, activated by ReLU functions, providing the model with the non-linear capabilities to capture complex patterns in the data. The output layer employed a Sigmoid activation function which is suitable for binary classification tasks, and the Binary Cross-Entropy loss function aligned with this setup by measuring the disparity between the predicted and actual class probabilities.

Pseudo-code for TML
1. Data Preparation and Splitting - Load dataset, update insider' column - Separate features (x) and target (y) - Split the data into training and testing sets using 80-20 ratio 2. Model Building and Compilation - Standardize features using StandardScaler - Build DNN model with input layer, two hidden layers, and output layer - Compile model with binary cross-entropy loss and Adam optimizer 3. Train model on training data 4. Evaluate model on test set 5. Compute and plot: Accuracy, Precision, F1-score

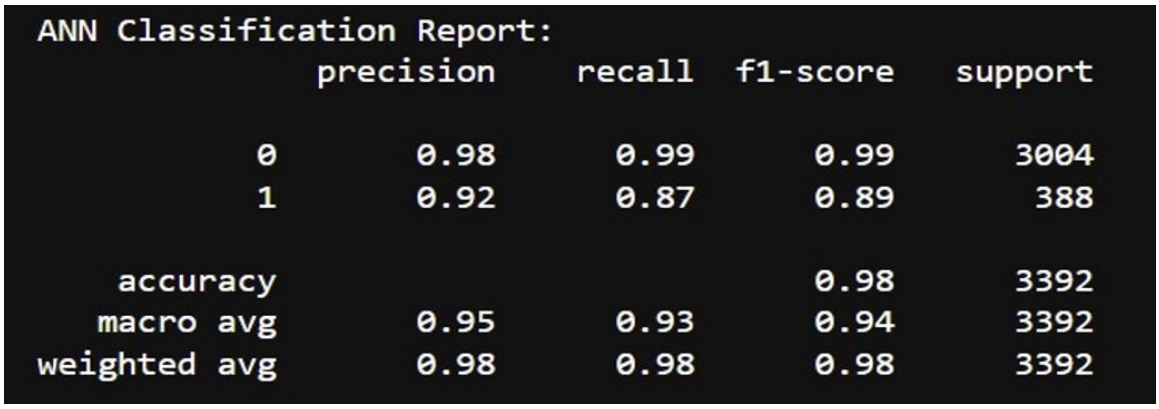
Table 8. Pseudo-code for TML

Source: (Sze et al., 2017).

4.2.2. Modern Machine Learning Model Results and Performance

In terms of performance results, The ANN model performed well in identifying insider and non-insider threats, achieving a high accuracy of 98% and a weighted average F1-score of 98% as shown in Figure 18. The precision and recall for non-insider threats are 98% and 99%, respectively. Despite these outstanding outcomes, ANN model architectures pose significant privacy risks, particularly in healthcare settings where sensitive, personal, and financial information is involved. This concentration of data in one place amplifies the risk to privacy (Iyengar, et al., 2024). The results highlight the TML model's performance, thanks to the availability of extensive datasets. However, centralizing such private information introduces serious privacy issues (Dinov, 2023; Bin Goh et al., 2024).

In order to improve privacy without sacrificing speed, alternative approaches, including the centralized federated learning (CFL) approach, are being investigated.



ANN Classification Report:

	precision	recall	f1-score	support
0	0.98	0.99	0.99	3004
1	0.92	0.87	0.89	388
accuracy			0.98	3392
macro avg	0.95	0.93	0.94	3392
weighted avg	0.98	0.98	0.98	3392

Figure 18. Modern Machine Learning Model Results.

CFL allows data to remain decentralized, thus enhancing a degree of privacy while maintaining superior model accuracy with some risk, as discussed in Chapter 2. In the next section, we will summarize the implementation process and results of the centralized federated learning (CFL) method tested on the same dataset versions as those used for the traditional, modern, and our proposed APPDFL system solutions.

4.3 Centralized Federated Learning Method

As discussed in Chapter 2, CFL aggregates model updates from multiple clients, allowing for collaborative learning without the need to share raw data. This method leverages varied datasets from different institutions, which can enhance the learning process. However, it introduces a single point of failure, the central server, which is not ideal in environments requiring robust privacy preservation, such as healthcare. This poses a significant challenge if adopted, highlighting the necessity to address certain privacy concerns. While CFL significantly enhances privacy by keeping data decentralized and not sharing raw data, the central server's vulnerability means some residual privacy risks remain. Although CFL facilitates data integration and collaborative learning, these privacy issues, while reduced, are not entirely eliminated.

4.3.1 Centralized Federated Learning Implementation Process

1. Initialization: We made the architecture of the global model available to all involved clients and initialized it using random weights. For this task, we used a neural network designed for identifying insider threats as the initial global model. To ensure that every client started with the same model parameters, we provided the same initial weights for this global model to each client (Al-Mhiqani et al., 2020).

2. Local Training at Clients: For each client,

- **Data Loading:** Each client loaded its local dataset into memory for training. This dataset was used to compute the gradients required to update the local model. is used to compute the gradients required to update the local model.
- **Model Training:** We trained the local model for a predefined number of local epochs using Adam as an optimizer and a loss function such as cross-entropy loss. The local model weights were updated using gradient descent:

$$w_k^t = w_k^{t-1} - \eta \nabla \ell(w_k^{t-1}; D_k)$$

Equation 15. The local model weights were updated using gradient descent in CFL (McMahan et al., 2017)

where w_k^t are the weights of the model at client k at time t, η is the learning rate, and $\nabla \ell$ is the gradient of the loss function with respect to the local data D_k.

According to McMahan et al. (2017), the following are the steps and equation:

1. **Aggregation at Central Server:** After local training, clients send their up- dated model weights to the central server. This step is crucial because "the central server aggregates these updates by averaging them to form a new global model" (McMahan et al., 2017):

$$\boldsymbol{w^t} = \frac{1}{K}\sum_{k=1}^{K} w_k^t$$

Equation 16. Aggregation at Central Server in CFL

(McMahan et al., 2017)

where K is the number of clients. This aggregation step ensures that the global model benefits from the data and learning progress of all participating clients. The server averages the updated weights w_k^t received from K client to form the new global model w^t. This step is crucial for updating the global model with contributions from all clients.

2. **Model Redistribution:** For the next round of local training, the new global model is sent to all clients again. This process repeats itself a set number of times, called" global iterations." In each iteration, clients change their local models based on the global model they receive. They then send their updated weights back to the central computer so that they can be added together.
3. **Evaluation:** The global model is tested on a validation sample every time iteration is finished to evaluate its performance. To keep track of the model's growth and make sure it is learning well, metrics include accuracy, and F1-score are calculated. This enable us to observe if the model is convergent and gives information about how well it did in different training runs.

4.3.2 Centralized Federated Learning Model Results and Performance

For this experiment, we simulated three participant clients and test the models for centralized federated learning (CFL). We used the same dataset we used for other model developments include our proposed solution. We also applied the data-balancing techniques to address the underlying class imbalance. SMOTE, ADASYN, Near Miss, Tomek Links, SMOTE-ENN, and SMOTE-Tomek are among the most widely used to

balancing imbalanced data that have been explored based on their ability to enhance the identification of minority class occurrences. With these techniques, we want to improve the models' theft detection capabilities without benefiting the majority class.

Pseudo-code for Centralized Federated Learning
1. Initialization - Initialize global model with random weights - Share model architecture and weights with all clients 2. Local Training at Clients - For each client: - Load local dataset - Train local model for predefined local epochs - Update local model weights using gradient descent - Send updated weights to central server 3. Aggregation at Central Server - Aggregate updates from all clients by averaging weights - Form new global model with aggregated weights 4. Model Redistribution - Redistribute updated global model to all clients - Repeat local training and aggregation for predefined global iterations 5. Evaluation - Evaluate global model on validation dataset after each global iteration - Compute accuracy, Accuracy, F1-Score and monitor performance

Table 9. Pseudo-code for Centralized Federated Learning

Adapted from (Pelikan et al., 2023; Piran et al., 2024)

Moreover, in real-world environment of health care most of the data is imbalanced and to overcome this situation we include the balancing techniques especially in federated learning which enables the participants to adopt the data for insider theft detection. The model performance is evaluated with the Precision and F1-Scores for each customer which are shown here, together with the average score for all clients. In particular, this research is important to understand how each balancing technique contributes to the accuracy verses privacy trade-off in centralized federated learning (CFL) model, where input data stays local, but model updates are centralized.

4.3.3 Balancing Techniques in Centralized Federated Learning

The analysis of the accuracy results shows that the SMOTE-ENN technique yielded the highest total average accuracy at 0.8586 (85.86%), demonstrating its effectiveness in handling imbalanced data within the context of Centralized Federated Learning (CFL). Following closely is the Tomek Links technique with an average accuracy of 0.8442 (84.42%), indicating strong performance in improving the model's robustness. The SMOTE-Tomek method also performed well, achieving an average accuracy of 0.8444 (84.44%), slightly higher than the standalone Tomek Links technique.

Next in line is the SMOTE technique, which achieved a commendable average accuracy of 0.8431 (84.31%), showcasing its capability in balancing datasets effectively. Adasyn had a total average accuracy of 0.8320 (83.20%), indicating its reasonable effectiveness but not quite matching up to the performance of the top techniques. Finally, Near Miss achieved the lowest total average accuracy at 0.8116 (81.16%), suggesting that it might not be as effective in improving the accuracy results within this specific CFL context. This performance is particularly significant considering the complexity and high class imbalance present in the CERT's insider threat dataset (version r4.2, r5.2, and r6.2) we used. For the detailed results, refer to Table X:

For participating healthcare client 1, the Tomek Links technique showed the highest accuracy at 0.8965 (89.65%), followed closely by SMOTE at 0.8939 (89.39%). This demonstrates the capability of Tomek Links in handling participating client 1's imbalanced data effectively. The SMOTE-ENN technique also yielded high accuracy at 0.8924 (89.24%), underscoring its effectiveness in improving the model's performance. Adasyn achieved an accuracy of 0.8919 (89.19%), slightly lower than SMOTE-ENN but

Techniques	Client 1	Client 2	Client 3	Total Average
SMOTE	0.8939	0.74155	0.8939	0.8431
Adasyn	0.8919	0.7122	0.8919	0.8320
Near Miss	0.8692	0.6963	0.8692	0.8116
Tomek Links	0.8965	0.7398	0.8965	0.8442
SMOTE-ENN	0.8924	0.7903	0.8924	0.8586
SMOTE-Tomek	0.87773	0.7841	0.8773	0.8444

Table 10. Accuracy Results using Balancing Techniques (CFL)

still commendable. Participating healthcare client 2 exhibited the highest accuracy with the SMOTE-ENN technique at 0.7903 (79.03%), showcasing its strong performance in this client's data balancing. SMOTE-Tomek followed with an accuracy of 0.7841 (78.41%), indicating its substantial efficacy. Tomek Links and SMOTE techniques also performed well, with accuracies of 0.7398 (73.98%) and 0.7415 (74.15%), respectively. Near Miss showed the lowest accuracy at 0.6963 (69.63%), suggesting its lesser effectiveness for client 2.

For participating healthcare client 3, both the SMOTE and Tomek Links techniques achieved the highest accuracy at 0.8939 (89.39%) and 0.8965 (89.65%), respectively, demonstrating their strong performance in handling this client's data. SMOTE-ENN followed closely with an accuracy of 0.8924 (89.24%), indicating its robust effectiveness. Adasyn also performed well with an accuracy of 0.8919 (89.19%). Near Miss and SMOTE-Tomek had lower accuracies of 0.8692 (86.92%) and 0.8773 (87.73%), respectively, suggesting these techniques were less effective for client 3 compared to other.

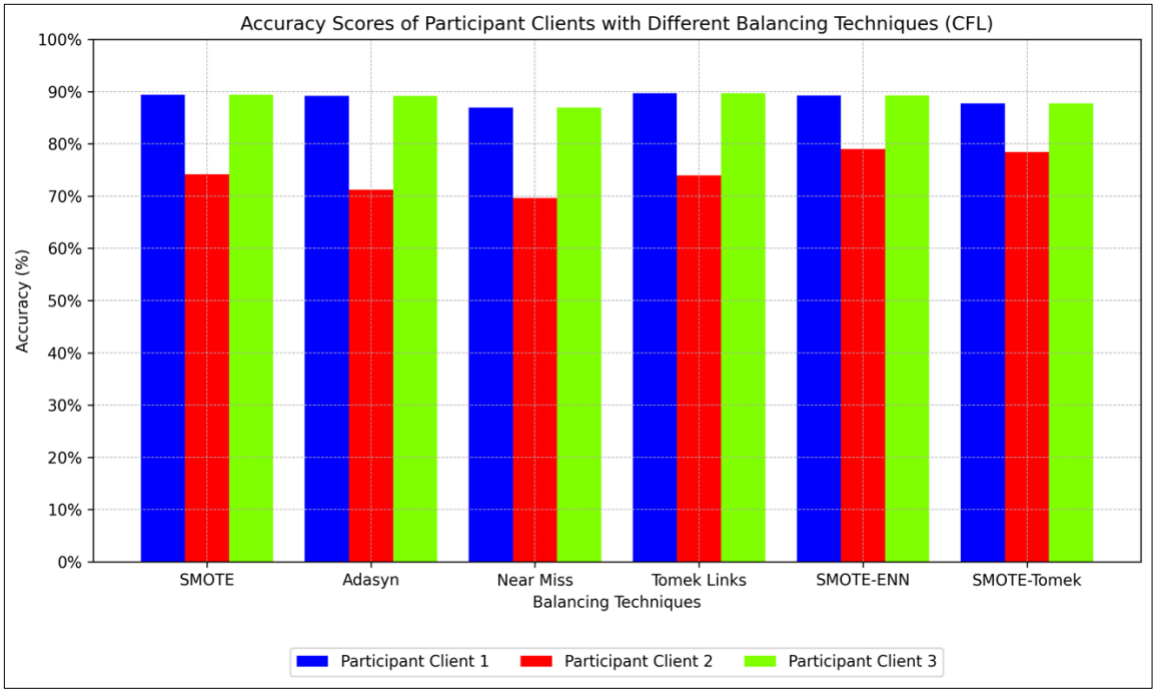


Figure 19. Accuracy Results suing Balancing Techniques (CFL).

This performance is particularly significant considering the complexity and high class imbalance present in the CERT's insider threat dataset (version r4.2, r5.2, and r6.2). For the detailed visual representation, refer to Figure 19.

The analysis of the F1-Scores shows that the SMOTE technique yielded the highest total average F1-Score at 0.8552 (85.52%), indicating its strong performance in handling imbalanced data within the context of Centralized Federated Learning (CFL). Adasyn follows with an average F1-Score of 0.8148 (81.48%), demonstrating its considerable effectiveness, although not as high as SMOTE. The Tomek Links technique also performed well, achieving an average F1-Score of 0.8115 (81.15%), showing its capability to enhance model performance.

SMOTE-Tomek achieved an average F1-Score of 0.8047 (80.47%), indicating its effectiveness, although slightly lower than Tomek Links. The SMOTE-ENN technique

Techniques	Client 1	Client 2	Client 3	Total Average
SMOTE	0.8773	0.7219	0.8552	0.8552
Adasyn	0.8786	0.4680	0.8148	0.8148
Near Miss	0.8655	0.4848	0.7393	0.7393
Tomek Links	0.8666	0.7015	0.8115	0.8115
SMOTE-ENN	0.8714	0.6536	0.7988	0.7988
SMOTE-Tomek	0.8738	0.6664	0.8047	0.8047

Table 11. F1-Score Comparison with Balancing Techniques in (CFL)

had a total average F1-Score of 0.7988 (79.88%), showcasing its ability to improve model accuracy, but not at the same level as SMOTE or Adasyn. Finally, Near Miss had the lowest total average F1-Score at 0.7393 (73.93%), suggesting its lesser effectiveness in enhancing the F1-Scores within this specific CFL context. This performance is particularly significant considering the complexity and high class imbalance present in the CERT's insider threat dataset (version r4.2, r5.2, and r6.2). For more details, refer to Table 11.

For participating healthcare client 1, Adasyn showed the highest F1-Score at 0.8786 (87.86%), followed closely by SMOTE at 0.8773 (87.73%). This demonstrates Adasyn's capability in handling participating client 1's imbalanced data effectively. The SMOTE-ENN technique also yielded a high F1-Score at 0.8714 (87.14%), underscoring its effectiveness in improving the model's performance. Tomek Links achieved an F1-Score of 0.8666 (86.66%), slightly lower than SMOTE-ENN but still commendable. Participating healthcare client 2 exhibited the highest F1-Score with the SMOTE technique

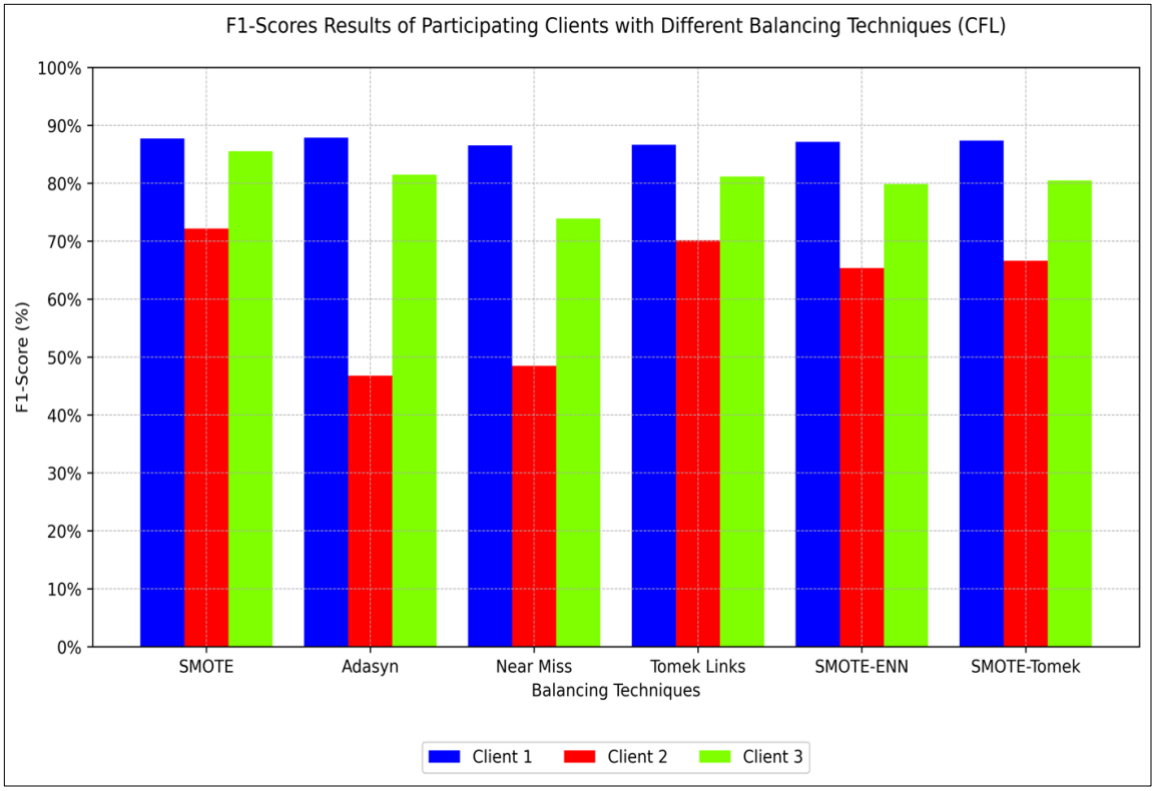


Figure 20. F1-Score Comparison with Balancing Techniques in (CFL).

at 0.7219 (72.19%), showcasing its strong performance in this client's data balancing. Tomek Links followed with an F1-Score of 0.7015 (70.15%), indicating its substantial efficacy. SMOTE-ENN and SMOTE-Tomek also performed well, with F1-Scores of 0.6536 (65.36%) and 0.6664 (66.64%), respectively. Near Miss showed the lowest F1-Score at 0.4848 (48.48%), suggesting its lesser effectiveness for client 2.

For participating healthcare client 3, the SMOTE technique achieved the highest F1-Score at 0.8552 (85.52%), demonstrating its strong performance in handling this client's data. SMOTE-Tomek followed closely with an F1-Score of 0.8047 (80.47%), indicating its robust effectiveness. Adasyn also performed well with an F1-Score of 0.8148 (81.48%). Tomek Links had a lower F1-Score of 0.8115 (81.15%), suggesting these techniques were effective but not as much as SMOTE. Near Miss and SMOTE-ENN had

lower F1-Scores of 0.7393 (73.93%) and 0.7988 (79.88%), respectively, suggesting these techniques were less effective for client 3 compared to others. This performance is particularly significant considering the complexity and high class imbalance present in the CERT's insider threat dataset (Version: r4.2, r5.2, and r6.2). For the detailed visual representation, refer to Figure 20.

As centralizing sensitive data has considerable privacy issues, CFL is not advised for use in healthcare settings, even if it enhances model performance by aggregating data across institutions or improved model performance, CFL is less suited than privacy-preserving methods such as proposed APPDFL. For the overall performance of CFL, refer to the Figure 21.

```
Mean accuracy for client 0: 0.8789285714285715
Mean accuracy for client 1: 0.7898659966499163
Mean accuracy for client 2: 0.8789285714285715
Mean accuracy for client avg: 0.849241046502353
Mean F1 score for client 0: 0.8822105181744245
Mean F1 score for client 1: 0.7957336280779064
Mean F1 score for client 2: 0.8822105181744245
Mean F1 score for client avg: 0.8533848881422518
```

Figure 21. Performance Report of CFL.

4.4 Advanced Privacy-Preserving Decentralized Federated Learning

Based on our experiments, the accuracy for the proposed Advanced Privacy-Preserving Decentralized Federated Learning (APPDFL) system using different data-balancing techniques are shown below. The analysis of the F1-Scores indicates that the Tomek Links technique achieved the highest total average accuracy of 0.9012 (90.12%), showcasing its superior performance in handling imbalanced data within the context of

Techniques	Client 1	Client 2	Client 3	Total Average
SMOTE	0.8119	0.7152	0.8222	0.777
Adasyn	0.7938	0.7421	0.8535	0.7964
Near Miss	0.6550	0.5136	0.5900	0.5865
Tomek Links	**0.8716**	**0.8450**	**0.9746**	**0.9012**
SMOTE-ENN	0.7497	0.6341	0.8253	0.7363
SMOTE-Tomek	0.7930	0.7004	0.8507	0.7814

Table 12. APPDFL Accuracy Comparison with Balancing Techniques

Advanced Privacy-Preserving Decentralized Federated Learning (APPDFL). Adasyn follows with an average accuracy of 0.7964 (79.64%), demonstrating its strong effectiveness but not quite reaching the level of Tomek Links. SMOTE-Tomek also performed well, achieving an average accuracy of 0.7814 (78.14%), indicating its capability to enhance model performance.

SMOTE achieved a total average accuracy of 0.777 (77.7%), showcasing its reasonable effectiveness but falling short compared to Tomek Links and Adasyn. The SMOTE-ENN technique had an average accuracy of 0.7363 (73.63%), highlighting its ability to improve model accuracy, though not as effectively as the top techniques. Finally, Near Miss had the lowest total average accuracy at 0.5865 (58.65%), suggesting its lesser effectiveness in enhancing the accuracy within this specific APPDFL context. For more details, refer to Table 12. For participating healthcare client 1, the Tomek Links technique showed the highest accuracy of 0.8716 (87.16%), followed closely by SMOTE at 0.8119 (81.19%). This demonstrates the capability of Tomek Links in handling participating client 1's imbalanced data effectively.

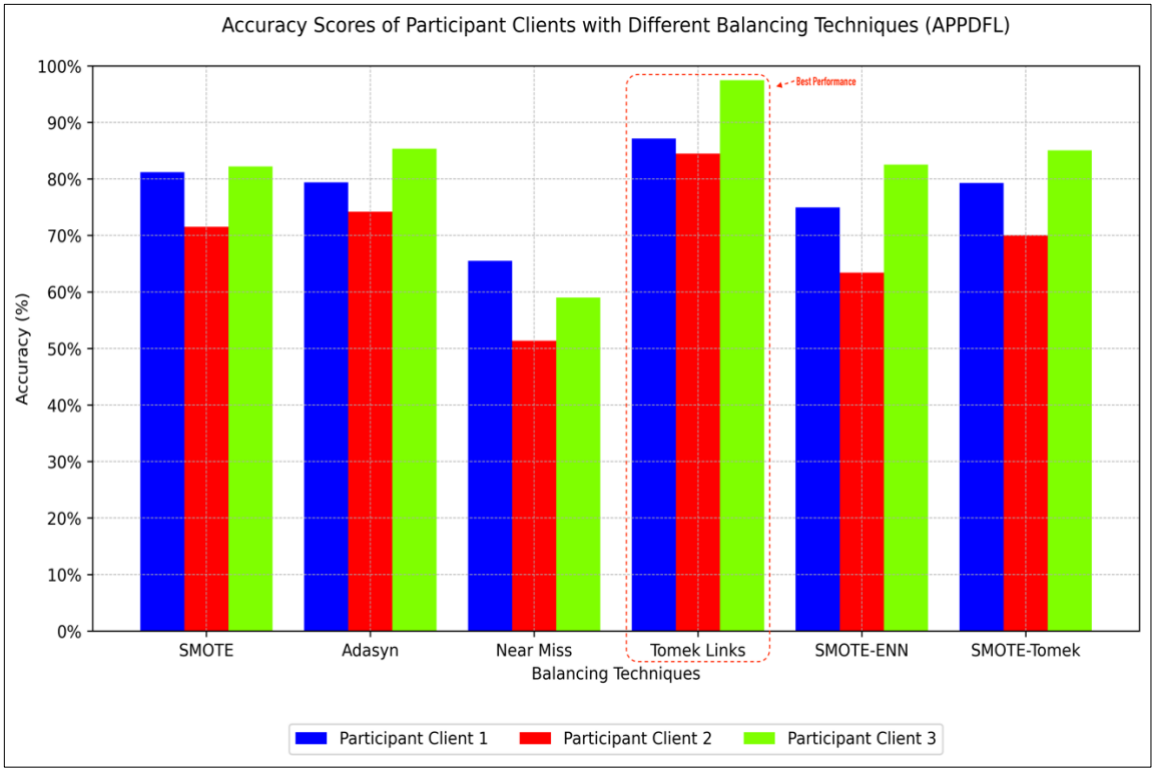


Figure 22. APPDFL Accuracy Comparison with Balancing Techniques.

The Adasyn technique also yielded a high accuracy of 0.7938 (79.38%), underscoring its effectiveness in improving the model's performance. SMOTE-Tomek achieved an accuracy of 0.7930 (79.3%), slightly lower than Adasyn but still commendable. Participating healthcare client 2 exhibited the highest accuracy with the Tomek Links technique at 0.8450 (84.50%), showcasing its strong performance in this client's data balancing. Adasyn followed with an accuracy of 0.7421 (74.21%), indicating its substantial efficacy. SMOTE and SMOTE-Tomek also performed well, with accuracy of 0.7152 (71.52%) and 0.7004 (70.04%), respectively. Near Miss showed the lowest accuracy of 0.5136 (51.36%), suggesting its lesser effectiveness for client 2.

For participating healthcare client 3, the Tomek Links technique achieved the highest accuracy of 0.9746 (97.46%), demonstrating its exceptional performance in

Techniques	Client 1	Client 2	Client 3	Total Average
SMOTE	0.7994	0.7221	0.8011	0.7562
Adasyn	0.7964	0.7221	0.8357	0.7777
Near Miss	0.5892	0.3651	0.5003	0.4849
Tomek Links	0.8411	0.8411	0.9823	0.8731
SMOTE-ENN	0.7250	0.5888	0.8062	0.7066
SMOTE-Tomek	0.7747	0.6769	0.8337	0.7618

Table 13. F1-Scores for APPDFL Models with Different Balancing Techniques

handling this client's data. Adasyn followed closely with an accuracy of 0.8535 (85.35%), indicating its robust effectiveness. SMOTE-Tomek and SMOTE also performed well, with accuracy of 0.8507 (85.07%) and 0.8222 (82.22%), respectively. SMOTE-ENN had an accuracy of 0.8253 (82.53%), suggesting these techniques were effective but not as much as Tomek Links. Near Miss had a lower accuracy of 0.5900 (59.00%), indicating it is the least effective for client 3 compared to others. For the detailed visual representation, refer to Figure 22 and Figure 23.

Based on our experiments findings, the F1 scores for the proposed Advanced Privacy-Preserving Decentralized Federated Learning (APPDFL) system using different data-balancing techniques are shown below. These findings demonstrate how various balancing strategies impact model performance across multiple clients. The analysis of the F1-Scores indicates that the Tomek Links technique achieved the highest total average F1-Score at 0.8731 (87.31%), highlighting its superior performance in handling imbalanced data

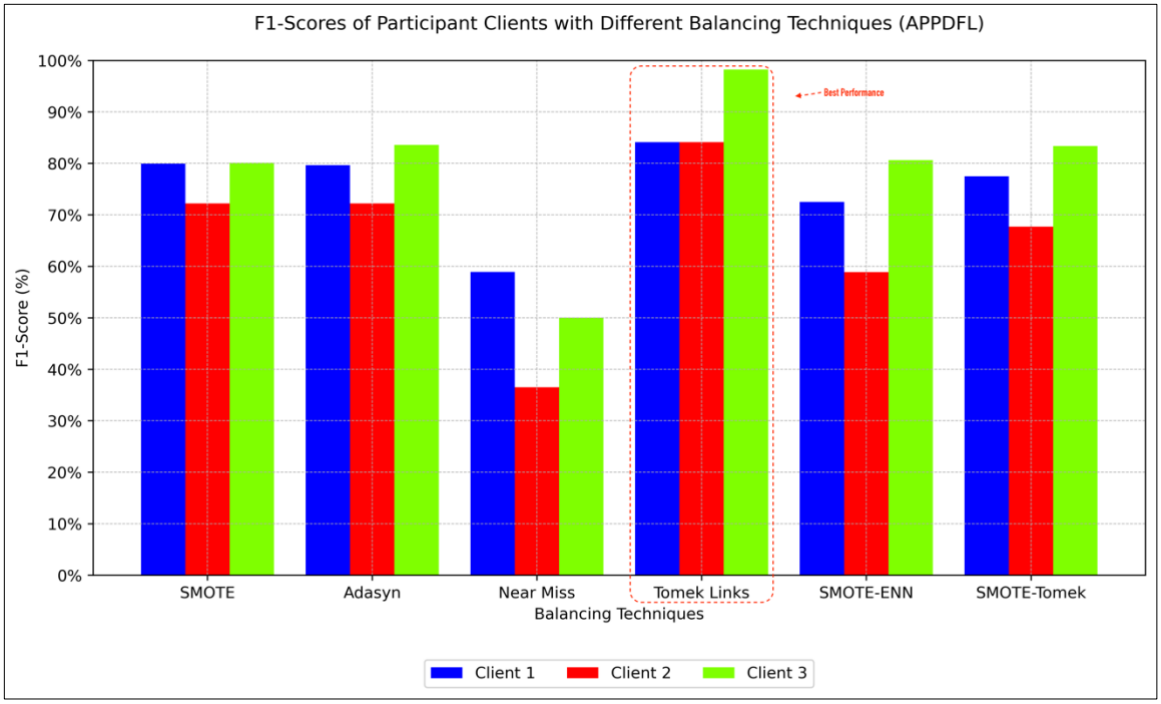


Figure 23. F1-Score Comparison Balancing Techniques in APPDFL.

within the context of APPDFL. Adasyn follows with an average F1-Score of 0.7777 (77.77%), showing strong effectiveness but not reaching the level of Tomek Links. SMOTE-Tomek also performed well, with an average F1-Score of 0.7618 (76.18%), indicating its capability to enhance model performance.

The SMOTE technique achieved a total average F1-Score of 0.7562 (75.62%), showcasing reasonable effectiveness but falling short compared to Tomek Links and Adasyn. The SMOTE-ENN technique had an average F1-Score of 0.7066 (70.66%), highlighting its ability to improve model accuracy, though not as effectively as the top techniques. Finally, Near Miss had the lowest total average F1-Score at 0.4849 (48.49%), suggesting its lesser effectiveness in enhancing the F1-Scores within this specific APPDFL context. For more details, refer to Table 13.

4.4.1 Effect of Noise Levels Budget on Accuracy of Advanced Privacy Preserving Decentralized Federated Learning (APPDFL) System

The client accuracy relative to privacy budgets reveals important trends and insights into the optimal balancing of privacy and performance. At a privacy budget of 0.4 epsilon (ε), Client 1 achieved an accuracy of 82.58%, Client 2 reached 72.41%, and Client 3 attained 88.05%, with an average accuracy of 81.01%. These results underscore the efficacy of a 0.4 epsilon privacy budget in balancing accuracy and privacy. The presence of a threshold at 0.75 highlights the significance of achieving high performance while maintaining privacy, indicating that clients' accuracy above this threshold, such as Client 1 and Client 3 at 0.4 epsilon, is particularly noteworthy. Similarly, the client F1 scores reveal significant findings at a privacy budget of 0.4 epsilon. Client 1 achieved an F1 score of 82.44%, Client 2 reached 71.29%, and Client 3 had 87.58%, with an average F1 score of 80.44%. This indicates that a 0.4 epsilon privacy budget optimally balances precision and recall, enhancing overall model performance. The threshold of 0.75 also serves as a benchmark for F1 scores, and both Client 1 and Client 3 surpass threshold at 0.4 epsilon, demonstrating the effectiveness of this privacy budget in maintaining high F1 scores.

4.4.2 The Optimal Privacy Budget for Balancing Accuracy and F1-Score in APPDFL System

The concept of the trade-off between accuracy and privacy is crucial in this analysis. As noted by Dwork & Roth (2014), "the lower epsilon values (indicating higher privacy) typically introduce more noise, potentially reducing accuracy". Interestingly, our results demonstrate that an epsilon value of 0.4, which considered to be in the lower range achieves an optimal balance, providing high accuracy and F1 scores while maintaining privacy standards. For more details, refer to Table 14 and Table 15.

Privacy Budget	Client 1: Accuracy	Client 2 : Accuracy	Client 3: Accuracy	Client Avg : Accuracy
0	79.30%	70.04%	85.08%	78.14%
0.1	79.02%	68.58%	82.65%	76.75%
0.2	78.96%	69.29%	84.70%	77.65%
0.3	77.54%	69.98%	86.65%	78.06%
0.4	82.58%	72.41%	88.05%	81.01%
0.5	80.25%	69.17%	84.67%	78.03%
0.6	79.92%	69.88%	80.78%	76.86%
0.7	78.67%	71.92%	87.59%	79.39%
0.8	75.74%	67.88%	83.55%	75.72%
0.9	80.59%	73.30%	86.48%	80.12%
1	81.59%	68.96%	81.21%	77.25%
1.1	78.27%	69.75%	85.30%	77.77%
1.2	81.40%	74.82%	86.34%	80.85%
1.3	78.31%	71.59%	85.21%	78.37%
1.4	78.84%	70.71%	84.04%	77.86%
1.5	81.10%	72.22%	85.44%	79.59%
1.6	80.35%	69.37%	77.93%	75.88%

Table 14. Accuracy with Different Privacy Budget Levels

This specific value of 0.4 epsilon, identified through our experiments, signifies an optimal privacy-preserving tolerance level that maximizes the performance metrics while ensuring data privacy. Clients' performances above the 0.75 threshold, particularly at the 0.4 epsilon level, validate this optimal balance. While Client 1 and Client 3 consistently

perform above the threshold of 0.75, Client 2 falls below it with an accuracy of 72.41% and an F1 score of 71.29% at a 0.4 epsilon privacy budget. Despite Client 2's scores being below the threshold, the overall average for both accuracy (81.01%) and F1 score (80.44%) remains above the threshold, highlighting the robustness of the APPDFL approach. The trade-off here demonstrates the necessity of calibrating privacy budgets to balance the needs of different clients while ensuring collective high performance.

The maximum acceptable accuracy observed is 88.05% for Client 3 at a 0.4 epsilon privacy budget, which is significantly above the 0.75 threshold, marking it as top-performing. Similarly, the top acceptable F1 score recorded is 87.58% for Client 3 at the same privacy budget, highlighting the robustness of the APPDFL approach in achieving high performance while preserving privacy. These top scores exemplify how the 0.4 epsilon privacy budget strikes an effective balance between maintaining accuracy and precision while introducing a tolerable level of noise for privacy protection. The analysis of clients' accuracy and F1 scores relative to privacy budgets provides critical insights into optimizing the balance between privacy and performance.

The threshold of 0.75 serves as a pivotal marker for acceptable performance levels, and the 0.4 epsilon privacy budget emerges as an optimal choice, maximizing both accuracy and F1 scores across different clients. These findings validate our Research Hypothesis 2 (RH2) and emphasize the significance of targeted privacy-budgeting strategies in enhancing the robustness and effectiveness of federated learning models in diverse settings. This performance is particularly noteworthy given the complexity and high class imbalance present in the CERT's insider threat dataset (version r4.2, r5.2, and r6.2; Dwork & Roth, 2014). For the detailed visual representation, refer to Figure 24.

Privacy Budget	Client 1: F1-Score	Client 2: F1-Score	Client 3: F1-Score	Client Avg: F1-Score
0	77.48%	67.69%	83.37%	76.18%
0.1	77.14%	65.65%	80.90%	74.56%
0.2	77.07%	66.70%	83.01%	75.59%
0.3	77.04%	68.32%	86.14%	77.17%
0.4	82.44%	71.29%	87.58%	80.44%
0.5	78.51%	66.13%	82.42%	75.69%
0.6	78.14%	66.86%	78.16%	74.39%
0.7	76.77%	69.44%	85.91%	77.37%
0.8	73.51%	65.22%	81.59%	73.44%
0.9	78.82%	71.19%	84.78%	78.26%
1	79.89%	66.69%	79.07%	75.21%
1.1	76.83%	67.63%	83.73%	76.06%
1.2	79.71%	72.89%	84.64%	79.08%
1.3	76.44%	69.83%	83.50%	76.59%
1.4	77.05%	68.16%	82.26%	75.82%
1.5	79.39%	70.43%	83.74%	77.86%
1.6	78.59%	66.26%	75.02%	73.29%

Table 15. F1 Score with Different Privacy Budget Level

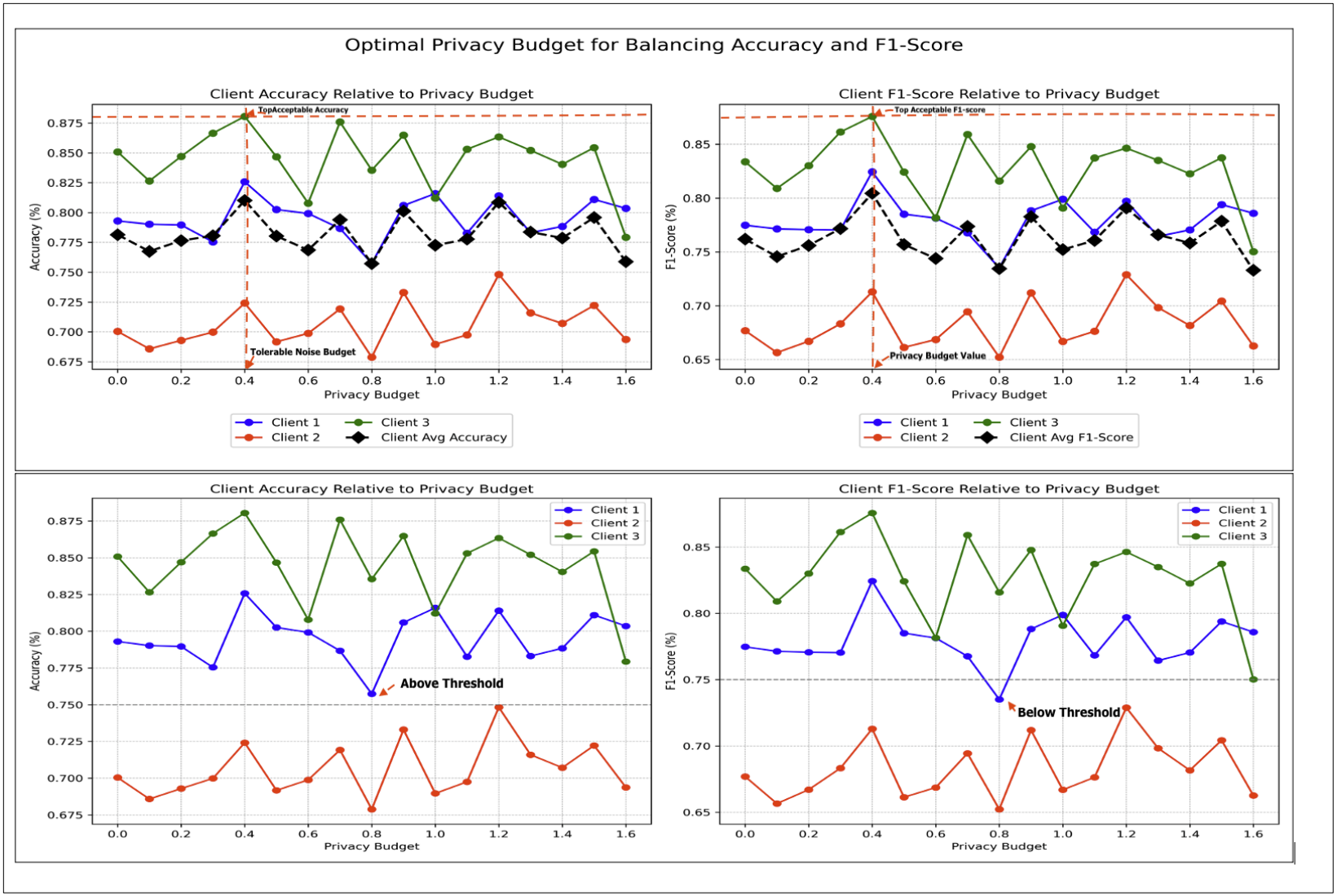


Figure 24. Optimal Privacy-Accuracy Trade-Off in APPDFL.

4.5 Comparative Analysis of TML, CFL and APPDFL testing evaluations

This section presents a comprehensive comparison of Traditional Machine Learning (TML), Centralized Federated Learning (CFL), and Advanced Differential Privacy-Preserving Decentralized Federated Learning (APPDFL) to evaluate their effectiveness in detecting insider threats in healthcare environments. Each model's performance is evaluated using key performance measures including accuracy, F1 scores, and privacy preservation, with an importance on achieving the optimal balance between privacy protection and detection accuracy. The influence of noise introduced for privacy protection is thoroughly examined in advanced APPDFL, illustrating the trade-offs between guaranteeing privacy and preserving robust detection performance.

4.5.1 Traditional Machine Learning (TML)

Traditional Machine Learning (TML) demonstrates impressive capabilities with 98% accuracy and a weighted average F1 score of 0.98, benefiting from the centralized aggregation of large datasets across multiple institutions. This method improves the model's capacity to identify insider threats by employing a large and varied dataset, leading to high recall and precision, especially in situations when a large amount of data is available for training.

However, this centralized approach has serious drawbacks, as it creates a single point of failure for unauthorized access and raises the possibility of privacy violations by storing all sensitive data in one place.. Although TML is most of the time get high performance, it is not ideally recommended in sensitive sector due to privacy-risks that can be introduced in sectors such as the healthcare industry, where maintaining confidentiality is essential, because to its dependency on central data storage. Its

applicability is limited in environments that value data secrecy due to the trade-off between high performance and privacy issues.

4.5.2 Centralized Federated Learning evaluation

Centralized Federated Learning (CFL) improves model performance with data balancing techniques such as Tomek Links and SMOTE-ENN, achieving high accuracy scores of 0.8965 for Clients 1 and 3, and 0.7398 for Client 2, with an overall accuracy of 0.8442. These methods also improve F1 scores; in particular, CFL produces good detection performance by addressing the problem of class imbalance.

Although CFL maintains decentralized raw data, which enhances privacy compared to TML, it still centralizes model updates, leaving it vulnerable to inference attacks. There are privacy challenges associated with this, particularly in hospital contexts where maintaining data secrecy is crucial. The main advantage of CFL is that it can increase model performance; nevertheless, the risk to privacy associated with centralized updates is a drawback.

4.5.3 Proposed Advanced Privacy Preserving Decentralized Federated Learning approach evaluation

Advanced Privacy Preserving Decentralized Federated Learning with Differential Privacy (APPDFL) facilitates direct client-to-client communication without requiring a central server, ensuring privacy preservation. Compared to centralized models to the traditional machine learning (TML) and Centralized Federated Learning (CFL), this decentralized technique greatly reduces privacy issues by guaranteeing that raw data stays inside the local context. Controlled noise incorporated into the model updates by using differential privacy (DP) technique, which improves privacy protection but may have an

effect on the accuracy of the model. However, the adoption of data balancing techniques like SMOTE, Tomek Links, and SMOTE-ENN efficiently mitigates this impact.

For instance, advanced APPDFL with Tomek Links demonstrates strong performance while upholding strict privacy norms, with an accuracy of 90% and an F1-Score of 0.8733. These findings demonstrate how well data balancing strategies work when combined with DP to maintain accuracy levels that are competitive while improving privacy. Advanced APPDFL eliminates the need for a central server, enabling clients to communicate model changes directly with one another, in contrast to CFL, which centralizes model updates. By introducing noise to the model updates, peer-to-peer communication and distributed persistence guarantee that raw data never leaves the local environment, preserving individual data points even more. Although differential privacy (DP) can add noise that degrades accuracy, techniques such as Tomek Links contribute to good performance, F1-Scores and great accuracy are routinely achieved using Advanced Privacy Preserving Decentralized Federated Learning with Differential Privacy (APPDFL).

Based on results evaluated from experiments, advanced APPDFL is superior to traditional machine learning and centralized federated learning. It is the best option for privacy-sensitive situations like healthcare since it not only avoids the threats related to centralizing data or updating models, but it also maintains high detection accuracy. Furthermore, APPDFL offers a scalable solution for collaborative environments, enabling multiple institutions to benefit from collective insights without compromising the confidentiality of their data. This approach ensures that each participant maintains control over their local datasets while contributing to an improved and resilient global model.

The implementation of APPDFL with differential privacy and effective data balancing techniques like Tomek Links and SMOTE-ENN not only addresses privacy concerns but also improve the overall performance of the learning model. This dual benefit of privacy preservation and high accuracy makes APPDFL a compelling choice for environments where data sensitivity is paramount. The advanced capabilities of APPDFL highlight its potential to revolutionize privacy-preserving machine learning practices in various domains, particularly in sectors requiring stringent data confidentiality, such as healthcare.

4.6 Analysis of APPDFL Using Optimized DNN and Optimized LSTM

This section compares the efficiency of optimized Deep Neural Networks (DNN) with optimized Long Short-Term Memory (LSTM) networks for advanced decentralized privacy-preserving federated learning (APPDFL). As discussed Chapter 3, we assessed how well each optimized model detects insider threats by examining the mean accuracy and mean F1 scores among various clients. This comparison is especially relevant to Hypothesis 3 (RH 3), which states that optimized DNN's ability to handle complicated data patterns would often result in better classification performance when integrated with advance APPDFL than optimized LSTM.

A comprehensive comparison of APPDFL using optimized DNN and LSTM models is presented in Figure 25, as well as in Tables 16 and 17. The findings demonstrate that, in comparison to advanced DFL using optimized LSTM, advanced DFL using optimized DNN consistently obtains greater mean accuracy and F1 scores across a variety of customers. These performance metrics imply that optimized DNN, even with the additional noise for privacy protection, is more adaptive at handling the behavior of trained data for

		Accuracy	F1-Score
DNN	**APPDFL**	90.12 %	87.89%
	CFL	84.92%	85.33%
LSTM	**APPDFL**	76.96%	75.02%
	CFL	68.96%	71.49%

Table 16. APPDFL vs. CFL: DNN and LSTM Results

decentralized federated learning with differential privacy, making it a suitable architecture fit for insider threat detection in collaborative healthcare institutions.

By verifying that advanced DFL using optimized DNN provides better detection accuracy and overall performance than advanced DFL using optimized LSTM, this investigation validates Hypothesis 3. This comparison is improved by the use of optimized models, which demonstrate that optimized DNN is a more dependable solution for usage in settings like healthcare where data privacy and detection accuracy must be maintained.

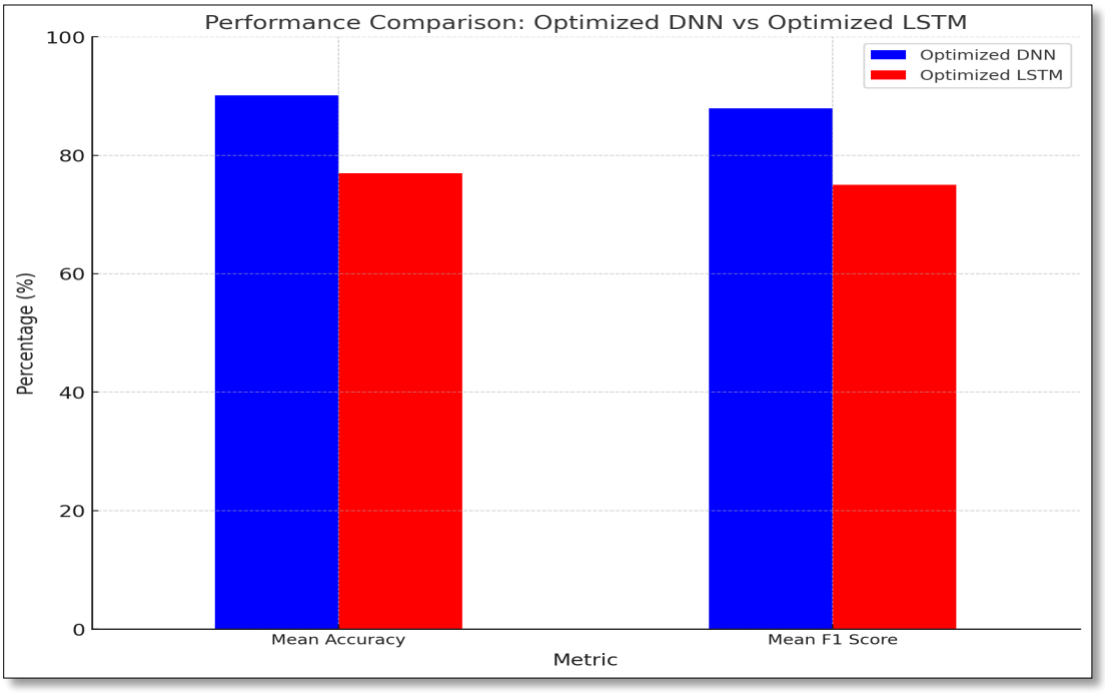


Figure 25. APPDFL-DNN vs. APPDFL-LSTM Performance.

Metrics	APPDFL-DNN	APPDFL-LSTM
Mean Accuracy	**90.12%**	**76.96%**
Client 1 Accuracy	87.21%	79.87%
Client 2 Accuracy	85.93%	72.32%
Client 3 Accuracy	97.22%	78.68%
Mean F1 Score	**87.89%**	**75.02%**
Client 1 F1 Score	84.41%	78.19%
Client 2 F1 Score	81.14%	70.31%
Client 3 F1 Score	98.12%	76.57%

Table 17. APPDFL-DNN vs. APPDFL-LSTM Clients Performance

Optimized APPDFL-DNN performs better than APPDFL-LSTM in identifying insider threats, according to the study of mean accuracy and mean F1 scores. In particular, APPDFL-LSTM generated a mean accuracy of 76.96% and a mean F1 score of 75.02%, whereas APPDFL-DNN attained a mean accuracy of 90.12% and a mean F1 score of 87.89% of data. DFL-DNN clearly outperforms other methods when comparing the performance of individual clients. Regarding the first client, APPDFL-DNN achieved an accuracy of 87.21% and an F1 score of 84.41%, whereas APPDFL-LSTM achieved 79.87% accuracy and a 78.19% F1 score. With an accuracy of 85.93% and an F1 score of 81.14% for Client 2, APPDFL-DNN outperformed APPDFL-LSTM, which came in last

with an accuracy of 72.32% and an F1 score of 70.31%. With regard to Client 3, there is the biggest disparity between APPDFL-DNN's accuracy of 97.22% and F1 score of 98.12% and APPDFL-LSTM's lower accuracy of 78.68% and F1 score of 76.57%.

These findings demonstrate the endurance of optimized APPDFL-DNN in handling intricate patterns in decentralized data, leading to an enhanced ability to identify insider threats. APPDFL-LSTM does not always outperform APPDFL-DNN in relation to accuracy and F1 scores, despite its reasonable performance. This application finds that DNN is more appropriate in the setting of Decentralized Federated Learning (APPDFL) when optimized. APPDFL-DNN is a better option than APPDFL-LSTM because it offers stronger data privacy protection through differential privacy methods and a more accurate and dependable model for insider threat detection in a decentralized learning environment.

4.7 Hypothesis Evaluation and Validation

Through the analysis of performance metrics and data derived from the models employed in this study, this part assesses and validates the offered hypotheses. The effectiveness of various strategies, including data balancing tactics, privacy-preserving strategies, and model designs, in detecting insider threats is evaluated by looking at the experiment outcomes and evaluating each hypothesis.

The assessment sheds light on the ways that various approaches—including Traditional Machine Learning (TML), Centralized Federated Learning (CFL) and Advanced Decentralized Federated Learning (APPDFL)—affect the models' capacity to detect insider threats. Analysis of these techniques' effects on data privacy and detection performance is provided. In healthcare settings, where patient privacy is extremely important, this is especially crucial. The findings support the significance of the suggested

strategies by demonstrating that federated learning architectures and balancing techniques are important in improving the security, precision, and privacy of machine learning models in healthcare environments.

- **Research Question 1 (RQ1):** How does the proposed Advanced Privacy-Preserving Decentralized Federated Learning System perform in terms of detecting insider threats, compared to centralized federated learning?
- **Research Hypothesis 1 (RH1):** The proposed Advanced Privacy-Preserving Decentralized Federated Learning System will detect insider threats with greater accuracy than centralized federated learning.
- **Research Question 2 (RQ2):** In the proposed Advanced Privacy-Preserving Decentralized Federated Learning System, what is the optimal trade-off point value to ensure sufficient protection and acceptable accuracy insider threat detection across collaborative healthcare institutions?
- **Research Hypothesis 2 (RH2):** The analysis of the proposed Advanced Privacy-Preserving Decentralized Federated Learning System will identify an optimal trade-off point value that ensures sufficient privacy protection while maintaining acceptable accuracy in insider threat detection across collaborative healthcare institutions.
- **Research Question 3 (RQ3):** Which model architecture, Deep Neural Network (DNN) or Long Short-Term Memory (LSTM), yields higher classification performance for insider threat detection when integrated with selected balancing techniques (SMOTE, ADASYN, Near Miss, SMOTE-ENN, Tomek Links, or

SMOTE-Tomek Links) within the Advanced Privacy-Preserving Decentralized Federated Learning System?

- **Research Hypothesis 3 (RH3):** The optimized Deep Neural Network (DNN) model architecture, integrated with the Tomek Links balancing technique in the proposed Advanced Privacy-Preserving Decentralized Federated Learning System, yields higher classification performance for insider threat detection than the optimized Long Short-Term Memory (LSTM) model architecture.

4.7.1 Hypothesis 1 (RH1) Validation

The first hypothesis (H1) suggests that by enabling effective information sharing across several healthcare organizations, the Advanced Decentralized Federated Learning (APPDFL) model can raise the accuracy of insider threat detection. The results presented in this Chapter 4, show that advanced APPDFL improves detection performance across all clients when integrated with Tomek Links for data balancing, provide strong support for this idea.

APPDFL enables each institution to maintain its local data while adding to a global model, increasing the effectiveness of the learning process. By leveraging a wider range of information, this collaborative method overcomes the drawbacks of isolated data and improves the model's capacity to identify insider threats.

The use of Tomek Links, as elaborated in Chapter 2, enhances the model's efficacy by eliminating superfluous instances, guaranteeing more distinct differentiations across classes. Because of the class imbalance, it might be difficult to discern between benign and malicious behaviors in insider threat detection, which is where this is very helpful. Tomek Links produces a more balanced dataset by removing unnecessary or

unclear data points, which enables the model to concentrate on important trends. Consequently, As demonstrated by the results in Table 16 and Figure 25, the proposed APPDFL with Tomek Links shows significant performance, achieving an accuracy of 90.12% and an F1-score of 87.833%. This approach not only outperforms the centralized federated learning method tested under the same dataset and configuration but also delivers results comparable to centralized methods. Additionally, it ensures the privacy of each healthcare participant through a peer-to-peer architecture that enhances scalability and fosters collaborative threat intelligence sharing.

The scalability and security improvements from the centralized learning method to a decentralized and secure approach with serverless architecture and data balancing are the main emphases of this proposal as a detection solution for insider threats in collaborative healthcare institutions. These enhancements address common limitations like single points of failure and data privacy concerns as we highlighted throughout this Praxis, fostering a more resilient and efficient system.

Approach	Model Used	CMU CERT Dataset(s)	Machine Learning Approach			Performance Metrics		Collaborative Approach
			TML	CFL	DFL	Accuracy	Privacy Preserving	
TML Approach Tested	DNN	r4.2; r5.2; r6.2	✓	✗	✗	98%	✗	✗
CFL Approach Tested	DNN	r4.2; r5.2; r6.2	✗	✓	✗	84%	◑	✓
Proposed APPDFL Model	Optimized DNN	r4.2; r5.2; r6.2	✓	✓	✓	90.1%	✓	✓

Legend: ✓*Fully addressed ;* ◑ *Partially addressed ;* ✗ *Not addressed by the work.*

Table 18. APPDFL vs Tested Methods

Papers	Citation	Model Used	CMU CERT Dataset(s)	Machine Learning Approach			Performance Metrics		Collaborative Approach
				TML	CFL	DFL	Accuracy	Privacy Preserving	
Securing the Internet of Health Things: Embedded Federated Learning-Driven Long Short-Term Memory for Cyberattack Detection	(Kumar and Kim, 2024)	LSTM	✗	✗	✓	✗	97.16%	◑	✓
FedAT: Federated Adversarial Training for Distributed Insider Threat Detection	(Gayathri et al., 2024)	SNN-MLP	r4.2; r5.2	✗	✓	✗	🔍	◑	✓
A Federated leaning and explainable approach for insider threat detection in IoT	(Amir-Zarandi et al., 2023)	DE	r6.2	✗	✓	✗	🔍	◑	✓
CFL Approach Tested		DNN	r4.2; r5.2; r6.2	✗	✓	✗	84%	◑	✓
Proposed APPDFL Model		**Optimized DNN**	**r4.2; r5.2; r6.2**	✓	✓	✓	**90.1%**	✓	✓

Legend: ✓ *Fully addressed ;* ◑ *Partially addressed ;* ✗ *Not addressed by the work;* 🔍 *Focused on Different Metric.*

Table 19. APPDFL vs. Centralized Federated Learning Methods include tested CFL

Papers	Citation	Model Used	CMU CERT Dataset(s)	Machine Learning Approach			Performance Metrics		Collaborative Approach
				TML	CFL	DFL	Accuracy	Privacy Preserving	
User Behavior Analytics for Anomaly Detection Using LSTM Autoencoder - Insider Threat Detection	(Sharma et al., 2020)	AE	r4.2	✓	✗	✗	90.17%	✗	✗
Insider Detection using Deep Autoencoder and Variational Autoencoder Neural Networks	(Pantelidis et al., 2021)	AE VAE	r4.2	✓	✗	✗	96% 95%	✗	✗
Detecting insider threat within institutions using CERT dataset and different ML techniques	(Dosh (2021)	1 NN	r4.1; r4.2	✓	✗	✗	94.6%	✗	✗
Insider Threat Detection Model Using Anomaly-Based Isolation Forest Algorithm	(Al-Shehari et al., 2023)	IF	r4.2	✓	✗	✗	98%	✗	✗
Mitigating Insider Threat: A Neural Network Approach for Enhanced Security	(Lavanya et al. ,2024)	DNN	r6.1; r6.2	✓	✗	✗	98.8%	✗	✗
Insider Threat Detection Using Machine Learning Approach	(Bin Sarhan and Altwaijry, 2023)	SVM OCSVM iForest	r4.2	✓	✗	✗	100% 91%	✗	✗
Insider threat detection using supervised machine learning algorithms	(Manoharan et al., 2023)	RF	r4.2	✓	✗	✗	95.9%	✗	✗
Securing the Internet of Health Things: Embedded Federated Learning-Driven Long Short-Term Memory for Cyberattack Detection	(Kumar and Kim, 2024)	LSTM	✗	✗	✓	✗	97.16%	◑	✓
FedAT: Federated Adversarial Training for Distributed Insider Threat Detection	(Gayathri et al., 2024)	SNN-MLP	r4.2; r5.2	✗	✓	✗	🔍	◑	✓
A Federated leaning and explainable approach for insider threat detection in IoT	(Amir-Zarandi et al., 2023)	DE	r6.2	✗	✓	✗	🔍	◑	✓
TML Approach Tested		DNN	r4.2; r5.2; r6.2	✓	✗	✗	98%	✗	✗
CFL Approach Tested		DNN	r4.2; r5.2; r6.2	✗	✓	✗	84%	◑	✓
Proposed APPDFL Model		Optimized DNN	r4.2; r5.2; r6.2	✓	✓	✓	90.1%	✓	✓

Legend: ✓ *Fully addressed ;* ◑ *Partially addressed ;* ✗ *Not addressed by the work;* 🔍 *Focused on Different Metric.*

Table 20. Results and Strategies: APPDFL vs. Tested Methods and Previous Works

Hypothesis	Statement	Validation	Answer Reasons
Hypothesis 1 (RH1)	The proposed Advanced Privacy-Preserving Decentralized Federated Learning system will detect insider threats with greater accuracy than centralized federated learning.	Proven	The proposed Advanced Privacy-Preserving Decentralized Federated Learning (APPDFL) system achieved an outstanding accuracy of 90.12% and an F1-score of 87.33% (see Table 16 and Figure 19), surpassing the performance of the tested centralized federated learning methods, which achieved an accuracy of 84.92% and an F1-score of 85.33%. Additionally, our serverless solution addresses the common problem faced by previous centralized methods, which rely on servers (see, Table 18, Table 19, Table 20). Through innovative strategies, our approach advances insider threat detection in a secure and collaborative manner.

Table 21. Hypothesis 1 (RH1) Evaluation Summary

4.7.2 Hypothesis 2 (RH2) Validation

Hypothesis 2 states that an optimal trade-off point between adequate privacy protection and acceptable insider threat detection accuracy will be found through analysis of the proposed Advanced Decentralized Federated Learning (APPDFL) approach among collaborating healthcare institutions. The concept is supported by the study's findings, which show that APPDFL can maintain sufficient privacy protection while maintaining acceptable accuracy in insider threat detection across collaborative healthcare institutions

To verify the impact on model performance, we assessed the APPDFL's performance at various noise levels, from 0 ε (no noise) to 1.6 ε in sequential increments of 0.1 ε. This increment size is chosen for several reasons: to provide a detailed and nuanced analysis (granularity) (Dagan & Kur, 2022); to strike a balance between having enough data points and keeping the number of experiments manageable (balance) (Yin et al., 2022); to align with existing literature (precedent) (Dwork et al., 2006a, 2006b); and to identify specific thresholds where performance metrics significantly change (thresholds) (Feretzakis et al, 2024). APPDFL managed to strike a compromise between privacy and accuracy at an acceptable noise level budget of 0.4 ε, preserving 81.01% accuracy and an F1-score of 80.44% while still offering robust privacy safeguards for healthcare institutions participants data.

Finding the best balance between data privacy and the accuracy of insider threat detection is especially crucial in healthcare settings. The APPDFL approach validates the hypothesis by pointing out this trade-off point and showing that strong privacy criteria is achieved with optimal model performance.

Hypothesis	Statement	Validation	Answer Reasons
Hypothesis 2 (RH2)	The analysis of the proposed Advanced Privacy-Preserving Decentralized Federated Learning approach will identify an optimal trade-off point value that ensures sufficient privacy protection while maintaining acceptable accuracy in insider threat detection across collaborative healthcare institutions.	Proven	APPDFL as a proposed solution identifies an optimal trade-off between privacy and accuracy at noise budget epsilon of 0.4 (0.4 ε), achieving 81.01% accuracy and 80.44% F1-Score while maintaining privacy protections through differential privacy techniques. Refer to Tables 14 and Table 15, as well as Figure 24, for detailed insights.

Table 22. Hypothesis 2 (RH2) Evaluation Summary

4.7.3 Hypothesis 3 (RH3) Validation

Research Question 3 (RQ3) examines that in the Advanced Privacy-Preserving Decentralized Federated Learning (APPDFL) system whether—optimized Deep Neural Network (DNN) or optimized Long Short-Term Memory (LSTM)—yields greater classification performance for insider threat detection when combined with Tomek Links. According to the Research Hypothesis 3 (RH3), in this architecture, the optimized Deep

Neural Network (DNN) architecture with the integration of Tomek Links will produce better classification performance than optimized Long Short-Term Memory (LSTM).

This hypothesis is substantially supported by the results illustrated in Figure 25 and Tables 16 and 17. When integrated with Tomek Links for data balancing, the optimized DNN consistently outperformed the optimized LSTM in terms of mean accuracy and mean F1 scores. The implementation results from both tests indicate that the optimized DNN achieved a mean accuracy of 90.12% and a mean F1-score of 87.89%. In contrast, the optimized LSTM recorded a mean accuracy of 76.96% and a mean F1-score of 75.02%. These findings underscore the enhanced performance of the optimized DNN, validating the proposed approach's effectiveness in insider threat detection within collaborative healthcare institutions.

In this particular context, optimizing DNN architecture did exceptional than LSTM . Based on both architecture advantages , our theory and prediction are supported by the findings as we predicted that the optimized DNN may identify more intricate patterns in the data by carefully choosing hyperparameters such as the number of layers, neurons, and activation functions. To further improve the model, balancing techniques played a crucial role in these experiments . for example, among six different balancing techniques we tested Tomek Link outperform all of them. Form the theory triangulation prospective, the integration of helps to cleans the data, eliminating overlapping occurrences that might mislead the classifier and produce more definite class differences, ultimately enables our process not to limit our imagination on one single tool or solution lather test many and use the best among them to improve our qualitative research using theory triangulation mindset.

From the experiment stand point, the superior classification performance for optimized DNN is achieved by the combination of an efficient data-balancing approach and an optimized architecture. However, while optimized LSTM networks perform well in handling sequential and time-dependent data, they are less effective in instances involving static data, such as identifying insider threats in the healthcare industry. From a theoretical viewpoint, the optimized DNN should better identify intricate, non-sequential patterns—a crucial skill for precisely differentiating between harmful and benign activities. CERT's insider threat dataset includes both sequential and non-sequential data, which further emphasizes this need. The benefit of designing the optimized Deep Neural Network (DNN) model architecture with a dropout rate of 50% at each stage proves that architecture design plays an important role in the success of engineering solutions. Specifically in healthcare, this application is validated by the design introduced in our methodology, as illustrated in Figures 15 and 16 in Chapter 3, and their performance, as shown in Table 16, Table 17, and Table 23.

The results therefore support Research Hypothesis 3 (RH3), demonstrating that the optimized DNN architecture integrated with Tomek Links can achieve superior insider threat detection classification performance compared to the optimized Long Short-Term Memory (LSTM) architecture within the Advanced Privacy-Preserving Decentralized Federated Learning (APPDFL) system.

	Optimized DNN	Optimized LSTM
Mean Accuracy	**90.12%**	76.96%
Mean F1 Score	**87.89%**	75.02%

Table 23. Mean Performance comparison- Optimized-DNN and optimized- LSTM

Hypothesis	Statement	Validation	Answer Reasons
Hypothesis 3 (RH3)	The optimized Deep Neural Network (DNN) Model Architecture, integrated with the Tomek Links balancing technique in the proposed Advanced Privacy-Preserving Decentralized Federated Learning approach, yields higher classification performance for insider threat detection than the optimized Long Short-Term Memory (LSTM) model architecture.	Proven	The optimized DNN integrated with Tomek Links outperforms the optimized LSTM, achieving a mean accuracy of 90.12% and a mean F1 score of 87.89%. This is due to the optimized DNN's ability to model complex patterns in static data, enhanced by the Tomek Links technique for class distinction.

Table 24. Hypothesis 3 (RH3) Evaluation Summary

Chapter 5—Discussion and Conclusions

5.1 Discussion

Insider threats are a growing concern across various sectors, with healthcare institutions being no exception. As the sophistication and frequency of these threats increase, healthcare institutions face the enormous challenge of protecting themselves from these internal network threats as their sophistication and frequency increase. A report from the Ponemon Institute reveals that insider threat incidents increased by 44% between 2020 and 2022, with costs per incident rising to as much as USD $15.38 million (Ponemon Institute, 2022).

The increasing frequency of insider threats within healthcare institutions presents critical and growing concerns over safeguarding sensitive patient data, especially in light of the rising number of security incidents in the healthcare sector. Healthcare companies are increasingly falling victim to insider threats and sophisticated hacking efforts despite the highly confidential nature of patient data. In fact, patient data breaches doubled in 2023, affecting 87 million individuals (Transforming Data With Intelligence, 2023).

Traditional centralized machine learning methods not only fall short in balancing high accuracy with robust privacy guarantees but also fail to incorporate the necessary collaborative approach for insider threat intelligence sharing among healthcare institution participants. This lack of collaborative engagement significantly impedes the effectiveness of detection efforts, leaving healthcare institutions vulnerable to insider threat breaches and sophisticated hacking attacks.

Due to the evolving privacy demands in healthcare, driven by the critical importance of patient confidentiality, it has become imperative to develop advanced privacy-preserving solutions. Such solutions are essential for strengthening insider threat detection, maintaining operational efficiency, and promoting the adoption of insider threat intelligence sharing among healthcare institution participants to foster collaboration. These solutions also enable institutions to adhere to increasingly complex regulatory requirements while safeguarding sensitive patient information.

Motivated by these urgent needs, and recognizing that healthcare is vital to human well-being, this Praxis proposed and implemented the Advanced Privacy-Preserving Decentralized Federated Learning (APPDFL) approach. We designed this solution to address insider threat detection in collaborative healthcare institutions by balancing the trade-offs between privacy preservation and classification accuracy. By using secure by design as a security architecture principal we designed a horizontal server-less peer-to-peer (P2P) decentralized federated learning communication process where each healthcare institutions participants' node is protected to allow join and share intelligence through sharing model updates instead of raw data, our system method fosters collaboration between healthcare institutions without compromising sensitive patient information.

The main objectives is to advance the detection approach from centralized method to server less or server free privacy-preserving decentralized federated learning to improve classification accuracy by integrating optimized architectures with data balancing techniques while ensuring acceptable privacy preservation.

5.2 Conclusions

As detailed throughout this Praxis, the novelty of our approach is threefold. First, this Praxis is the first to implement a collaborative insider threat detection system specifically designed for collaborative healthcare institutions. This approach enables advanced privacy-preserving decentralized federated learning for insider threat detection among participating healthcare institutions, fostering the sharing of insider threat intelligence without compromising sensitive patient data. By addressing the challenges inherent with data sharing and data security in healthcare, our method enhances insider threat detection in collaborative healthcare institutions, a highly regulated sector. Additionally, our method addresses privacy concerns, security risks, and regulatory compliance related to the Health Insurance Portability and Accountability Act (HIPAA) and the General Data Protection Regulation (GDPR), among other regulatory requirements.

Second, this Praxis introduces the first integration of different data balancing techniques to enhance classification performance in insider threat detection in collaborative healthcare institutions. The optimized data balancing techniques address the inherent class imbalance in insider threat datasets, resulting in improved detection accuracy for both insider and non-insider threats across participating healthcare institutions.

Third, this Praxis uniquely incorporates all three most commonly and widely used versions of the CERT insider threat datasets: r4.2, r5.2, and r6.2, which are popular datasets used for evaluating insider threat detection. The comprehensive use of these datasets provides a more realistic simulation of insider threat scenarios, offering a broader and more reliable assessment of real-world insider threat detection in collaborative healthcare institutions.

These contributions form the foundation of the Advanced Privacy-Preserving Decentralized Federated Learning (APPDFL) approach. The proposed system demonstrates superior ability to improve classification accuracy while maintaining acceptable levels of privacy preservation in collaborative healthcare institutions. By enabling collaboration without risking data breaches, This approach represents a significant and transformative advancement in insider threat detection for healthcare institutions. This system, implemented following the principal of secure-by-design and zero trust architecture (Munuo, 2023; National Institute of Standards and Technology, 2020), guarantees adherence to the fundamental principles of information security, such as confidentiality, integrity, and availability (Cybersecurity and Infrastructure Security Agency, n.d.), thereby strengthening overall system robustness and compliance with regulatory standards.

5.3 Contributions to Body of Knowledge

This Praxis is, to our knowledge, the first study effort at the intersection of several key domains to implement a collaborative approach for insider threat detection specifically designed for healthcare institutions using an advanced privacy-preserving federated learning solution. These key domains include:

- Decentralized Federated Learning
- Differential Privacy Mechanism
- Insider Threat Detection
- Collaborative Healthcare Institutions
- Server-less Technologies

The research made significant contributions to the fields of insider threat detection and federated learning to facilitate the adoption of a collaborative approach for insider threat intelligence sharing among healthcare institution participants. By focusing on collaborative healthcare institutions, this study advances understanding in these intersecting domains by demonstrating how a decentralized federated learning approach with differential privacy mechanisms can enhance collaborative insider threat intelligence sharing.

This addresses challenges that centralized federated learning tested and existing work on traditional centralized machine learning approaches face while facilitating insider threat intelligence sharing and collaborative adoption, all while maintaining data privacy and security in compliance with regulations like GDPR and HIPAA. By addressing these interconnected domains, this Praxis advances the understanding of how decentralized federated learning can facilitate secure insider threat intelligence sharing, overcoming the limitations of traditional centralized machine learning and federated learning approaches.

Our proposed Advanced Privacy-Preserving Decentralized Federated Learning (APPDFL) system, as a collaborative solution, highlights its superiority over existing centralized machine learning and centralized federated learning approaches tested in this study. It provides a secure, efficient solution for insider threat detection while mitigating key data privacy and security challenges related to insider attacks in healthcare institutions. This research's primary contributions are summarized below:

1. **A collaborative Approach to Insider Threat Detection**: This Praxis is the first to implement a collaborative insider threat detection approach specifically designed for healthcare institutions. It leverages an advanced privacy-preserving decentralized federated learning solution to facilitate

secure insider threat intelligence sharing among healthcare institution participants as a tactical actions and collaborative strategic implication as a solution for IA without compromising sensitive patient data. This approach addresses the limitations of centralized approaches and demonstrates its superiority over existing centralized machine learning and centralized federated learning approaches tested in this study.

2. **Identifying the Optimal Architecture and Data Balancing Techniques for Insider Threat Detection:** The study introduces and tested an optimized deep neural network (DNN) architecture integrated with the Tomek links balancing technique within the proposed APPDFL approach. This innovation significantly improves the classification performance for insider threat detection in healthcare institutions, outperforming the optimized Long Short-Term Memory (LSTM) architecture tested in this study. The combination of data balancing and optimized architecture effectively addresses the issue of class imbalance in insider threat datasets, leading to higher accuracy in detecting both insider threat and non-insider threat behaviors.
3. **Comprehensive Use of CERT Insider Threat Datasets** : This Praxis is the first to incorporate all three most commonly used versions of the CERT's insider threat datasets (R4.2, R5.2, and R6.2). These datasets provide a thorough evaluation of insider threat detection in collaborative healthcare institutions, allowing for a broader and more realistic simulation of insider threat scenarios not explored in previous studies. As highlighted

by National Insider Threat Task Force (2016), in their mission of protecting organization from the insider out, they acknowledge CERT's insider threats dataset as "a well-recognized source for insider threat data". The use of these datasets ensures a thorough assessment of privacy preservation and classification accuracy across diverse insider threat behaviors.

4. **Advancing the Adoption of Collaborative Intelligence Sharing in Regulated Sectors**: The research demonstrates that a decentralized federated learning approach with privacy-preserving mechanisms can significantly enhance collaborative intelligence sharing in regulated sectors, particularly in healthcare. This contribution paves the way for future researchers to explore the integration of federated learning, privacy preservation, and collaborative threat detection beyond healthcare, addressing critical challenges in other regulated sectors such as finance and government.

Overall, this Praxis introduces a novel approach that specifically addresses the unique challenges faced by collaborative healthcare institutions. Building on the work of previous researchers while advancing the field, our methodology and findings have also contributed to the field by paving the way for others to follow in similar footsteps. Future researchers interested in conducting research at the intersection of federated learning, insider threat, and collaborative healthcare institutions can use this study as a foundation to adopt a collaborative approach and enhance privacy-preserving detection approaches, particularly in healthcare institutions and other regulated areas in general.

5.4 Recommendations for Future Research

We recognize that as insider threat landscapes continue to evolve, so too must the methods used to detect and mitigate these threats. Traditional centralized machine learning approaches have struggled with privacy concerns due to data centralization, as highlighted in our literature review. Federated Learning (FL) offers a promising solution by addressing these challenges; however, it is not without limitations, particularly when applied to safeguarding healthcare information in collaborative healthcare institutions. For example, the Centralized Federated Learning (CFL) approach, while facilitating the adoption of a collaborative approach, faces challenges related to its reliance on a central server, which can create a single (Lu et al., 2022).

Numerous academic studies have explored various centralized approaches to detect insider threats. However, to the best of our knowledge, beside this Praxis research, none have comprehensively tackled the unique challenges presented by insider threat detection in collaborative healthcare institutions by adopting decentralized federated learning approach with different privacy-preserving techniques. This remains a sophisticated and complex problem to fully resolve, particularly within the constraints of academic research, limited time, budget, and the lack of real-world data from healthcare institutions.

The Federated Learning (FL) approach for insider threat detection is still emerging. While much of the existing work focuses on centralized approaches for insider threat detection or applies CFL to insider threat detection on the Internet of Things (IoT), other studies have explored DFL in cybersecurity areas, including intrusion detection systems (IDS) and related fields. DFL's potential for insider threat detection, particularly in collaborative healthcare environments, is relatively new because much of the previous research has predominantly focused on centralized machine learning approaches. DFL,

especially in highly regulated sectors like healthcare, represents an area of research with room for further exploration and innovation.

Organizations must confront both malicious and accidental insider threats, given their subtle nature. Like most of previous works on insider threat, we can also confirm that there are many limitations to the quality of the dataset. For example, the recently created CERT dataset focuses only on known attack scenarios (Kim et al., 2019), which are not the same as the real-world scenario. Not only is the lack of real-world datasets from healthcare institutions a challenge, but it also poses a significant obstacle to the advancement of insider threat detection, impeding the development of effective solutions for real-world environments. In this study, like much of the previous work on insider threat detection, we recognize that relying on synthetic datasets, such as CERT's Insider Threat dataset, while valuable, does not fully capture the complexities of real-world environments.

Despite advanced research on insider threats, the lack of real-world data from organizations underscores “the challenges in validating and refining detection models” (Zaytsev et al., 2017; Aditham et al., 2016; Gheyas et al., 2016; Al-Mhiqani et al., 2020). This gap in real-world data makes it difficult to validate the practical implementation of collaborative insider threat solutions. For insider threat detection to progress, particularly in collaborative healthcare environments, obtaining real-world data is essential to assess scalability and effectiveness (Salem & Stolfo, 2011; Kim et al., 2019; Amiri-Zarandi et al., 2023). Other studies, such as those by Sarhan and Altwajiry (2023); Al-Mhiqani et al. (2020), have similarly emphasized the critical need for real-world data to validate insider threat systems, highlighting that without such data, solutions will remain constrained in their practical applicability and robustness.

Based on our implementation experience and previous research, we note that future solutions can be trained, tested, and evaluated using real-world datasets from collaborative healthcare environments to detect suspicious activities in real-time scenarios, enabling the development of effective solutions for real-world environments. We strongly recommend that future research prioritize acquiring real-world datasets, as they would provide a more comprehensive and meaningful assessment of insider threat detection systems.

References

Abadi, M., Chu, A., Goodfellow, I., McMahan, HB., Mironov, I., Talwar, K., & Zhang, L. (2016) Deep learning with differential privacy. In: Proceedings of the 2016 ACM SIGSAC conference on computer and communications security. ACM, pp 308–318

Aditham, S., Ranganathan, N., & Katkoori, S. Memory access pattern based insider threat detection in big data systems. In Proceedings of the 2016 IEEE International Conference on Big Data (Big Data), IEEE, Washington, DC, USA, 5–8 December 2016; pp. 3625–3628

Adkins, H. [Google]. (October 22, 2024). 7 ways we're incorporating security by design into our products and services. https://blog.google/technology/safety-security/google-secure-by-design-pledge

Agarap, A. F. (2018). Deep learning using rectified linear units (ReLU). arXiv. https://arxiv.org/abs/1803.08375

Ahamed, S. F., Banerjee, S., Roy, S., Quinn, D., Vucovich, M., Choi, K., Rahman, A., Hu, A., Bowen, E., & Shetty, S. (2024). Accuracy-privacy trade-off in the mitigation of membership inference attack in federated learning. arXiv. https://arxiv.org/abs/2407.19119

AIDcmmdamuj0pc_SEM__k_1a2d99fc72451b5eaf8efecde0d4e642_k_&msclkid=1a2d99fc72451b5eaf8efecde0d4e642

Al-Mhiqani, M. N. , Ahmad, R., Abdin, Z. Z., Yassin, W., Hassan, A., & Mohammed, A. N. (2020). New insider threat detection method based on recurrent neural networks. *Indonesian Journal of Electrical Engineering and Computer Science. 17*(3), 1479. https://doi.org/10.1159/ijeeecs.v17.i3.pp1474-1479

Al-Mhiqani, M. N., Ahmad, R., Abidin, Z. Z., Yassin, W., Hassan, A., Abdulkareem, K. H., & Ali, N. S., and Yunos. Z. (2020). A review of insider threat detection: Classification, machine learning techniques, datasets, open challenges, and recommendations. *Applied Sciences*, 2020, *10*(15), 5208. https://doi.org/10.3390/app10155208

Al-Shehari, T., & Alsowail, R. A. Techniques and countermeasures for preventing insider threats. PeerJ Comput Sci. 2022 Apr 1;8:e938. doi: 10.7717/peerj-cs.938. PMID: 35494800; PMCID: PMC9044369. [National Library of Medicine]. National Center for Biotechnology Information.

Al-Shehari, T., Al-Razgan, M., Alfakih, T., Alsowail, R. A., & Pandiaraj, S. (2023). Insider threat detection model using anomaly-based isolation forest algorithm," in *IEEE Access*, vol. 11, pp. 118170-118185, 2023, doi: 10.1109/ACCESS.2023.3326750.

Alder, S. (2017, November 15). 5 Year Jail Term Upheld for Clinic Worker Who Stole PHI. HIPAA Journal. https://www.hipaajournal.com/5-year-jail-term-upheld-clinic-worker-phi

Alder, S. (2018, April 26). How to defend against insider threats in healthcare. HIPAA Journal. https://www.hipaajournal.com/how-to-defend-against-insider-threats-in-healthcare/

Alder, S. (2019, June 17). Patient care coordinator gets 1-year jail term for HIPAA violation. HIPAA Journal. https://www.hipaajournal.com/patient-care-coordinator-gets-1-year-jail-term-for-hipaa-violation

Alder, S. (2020, November 18). Failure to terminate former employee's access rights results in $202,000 HIPAA fine for New Haven, CT. HIPAA Journal. https://www.hipaajournal.com/failure-to-terminate-former-employees-access-rights-results-in-202000-hipaa-fine-for-new-haven-ct

Alder, S. (2022, April 25). HHS warns HPH sector about insider threats in healthcare. HIPAA Journal. https://www.hipaajournal.com/hhs-warns-hph-sector-about-insider-threats-in-healthcare

Alder, S. (2023, June 15). Washington Hospital Pays $240,000 HIPAA Penalty After Security Guards Access Medical Records. HIPAA Journal. https://www.hipaajournal.com/washington-hospital-security-guard-snooping-240000-hipaa-penalty

Alder, S. (2023, October 5). Editorial: Insider Threats to Healthcare Records. HIPAA Journal.

Alder, S. (2024, October 24). Healthcare Data Breach Statistics. https://www.hipaajournal.com/healthcare-data-breach-statistics

Alpaydin, E. (2014). *Introduction to machine learning.* Third Edition. The MIT Press.

Amao, M. (2023, August 15). The Hidden Risk of Insider Threat in Healthcare. [Healthcare Technology]. https://www.healthtechzone.com/topics/healthcare/articles/2023/08/15/456791-hidden-risk-insider-threat-healthcare.htm

Amao, T. (2023). Federated learning: A new paradigm for decentralized data processing. *Journal of Privacy and Security in Computing, 15*(4), 321-335.

American Medical Association [AMA]. (2022, July 25). Patient survey shows unresolved tension over health data privacy. https://www.ama-assn.org/press-center/press-releases/patient-survey-shows-unresolved-tension-over-health-data-privacy

Amiri-Zarandi, M., Karimipour, H., & Dara, R. A. A federated and explainable approach for insider threat detection in IoT. Internet of Things 24 (2023) 100965. https://doi.org/10.1016/j.iot.2023.100965

Aouedi, O., Sacco, A., Piamrat, K., & Marchetto, G. Handling Privacy-Sensitive Medical Data With Federated Learning: Challenges and Future Directions. IEEE J Biomed Health Inform. 2023 Feb;27(2):790-803. doi: 10.1109/JBHI.2022.3185673. Epub 2023 Feb 3. PMID: 35737624.

Appari, A., & Johnson, M. E. (2010). Information security and privacy in healthcare: Current state of research. *International Journal of Internet and Enterprise Management*, 6(4), 279–314.

Asad, M., Moustafa, A., & Ito, T. (2022, November 24). Federated Learning Versus Classical Machine Learning: A Convergence Comparison. CoRR, abs/2107.10976, https://dblp.org/rec/journals/corr/abs-2107-10976.bib

Atalay, H.N., Yücel, Ş. Decoding privacy concerns: the role of perceived risk and benefits in personal health data disclosure. *Arch Public Health* 82, 180 (2024). https://doi.org/10.1186/s13690-024-01416-z

Ataman, A. (2024, October 26). Cybersecurity in Healthcare: 7 Challenges & 10 Best Practices. https://research.aimultiple.com/cybersecurity-in-healthcare

Banse, A., Kreischer, J., & Jurgens, X.O.I. (2024, February 3). Federated Learning with Differential Privacy. https://arxiv.org/pdf/2402.02230

Bartoszewski. F.W., Just. M., Lones. M.A., and Mandrychenko O. (202). Anomaly detection for insider threats: an objective comparison of machine learning models and ensembles. ICT Systems Security and Privacy Protection: 36th IFIP TC 11 International Conference, Proceedings, SEC 2021, Oslo, Norway, June 22–24, 2021, Springer (2021), pp. 367-381

Batra, N., Betts, D., & Davis, S. (2019, April 30). Forces of change: The future of health. https://www2.deloitte.com/us/en/insights/industry/health-care/forces-of-change-health-care.html?icid=dcom_promo_featured%7Cus;en

Bin Goh, W. W., Kabir, M.N, Yoo, S., & Wong L. (2024, September 19). Ten quick tips for ensuring machine learning model validity. https://doi.org/10.1371/journal.pcbi.1012402

Biswal, A. (2024, September 23). 24 Cutting-Edge Artificial Intelligence Applications in 2024. https://www.simplilearn.com/tutorials/artificial-intelligence-tutorial/artificial-intelligence-applications

Bobbitt, Z. (2021, September 8). F1 Score vs. Accuracy: Which Should You Use?. https://www.statology.org/f1-score-vs-accuracy

Boenisch, F., Dziedzic, A., Schuster, R., Shamsabadi,A. S., Shumailov, L., & Papernot, N. When the curious abandon honesty: Federated learning is not private. In 2023 IEEE 8th European Symposium on Security and Privacy (EuroS&P), pp. 175–199. IEEE, 2023. 1, 3, 18

Bonawitz, K., Ivanov, V., Kreuter, B., Marcedone, A., McMahan, H. B., Patel, S., … & Seth, K. (2016). Practical secure aggregation for federated learning on user-held data. CoRR, abs/1611.04482, 2016. URL http://arxiv. org/abs/1611.04482. 1

Bonnet, A. (2023, November 23). Accuracy vs. Precision vs Recall in Machine Learning: What is the difference? https://encord.com/blog/classification-metrics-accuracy-precision-recall

Brauneck, A., Müller, F., & Schmidt, H. (2023). Privacy challenges in collaborative healthcare environments: The role of machine learning. *Healthcare Data Science, 12*(2), 45-58.

Brownlee, J. (2021, January 13). Gentle Introduction to the Adam Optimization Algorithm for Deep Learning. https://machinelearningmastery.com/adam-optimization-algorithm-for-deep-learning

Buba, M., Maki, A., & Mazurowski, M. A. A. (2018). Systematic study of the class imbalance problem in convolutional neural networks, Neural networks, vol.106, pp. 249-259, 2018.

Burrell, D. N., Nobles, C., Cusak, A., Omar, M., & Gillesania, L. (2022).Cybercrime and the Nature of Insider Threat Complexities in Healthcare and Biotechnology Engineering Organizations. Journal of Crime and Criminal Behavior, 2: 2, pp. 131-144. https://doi. org/10.47509 /JCCB.2022. v02i02.0

Cai, X., Wang, Y., Xu, S., Li, H., Zhang, Y., Liu, Z., & Yuan, X. (2024, March 17). LAN: Learning Adaptive Neighbors for Real-Time Insider Threat Detection. http://arxiv.org/pdf/2403.09209v2

Cappelli D M, Moore A P, Trzeciak R F. The CERT guide to insider threats: how to prevent, detect, and respond to information technology crimes (Theft, Sabotage, Fraud). Addison-Wesley, 2012.

Carlini, N., Ippolito, D., Jagielski, M., Lee, K., Tramer, F., & Zhang, C. Quantifying memorization across neural language models. In The Eleventh International Conference on Learning Representations, 2023. URL https://openreview.net/forum? id=TatRHT_1cK. 1, 18

Centers for Medicare & Medicaid Services. (2018, December 6). CMS Office of the Actuary Releases 2017 National Health Expenditures: National Health Expenditures 2017 Highlights. https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/NationalHealthExpendData/Downloads/highlights.pdf

Chalamala, S.R., Kummari, N.K., Singh, A.K. *et al.* Federated learning to comply with data protection regulations. *CSIT* 10, 47–60 (2022). https://doi.org/10.1007/s40012-022-00351-0

Chan, P. K., Ho, T. J., & Wang, C. K. (2019). Handling Incomplete Records in Machine Learnin g. Journal of Machine Learning Research, 20(65), 1-27. http://www.jmlr.org/papers/v20/19-109.html

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002, June). SMOTE: Synthetic Minority Over-sampling Technique. Journal of Artificial Intelligence Research, 16, 321-357. https://doi.org/10.1613/jair.953

Chen, S., Yu, D., Zou, Y., Yu, J., & Cheng, X. (2015). Decentralized Wireless Federated Learning with Differential Privacy. Journal of LaTeX Class Files, 14(8), 1-15.

Chinchani, R., & Iyer, A. (2005). A semantic based model for the specification, verification and enforcement of security protocols. Journal of Computer Security, 13(5), 813-868.

Choi, S., & Johnson, M. E. (2019, April). Do Hospital Data Breaches Reduce Patient Care Quality? https://arxiv.org/pdf/1904.02058

Compliancy Group. (n.d.). The CIA Triad: Confidentiality, Integrity, Availability for HIPAA. The CIA Triad: Confidentiality, Integrity, Availability for HIPAA. https://compliancy-group.com/the-cia-triad-confidentiality-integrity-availability-for-hipaa

Constable, S. D., Tang, Y., Wang, S., Jiang, X., & Chapin, S. (2015). Privacy-preserving GWAS analysis on federated genomic datasets. BMC medical informatics and decision making, 15 Suppl 5(Suppl 5), S2. https://doi.org/10.1186/1472-6947-15-S5-S2

Constable, S.D., Tang, Y., Wang, S. *et al.* Privacy-preserving GWAS analysis on federated genomic datasets. *BMC Med Inform Decis Mak* 15 (Suppl 5), S2 (2015). https://doi.org/10.1186/1472-6947-15-S5-S2

Cybersecurity and Infrastructure Security Agency [CISA]. (2023, October 25). Cybersecurity Best Practice: Secure-by-Design. https://www.cisa.gov/resources-tools/resources/secure-by-design

Cybersecurity and Infrastructure Security Agency [CISA]. (n.d.). Secure by Design: CISA, U.S. and International Partners Announce Updated Secure by Design Principles Joint Guide. https://www.cisa.gov/securebydesign

Dagan, Y., & Kur, G. (2022). A bounded-noise mechanism for differential privacy. Proceedings of Machine Learning Research vol 178:1–37, 202. https://proceedings.mlr.press/v178/dagan22a/dagan22a.pdf

Danger, A. (2022, May 19). Differential Privacy: What is all the noise about? https://arxiv.org/abs/2205.09453v1

Dinov, I. D. (2023). Model Performance Assessment, Validation, and Improvement. In: Data Science and Predictive Analytics. The Springer Series in Applied Machine Learning. Springer, Cham. https://doi.org/10.1007/978-3-031-17483-4_9

Dosh, M. (2021, May). Detecting insider threat within institutions using CERT dataset and different ML techniques. Periodicals of Engineering and Natural Sciences (PEN) 9(2):873. DOI:10.21533/pen.v9i2.1911

Douzas, G., Bação, F., & Last, F. (2018). Improving imbalanced learning through a heuristic oversampling method based on k-means and SMOTE. Information Sciences, 465, 1-20. https://doi.org/10.1016/j.ins.2018.06.056

Dwork, C., & Roth, A. (2014). The Algorithmic Foundations of Differential Privacy. *Foundations and Trends® in Theoretical Computer Science, 9*(3–4), 211–407. https://doi.org/10.1561/0400000042

Dwork, C., Kenthapadi, K., McSherry, F., Mironov, I., Naor, M. (2006). Our Data, Ourselves: Privacy Via Distributed Noise Generation. In: Vaudenay, S. (eds) Advances in Cryptology - EUROCRYPT 2006. EUROCRYPT 2006. Lecture Notes in Computer Science, vol 4004. Springer, Berlin, Heidelberg. https://doi.org/10.1007/11761679_29

Dwork, C., McSherry, F., Nissim, K., & Smith, A. (2014). Calibrating noise to sensitivity in private data analysis. Journal of Privacy and Confidentiality.

Dwork, C., McSherry, F., Nissim, K., Smith, A. (2006). Calibrating Noise to Sensitivity in Private Data Analysis. In: Halevi, S., Rabin, T. (eds) Theory of Cryptography. TCC 2006. Lecture Notes in Computer Science, vol 3876. Springer, Berlin, Heidelberg. https://doi.org/10.1007/11681878_14

Dwork, C., McSherry, F., Nissim, K., Smith, A. (2006). Calibrating Noise to Sensitivity in Private Data Analysis. In: Halevi, S., Rabin, T. (eds) Theory of Cryptography. TCC 2006. Lecture Notes in Computer Science, vol 3876. Springer, Berlin, Heidelberg. https://doi.org/10.1007/11681878_14

Fahmy, M.M. (2022, November 5). Confusion Matrix in Binary Classification Problems: A Step-by-Step Tutorial.Vol. 6, No. 5 – 2022 Journal of Engineering Research (ERJ). https://pdfs.semanticscholar.org/1af4/de16afcb0a5a0034beb26ac49bf16c7d2fd0.pdf

Federal Bureau of Investigation (FBI). (2021, January 6). Medical Equipment Packaging Company Hacker Sentenced. https://www.fbi.gov/news/stories/hacker-who-disrupted-ppe-shipments-sentenced-010621

Feretzakis, G., Papaspyridis, K., Gkoulalas-Divanis, A., Verykios, V.S. Privacy-Preserving Techniques in Generative AI and Large Language Models: A Narrative Review. *Information* 2024, *15*, 697. https://doi.org/10.3390/info15110697

Fisher & Phillips. (2024, May 21). Insider Threats to Healthcare Data: What You Need to Know and 5 Steps You Can Take Now. https://www.fisherphillips.com/en/news-insights/insider-threats-to-healthcare-data.html

Fitzgerald, A. (April, 2024). Secure by Design: What Does It Mean & How to Reasonably Implement It. https://secureframe.com/blog/secure-by-design

Flywheel. (2024). Federated learning in medical imaging: Project connects pharma with university to train AI model. https://flywheel.io/insights/blog/federated-learning-project-example

Fortinet. (n.d.). What Is An Insider Threat? https://www.fortinet.com/resources/cyberglossary/insider-threats

Gayathri, R.G., Sajjanhar, A, Uddin, M.P., & Xiang, Y. (2024, September). FedAT: Federated Adversarial Training for Distributed Insider Threat Detection. https://arxiv.org/abs/2409.13083

Ge, D., Zhong, S., & Chen, K. (2022). Multi-source data fusion for insider threat detection using residual networks. *2022 3rd International Conference on Electronics, Communications and Information Technology (CECIT)*,DOI: 10.1109/CECIT58139.2022.00069 . https://ieeexplore.ieee.org/abstract/document/10086398/authors

GeeksforGeeks. (2024, May 27). GeeksforGeeks. https://www.geeksforgeeks.org/types-of-federated-learning-in-machine-learning/?ref=gcse_outind

George Washington University. (2024). Doctor of engineering- Artificial intelligence and machine learning.https://online.engineering.gwu.edu/sites/g/files/zaxdzs5816/files/2024-04/gwu_program_flyers_-_april_2024_deng_ai.pdf

George Washington University.[Python Application in Cyber Analytics-SEAS 6800 DC4]. (2023). Introduction to python, ML, and Apps in cybersecurity. https://blackboard.gwu.edu

Gheyas, I.A., Abdallah, A.E. Detection and prediction of insider threats to cyber security: a systematic literature review and meta-analysis. Big Data Anal 1, 6 (2016). https://doi.org/10.1186/s41044-016-0006-0

Glasser, J.; Lindauer, B. Bridging the gap: A pragmatic approach to generating insider threat data. In Proceedings of the 2013 IEEE Security and Privacy Workshops, San Francisco, CA, USA, 23–24 May 2013; pp. 98–104.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT Press.

Google Trends. (n.d.). PyTorch vs TensorFlow. https://trends.google.com/trends

Goyal, P., & Malviya R. (2023, October 4). Healthcare Science: Challenges and opportunities of big data analytics in healthcare. https://doi.org/10.1002/hcs2.66

Goyal, P., & Malviya, R. (2023, August 17). Challenges and opportunities of big data analytics in healthcare. Healthcare Science. DOI: 10.1002/hcs2.66. https://onlinelibrary.wiley.com/doi/pdfdirect/10.1002/hcs2.66

Grafton, J. (2023, August 2). UEBA vs SIEM: The key differences of each solution. Gurucul. https://gurucul.com/blog/ueba-vs-siem

Grand View Research. (2023). Artificial intelligence market size, share, growth report 2030. https://www.grandviewresearch.com/industry-analysis/artificial-intelligence-ai-market

Greitzer, F. L., Moore, A. P., Cappelli, D. M., Andrews, D., Carroll. L. A., & Hull, T. (2008, February). Combating the insider cyber threat. IEEE Security and Privacy Magazine 6(1):61 - 64. DOI:10.1109/MSP.2008.8

Gustavo, E., A., P., A., Batista., Ronaldo, C., Prati., Maria, Carolina, Monard. (2004). A study of the behavior of several methods for balancing machine learning training data. Sigkdd Explorations, 6(1):20-29. doi: 10.1145/1007730.1007735

He, C., Tan, C., Tang, H., Qiu, S., & Liu, J. (2020) Central server free federated learning over single-sided trust social networks. arXiv:1910.04956

He, H., Bai, Y., Garcia, E. A., & Li, S. (2008). ADASYN: Adaptive Synthetic Sampling Approach for Imbalanced Learning. IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence), 1322-1328.

Hinton , G. E., Srivastava, N., Krizhevsky, N., Sutskever, I, & Salakhutdinov, R. (2012). Improving neural networks by preventing co-adaptation of feature detectors. https://arxiv.org/abs/1207.0580

HIPAA Journal. (2024). What are the Penalties for HIPAA Violations? https://www.hipaajournal.com/what-are-the-penalties-for-hipaa-violations-7096

Hochreiter, S., & Schmidhuber, J. (1997, December). Long Short-term Memory. Neural Computation 9(8):1735-80. DOI: 10.1162/neco.1997.9.8.1735

https://arxiv.org/abs/2107.10976 , arXiv:2107.10976

https://mdpires.com/bookfiles/book/5903/Digital_Transformation_in_Healthcare.pdf?v=1707844831

https://www.hhs.gov/sites/default/files/insider-threats-in-healthcare.pdf

https://www.hipaajournal.com/insider-threats-to-healthcare-records

IBM. (2024.). Cost of a Data Breach Report 2024. https://www.ibm.com/reports/data-breach

Ibrahim, H. E., Badr, S. M., & Shaheen, M., A. (2012, October). Adaptive Layered Approach using Machine Learning Techniques with Gain Ratio for Intrusion Detection Systems. International Journal of Computer Applications (0975 – 8887) Volume 56– No.7, October 2012

InfoSec Champ. (n.d). 25 PII Security Best Practices: A Guide to Keeping Personal Information Safe | The Role of a PII Security Policy in Protecting Personally Identifiable Information. https://infosecchamp.com/pii-security-best-practices-pii-security-policy/

Intel. (2024). OpenFL: What is Federated Learning?. https://openfl.readthedocs.io/en/latest/about/overview.html#what-is-federated-learning

Islam, T. U., Ghasemi, R., & Mohammed, N. (2022). Privacy-Preserving Federated Learning Model for Healthcare Data, *2022 IEEE 12th Annual Computing and Communication Workshop and Conference (CCWC)*, Las Vegas, NV, USA, 2022, pp. 0281-0287, doi: 10.1109/CCWC54503.2022.9720752.

Iyengar, G., Lam, H., $ Wang T. (2024, August 20). Is Cross-Validation the Gold Standard to Evaluate Model Performance?. https://arxiv.org/pdf/2407.02754

Jacobson, R. S., Becich, M. J., Bollag, R. J., Chavan, G., Corrigan, J., Dhir, R., Feldman, M. D., Gaudioso, C., Legowski, E., Maihle, N. J., Mitchell, K., Murphy, M., Sakthivel, M., Tseytlin, E., & Weaver, J. (2015). A Federated Network for Translational Cancer Research Using Clinical Data and Biospecimens. Cancer research, 75(24), 5194–5201. https://doi.org/10.1158/0008-5472.CAN-15-1973

Jawad, L. A. (2024). Security and Privacy in Digital Healthcare Systems: Challenges and Mitigation Strategies. Abhigyan, 42(1), 23-31. https://doi.org/10.1177/09702385241233073

Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. Science, 349(6245), 255-260. https://doi.org/10.1126/science.aaa8415

Joshi, M., Pal, A., & Sankarasubbu, M. (2022). Federated Learning for Healthcare Domain - Pipeline, Ap_plications and Challenges. ACM Trans. Comput. Healthcare 3, 4, Article 40 (October 2022), 36 pages. https://doi.org/10.1145/3533708

Kan, E. (2020, November 6). *How to implement an Adam Optimizer from Scratch.* https://medium.com/the-ml-practitioner/how-to-implement-an-adam-optimizer-from-scratch-76e7b217f1cc

Karthikeyan, A., & Priyakumar, U. D. (2022, March). *Artificial intelligence: machine learning for chemical sciences*. Journal of Chemical Sciences. 134. 10.1007/s12039-021-01995-2. DOI:10.1007/s12039-021-01995-2

Kim, A., Oh, J., Ryu, J., & Lee, K. (2020). A review of insider threat detection approaches with IoT perspective. *IEEE Internet of Things Journal*, 7(5), 4507-4519. https://ieeexplore.ieee.org/document/9078082

Kim, J., Park, M., Kim, H., Cho, S., Kang, P. Insider Threat Detection Based on User Behavior Modeling and Anomaly Detection Algorithms. *Appl. Sci.* 2019, *9*, 4018. https://doi.org/10.3390/app9194018

Konečný, J.; McMahan, H.B.; Ramage, D.; Richtárik, P. Federated Optimization: Distributed Machine Learning for On-Device Intelligence. *arXiv* 2016, arXiv:1610.02527.

Kumar, M.; Kim, S. Securing the Internet of Health Things: Embedded Federated Learning-Driven Long Short-Term Memory for Cyberattack Detection. *Electronics* 2024,*13*,3461. https:// doi.org/10.3390/electronics13173461

Lagasse, J. (2022, July 25). Vulnerabilities in health data privacy causing tension among patients. https://www.healthcarefinancenews.com/news/vulnerabilities-health-data-privacy-causing-tension-among-patients

Lavanya, P., Glory H. A., & Shankar Sriram V. S. (2024, May) Mitigating Insider Threat: A Neural Network Approach for Enhanced Security. IEEE Access. 2024;12:73752-73768. DOI:10.1109/ACCESS.2024.3404814

Le, D. C. (2021, August). Machine Learning based Framework for User-Centered Insider Threat Detection. https://web.cs.dal.ca/~mheywood/Thesis/DucLe2021.pdf

Le, D. C., & Zincir-Heywood, N. (2021). Anomaly detection for insider threats using unsupervised ensembles. IEEE Transactions on Network and Service Management, 18(2), 1152–1164. https://doi.org/10.1109/TNSM.2021.3071928

Le, D. C., Zincir-Heywood, N., & Heywood, M. I. (2020). Analyzing data granularity levels for insider threat detection using machine learning. IEEE Transactions on Network and Service Management, 17(1), 30–44. https://doi.org/10.1109/TNSM.2020.2967721

LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* 521, 436–444 (2015). https://doi.org/10.1038/nature14539

Li, T., Sahu, A. K., Talwalkar, A., & Smith, V. (2020). Federated learning: Challenges, methods, and future directions. IEEE Signal Processing Magazine, 37(3), 50-60. https://doi.org/10.1109/MSP.2020.297574

Lian, Z., Yang, Q., Wang, W., Zeng, Q., Alazab, M., Zhao, H., & Su, C. DEEP-FEL: Decentralized, Efficient and Privacy-Enhanced Federated Edge Learning for Healthcare Cyber Physical Systems," in *IEEE Transactions on Network Science and Engineering*, vol. 9, no. 5, pp. 3558-3569, 1 Sept.-Oct. 2022, doi: 10.1109/TNSE.2022.3175945.

Lindauer, B. (2020). Insider Threat Test Dataset. https://kilthub.cmu.edu/articles/dataset/Insider_Threat_Test_Dataset/12841247

Liu, X., Li, H., Xu, G. *et al.* Adaptive privacy-preserving federated learning. *Peer-to-Peer Netw. Appl.* 13, 2356–2366 (2020). https://doi.org/10.1007/s12083-019-00869-2

Lu, Y., Yu, Z., & Suri, N. (2022). Privacy-preserving decentralized federated learning over time-varying communication graph. arXiv. https://arxiv.org/pdf/2210.00325

Manoharan, P., Yin, J., Wang, H. *et al.* Insider threat detection using supervised machine learning algorithms. *Telecommun Syst* (2023). https://doi.org/10.1007/s11235-023-01085-3

Martínez Beltrán, E.T., Pérez, M.Q., S'anchez, P.M., Bernal, S.L., Bovet, G., Pérez, M.G., P'erez, G.M., & Celdr'an, A.H. (2022). Decentralized Federated Learning: Fundamentals, State of the Art, Frameworks, Trends, and Challenges. IEEE Communications Surveys & Tutorials, 25, 2983-3013.

Mazzarolo, G., & Jurcut, A. D. (2019). Insider threats in Cyber Security: The enemy within the gates. arXiv. https://arxiv.org/abs/1911.09575

McKeon, J. (2024, May 08). Insider threats in healthcare remain prevalent. https://www.techtarget.com/healthtechsecurity/news/366594060/Insider-threats-in-healthcare-remain-prevalent

McKinsey & Company. (2022). The state of AI in 2022—and a half decade in review. https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai-in-2022-and-a-half-decade-in-review

McMahan, B., Moore, E., Ramage, D., Hampson, S. & Arcas, B.A.y.. (2023). Communication-Efficient Learning of Deep Networks from Decentralized Data. Proceedings of the 20th International Conference on Artificial Intelligence and Statistics, in Proceedings of Machine Learning Research 54:1273-1282. https://proceedings.mlr.press/v54/mcmahan17a.html.

McMahan, H. B, Ramage D, Talwar K, Zhang L (2017) Learning differentially private recurrent language models. arXiv:1710.06963

McMahan, H. B., Moore, E., Ramage, D., & Hampson, S. (2017). Communication-efficient learning of deep networks from decentralized data. In Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS 2017), 1273-1282. https://arxiv.org/abs/1602.05629

Microsoft Security. (n.d.). What is identity access management (IAM)? Microsoft. https://www.microsoft.com/en-us/security/business/security-101/what-is-identity-access-management-iam?ef_id=_k_1a2d99fc72451b5eaf8efecde0d4e642_k_&OCID=

Munuo, J. (2023). Protecting Personal Identifiable Information in the Financial Services Sector with Zero Trust Architecture. https://scholarspace.library.gwu.edu/etd/4x51hj87d

Murdoch, B. Privacy and artificial intelligence: challenges for protecting health information in a new era. BMC Medical Ethics 22, 122 (2021). https://doi.org/10.1186/s12910-021-00687-3

Naresh, V.S., Thamarai, M. & Allavarpu, V.V.L.D. Privacy-preserving deep learning in medical informatics: applications, challenges, and solutions. *Artif Intell Rev* 56 (Suppl 1), 1199–1241 (2023). https://doi.org/10.1007/s10462-023-10556-7

National Insider Threat Task Force. (2016). Protect your organization from the inside out: Government best practice. https://www.dni.gov/files/NCSC/documents/products/Govt_Best_Practices_Guide_Insider_Threat.pdf

National Institute of Standards and Technology (NIST). (July 21, 2022). NIST Updates Guidance for Health Care Cybersecurity-Revised draft publication aims to help organizations comply with HIPAA Security Rule. https://www.nist.gov/news-events/news/2022/07/nist-updates-guidance-health-care-cybersecurity

National Institute of Standards and Technology [NIST]. . Engineering Trustworthy Secure. NIST Special Publication. NIST SP 800-160v1r1. Systems. https://doi.org/10.6028/NIST.SP.800-160v1r1

National Institute of Standards and Technology [NIST]. Engineering Principles for Information Technology Security. Special Publication 800-27 Rev A. (A Baseline for Achieving Security), Revision A. National Institute of Standards and Technology Special Publication 800-27 Rev A Natl. Inst. Stand. Technol. Spec. Publ. 800-27 Rev A, 33 pages (June, 2004) CODEN: NSPUE2. https://tsapps.nist.gov/publication/get_pdf.cfm?pub_id=151294

National Institute of Standards and Technology [NIST]. NIST Special Publication 800-207. Zero Trust Architecture. https://doi.org/10.6028/NIST.SP.800-207

National Security Agency/ Central Security Service [NSA]. (2023). https://www.nsa.gov/Press-Room/Press-Releases-Statements/Press-Release-View/Article/3558946/nsa-and-partners-issue-additional-guidance-for-secure-by-design-software

Nepal, B. (2024).Threat Detection and Intelligence Sharing Using a Decentralized Federated Learning. https://scholarspace.library.gwu.edu/etd/rv042t86x

Nguyen, T. V., Dakka, M. A., Diakiw, S. M., VerMilyea, M. D., Perugini, M., Hall, J. M. M., and Perugini, D. (2022, May 25). https://doi.org/10.1038/s41598-022-12833-x

North Atlantic Treaty Organization [NATO] Cooperative Cyber Defense Centre of Excellence. (n.d.). *Insider Threat Study.* https://www.nationalinsiderthreatsig.org/itrmresources/Insider%20Threat%20Study%20By%20NATO.pdf

Office of Public Affairs-U.S. Department of Justice. (February 1, 2021). Hospital Researcher Sentenced to Prison for Conspiring to Steal Trade Secrets, Sell Them in China. https://www.justice.gov/opa/pr/hospital-researcher-sentenced-prison-conspiring-steal-trade-secrets-sell-them-china

Oh, S.-R.; Seo, Y.-D.; Lee, E.; Kim, Y.-G. A Comprehensive Survey on Security and Privacy for Electronic Health Data. *Int. J. Environ. Res. Public Health* 2021, *18*, 9668. https://doi.org/10.3390/ijerph18189668

Olaoye, F., & Egon, A. (2024, August). Artificial Intelligence: Explainable AI for Security Decision Making. https://www.researchgate.net/publication/383565465_Explainable_AI_for_Security_Decision_Making

Ostendorf, C. (2023, December 8). 11 Real-Life Insider Threat Examples. Code42. https://www.code42.com/blog/insider-threat-examples-in-real-life

Owkin. (2022). Privacy-preserving machine learning to protect health data. https://www.owkin.com/blogs-case-studies/privacy-preserving-machine-learning-to-protect-health-data

Pantelidis, E., Bendiab, G., Shiaeles, S., & Kolokotronis, N. (July, 2021). Insider detection using deep autoencoder and neural networks. *2021 IEEE International Conference on Cyber Security and Resilience (CSR)*, Rhodes, Greece, 2021, pp. 129-134, doi: 10.1109/CSR51186.2021.9527925. https://ieeexplore.ieee.org/document/9527925

Paracha, A., Arshad, J., Farah, M. et al. Machine learning security and privacy: a review of threats and countermeasures. EURASIP J. on Info. Security 2024, 10 (2024). https://doi.org/10.1186/s13635-024-00158-3

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12, 2825–2830. https://www.jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf

Pelikan, M., Azam, S. S., Feldman, V., Silovsky, J., Talwar, K., & Likhomanenko, T. (2023). Federated learning with differential privacy for end-to-end speech recognition. https://arxiv.org/pdf/2310.00098

Peng, L., Qiu, M. (2024). AI in Healthcare Data Privacy-Preserving: Enhanced Trade-Off Between Security and Utility. In: Cao, C., Chen, H., Zhao, L., Arshad, J., Asyhari, T., Wang, Y. (eds) Knowledge Science, Engineering and Management. KSEM 2024. Lecture

Pfitzner, B., Steckhan, N & Arnrich, B. (2021, June 2). Federated Learning in a Medical Context: A Systematic Literature Review. https://doi.org/10.1145/3412357

Piran, F.J., Chen, Z., Imani, M., & Imani, F. (2024, November 5). Privacy-Preserving Federated Learning with Differentially Privacy Hyperdimensional Computing. Notes in Computer Science, vol 14886. Springer, Singapore. https://doi.org/10.1007/978-981-97-5498-4_27

Ponemon Institute [Proofpoint]. (2022). 2022 Ponemon Cost of Insider Threats- Global Report. https://www.proofpoint.com/us/resources/threat-reports/cost-of-insider-threats

Ponemon Institute [Proofpoint]. (2023, October 11). Second Annual Ponemon Institute Report Finds That Two-Thirds of Healthcare Organizations Surveyed Experienced Disruption to Patient Care Due to Cyber Attacks. https://www.proofpoint.com/us/newsroom/press-releases/second-annual-ponemon-institute-report-finds-two-thirds-healthcare

Ponemon Institute [Proofpoint]. (2024, October 08). Third Annual Ponemon Institute Report: Nearly Seven in 10 Healthcare Organizations Experienced Disruption to Patient Care Due to Cyber Attacks. https://www.proofpoint.com/us/newsroom/press-releases/third-annual-ponemon-institute-report-nearly-seven-10-healthcare

Probst, C. W., Hunker, J., Gollmann, D. & Bishop, M. (2010). Insider threats in cyber security. In Insider threats in cyber security (pp. 1-6). Springer, Boston, MA.

QDacity (n.d.). What is Theory Triangulation?: How to improve your qualitative research using Theory Triangulation. https://qdacity.com/theory-triangulation

Qi, J., Zhou, Q., Lei, L., & Zheng, K. Federated reinforcement learning: techniques, applications, and open chal_lenges. Intell Robot 2021;1(1):18-57. http://dx.doi.org/10.20517/ir.2021.02

Rajagopal, A., Ayanian, S., Ryu, A. J., Qian, R., Legler, S. R., Peeler, E. A., Issa, M., Coons, T. J., & Kawamoto, K. (2024). Machine learning operations in health care: A scoping review. Mayo Clinic Proceedings: Digital Health, 2(3), 421–437. https://doi.org/10.1016/j.mcpdig.2024.06.009

Rieke, N., Hancox, J., Li, W., Milletari, F., Roth, H. R., Albarqouni, S., ... & Kaissis, G. (2020). The future of digital health with federated learning. *npj Digital Medicine, 3*(1), 1-7. https://doi.org/10.1038/s41746-020-00323-1

Rivera, D. J. (2023). A Collaborative Ensemble of Reinforcement Learning Agents for Automated Network Penetration Testing. https://scholarspace.library.gwu.edu/etd/r781wg776

Rodriguez, L. M. (2023). Predictive Model for Electrical Assembly Manufacturing Defect Dispositions. https://scholarspace.library.gwu.edu/etd/5712m7512

Roy A. G, Siddiqui S, Pölsterl S, Navab N, Wachinger C (May 16, 2019) Braintorrent: a peer-to-peer environment for decentralized federated learning. arXiv:1905.06731.

Saha, S., Hota, A., Chattopadhyay, A.K. et al. A multifaceted survey on privacy preservation of federated learning: progress, challenges, and opportunities. Artif Intell Rev 57, 184 (2024). https://doi.org/10.1007/s10462-024-10766-7

Salem, M. B., & Stolfo, S. J. (2011). Modeling user search behavior for masquerade detection. https://link.springer.com/chapter/10.1007/978-3-642-23644-0_10

Salmi, M., Atif, D., Oliva, D. et al. Handling imbalanced medical datasets: review of a decade of research. Artif Intell Rev 57, 273 (2024). https://doi.org/10.1007/s10462-024-10884-2

Sarhan, B. B., & Altwaijry, N. Insider Threat Detection Using Machine Learning Approach. *Appl. Sci.*2023,*13*,259. https://doi.org/ 10.3390/app13010259

Sarhan, B. B., & Altwajiry A. (2023). Insider threat detection using machine learning approaches. *Applied Sciences*, 2023, *13*(1), 259; https://doi.org/10.3390/app13010259

Sharma, B., Pokharel, P., & Joshi, B. (2020). User behavior analytics for anomaly detection using lstm autoencoder - insider threat detection. Proceedings of the 11th International Conference on Advances in Information Technology. https://doi.org/10.1145/3406601.3406610

Sharma, H., Gosain, A. (2023). Oversampling Methods to Handle the Class Imbalance Problem: A Review. In: Patel, K.K., Santosh, K.C., Patel, A., Ghosh, A. (eds) Soft Computing and Its Engineering Applications. icSoftComp 2022. Communications in Computer and Information Science, vol 1788. Springer, Cham. https://doi.org/10.1007/978-3-031-27609-5_8

Sharma, V., & Kumar, S. "Role of Artificial Intelligence (AI) to Enhance the Security and Privacy of Data in Smart Cities," 2023 3rd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), Greater Noida, India, 2023, pp. 596-599, doi: 10.1109/ICACITE57410.2023.10182455.

Shayan M, Fung C, Yoon CJ, Beschastnikh I (2019) Biscotti: a ledger for private and secure peer-to-peer machine learning. arXiv:1811.09904

Shivanandhan, M. (2024, January 10). Deep learning: PyTorch vs TensorFlow – Which is Better for Deep Learning Projects? https://www.freecodecamp.org/news/pytorch-vs-tensorflow-for-deep-learning-projects

Sindhwani, N., Anand, R., George, A.S. & Pandey, D. Robotics and Automation in Industry 4.0 –

Singh, P., Singh, M.K., Singh, R., Singh, N. (2022). Federated Learning: Challenges, Methods, and Future Directions. In: Yadav, S.P., Bhati, B.S., Mahato, D.P., Kumar, S. (eds)

Federated Learning for IoT Applications. EAI/Springer Innovations in Communication and Computing. Springer, Cham. https://doi.org/10.1007/978-3-030-85559-8_13

Skahill, E & West, D. M. (2021, August 9). Why hospitals and healthcare organizations need to take cybersecurity more seriously. https://www.brookings.edu/articles/why-hospitals-and-healthcare-organizations-need-to-take-cybersecurity-more-seriously

Smart Industries and Intelligent Technologies", CRC Press, 2024

Smith, A., & Nguyen, T. (2021). Privacy in Federated Learning: A Comprehensive Review. Journal of Machine Learning, 9(2), 105–130

Soares dos Santos, M. P. (2022). Digital Transformation in Healthcare. MDPI.

Soleymanzadeh, R., Kashef, R. Efficient intrusion detection using multi-player generative adversarial networks (GANs): an ensemble-based deep learning architecture. *Neural Comput & Applic* 35, 12545–12563 (2023). https://doi.org/10.1007/s00521-023-08398-z

Srivastava, N., Geoffrey E. Hinton., Srivastava, S., Krizhevsky,A., Sutskever, I Salakhutdinov, R.R. (2014). Journal of Machine Learning Research 15 (2014) 1929-1958. https://jmlr.org/papers/volume15/srivastava14a/srivastava14a.pdf

Subramanian, R. Have the cake and eat it too: Differential Privacy enables privacy and precise analytics. *J Big Data* 10, 117 (2023). https://doi.org/10.1186/s40537-023-00712-9

Sun, Y., Shen, Li., & Tao, D. (2023). WHICH MODE IS BETTER FOR FEDERATED LEARNING? CENTRALIZED OR DECENTRALIZED. 10.13140/RG.2.2.22032.79362.

Sze, V., Chen, Y.H., Yang, T.J. Emer, J. S. (2017). Efficient Processing of Deep Neural Networks: A Tutorial and Survey. Digital Object Identifier: 10.1109/JPROC.2017.2761740. Vol. 105, No. 12, December 2017

Tabassum, M., Mahmood, S., Bukhari, A. *et al.* Anomaly-based threat detection in smart health using machine learning. *BMC Med Inform Decis Mak* 24, 347 (2024). https://doi.org/10.1186/s12911-024-02760-4

Teo, Z. L., Jin, L., Liu, N., Li, S., Miao, D., Zhang X., Ng, W. Y., Tan, T, F., Lee, D. M., Chua, K.J., Heng, J., Liu., Goh, R. S. M., and Ting, D.S.W. (2024, February 20) Cell Reports Medicine, Volume 5, Issue 2, 101419. https://www.cell.com/cell-reports-medicine/fulltext/S2666-3791(24)00042-9

Thapa, C., Chamikara, M.A.P. , Camtepe, S., & Sun, L. (2020). Research on Data Splitting Techniques in Federated Learning Environments.

The Center for Internet Security. (n.d.). Insider Threats: In the Healthcare Sector. https://www.cisecurity.org/insights/blog/insider-threats-in-the-healthcare-sector

Tian, Z., Shi, W., Tan, Z. et al. Deep LeNetw Appld Dempster-Shafer Theory Based Insider Threat Detection. Mobile Netw Appl (2020). https://doi.org/10.1007/11036-020-01656-7

Tomek, I. (1976). Two Modifications of CNN. IEEE Transactions on Systems, Man, and Cybernetics, SMC-6(11), 769-772. https://ieeexplore.ieee.org/document/4309452

Transforming Data With Intelligence. (2023, October 19). *Patient data breaches doubled in 2023, reaching 87M*. https://tdwi.org/articles/2023/10/19/patient-data-breaches-doubled-in-2023.aspx

Truong, N., Sun, K., Wang, S., Guitton, F., & Guo, Y. (March, 2021). Privacy Preservation in Federated Learning: An insightful survey from the GDPR Perspective. https://arxiv.org/pdf/2011.05411

Tuychiev, B. (2024, August 29). Adam Optimizer Tutorial: Intuition and Implementation in Python. https://www.datacamp.com/tutorial/adam-optimizer-tutorial

U.S. Attorney's Public Affairs Office. (2020, October 20). Former employee of medical packaging company sentenced to federal prison for disrupting PPE shipments. https://www.justice.gov/usao-ndga/pr/former-employee-medical-packaging-company-sentenced-federal-prison-disrupting-ppe

U.S. Department of Health & Human Services. (2022, April 21). Insider threats in healthcare.

UPMC HealthBeat. (2024, October 8). Policy and Advocacy: Why Cybersecurity Is Crucial for Health Care Systems. https://share.upmc.com/2024/10/cybersecurity-for-health-care

Vaideeswaran, N. (2023, November 17). Identity access management (IAM). CrowdStrike. https://www.crowdstrike.com/en-us/cybersecurity-101/identity-protection/identity-access-management-iam

Vasko, J., & Gardner, C. (2019, February 21). Insider Threats in Healthcare (Part 7 of 9: Insider Threats Across Industry Sectors). Retrieved October 30, 2024, from https://insights.sei.cmu.edu/blog/insider-threats-in-healthcare-part-7-of-9-insider-threats-across-industry-sectors/.

Vishwakarma, N. (2024, October 15). What is Adam Optimizer? https://www.analyticsvidhya.com/blog/2023/09/what-is-adam-optimizer

Wang, L., Xu, S., Wang, X., & Zhu, Q. (2021). Addressing Class Imbalance in Federated Learning. *Proceedings of the AAAI Conference on Artificial Intelligence*, *35*(11), 10165-10173. https://doi.org/10.1609/aaai.v35i11.17219

Wilder, J. (2022, April 27). HHS Warns of Insider Threats in the Healthcare Sector. Healthcare Innovation. https://www.hcinnovationgroup.com/cybersecurity/data-breaches/news/21265719/hhs-warns-of-insider-threats-in-the-healthcare-sector

World Economic Forum. (2019, December 5). Healthcare and Systems: 4 ways data is improving healthcare. https://www.weforum.org/stories/2019/12/four-ways-data-is-improving-healthcare/

Xiang, D., & Cai, W. Privacy Protection and Secondary Use of Health Data: Strategies and Methods. Biomed Res Int. 2021 Oct 7;2021:6967166. doi: 10.1155/2021/6967166. PMID: 34660798; PMCID: PMC8516535.

Xu, J., Glicksberg, B.S., Su, C. *et al.* Federated Learning for Healthcare Informatics. *J Healthc Inform Res* 5, 1–19 (2021). https://doi.org/10.1007/s41666-020-00082-4

Yan, Z., & Li, D. (2024, March). Adaptive Decentralized Federated Learning in Energy and Latency Constrained Wireless Networks. https://arxiv.org/pdf/2403.20075

Yang Q., Liu, Y., Chen, T. , & Tong, Y. (2019, January 28). Federated Machine Learning: Concept and Applications. https://doi.org/10.1145/3298981

Yang, X.; Dong, Z. Kalman Filter-Based Differential Privacy Federated Learning Method. *Appl. Sci.*2022,*12*,7787. https://doi.org/ 10.3390/app12157787

Yin, L. ., Lv, Y., Sun, Z. , Tao, F., & Li, R. "A Quantifying Approach for Evaluating Differential Privacy in Deep Learning," *2022 IEEE Smartworld, Ubiquitous Intelligence & Computing, Scalable Computing & Communications, Digital Twin, Privacy Computing, Metaverse, Autonomous & Trusted Vehicles (SmartWorld/UIC/ScalCom/DigitalTwin/PriComp/Meta)*, Haikou, China, 2022, pp. 2272-2278, doi: 10.1109/SmartWorld-UIC-ATC-ScalCom-DigitalTwin-PriComp-Metaverse56740.2022.00323.

Yousefpour, A., Shilov, I., Sablayrolles, A., Testuggine, D., Prasad, K., Malek, M., Nguyen, J., Ghosh, S., Bharadwaj, A., Zhao, J., Cormode, G., & Mironov, I. (2022). Opacus: User-

Friendly Differential Privacy Library in PyTorch. arXiv preprint arXiv:2109.12298. https://arxiv.org/abs/2109.12298

Youssef, A., Ng, M. Y., Long, J., Hernandez-Boussard, T., Shah, N., Miner, A., Larson, D., & Langlotz, C. P. (2023). Organizational Factors in Clinical Data Sharing for Artificial Intelligence in Health Care. JAMA Network Open, 6(12), e2348422

Yuan S, Wu X. Deep learning for insider threat detection: Review, challenges and opportunities. Computers & Security, 2021, 104: 102221.

Yuan, L. (May, 2024). Decentralized Federated Learning: A Survey and Perspective. https://arxiv.org/pdf/2306.01603

Zaytsev, A., Malyuk, A.; Miloslavskaya, N. Critical Analysis in the Research Area of Insider Threats. In Proceedings of the 2017 IEEE 5th International Conference on Future Internet of Things and Cloud (FiCloud), Prague, Czech Republic, 21–23 August 2017; pp. 288–296.

Zhang, X., Yin, W., Hong, M., & Chen, T. (2021). Hybrid federated learning algorithms and implementation. https://arxiv.org/abs/2012.12420

Zhao, B. Z. H., Kaafar, M. A., & Kourtellis, N. (2020). Not one but many trade-offs: Privacy vs. utility in differentially private machine learning. arXiv. https://arxiv.org/abs/2008.08807

Zhauniarovich, Y., Khalil, I., Yu, T., and Dacier, M. (2018). A survey on malicious domains detection through dns data analysis. ACM Computing Surveys, 51(4):Article 67.

Zibak, A., & Simpson, A. (2019). A survey on machine learning for insider threat detection. ACM Computing Surveys (CSUR), 52(5), 1-36.

Appendix A

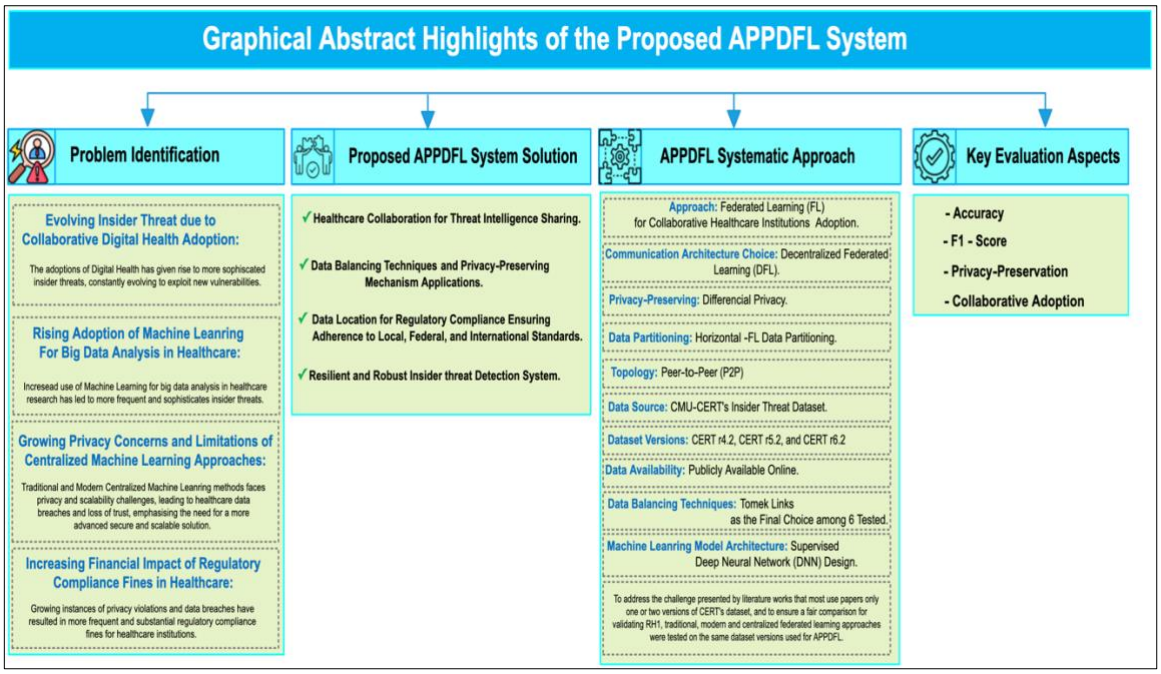


Figure A- 1. APPDFL System Graphical Abstract Summary.

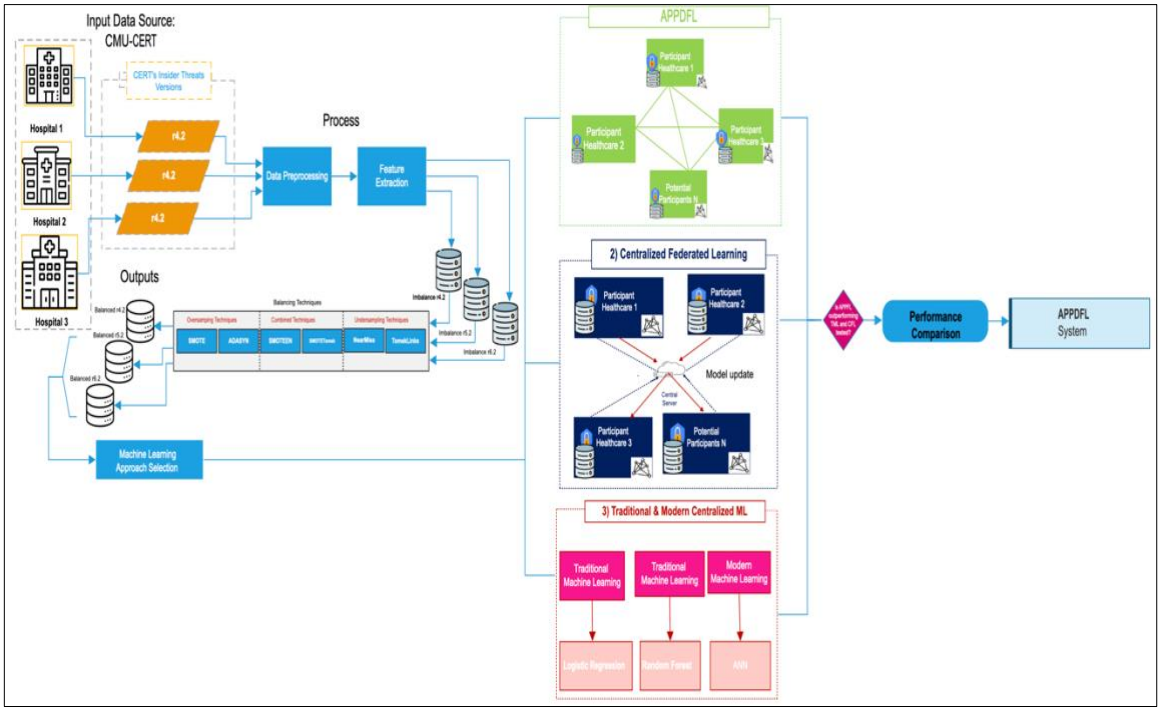


Figure A- 2. High level workflow diagram for the implementation process.

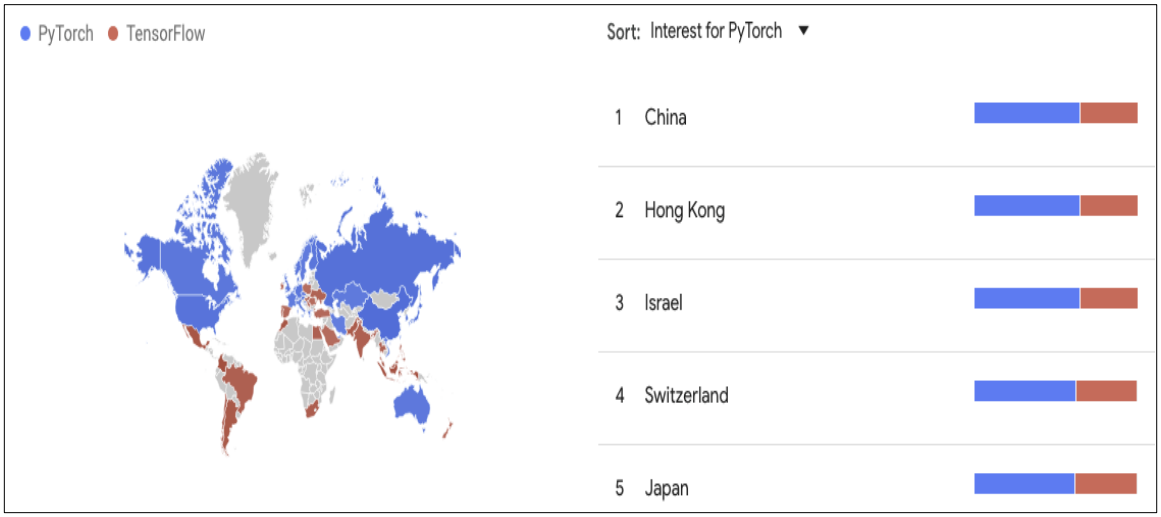


Figure A- 3. Interest for PyTorch vs TensorFlow by regions.

Source: (Google Trends, n.d.).

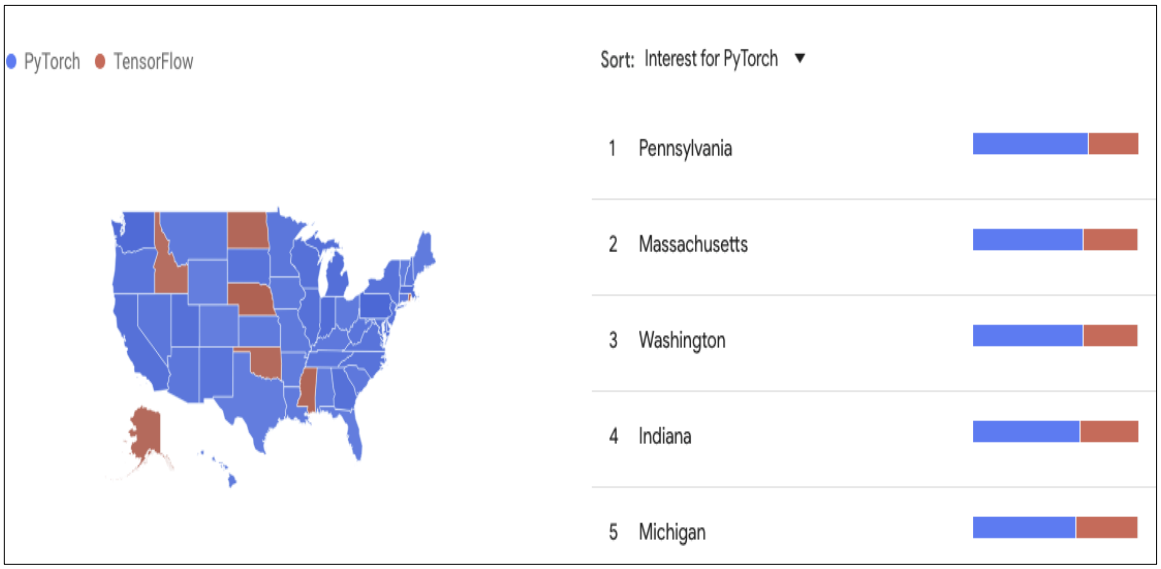


Figure A- 4. Interest for PyTorch vs TensorFlow by sub-region.

Source: (Google Trends, n.d.).

Table A- 1. Potential Consequences of Insider Threat in Healthcare Institutions.

Impacts	Categories	Specifics	Examples
Impact on Patients	**Erosion of Privacy**	Unauthorized Entries	Patient records, Personal details
		Information Heists	Sensitive data, Medical histories
		Data Exposure	Confidential data, Identifiable information
	Health Detriments	Data Manipulation	Misdiagnosis, Improper treatment
		Delayed Medical Actions	Altered records, Incorrect data
	Care Disruptions	Access Complications	Medical records, Clinical data
		Service Interruptions	Medical procedures, Critical care services
	Psychological Effects	Mental Health Issues	Anxiety, Depression
		Stress Inducers	Fear of identity theft, Privacy concerns
Impact on the Institution	**Financial Repercussions**	Regulatory Penalties	HIPAA violations, Data protection breaches
		Breach Response Costs	IT repairs, Investigation expenses
		Revenue Declines	Reduced patient trust, Loss of patients
	Reputational Fallout	Loss of Patient Confidence	Trust erosion, Patient dissatisfaction
		Negative Publicity	Negative press, Public backlash
		Brand Damage	Reputation hit, Institutional credibility
	Operational Challenges	System Downtime	Clinical operations, Medical systems
		Service Disruptions	Scheduling, Billing, Pharmacy operations
		Increased Workload	IT teams, Security teams
	Legal Consequences	Litigations	Affected patients, Class action suits

		Regulatory Investigations	Compliance issues, Government scrutiny
		Legal Expenses	Court costs, Settlements
	Staff Morale and Efficiency	Workplace Atmosphere	Suspicious atmosphere, Stressful conditions
		Morale Decline	Decreased morale, Job dissatisfaction
		Productivity Issues	Reduced efficiency, Lower output
Public Health Impact	**Trust Erosion in Healthcare System**	Patient Reluctance	Reluctance to share information, Avoidance of care
		Public Perception Challenges	Negative outlook, Distrust in medical services
	Service Disruption	Community Impact	Service disruption, Healthcare availability issues
		Delivery Interruptions	Interruptions, Delays in healthcare service
	Spread of Misinformation or Disease	Data Tampering	Incorrect information dissemination, Public health data issues
		Misleading Communication	Health directive manipulation, Spread of misinformation
Additional Considerations	**Fraud and System Sabotage**	Fraudulent Activities	Insurance claims, Prescription misuse
		System Tampering	Unauthorized changes, Sabotage operations
		Medical Record Issues	Data alteration, Integrity issues
	Intellectual Property (IP) Theft	Research Theft	R&D investments, Loss of innovation
		Competitive Edge Loss	Unfair advantage, Competitiveness impact
	Regulatory Fines	Data Protection Penalties	Non-compliance fines, Regulatory penalties
		Privacy Breaches	Legal consequences, Financial impact
Incident Response Challenges	**Balancing Transparency and Privacy**	Transparency Efforts	Ensuring transparency, Data protection
		Disclosure Issues	Maintaining trust, Regulatory compliance
	Investigating Insider Threats	Complex Investigations	Insider knowledge, Sophisticated threats

		Detection Efforts	Identifying source, Preventing future breaches
	Collaborative Intelligence Sharing	Threat Detection Collaboration	Sharing practices, Collective threat detection
		Response Implementation	Network building, Coordinated response efforts
	Continuity of Care	Ensuring Continuous Care	Care assurance, Patient safety

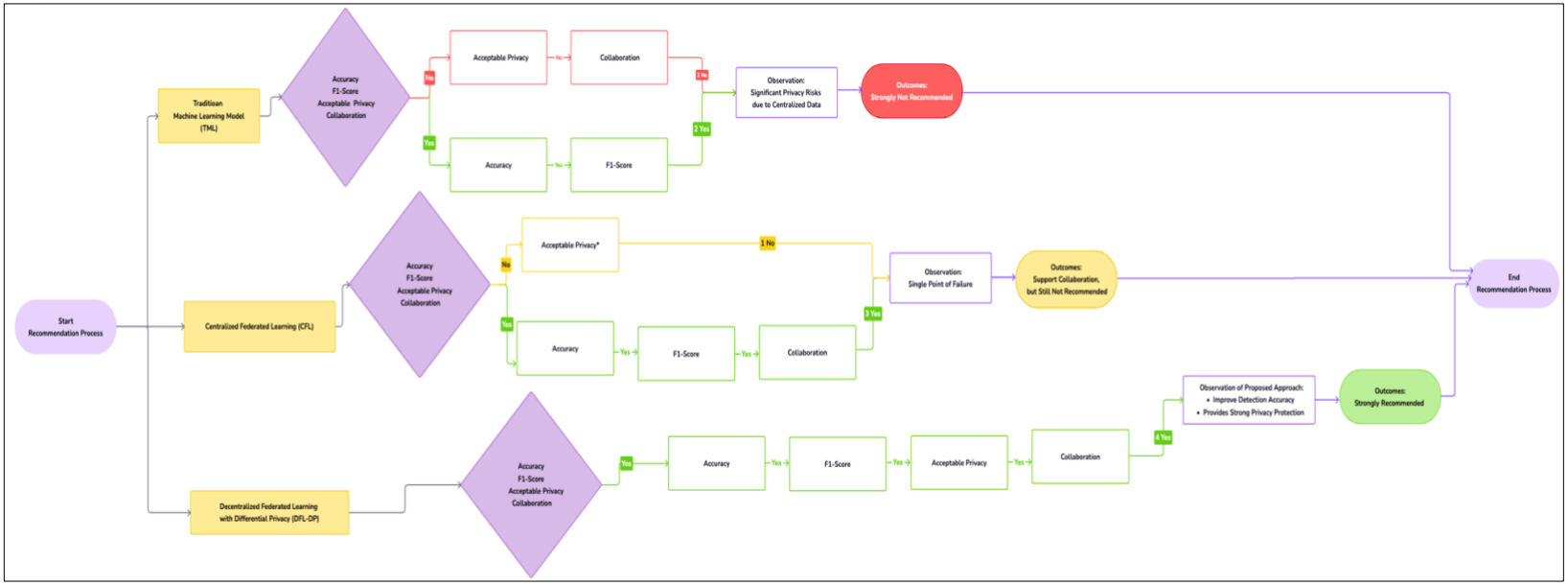


Figure A- 5. Comparative Analysis of Metrics and Strategies used in this Praxis.