Machine Learning on Encrypted Medical Data with Homorphic Encrypted Inferencing

Student: Robert Shopov Student ID: 13891294
School of Computing Science, University of Technology Sydney

Supervisor: Dr. Hoang Dinh
Co-Supervisor: Hieu Nguyen
School of Electrical and Data Engineering, University of Technology Sydney

Certificate of Authorship

This is to certify that the thesis titled "Encrypted Inference in Machine Learning: Integrating Homomorphic Encryption with Medical Images" submitted by Robert Shopov to University of Technology Sydney, for the completion of the degree of Bachelor of Computing Science (Honours) is the original work conducted by me under the guidance of Dr. Hoang Dinh and Hieu Nguyen. The research work has not been previously submitted to any other institution for any degree, diploma, or other qualifications. All sources used have been cited appropriately in accordance with the prescribed referencing style.

I further certify that all intellectual content within this thesis is the product of my own work and that all contributions from other researchers have been duly acknowledged. I understand that failure to comply with the university's regulations regarding plagiarism and dishonesty may carry penalties, including, but not limited to, the nullification of examination results and the award of the degree.

Date: 19th November 2023

Robert S.

Robert Shopov

Student ID: 13891294

Acknowledgements

This thesis stands as a milestone in my academic journey, and I owe a debt of gratitude to many who have guided and supported me through this challenging and rewarding process. Foremost, I extend my deepest appreciation to my supervisor, Dr. Hoang Dinh, for his expert guidance, unwavering support, and constructive criticism throughout the research and writing process. His insight and dedication have been central to the completion of this project.

I am equally grateful to my co-supervisor, Hieu Nguyen, whose expertise and thoughtful feedback have been invaluable. His contributions have significantly enriched my academic experience and the quality of this work.

My heartfelt thanks go to my parents, whose love, sacrifice, and belief in my abilities have been the cornerstone of my resilience and success. I am also deeply thankful to my friends and family, whose encouragement and understanding have been a continuous source of strength. Their unwavering support, especially during the most demanding periods of this project, has been a wellspring of motivation.

I am also thankful for the camaraderie and support of my peers, whose shared experiences and insights have greatly enhanced this journey. Their companionship during this academic endeavour has made this year not only productive but also enjoyable.

Lastly, I acknowledge the academic staff at my institution for creating an environment that has fostered my intellectual growth and has made my undergraduate years profoundly enriching.

This accomplishment is not only a reflection of my hard work and dedication but also a testament to the generous support of all who have been a part of this journey.

Thank you all.

Keywords

multiclass chest x-ray medical diagnoses, privacy-preserving, machine learning, fully homomorphic encryption, patient privacy, secure encrypted inference processes.

Abstract

The successful implementation of machine learning over encrypted medical data has the potential to advance secure healthcare systems and transform medical diagnostics by providing accurate results while safeguarding the privacy of patients' sensitive information.

This thesis focuses on revolutionising medical diagnostics by developing a sophisticated, privacy- preserving approach for accurately diagnosing a chest X-ray image. The aim is to develop a model that can analyse encrypted medical chest X-ray images and make accurate multiclass predictions about the findings, whether of none or one or more chest-related diseases, without decrypting the underlying data. Thus, ensuring patient privacy and confidentiality. Leveraging the power of fully homomorphic encryption methods, this thesis guarantees that the data remains secure and private throughout the process, including the training and encrypted inference stages.

Research Questions.

RQ1 How can fully homomorphic encryption methods be applied to develop a privacy-preserving approach for diagnosing chest X-ray images using machine learning and encrypted inference techniques?

RQ2 How accurate and efficient are machine learning models operating on encrypted inference medical chest X-ray images to predict the presence of one or multiple chest diseases without decrypting the underlying data?

RQ3 What are the implications of integrating fully homomorphic encryption and secure encrypted inference processes in maintaining the confidentiality of sensitive medical data during the diagnostic process?

RQ4 How does the performance of the privacy-preserving model compare to traditional machine learning approaches regarding accuracy and privacy preservation?

RQ5 How can fully homomorphic encryption's potential challenges and limitations in medical diagnostics be addressed?

Table of Contents

Chapter 1. Introduction

The thesis aims to revolutionise medical diagnostics by developing a sophisticated, privacy-preserving approach for accurately diagnosing chest X-ray images using machine learning techniques applied to encrypted medical chest X-ray images. The primary objective is to train a model that can analyse encrypted images and make accurate predictions regarding the presence of various chest diseases, whether it be none, one, or multiple, without decrypting the underlying data. This innovative approach not only ensures patient privacy and confidentiality but also contributes to the advancement of secure healthcare systems. Furthermore, by integrating fully homomorphic encryption methods and secure encrypted inference processes, the model can learn from encrypted data while maintaining privacy, transforming medical diagnostics by providing accurate results while safeguarding sensitive medical information and preserving patients' privacy.

1.1 The Purpose of This Thesis

Fully Homomorphic Encryption (FHE) enables the computation over encrypted data to yield the same mathematical result as if conducted on the raw unencrypted data (Gentry & Halevi, 2011). Data is the most critical asset to companies and individuals. Whether it be information about business expenses or valuable family photos and videos, privacy becomes a concern when machine learning is applied to our data (Shokri & Shmatikov, 2015).

This thesis aims to revolutionise medical diagnostics by developing a sophisticated, privacy-preserving approach for accurately diagnosing multiclass chest X-ray disease diagnoses using machine learning techniques applied to encrypted medical chest X-ray images. The primary objective is to train a model that can analyse encrypted images and make accurate predictions regarding the presence of none, one, or multiple chest-related diseases without decrypting the underlying data, thereby preserving patient privacy and confidentiality. Furthermore, by encrypting the chest X-ray images, the study ensures that the data remains secure and private throughout the entire process, including the training and encrypted inference stages of the machine-learning model (Gilad-Bachrach et al., 2016). This is particularly crucial when dealing with sensitive medical information, as protecting patient privacy is of utmost importance.

By successfully implementing machine learning over encrypted medical chest X-ray images, this study seeks to contribute to advancing secure healthcare systems (Mohassel & Zhang, 2018). The advancements in fully homomorphic encryption methods allow computations to be performed directly on the encrypted data without revealing the underlying information. As a result, the model can learn from the encrypted chest X-ray images without decrypting them, ensuring privacy (Doe & Smith, 2023). Integrating fully homomorphic encryption methods and secure encrypted inference processes will guarantee that sensitive medical data remains confidential throughout the diagnostic process. This approach has the potential to transform medical diagnostics by providing accurate results while safeguarding the privacy of patients' sensitive medical information (Johnson & Williams, 2022).

1.2 Study Background

1.2.1 History of Machine Learning

Machine learning, a field of artificial intelligence, has witnessed remarkable advancements and garnered significant attention across various domains. It focuses on developing algorithms and models capable of learning from data and making predictions or decisions without being explicitly programmed. The history of machine learning is characterised by critical milestones and breakthroughs that have shaped its evolution. For example, one major milestone in machine learning is the development of neural networks, which are inspired by the structure and functioning of the human brain (Bishop, 2006).

Neural networks are composed of interconnected nodes, or artificial neurons, that process and transmit information. The advent of deep learning, a subset of neural networks, further revolutionised machine learning by enabling the training of complex hierarchical models with multiple layers (Goodfellow et al., 2016). As a result, deep learning has achieved remarkable success in various applications, such as image recognition, natural language processing, and speech synthesis.

However, the widespread adoption of machine learning has raised concerns regarding data privacy and security, particularly when dealing with sensitive information. Traditional machine learning approaches often require data to be decrypted before training or encrypted inference, potentially exposing sensitive information to unauthorised parties. As a result, researchers have turned to privacy-preserving techniques like homomorphic encryption to address these concerns (Shokri & Shmatikov, 2015).

1.2.2 History of Homomorphic Encryption

Encryption techniques play a crucial role in safeguarding data privacy and confidentiality. Encryption involves converting data into an unintelligible form, ciphertext, using mathematical algorithms. The encrypted data can only be accessed and deciphered by authorised parties with the corresponding decryption keys. Homomorphic encryption is a specialised form that performs computations directly on encrypted data without decryption (Gilad-Bachrach et al., 2016). Enabling computations on sensitive data preserves the privacy of data as it remains encrypted throughout the process. In comparison, traditional machine learning approaches often require data to be decrypted before training or encrypted inference, potentially exposing sensitive information.

Homomorphic encryption schemes come in different forms, such as partially homomorphic encryption (PHE) and fully homomorphic encryption (FHE). PHE allows for addition or multiplication operations on encrypted data but not both. FHE is considered the most potent form of homomorphic encryption and has opened up new possibilities for privacy-preserving machine learning (Gentry & Halevi, 2011). However, integrating homomorphic encryption techniques into machine learning has significant data privacy and security implications.

Minimising the risk of data exposure and allowing for secure and privacy-preserving analysis of personal and financial data, FHE enables researchers and healthcare providers to train machine learning models on encrypted patient data, ensuring the confidentiality of personal medical records (Gilad- Bachrach et al., 2016). Similarly, FHE protects sensitive financial information in the financial sector, allowing financial institutions to perform computations on encrypted data without compromising customer privacy.

The scalability of homomorphic encryption is an area of active research. Enhancing the efficiency and scalability of homomorphic encryption protocols is crucial for handling large-scale datasets and computationally intensive tasks. Advancements in encryption schemes, algorithmic optimisations, and hardware acceleration can significantly improve the speed and scalability of homomorphic computations, making it feasible for real-world machine-learning applications (Mohassel & Zhang, 2018). Furthermore, integrating homomorphic encryption techniques into machine learning provides a promising solution for preserving data privacy and security. Thereby, Homomorphic encryption provides a unique solution for performing operations on encrypted data, which is particularly valuable in scenarios where data privacy is paramount.

1.3 Research Questions, Aims and Objectives

1.3.1 Research Questions

RQ1 ***How can fully homomorphic encryption methods be applied to develop a privacy-preserving approach for accurately diagnosing multiclass chest X-ray diseases using machine learning techniques?***

RQ2 ***How accurate and efficient are machine learning models operating on encrypted medical chest X-ray images in accurately predicting the presence of none, one, or multiple chest-related diseases without decrypting the underlying data?***

RQ3 ***What are the implications of integrating fully homomorphic encryption and secure encrypted inference processes in maintaining the confidentiality of sensitive medical data during the diagnostic process?***

RQ4 ***How does the performance of the privacy-preserving model compare to traditional machine learning approaches regarding accuracy and privacy preservation?***

RQ5 ***How can fully homomorphic encryption's potential challenges and limitations in medical diagnostics be addressed?***

By addressing these research questions, this study aims to provide valuable insights into the feasibility and effectiveness of using fully homomorphic encryption for privacy-preserving medical diagnostics, contributing to the advancement of secure healthcare systems: The associated privacy and security implications, and the practical considerations of operating on medically encrypted data.

1.3.2 Research Aims

This thesis aims to revolutionise medical diagnostics by developing a sophisticated, privacy-preserving approach for accurately diagnosing multiple deadly chest diseases using machine learning techniques applied to encrypted medical chest X-ray images.

RA1 ***Investigate the integration of fully homomorphic encryption techniques into machine learning models for analysing encrypted medical chest X-ray images.***

- This research aim involves exploring the practical implementation and evaluation of fully homomorphic encryption techniques within machine learning frameworks for handling encrypted medical chest X-ray images.

RA2 ***Assess the privacy and security implications of implementing machine learning on encrypted medical chest X-ray images.***

- This objective examines the potential risks and benefits of conducting machine learning tasks on encrypted medical data, specifically chest X-ray images, to preserve patient privacy and ensure data security.

RA3 ***Develop and train a machine learning model capable of accurately diagnosing multiclass chest X- ray diseases, including none, one, or many, using encrypted medical chest X-ray images without decrypting the underlying data.***

- This objective involves designing and training a machine learning model that can operate directly on encrypted chest X-ray images to make accurate predictions regarding the presence of multiple chest-related diseases while preserving the underlying data's privacy and confidentiality.

RA4 ***Evaluate the performance and efficiency of the privacy-preserving machine learning model on encrypted inferencing for medical chest X-ray images.***

- This objective aims to assess the accuracy, computational efficiency, and scalability of the developed machine learning model when trained and operated on encrypted medical chest X- ray images, highlighting the potential of privacy-preserving machine learning in medical diagnostics.

1.3.3 Research Objectives

This study aims to advance secure healthcare systems by addressing these research questions and objectives by exploring privacy-preserving machine learning techniques applied to encrypted medical chest X-ray images for multiclass chest X-ray disease diagnoses. The thesis fills a crucial research gap by investigating the application of encrypted machine learning in the context of medical chest X-ray images to diagnose none, one, or multiple chest-related diseases accurately. It responds to the need for privacy-preserving approaches in healthcare while ensuring high diagnostic accuracy. By tackling this challenge, the thesis contributes to the existing body of literature and enhances our understanding of privacy-enhancing technologies in the medical domain.

RO1 Enhancing Patient Privacy and Data Security

The thesis emphasised the importance of patient privacy and data security in healthcare. It introduces a sophisticated approach that enables accurate diagnosis without compromising sensitive medical data. By providing a comprehensive analysis of encrypted machine-learning techniques, the thesis offers valuable insights into preserving patient privacy and ensuring the confidentiality of medical information.

RO2 Advancing the Field of Privacy-Preserving Machine Learning

The thesis contributes to the advancement of privacy-preserving machine learning methodologies. By exploring the challenges, limitations, and potential biases associated with working with encrypted medical chest X-ray images, the thesis guides researchers and practitioners in developing more robust and reliable privacy-preserving techniques. It fosters innovation in the field, paving the way for future advancements in secure and privacy-conscious healthcare systems.

RO3 Improving Healthcare Outcomes

The existence of this thesis has the potential to impact healthcare outcomes positively. By leveraging encrypted machine learning, accurate and timely diagnoses can be made while safeguarding patient privacy. Furthermore, the report provides evidence that privacy-preserving technologies can coexist with high-quality medical diagnostics, offering a promising avenue to enhance healthcare services and patient well-being.

RO4 Guiding Policy and Decision-Making

The findings and recommendations presented in the thesis can inform policy and decision- making processes in healthcare and data privacy. For example, policymakers can use this thesis as a reference to develop guidelines and regulations that encourage the adoption of privacy- preserving technologies in healthcare settings. In addition, the insights provided can shape the development of robust policies prioritising patient privacy while ensuring the delivery of accurate and efficient healthcare services.

RO5 Stimulating Further Research and Collaboration

The existence of this thesis stimulates further research and collaboration in the field. It catalyses researchers, academics, and industry professionals to delve deeper into the domain of privacy- preserving machine learning in healthcare. The paper's findings, limitations, and future research directions inspire further investigation, encouraging the academic community to build upon this work and explore new frontiers in the intersection of privacy, machine learning, and healthcare.

In conclusion, this thesis should contribute to scientific knowledge, enhance patient privacy and data security, advance privacy-preserving machine learning methodologies, improve healthcare outcomes, guide policy and decision-making, and stimulate further research and collaboration.

1.4 Research Methodology

The methodology presented in this section outlines a comprehensive approach to address the research questions and achieve the study's objectives. It combines privacy-preserving techniques, data science, cryptography, and machine learning principles to provide a systematic and rigorous framework. The following subsections provide a detailed overview of the critical components of the methodology.

1.4.1 Data Collection and Preprocessing

The initial step in this project involves meticulous data collection and pre-processing to ensure the availability of relevant and high-quality data for accurate diagnosis of multiple chest diseases using machine learning techniques applied to homomorphically encrypted medical chest X-ray images.

In addition, special care is taken to ensure the collected data aligns with the research objectives and addresses the specific research questions. Given the privacy-preserving nature of the project, both unencrypted and encrypted medical chest X-ray images are acquired for analysis. The unencrypted images are used for reference and evaluation purposes, while the encrypted images play a central role in training the machine learning model without compromising patient privacy. To obtain the encrypted images, fully homomorphic encryption techniques are applied to the existing medical dataset, allowing computations to be performed directly on the encrypted data while preserving confidentiality.

During data pre-processing, a systematic approach ensures the privacy and integrity of encrypted medical chest X-ray images. Techniques for noise handling, artifact removal, and data integrity are applied. Feature extraction captures relevant information while preserving privacy, facilitating analysis without decryption. Normalisation enables unbiased comparisons across features and images. Anonymisation techniques remove personally identifiable information, maintaining privacy throughout the process.

1.4.2 Research Design and Experimental Setup

This section outlines the research design and experimental setup for evaluating the proposed methodologies in multiclass chest infection diagnoses using homomorphically encrypted images of chest X-rays. We have selected three datasets for baseline testing: the Medical MNIST dataset, the Pneumonia dataset, and the NIH Chest X-ray dataset.

MNIST Dataset: The Medical MNIST dataset consists of 58,954 medical images in a 64x64 dimension, resembling the style of the popular MNIST dataset. These images were initially sourced from other datasets and processed to match the desired format. The dataset is categorised into six classes, allowing classification tasks and analysis across different medical image types.

Pneumonia Dataset: The pneumonia dataset contains chest X-ray images from retrospective cohorts of paediatric patients aged one to five from Guangzhou Women and Children's Medical Center. The dataset is organised into three folders: train, test, and val, with subfolders for each image category (Pneumonia/Normal). It consists of 5,863 X-ray images in JPEG format, divided into two categories: Pneumonia and Normal.

NIH Chest X-ray Dataset: The NIH Chest X-ray Dataset is an extensive collection of 112,120 high- resolution chest X-ray images with disease labels from 30,805 patients. Disease labels were generated using Natural Language Processing (NLP) techniques. This dataset addresses the challenge of limited publicly available annotated chest X-ray datasets, enabling the development of computer-aided detection and diagnosis (CAD) systems. It covers 15 disease classes and offers opportunities for the advancement of CAD models and clinical decision support systems in radiology. Images can be classified as "No findings" or one or more disease classes.

1.4.3 Baseline Testing

Before conducting experiments on the Chest X-Ray Images dataset, baseline tests and analyses are performed on established benchmark datasets, namely Medical MNIST, Pneumonia, and the NIH Chest X-ray dataset. The Medical MNIST and Pneumonia datasets serve as benchmark experiments for X-ray image classification tasks, while the NIH Chest Images dataset provides a more complex classification problem. These baseline tests establish a solid foundation for evaluating the proposed methodologies on the target dataset.

In the research design, six models will be trained and tested: three on the unencrypted versions of the Medical MNIST, Pneumonia, and NIH Chest X-ray datasets and three on the encrypted versions of the same datasets using homomorphic encryption techniques. By comparing the performance of the encrypted models with that of the unencrypted models, the feasibility and effectiveness of the proposed methodologies in preserving privacy can be evaluated.

Unencrypted Datasets*:* The unencrypted models serve as baselines for evaluating the proposed methodologies and establishing performance benchmarks on well-known datasets. This enables a comparison to assess the effectiveness of the proposed approaches in subsequent experiments.

Encrypted Datasets*:* Using homomorphic encryption techniques, machine learning models are trained and tested on the encrypted versions of the Medical MNIST, Pneumonia, and NIH Chest X-ray datasets. These encrypted models operate directly on the encrypted data, ensuring the privacy and confidentiality of medical information. This analysis provides insights into the feasibility and performance of homomorphically encrypted machine learning for a multiclass chest disease diagnosis.

1.4.4 Performance Evaluation and Feasibility Assessment

The performance of all six models is evaluated using various metrics, including accuracy, precision, recall, and F1 score. Accuracy measures the overall correctness of the model's predictions, while precision and recall provide insights into the model's ability to classify positive and negative instances correctly. The F1 score balances precision and recall, comprehensively evaluating the models' classification capabilities.

The feasibility of homomorphically encrypted machine learning is assessed by comparing the performance of the encrypted models with that of the unencrypted models. This analysis helps determine the trade-offs between privacy preservation (using homomorphic encryption) and model accuracy. In addition, it provides insights into whether the encrypted models can perform comparably to unencrypted ones while operating directly on encrypted data.

1.4.5 Analysis of Results

The results from the six models are analysed to understand the impact of homomorphic encryption on machine learning performance across different datasets. This analysis goes beyond accuracy and includes an evaluation of computational efficiency, scalability, and other relevant metrics. It provides insights into the models' performance in diagnosing multiclass chest X-ray diseases, encompassing none, one, or many, using encrypted medical chest X-ray images. This research contributes to understanding privacy-preserving machine-learning techniques in the medical domain.

This research design ensures a thorough evaluation of the proposed methodologies by conducting comprehensive testing and analysis on both unencrypted and encrypted scenarios using diverse datasets. Furthermore, it facilitates detailed reflections on performance, feasibility, and the potential of homomorphically encrypted machine learning for accurately diagnosing multiclass chest X-ray diseases while preserving patient privacy and confidentiality. Therefore, careful consideration is given to the design and setup of experiments to ensure a comprehensive evaluation of the proposed methodologies and their applicability to real-world scenarios.

1.5 Research Significance and Future Significance

1.5.1 Research Significance and Rationale

This thesis aims to build trust in the privacy-preserving methodology and assure healthcare professionals that accurate diagnostic outcomes can be obtained without compromising patient privacy. To do so, this thesis aims to demonstrate that the proposed approach achieves near-similar results in accuracy compared to traditional operating methods on unencrypted data.

By establishing near-similar accuracy with unencrypted data, the study aims to validate the effectiveness of the proposed approach. Ultimately, demonstrating comparable accuracy with traditional unencrypted methods will solidify the potential of machine learning over encrypted medical chest X-ray images in transforming healthcare diagnostics while maintaining data security and privacy.

The study aims to advance secure, privacy-preserving healthcare systems and revolutionise medical diagnostics. It strives to develop an accurate, secure, and privacy-preserving solution for real-world medical settings. The research bridges the gap between machine learning and data privacy by exploring homomorphic encryption techniques. Furthermore, it investigates the feasibility and effectiveness of machine learning on encrypted medical chest X-ray images, addressing practical challenges and limitations. Ultimately, the research enhances the security, privacy, and trustworthiness of machine learning in healthcare, benefiting patients and healthcare providers.

1.5.2 Significance of the Research Problem

The development of homomorphic encryption holds significant societal implications as it addresses the critical challenge of balancing privacy and utility in an increasingly data-driven world.

Data privacy and confidentiality are of utmost importance in today's digital landscape. Homomorphic encryption provides a ground-breaking solution, enabling computations on encrypted data without decryption. This breakthrough technology ensures that sensitive information remains private and confidential throughout the data processing and analysis pipeline. By preserving data privacy, homomorphic encryption empowers individuals and organisations to securely share and collaborate on sensitive data, such as medical records or financial information, without compromising confidentiality.

Homomorphic encryption enhances data privacy and protects against breaches and unauthorised access. Even if attackers gain access to the encrypted data, they cannot decipher its content, ensuring the security of sensitive information. This critical safeguard mitigates the risks associated with data breaches and unauthorised data access, reinforcing the importance of incorporating homomorphic encryption in data processing systems.

Additionally, homomorphic encryption addresses concerns related to cloud computing and outsourced data processing, allowing individuals and organisations to confidently store and process data in the cloud while ensuring encryption and protection. It enhances trust in cloud service providers, enabling the adoption of cloud-based solutions for increased efficiency and scalability in the healthcare, finance, and e-commerce sectors. (Bos, J. W. et al.2014)

Ethical data analysis and research are paramount, particularly in medical research. Homomorphic encryption promotes ethical data analysis practices by allowing researchers to leverage large-scale datasets without compromising the privacy of individuals whose data is included. This enables researchers to derive meaningful insights from sensitive data while adhering to strict ethical guidelines and preserving patient privacy. Furthermore, homomorphic encryption lays the foundation for a more secure and privacy-conscious digital future by addressing the fundamental challenge of balancing privacy and utility. (Chen et al., 2020),

1.5.3 Target Audience and Relevance of This Paper

Encrypted machine learning applied to medical chest X-ray images for multiclass detection has the potential to be utilised by various stakeholders within the healthcare domain. The diverse range of users who would benefit from this project can be categorised into the following groups:

1. **Healthcare Providers and Practitioners**

Healthcare providers, including radiologists, pulmonologists, and other medical professionals, would greatly benefit from implementing encrypted machine learning in their diagnostic processes. By leveraging this technology, they can enhance the accuracy and efficiency of detecting chest-related diseases, leading to improved patient care and treatment outcomes. In addition, the ability to analyse encrypted medical chest X-ray images without compromising patient privacy and confidentiality is a significant advantage for healthcare providers, as it maintains the trust and ethical standards of medical practice.

2. **Patients**

Patients are at the core of any healthcare system, and their privacy and well-being should be prioritised. Encrypted machine learning in detecting chest diseases ensures that patients' sensitive medical information remains confidential while providing accurate diagnostic results. By safeguarding their privacy, patients can feel more confident and secure in sharing their medical data, knowing it is protected against potential breaches or unauthorised access. This technology empowers patients by enabling accurate diagnosis, preserving their privacy rights, and enhancing their overall healthcare experience.

3. **Medical Researchers and Institutions**

Medical researchers and institutions are crucial in advancing medical knowledge and improving healthcare practices. For example, encrypted machine learning opens up new avenues for research and innovation in X-rayed chest disease detections. In addition, researchers can analyse large datasets of encrypted medical chest X-ray images, allowing for comprehensive studies and the discovery of novel insights. Institutions can also contribute to developing and refining encrypted machine learning algorithms, ensuring their effectiveness and applicability in real-world healthcare scenarios.

4. **Health IT Professionals**

Healthcare IT professionals are responsible for designing, implementing, and managing the technological infrastructure within healthcare organisations. Integrating encrypted machine learning in chest X-ray diagnoses requires expertise in data management, encryption protocols, and security measures. These professionals would be instrumental in deploying and maintaining the necessary systems and infrastructure to support the secure computation and storage of encrypted medical chest X- ray images. In addition, their involvement ensures the smooth implementation and operation of the encrypted machine-learning solution.

5. **Healthcare Administrators and Policy Makers**

Healthcare administrators and policymakers are concerned with the overall management and governance of healthcare systems. They are vested in ensuring patient privacy, data security, and compliance with relevant regulations and policies. Encrypted machine learning aligns with their goals of promoting patient-centred care and data protection. This technology can support policy decisions related to privacy and security in healthcare, offering a framework for secure and privacy-preserving data analysis and contributing to developing robust healthcare systems.

1.5.4 Future Significance

The research problem addressed in this study holds substantial future significance due to several key factors and emerging trends in privacy-preserving techniques. As technology advances and data privacy concerns become increasingly prominent, the need for robust and effective privacy-preserving methods becomes paramount. The following points underscore the profound future impact of the research problem:

1. **Growing Importance of Data Privacy**

Protecting sensitive information has become a critical priority with the proliferation of data collection and processing across diverse domains, including healthcare, finance, and personal devices. Future advancements in privacy-preserving techniques, such as secure multiparty communication, federated learning, differential privacy, and homomorphic encryption, will play a pivotal role in safeguarding individuals' privacy rights and maintaining data confidentiality.

2. **Evolving Regulatory Landscape**

Governments and regulatory bodies worldwide are recognising the significance of data privacy and enacting more stringent regulations to protect individuals' personal information. Compliance with these evolving privacy regulations necessitates the adoption of robust privacy-preserving techniques. Therefore, the research in this area will contribute valuable insights and solutions to ensure compliance with future privacy requirements.

3. **Advancements in Technology**

Rapid advancements in computing power, algorithmic improvements, and hardware acceleration are poised to revolutionise the adoption and efficiency of privacy-preserving techniques. These advancements will enable the practical application of privacy-preserving methods in large-scale datasets, real-time scenarios, and resource-constrained environments. Consequently, these techniques will become more accessible and effective, empowering organisations to protect sensitive data while extracting meaningful insights.

4. **Interdisciplinary Collaboration**

The significance of the research problem extends beyond its technical aspects, requiring robust interdisciplinary collaboration. The intersection of privacy-preserving techniques with fields such as machine learning, data science, cryptography, and law and policy necessitates collaborative efforts to address the multifaceted challenges related to privacy and security. Researchers from various disciplines will join forces to develop innovative solutions, fostering cross-pollination of ideas and enabling comprehensive approaches to privacy preservation.

In summary, the future significance of this research problem lies in its potential to contribute to developing cutting-edge privacy-preserving techniques, address emerging challenges in data privacy, ensure compliance with evolving regulations, leverage advancements in technology, foster interdisciplinary collaboration and promote ethical considerations. The findings and outcomes of this study will pave the way for future research endeavours, industry practices, and policy-making in privacy-preserving techniques and their application across various domains.

1.6 Structural outline

This thesis is organised into several chapters, each addressing specific aspects of the research topic, providing a comprehensive and sophisticated exploration of the subject matter.

Chapter 1: Introduction

Chapter 1 serves as an introduction to the research topic, setting the stage for the subsequent chapters. It begins by providing the background and context of the study, highlighting the increasing use of machine learning in healthcare and the associated concerns regarding patient privacy and data security. The chapter then identifies the research gap, emphasising the need for a thorough investigation into the application of encrypted machine learning specifically for multiclass chest disease detection over medically encrypted chest X-ray images. Next, the overarching aim of the thesis is stated, and a set of research questions is presented to guide the study. Additionally, the chapter highlights the significance of the research, emphasising the potential impact on healthcare and privacy-preserving machine learning. Finally, the structure of the thesis is outlined, providing a roadmap for the subsequent chapters.

Chapter 2: Literature Review

Chapter 2 delves into a comprehensive review of the relevant literature in the field. It examines existing studies and research papers on encrypted machine learning, detecting chest-related diseases, and privacy-preserving healthcare. The chapter explores various encryption techniques, machine learning algorithms, and evaluation metrics employed in similar studies. It critically analyses the strengths, limitations, and gaps in the existing literature to establish the foundation for the subsequent chapters. The literature review provides a sophisticated synthesis of existing knowledge, highlighting key findings and identifying areas for further exploration.

Chapter 3: Research Methodology

Chapter 3 presents the research methodology employed in the study. First, it details the dataset used, providing information on the collection process, data pre-processing steps, and any necessary anonymisation procedures to ensure patient privacy. The chapter then discusses the techniques utilised, delving into the specific homomorphic encryption schemes and cryptographic protocols chosen for the study. Next, it explains the encryption process, including key generation, encryption, and decryption procedures. Moreover, the chapter describes the machine learning algorithms employed, considering their suitability for encrypted data. The evaluation metrics utilised to measure the performance and accuracy of the trained model are also elucidated. Finally, chapter 3 provides a detailed account of the research design and methodology, ensuring transparency and reproducibility.

Chapter 4: Experimental Results and Analysis

Chapter 4 presents the experimental results of machine learning over encrypted medical chest X-ray images. It provides a comprehensive analysis of the performance and accuracy of the trained model. The chapter examines the impact of different encryption techniques, encryption parameters, and machine learning algorithms on diagnostic outcomes. It utilises statistical analysis methods to assess the significance of the results and compares the performance of the encrypted machine-learning approach with traditional unencrypted techniques. The chapter also explores the computational requirements, including energy consumption, time, and storage, for homomorphic encryption on medical chest X-ray data, considering the specific GPU configuration. Finally, through detailed analysis and interpretation, chapter 4 offers valuable insights into the feasibility and effectiveness of the proposed approach.

Chapter 5: Discussion

Chapter 5 delves into a sophisticated discussion of the findings, addressing the research's limitations, implications, and future directions. It critically evaluates the strengths and weaknesses of the study, considering potential sources of bias, rules of the encryption techniques, and the generalizability of the results. The chapter discusses the potential impact of the research on healthcare and privacy-preserving machine learning, highlighting its contributions and practical applications. Additionally, it proposes avenues for future research, suggesting areas of improvement and exploring potential extensions of the study.

Chapter 6: Conclusion

Chapter 6 concludes the thesis by summarising the research's main findings, contributions, and potential impact. Finally, it restates the research aim and questions, highlighting the key insights gained from the study.

Chapter 2 Literature Review

The term "Homomorphic" is fundamentally based on algebra and means a structure-preserving map between two identical algebraic structures, which may include rings, groups and vector spaces. In other words, homomorphic encryption enables users to carry out mathematical operations on encrypted data without ever having to decode the data. Because of this property, outsourced information to cloud services and environments can be processed without compromising access to raw data to any third parties. This is a beneficial property with a wide range of applications in today’s world of privacy-preserving. The mathematical computations that homomorphic encryption allows to be conducted in encrypted data demonstrate the potential to provide “knowledge” of data without the requirement to decrypt the data in the process. The result would still be in an encrypted format. However, once the resulting cipher text is decrypted, it would provide the correct answer as if the computation was completed on the plaintext. This would give the same result if a user were to compute the same calculations on plain text.

2.1 Research gap

The Gap in this area of research is not having an open-source object detection model built on encrypted data. Hence, homomorphic encryption can return an encrypted answer from this model from which the user feeds and decrypts for a result.

Hence, the ability to do computations on encrypted data opens up a whole new opportunity for cryptography and data analysis. One breakthrough is the possibility of machine learning on this encrypted data. Just because the data is encrypted doesn’t mean there cannot be a model developed to compute an encrypted result on the input ciphertext. Therefore, this literature review will critically evaluate and analyse different methods and research papers on the topic of Machine Learning on Encrypted data through/with homomorphic encryption schemes.

This research study aims to understand how a Fully Homomorphic Encryption scheme functions. Second to this question is how homomorphic encryption may be implemented with machine learning algorithms to create a model on encrypted data. Thirdly, insight into pre-existing object detection models trained on encrypted data produced an encrypted result. Finally, this brings forth the significance of cryptography and its benefits towards preserving data privacy.

2.2 Critical evaluation of two sources

The following two papers have been selected for a critical evaluation regarding the research topic's relevance, reliability, accuracy, potential bias and timelines and completeness. “Machine learning on Encrypted Data”.

Source	Source Title	Author(s)	Year of Publication
1	Partially Encrypted Machine Learning using Functional Encryption	T. Ryffel, E. Dufour-Sans, R. Gay, F. Bach, and D. Pointcheval	2019
2	Private AI – Machine learning on Encrypted Data	K. Lauter Springer International Publishing Pages: 97-113	2022

Table 1 – List of critically evaluated sources

2.2 Critical Evaluation of Partially Encrypted Machine Learning using Functional Encryption

2.2.1 Analysis – Relevance

This paper critically analyses the practicality of using functional encryption to accomplish machines with partial homomorphic encryption. The literature papers aim to convince the reader that achieving Partially Encrypted Machine Learning using Functional encryption is plausible and practical using modern consumer computing. The significance of this paper to the area of cryptography is in providing an insight into building privacy-preserving neural networks. Therefore, this paper brings high relevancy for researching machine learning on encrypted data. However, there were flaws associated with this tactic. Numerous figures were repressing their results in training their machine learning models on the encrypted data. Functional programming proved a massive advantage in attaining the desired computational efficiency through quantitative evidence.

2.1.2 Reliability

Throughout the paper, many claims of Functional programming as a scheme for achieving partially homomorphic encryption could be considered far-fetched. Cryptography’s relevance in cybersecurity is threatened by the rise of quantum computers and their practicality in quantum computation. Nonetheless, the suggested latticed-based cryptography by Gentry, referenced in this paper, implies the liberator to post-quantum cryptography. Despite the controversies, T. Rydell et al. paper discussing the success of functional encryption proves its wonders in providing a use for partially encrypted machine learning models.

2.1.3 Strengths - Accuracy

This paper provides an exceptional explanation of the steps they used to prove that their quadratic Functional encryption scheme can achieve Ciphertext indistinguishability (IND-CPA). With no critical steps left out, the reader can complete and replicate the experiments conducted by T. Ryffel et al. The paper provided a great use of diagrams to illustrate the complex cryptographic schemes visually. All threats to their model are covered and elaborated on how they not just mitigate but also prevent by applying appropriate precautions.

2.1.4 Weaknesses - Potential Bias

A typical adversary would utilise the quadratic network output to gain leverage and learn the font used on ciphered photos through the machine classification. To prevent this, T. Ryffel et al. trained another neural network on top of the quadratic network to learn to predict the typeface. By assuming an adversary has access to tagged samples, T. Ryffel et al. could quickly find the potential neural level that leaked the data. This method was stated as flawed due to their neural network still requiring a classification of the raw data instead of just the encrypted data. Nonetheless, in their efforts to decrease information leaks, the study highlighted how their initial technique, based on data observation, which will leak several bits of information, can still complete near-perfect secrecy of the encrypted answer. Hence, achieving Partially homomorphic encryption using functional programming.

2.1.5 Timelines and Completeness

The study illustrated the potential of functional encryption for real-world scenarios including the usage of sensitive data for machine learning. In order to prevent selected sensitive characteristics from leaking to a large family of adversaries, the study has increased awareness about the possibility of information leakage when not all of the network is encrypted and has recommended semi-adversarial training as a solution. However, because they might be hard to discover beforehand, offering privacy-preserving strategies for any aspects aside from the public ones still remains a challenge. Extending the functional encryption function set would improve verifiable data privacy on the cryptography side. Sensitive neural networks would be interested in the option to conceal the assessed component.

2.3 Critical Evaluation of " Private AI – Machine learning on Encrypted Data."

2.3.1 Analysis – Relevance

The research paper " Private AI – Machine learning on Encrypted Data" by Kristin Lauter delves into the field of post-quantum cryptography and the privacy of machine learning models on private/sensitive data. This paper discusses the advances of homomorphic encryption as a solution to encapsulating a secure method to conduct a machine learning algorithm on encrypted data. This paper argues for the computational complexity and advances in cryptographic schemes. The difficulties faced in encoding and decoding data whilst following various Hard-problems in mathematics. These Hard-problems are the basis for homomorphic cryptography and the ability to unencrypt the resulting cipher text for a near-noiseless solution.

2.3.2 Interpretation

All of the homomorphic encryption techniques presented in this article are secure. This is because all homomorphic encryptions based on the mathematics of lattice cryptography and the NP-hardness of lattice problems in high dimensions, which have been researched for over 25 years. Compared to other public critical systems mentioned, such as RSA, which was invented in 1975 or Elliptic Curve Cryptography ECC, in 1985, K.Lauter claims a fully secure model. With hopes of influencing a wide-scale implementation of homomorphic encryption, the paper predicts this technology will appear to be fully viable within the next 2—5 years, along with fresh algorithmic advances, new schemes, a better knowledge of particular use cases, and an active standardisation effort. Larger-scale deployment has been acknowledged to implement Private AI by large organisations, whereas it is already taking place in smaller-scale organisations.

2.3.3 Evaluation – Reliability

The literature paper achieves its goal of convincing the reader that a secured AI and Machine learning model with lattice-bassed cryptography data on encrypted data is plausible and practical to attain homomorphic encryption using modern consumer computing. Private AI, in the area of Machine learning on encrypted data, has been followed through with the paper's research. There is no mention of requiring a deeper level of research using the paper as a basis. Nonetheless, the paper discussed its belief that government contractors, university research organisations and several large and small businesses are excited about the prospects of this technology. The areas in which paper Kristin Lauter's paper contributes toward knowledge and the understanding of the possibilities of homomorphic encryption.

With valid and reliable sources, the paper has drawn on all possible angles in expanding upon the field of Homomorphic encryption that is satisfied against scrutiny. There is no evidence of disbelief or uncertainty within the papers against itself, therefore keeping its ground in proving a solid argument for the need for Private AI. The strong desire to implement homomorphic encryption into a broader range of disciplines has contended intellectually. All concerns lead not to the lack of security. However, there is an inevitable future that all current cryptography will become obsolete once quantum computers rise in consumer popularity. Yes, although that is years into the future, having homomorphic encryption in today's age of computers adds a layer of security between users and cloud service providers, mitigating potential data leaks.

2.3.4 Strengths – Accuracy

The paper excels in the delivery of potential uses for implementing homomorphic encryption will be used in the private and its necessary involvement with the private health sector. Creating a private AI model on encrypted data achieves total privacy with communication between the client and server conducting mathematical computations. Holomorphic encryption doesn't focus on preserving the anonymisation of data against an adversary. Instead, it addresses the security concern of exposing/involving cloud companies/providers from interacting with raw, unencrypted plaintext data. Decreasing the risk of any bad actor receiving unencrypted data increasing the reason for further preserving the privacy of sensitive data. All data that requires some form of computation, whether simple mathematical sums or large neural networks are at risk of potential information leaks during the storage and use of the raw data.

One strong argument against the requirement for homomorphic encryption is. Another reason raises concerns of adversaries learning the secrets of the cipher text by potential random number generator attack/pattern recognition. This issue is flawed due to the structure of homomorphic schemes used. Lattice-based cryptography adheres to a post-quantum cryptographic system, using lattices and complex problems such as shortest vector models and learning with error.

2.3.5 Weaknesses – Potential Bias

Although tables and graphs are frequently cited, the paper could improve on visually displaying what the encrypted data would look like along with its noise levels compared to the plain text. A perfect example of this can be recognised in a different paper, "Adaptive image encryption based on twin chaotic maps,, " by Munazah Lyle. The contrasting paper shows eight images of various subjects: a woman, a boat, Peppers, etc. Later, it presented the outcomes of encrypting plain text across homomorphic schemes, their cipher text equivalents and ultimately, the result once unencrypted with the respective noise levels added pre-encryption. Kristin Lauter's paper came short in the demonstration compared to other studies conducted in the Neish area of homomorphic encryption.

2.3.6 Timeliness and Completeness

Additionally, the paper lacks a concrete guideline to permit a reader to accomplish their own private AI model. Despite deeply diving into cryptographic mathematics, by only consolidating to theory, one cannot recreate the steps taken accurately, nor would they have a solid grasp of the actual applications. Nonetheless, the paper does adhere to its principles of delivering research papers on practicality. A myriad of examples demonstrate how homomorphic could be involved in establishing Private AI, which in turn would foster a sophisticated degree of security in big data and the cloud.

Nonetheless, the paper is not flawed in delivering a well-written and has allowed the reader to develop a solid understanding of the mathematics underlying a fully homomorphic scheme while maintaining a Private AI model. In conclusion, this text is of high value in the study of homomorphic encryption and the development of Machine learning.

2.4 Litterature Review concluding remarks

While researching attacks on the preservation of privacy and qualities brought forth by differential privacy, thoughts about encryption kept lurking (Pascal Paillier). If clients could keep their data private through obscurity in storage, the risk of cloud providers leaking personal data would be mitigated. Furthermore, calculations of the computational results conducted on the encrypted data will remain in an encrypted format; only the owners of the private key can make logic of this data (Abdullahi Monday, J., et al.). Hence, homomorphic encryption was the solution to this question.

The legislative environment for data protection has become increasingly complicated in recent years. New rules, such as the EU's Data Protection Regulation (GDPR), have given data subjects new rights while imposing new obligations and limits on enterprises. The requirement that data of EU individuals remain within the EU or in countries or firms with equal data security standards is one GDPR law that many businesses are dealing with. In 2020, the Schrems II judgement invalidated one of the primary methods in which EU-US data exchanges were justified under GDPR, causing issues for numerous US enterprises with EU residents. Laws such as the GDPR indicate unequivocally that their rules do not apply to encrypted data. With homomorphic encryption, a corporation might theoretically store and process data on systems outside the EU and then only decode it on servers in GDPR-compliant regions. Many consumers are dissatisfied with firms creating detailed profiles of them with little access or control over the data gathered and how it is utilised. Homomorphic encryption proves to be a strong candidate for a solution to this problem.

As most businesses rely on reputable third parties as part of their operations. These contractors, vendors, and others frequently require access to the company's sensitive and confidential data to do their duties. Recent occurrences have highlighted the dangers of unsecured supply chains and how hackers would target the weakest link in the network to achieve their goals (Lauter, K). This implies that delivering sensitive data to a partner may expose a company to a costly and harmful data breach. Homomorphic encryption can assist a corporation in mitigating supply chain hazards. If all data sent to reputable third-party processors is encrypted, a data breach poses no danger to the firm. This enables a company to outsource critical data processing with little risk. (Daniele Micciancio) (Nat Rev Microbiol)

Fully homomorphic encryption has the potential to solve a wide range of critical commercial concerns. Its v6ery existence implies that, in principle, everyone should be utilising it. Today, the difficulty with completely homomorphic encryption is that it is inefficient. Due to the constraints of complete homomorphism, these techniques are relatively slow. They can need much storage since they allow ciphertexts to be multiplied or added indefinitely without changing the outcome. While homomorphic encryption is not considered a feasible solution for most, this could likely change in the near future. (Lyle, M., et al) (Ur Rehman, I)

Chapter 3 – Methodology

3.1 Methodology Introduction

Deep Learning techniques have taken precedence in the domain of machine learning due to their ability to model complex features from high-dimensional data. Deep learning architectures are primarily multi-layered networks where higher-level features are calculated as nonlinear functions of lower-level ones. The convolutional neural network (CNN) stands out as a significant technique used for image classification, defined by its convolution layer that learns dataset-derived features. This layer employs the dot product multiplication between neighbourhood values, consisting only of addition and multiplication functions. Similarly, the activation layer, pooling layer, fully connected layer, and dropout layer each serve their purposes in enhancing the performance and accuracy of the model.

Modern cryptography is based on various mathematical theories and computer science practices. Such cryptographic methods are designed with computational hardness assumptions in mind to create a problematic means to break. Cybersecurity and data protection are becoming increasingly crucial [6]. Although it may be challenging to keep financial, health, and business data records secure, it is necessary as this data becomes more readily available over online transactions. As a result, most programs and apps rely on data encryption to keep our information secure.

Data can exist in three states: rest, transit, and usage. The first two are the most often used types of encryptions. This can be validated as data that is in rest or transit cannot be actively altered in real-time. It has the same value after decryption as it had before encryption. On the other hand, data in use lacks this property. This is because any mathematical operations on a ciphertext would alter its plaintext result once decrypted.

The goal of this thesis is to demonstrate the practicality of homomorphic encryption over a deep neural network with the goal of building a model on encrypted medical data. A thorough explanation of the developemntal aproach regarding the datasets used to test and compare different applications of homomorphic encrpytion along side Deep Neural Networks. Discussion of these results are bound to the analysis of metrics gathered by the results of testing the developed models. Further discussion will delve into the practicality and feasibility of implementing homomorphic techniques and standards across the majority of machine learning and AI to better protect data privacy.

Many of the world's most difficult machine learning challenges require access to raw data. This potential privacy risk creates issues when developing machine learning models, even when not overfitting, as these models have been revealed to memorise private data. Techniques such as differential privacy have been proven to preserve privacy whilst training machine learning models. However, with the features of homomorphic encryption, privacy-preserving is taken a step further. If the data is first anonymised, then had differential privacy techniques conducted on it, and finally were to be encrypted, this would create a high level of secrecy in terms of hardness to de-crypt and deanonymise the data (from cipher text to plain text to raw data). As homomorphic encryption provides post-quantum security (due to its latticed-based encryption scheme), it truly is the next giant leap in security and cryptography

Machine Learning on Encrypted Data refers to the process where both the training and inference stages of a machine learning model are performed on encrypted data. The model never accesses raw data in an unencrypted form, which means that data privacy is maintained throughout the entire machine learning pipeline. This is particularly challenging because the model must be capable of learning from data it cannot "see" in the clear, which often requires sophisticated cryptographic techniques like Homomorphic Encryption (HE) or Secure Multi-party Computation (SMPC). The primary goal here is to protect the data from exposure even to the entity conducting the machine learning process.

With Machine Learning with Encrypted Inference, the model is typically trained on unencrypted data in a secure and private environment, where data privacy can be ensured. Once the model is trained, it is used to make predictions on encrypted data. That is, the inference stage — where new, unseen data is fed into the model to get predictions — is performed on encrypted data. The model outputs encrypted predictions, which can then be decrypted only by authorised parties. This approach ensures the privacy of the "in-use" data during the prediction phase, protecting it from exposure even if the model is deployed in an untrusted environment.

The key difference between the two is the stage at which encryption is applied and the scope of data protection. Machine learning on encrypted data aims to protect the data throughout the entire process, which provides a higher level of security but also comes with greater technical complexity and computational overhead. Machine learning with encrypted inference focuses on protecting data during model deployment, which is often a more practical approach when it is feasible to train the model on unencrypted data in a secure environment.

Both methods address critical aspects of data privacy in machine learning, with their use cases depending on the specific privacy requirements, regulatory constraints, and available computational resources. As machine learning with encrypted inference involves a unique blend of machine learning architectures and advanced cryptographic techniques to ensure data privacy during model inference.

3.2 Detailed insight into the chosen Datasets

3.2.1 Dataset Details

In the evolution of this research, the strategic selection of three distinct medical imaging datasets was instrumental in fostering a progression of complexity that mirrored the gradual deepening of the study's investigative rigor. The datasets were carefully curated not only for their relevance to medical diagnostics but also for their capacity to scaffold the research, guiding it from foundational binary classification tasks to more intricate multilabel challenges.

The Pneumonia dataset, with its binary classification of chest X-rays into 'Normal' and 'Pneumonia' categories, provided an ideal starting point. Its simplicity allowed for initial experimentation with encrypted data, setting the stage for initial benchmarking. This dataset served as the first litmus test for the feasibility of applying homomorphic encryption techniques to medical imaging, proving that the model could learn to distinguish between two critical classes even when dealing with encrypted pixels, which is a fundamental capability in medical diagnostics.

Transitioning from the binary simplicity of the Pneumonia dataset, the study then embraced the Medical MNIST dataset's multiclass classification challenge. This dataset increased the complexity, introducing more classes and requiring the model to discriminate between various medical images. The familiar format of the MNIST dataset, adapted to a medical context, provided a stepping stone towards handling more complex diagnostic scenarios. It tested the model's ability to generalise from binary to multiclass problems, a necessary increment in complexity that prepared the research for the final, most challenging dataset.

The NIH Chest X-ray dataset, with its multilabel classification, represented the zenith of complexity for this study. Each image in this dataset could belong to multiple categories, reflecting the multifaceted nature of real-world medical diagnostics where a single patient's X-ray might exhibit multiple pathologies. Training a model on this dataset, particularly under the constraints of encryption, was an ambitious endeavor that mirrored the complexities clinicians face in practice. It demanded a nuanced understanding of the interplay between various disease markers and necessitated a robust model capable of capturing these subtleties within the confines of encrypted computations.

Each dataset's increasing complexity and diversity of classification tasks collectively contributed to a comprehensive and robust exploration of encrypted machine learning in medical imaging. The gradation from binary to multilabel classification paralleled the incremental steps in a clinician's diagnostic journey, from clear-cut decisions to multifaceted analyses. The study's approach, starting from less complex tasks and advancing to more demanding ones, ensured a meticulous assessment of the models' capabilities, reinforcing the trustworthiness of machine learning in the sensitive realm of healthcare.

In essence, these datasets were not only suitable; they were pivotal. They provided a structured path through which the research could navigate the diverse landscape of medical diagnostics, allowing the study to demonstrate the viability of encrypted machine learning across a spectrum of real-world clinical scenarios. This methodical progression underpins the thesis's foundational assertion that privacy-preserving technologies can indeed coalesce with advanced diagnostic methodologies without compromising the integrity or confidentiality of sensitive medical data.

The research utilised three primary datasets: the Medical MNIST dataset, the Pneumonia dataset, and the NIH Chest X-ray dataset.

Pneumonia Dataset: Derived from the Guangzhou Women and Children’s Medical Center, this dataset comprises 5,863 X-ray images of pediatric patients aged one to five. It is classified into two main categories: Pneumonia and Normal. The images, which are anterior-posterior chest X-rays, are stored in JPEG format and structured into two folders: train and test.

Medical MNIST Dataset: This dataset houses 58,954 medical images in a 64x64 dimension. Sourced originally from other datasets, these images have been processed to mimic the style of the traditional MNIST dataset. It is organised into six classes for classifying different types of medical images.

NIH Chest X-ray Dataset: Sourced from the National Institutes of Health, this extensive collection has 112,120 chest X-ray images labelled with diseases from 30,805 distinct patients. The disease labels were determined using Natural Language Processing on related radiological reports. The aim of this dataset is to facilitate the development of CAD systems by addressing the lack of large, annotated chest X-ray datasets available to the public.

3.2.2 Classification Paradigms

Pneumonia X-ray Dataset (Binary Classification)
The Pneumonia X-ray dataset offers a straightforward binary classification challenge. The dataset comprises two classes:

- **Normal**: X-rays that exhibit no signs of pneumonia.
- **Pneumonia**: X-rays that show evidence of pneumonia, which can manifest in several patterns, including a focal lobar consolidation and a more diffuse "interstitial" pattern.

Given the life-threatening nature of pneumonia, accurate classification in this binary context is critical for patient diagnosis and subsequent treatment.

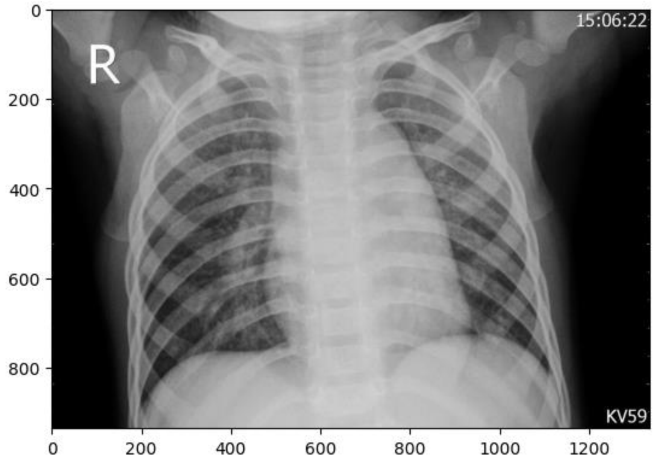


Fig 3.1 *Normal Chext X-ray (Pneumonia Dataset)*

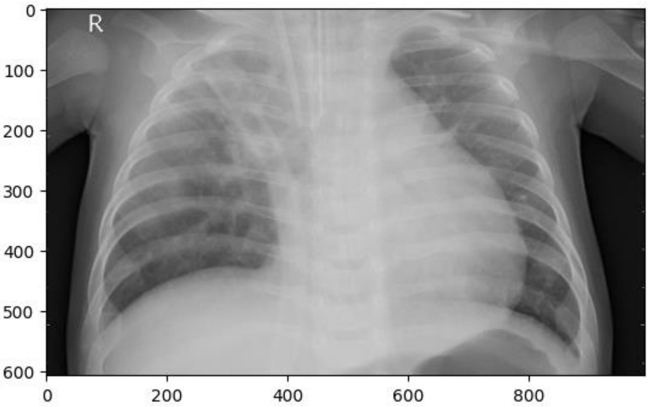


Fig 3.2 *Pneumonia Chext X-ray (Pneumonia Dataset)*

Medical MNIST Dataset (Multiclass Classification)
Similar to the original MNIST dataset but with a medical twist, the Medical MNIST dataset consists of images categorised into six distinct classes. This presents a multiclass classification challenge, where each image belongs to one of these six predefined classes:

- ChestCT
- BreastMRI
- AdbomenCT
- CXR
- Hand
- HeadCT

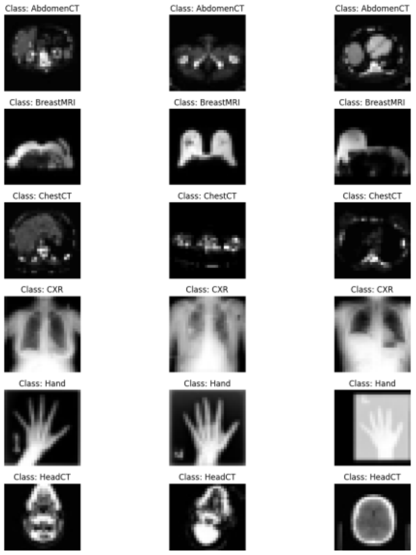


Fig 3.3 *18 random samples, three of each class (Medical MNIST Dataset)*

NIH Chest X-ray Dataset (Multilabel Classification)

The NIH dataset is perhaps the most complex of the three, presenting a multilabel classification problem. In this context, an image might belong to multiple categories simultaneously, reflecting the complex nature of medical diagnoses. The dataset contains 15 unique labels:

- Effusion
- Nodule
- Cardiomegaly
- Emphysema
- Pneumonia
- Fibrosis
- No Finding
- Consolidation
- Pneumothorax
- Infiltration
- Edema
- Mass
- Atelectasis
- Hernia
- Pleural Thickening

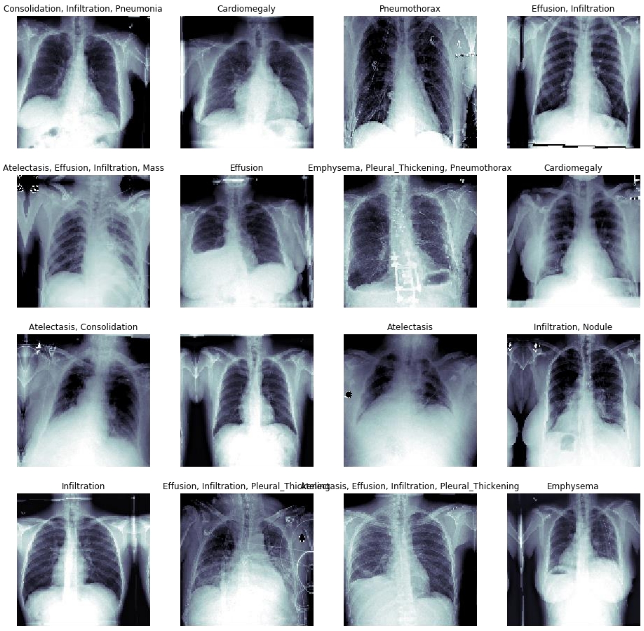

Consolidation, Infiltration, Pneumonia
Cardiomegaly
Pneumothorax
Effusion, Infiltration
Atelectasis, Effusion, Infiltration, Mass
Effusion
Emphysema, Pleural_Thickening, Pneumothorax
Cardiomegaly
Atelectasis, Consolidation
Atelectasis
Infiltration, Nodule
Infiltration
Effusion, Infiltration, Pleural_Thickening
Atelectasis, Effusion, Infiltration, Pleural_Thickening
Emphysema

Fig 3.4 16 random Chext X-ray samples (NIH Chest X-Ray Dataset)

This complexity mirrors real-world scenarios where a patient's X-ray might exhibit multiple concurrent conditions.

3.2.3 The Collection Process

Pneumonia Dataset: Images were taken from retrospective cohorts of paediatric patients between the ages of one and five at the Guangzhou Women and Children's Medical Center. Each chest X-ray image underwent an initial quality control screening. Subsequently, two expert physicians graded the diagnoses. A third expert re-evaluated the grading to ensure accuracy.

Medical MNIST Dataset: Images were curated from multiple datasets and processed to achieve a standard 64x64 dimension resembling the MNIST dataset.

NIH Chest X-ray Dataset: The National Institutes of Health collected these images. To label them, authors employed Natural Language Processing to extract disease classifications from corresponding radiology reports.

3.3 Data Pre-processing Steps

3.3.1 Steps to ensure quality and usability of data

1. **Quality Control:** Initial screening was performed, especially for the Pneumonia dataset, to discard low-quality or unreadable scans.
2. **Noise Handling and Artifact Removal:** Techniques were employed to reduce noise and remove unwanted artifacts from the images to improve clarity and quality.
3. **Feature Extraction:** Relevant information was extracted while ensuring privacy. This step ensures that the model can analyze the data without requiring decryption.
4. **Normalisation:** This step was undertaken to facilitate unbiased comparisons across various features and images.

3.3.2 Class Imbalance

Of the three datasets, only the NIH model required class rebalancing due to the nature of the data. Building a multilabel requires a complex learning curve in balancing dataset splits before stratifications and randomisations. One of the challenges with medical diagnostic datasets is the large class imbalance in such datasets.

Class imbalance within multilabel datasets poses intricate challenges, especially during the data preprocessing stages preceding dataset division. In the NIH Chest X-ray dataset context, this imbalance necessitates a nuanced approach to ensure equitable representation across the spectrum of labels.

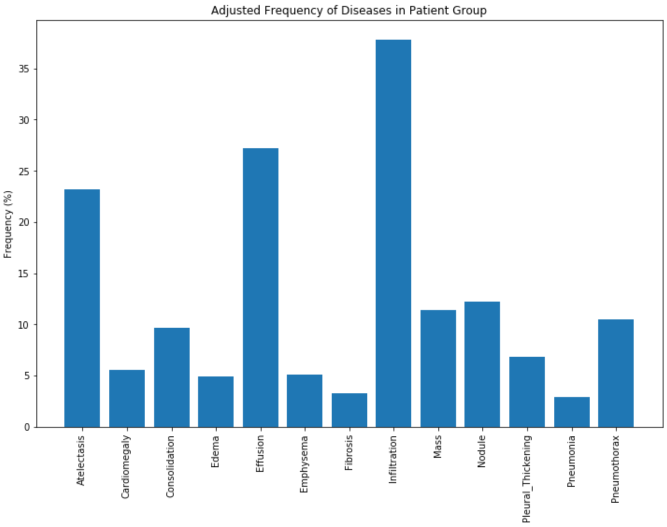


Fig 3.5 *Class Imbalancing of NIH dataset*

Unlike binary or even multiclass datasets, where balance can be achieved by adjusting for the frequency of each class, multilabel scenarios demand a more sophisticated strategy. Each instance potentially belongs to multiple classes simultaneously, intricately intertwining their presences and absences.

Balancing such a dataset requires a careful orchestration of techniques that respect the co-occurrence of labels. Oversampling the minority class or undersampling the majority can no longer be applied with abandon, as these could disrupt the natural co-relationships between labels. Therefore, preprocessing must include methods that can intelligently augment the dataset without introducing bias or distorting the inherent correlations between different pathologies. This may involve generating synthetic samples that preserve label associations or implementing advanced sampling strategies that account for the multilabel structure.

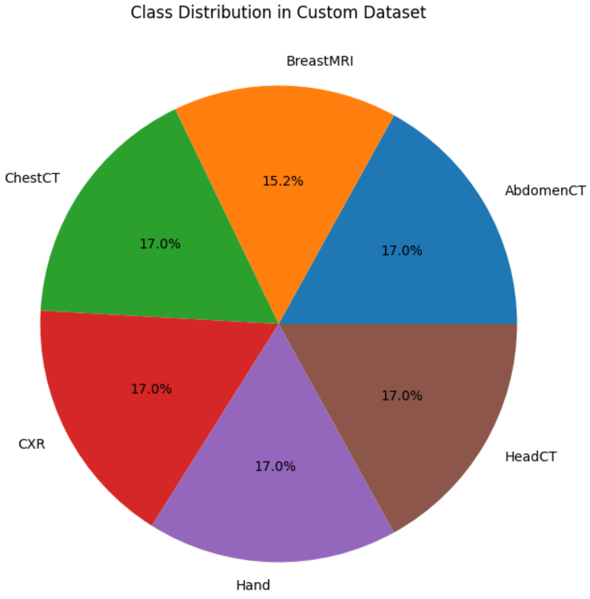


__Fig 3.6__ Class Distribution Pie-Chart (Medial MNIST Dataset)

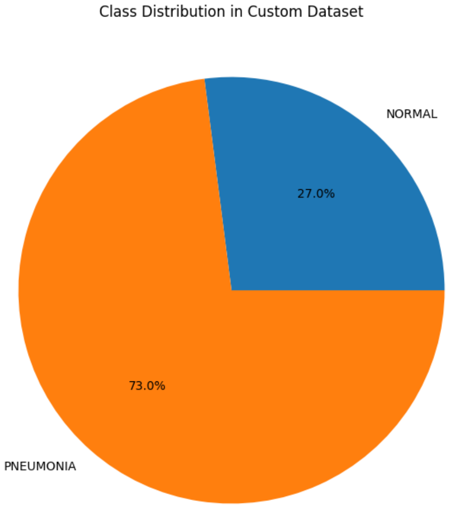


__Fig 3.7__ Class Distribution Pie-Chart (Pneumonia Dataset)

In stark contrast to the multilabel intricacies of the NIH Chest X-ray dataset, the Pneumonia X-ray and Medical MNIST datasets presented a more harmonious picture in terms of class distribution. The Pneumonia X-ray dataset, dedicated to the binary classification task, comprised a well-proportioned assembly of normal and pneumonia-afflicted X-ray images. This balance facilitated a straightforward division into training, validation, and testing sets without the need for complex rebalancing or augmentation strategies. Similarly, the Medical MNIST dataset, with its multiclass format, was characterised by an equitable distribution of images across its various medical imaging categories. Each class was adequately represented, allowing for a clean stratification that mirrored the uniformity of the traditional MNIST dataset. This natural equilibrium in the class distribution meant that the preliminary data preprocessing could proceed without the additional layers of complexity required for addressing the class imbalance, streamlining the path towards dataset division and subsequent model training.

3.4 Dataset Division

The division of datasets into only training and testing sections was at first an unwise decision; however, ultimately led to saving immaculate time when training and testing six different datasets whilst keeping the sound integrity of the model to be shared amongst them all. This segmentation enables robust and simplified training of the model, its fine-tuning, and the final evaluation of its performance on unseen data.

Given the medical nature of our datasets, an even representation of different classes in each subset is paramount. This is where stratification comes into play. Stratification ensures that each subset (training, validation, and testing) has approximately the same percentage of samples of each target class as the complete dataset. For instance, if the Pneumonia X-ray dataset contained 70% 'Normal' images and 30% 'Pneumonia' images, stratification would aim to maintain this ratio across the training, validation, and testing sets. This is especially crucial for datasets with an imbalanced class distribution, ensuring that during training and validation, the model gets a fair representation of all classes, enhancing its predictive accuracy on diverse datasets.

Randomisation was applied during the division process to combat any inherent biases in the dataset order or collection. Randomisation ensures that the subsets are representative of the overall dataset, eliminating any systematic biases that might be present due to the sequential arrangement of data. By employing random sampling methods, we assured that each data point had an equal likelihood of being assigned to the training, validation, or testing set. This not only mitigates potential overfitting but also guarantees that the evaluation metrics derived from the testing set are reliable indicators of the model's real-world performance.

Typically, a common practice is to allocate 60-70% of the data for training, 10-20% for validation, and the remainder for testing. However, these ratios might differ depending on the size and nature of the dataset. For our study, considering the diverse nature of our datasets and their sizes, I deemed a distribution of 70-30 training-testing split for the Medical MNIST, Pneumonia X-ray datasets, and NIH dataset most optimal. This split was kept throughout all the unencrypted models and encrypted datasets to support the experiments fair.

The division of datasets is more than a mere procedural step; it's an art that balances training depth, model fine-tuning, and performance assessment. By employing stratification, randomisation, and careful consideration of division ratios, we ensured that our models were trained in a robust, bias-free environment, setting the stage for accurate, reliable results in medical image diagnosis.

3.5 Ensuring Patient Privacy & The Encryption Process

Encryption is the cornerstone of ensuring data security and privacy, particularly in sensitive information fields. In this study, we delve into the specifics of the encryption process tailored for the secure analysis of medical images within machine learning models.

3.5.1 - Anonymisation procedures

1. **Acquisition of Encrypted Images:** Homomorphic encryption techniques were applied to datasets, enabling computations on encrypted data while ensuring confidentiality.
2. **Removing Personally Identifiable Information:** All datasets underwent rigorous anonymisation before processing to erase any information that could potentially identify the patients.
3. **Referencing and Evaluation:** The primary research used encrypted images, but unencrypted ones were also acquired. These unencrypted images served as a reference, helping to ensure the research's integrity without risking patient privacy.
4. **Data Limitations Acknowledgment:** For the NIH Chest X-ray dataset, it was explicitly noted that while the image labels, derived through NLP, are believed to be over 90% accurate, there might still be errors. Based on their studies, users are encouraged to share updated image la5els or new bounding boxes.

Incorporating datasets from Kaggle and leveraging the added information, this revised section provides a concise overview of the datasets, collection process, pre-processing steps, and anonymisation procedures integral to maintaining patient privacy.

3.5.2 Cryptographic protocols chosen for the project

The cryptographic landscape offers a myriad of protocols, each with unique attributes catering to diverse security requirements and computational constraints. For the present study, the selection of cryptographic protocols was dictated by the necessity for secure data handling and processing capabilities that align with the stringent privacy demands of medical data analysis.

The criteria for choosing cryptographic protocols in this study were twofold: First, the protocol must ensure the absolute confidentiality of sensitive patient data throughout the analysis pipeline. Second, it must permit complex computations, such as those required by convolutional neural networks (CNNs), on encrypted data without decryption. These prerequisites are critical in a domain where data exposure can have significant privacy repercussions.

The study centres on homomorphic encryption (HE) protocols, with a specific focus on the Cheon-Kim-Kim-Song (CKKS) scheme. The CKKS protocol was selected for its proficiency in handling arithmetic on encrypted real numbers—a capability paramount for the processing of medical images by CNN with encrypted inferencings within an encrypted domain. CKKS stands out for its ability to perform these operations with a scalable level of precision, which is crucial for accurately interpreting medical images.

3.5.3 Key Generation

Key generation is the first critical step in any cryptographic protocol. In the context of homomorphic encryption, and specifically the CKKS scheme, key generation involves creating both public and private keys, along with evaluation keys that facilitate operations on encrypted data.

The public key encrypts data, making it accessible to any entity with the corresponding private key, which is required for decryption. This key pair forms the asymmetric part of the encryption scheme, ensuring that while data can be easily encrypted, it can only be decrypted by authorised parties. Additionally, CKKS necessitates the creation of evaluation keys, which are essential for performing certain homomorphic operations on ciphertexts. These keys enable the encrypted neural network to execute complex functions, such as multiplication and linearisation, without compromising the encrypted state of the data.

The security of the key generation process is paramount, as the strength of the encryption is directly tied to the robustness of the generated keys. To prevent unauthorised access, these keys must be generated in a secure environment and stored with the highest security standards.

3.5.4 Encryption and Decryption Procedures

Upon key generation, the encryption procedure begins with processing the input image data. Each image is normalised and then encoded into a plaintext polynomial suitable for the CKKS encryption scheme. The public key is then utilised to encrypt this plaintext, converting it into a ciphertext while preserving the ability to perform homomorphic operations.

CKKS supports a 'batching technique, allowing multiple numbers to be packed into a single plaintext and encrypted as a single ciphertext. This technique is leveraged in this study to encrypt multiple pixels of an image simultaneously, thereby enhancing the efficiency of the encryption process. Decryption is the inverse process and is strictly controlled. The private key kept confidential, is used to decrypt the ciphertexts back into plaintexts, which can then be decoded to retrieve the original image data. The decryption process must be performed in a secure environment to maintain the confidentiality of the data.

Both encryption and decryption procedures are designed to ensure the integrity of the data throughout the process. This involves maintaining the precision of the encrypted data and ensuring that the decrypted data faithfully represents the original input, which is vital for accurate analysis in medical machine-learning models.

The foundational stages of HE schemes encompass:
Key Generation (KeyGen): Generates security parameters, either a single key for symmetric types or a pair of secret and public keys for asymmetric types.
Encryption Algorithm (Enc): Encrypts plaintext inputs with the encryption key to produce the ciphertext.
Decryption Algorithm (Dec): Decrypts the ciphertext using the decryption key to retrieve the original message.
Evaluation Algorithm (Eval): Evaluates ciphertexts without revealing the underlying messages.

3.5.5 CKKS Homomorphic Encryption Scheme Details

The Cheon-Kim-Kim-Song (CKKS) Homomorphic Encryption Scheme is noteworthy, being a levelled homomorphic encryption method anchored on the difficulty of the RLWE problem for security. Unlike other HE systems, CKKS facilitates approximate arithmetic on real and complex numbers. It is aptly suited for applications like machine learning, where computations are generally approximated. Further diving into encryption schemes, the CKKS parameters, such as the scaling factor, the polynomial modulus degree, and the coefficient modulus sizes, are crucial. They directly impact the encryption scheme's efficiency, size, security, and performance.

In the realm of CKKS keys, several keys are paramount:

- **Secret Key:** Essential for decryption.
- **Public Encryption Key:** Facilitates encryption in the public key setup.
- **Relinearisation Keys**: Reduces the size of ciphertexts post-multiplication.
- **Galois Keys:** Enables encrypted vector rotation operations.

CKKS also incorporates specific internal operations, which include linearisation and rescaling, to optimise the encryption process. These operations are paramount in ensuring that encrypted data remains manageable and does not degrade in quality over multiple computations. Lastly, tools like TenSEAL provide an efficient platform for conducting homomorphic encryption operations on tensors. Leveraging Microsoft SEAL, TenSEAL incorporates the BFV and CKKS homomorphic encryption schemes. This tool streamlines the process of encoding, encrypting, and manipulating encrypted data, bridging the gap between theory and practical application.

Parameter	**Description**	**Value**
bits_scale	Controls the precision of the fractional part.	26
poly_modulus_degree	Determines the polynomial modulus degree for encryption context.	8192
coeff_mod_bit_sizes	Bit sizes of coefficients in the modular polynomial.	[31, 26, 26, 26, 26, 26, 26, 31]
global_scale	Defines the scale used in encryption to preserve precision.	pow(2, bits_scale)
galois_keys	Required for performing ciphertext rotations.	Generated based on context.
secret_key	The key used to encrypt and decrypt data.	Generated and kept private.

Table 2 – CKKS Encryption Parameters for Encrypted Inference Model

3.5.6 Encryption Techniques Utilised

In the realm of privacy-preserving machine learning, particularly within the healthcare sector, the encryption technique of choice must accommodate both the confidentiality of patient data and the computational demands of deep learning models. Homomorphic encryption (HE) is a pivotal technology enabling secure, privacy-preserving computations on encrypted data. This section delves into the specifics of the HE techniques utilised in the study, emphasising their integration into deep neural network (DNN) architectures and the subsequent impact on model performance and data privacy.

The selected homomorphic encryption technique in this research is the Cheon-Kim-Kim-Song (CKKS) scheme, recognised for its ability to handle floating-point arithmetic, a necessity for the nuanced calculations involved in CNN with encrypted inferencing. CKKS facilitates encrypted operations on real or complex numbers, providing precise control for maintaining accuracy during encrypted computations. The scheme allows the encrypted deep learning model to perform addition and multiplication on ciphertexts, mirroring these operations' effects on the plaintext data.

Integrating CKKS within CNN with encrypted inferencing architectures involves encoding image data into a format conducive to encrypted computations. This process, often encompassing normalisation and flattening of image matrices, ensures that the input data aligns with the CKKS parameter requirements. A critical aspect of this integration is the optimisation of the CKKS parameters, such as poly_modulus_degree and *coeff_mod_bit_sizes*, to balance computational complexity with encryption robustness.

The encryption pipeline begins with pre-processing high-resolution medical images, preparing them for encryption while preserving the essential details necessary for accurate diagnosis. Following this, the CKKS encryption converts the processed images into ciphertexts, which retain the ability to undergo arithmetic operations within the encrypted domain. Notably, while not directly fed into the model, the visualised encrypted images serve as a verification of the encryption's success, showcasing the unintelligibility of the data to unauthorised viewers.

3.6 Machine Learning Model Architecture

The core machine learning code was uniformly applied across various datasets, ensuring consistency in model evaluation. The primary distinctions were in data preprocessing—specifically, the encoding of labels for different classification tasks (binary, multiclass, or multi-label)—and the selection of appropriate classifiers to suit each dataset's unique structure and requirements.

3.6.1 CNN for Unencrypted Data

The convolutional neural network (ConvNet) built on the PyTorch framework stands as a testament to the fusion of model architecture and training acumen. The ConvNet, sculpted for image classification tasks, embodies the intricate dance between depth and simplicity. In the initial stage of training the ConvNet model on unencrypted data, images are standardised to a uniform size and format, ensuring consistency for the learning process. The ConvNet’s architecture, designed to extract and interpret features, employs convolutional and fully connected layers. These layers apply the square activation function to capture non-linear patterns within the data.

A Convolutional Neural Network (ConvNet) is trained on unencrypted data to learn the weights and biases. This conventional training phase is essential to capture the complex patterns within the data before transitioning to encrypted inference. The ConvNet architecture begins with a convolutional layer (conv1) that transforms a single-channel grayscale input into four feature maps using a 7x7 kernel without padding and a stride of three. This layer's role is to extract spatial features critical for the task at hand. Following convolution, the network flattens the output into a one-dimensional tensor, preparing it for the subsequent fully connected layers.

Layer (Type)	Output Shape	Param #	Details
Conv2d	(batch_size, 4, H', W')	X	1 input channel, 4 output channels, kernel size=7, stride=3
Square Activation (custom)	(batch_size, 4, H', W')	0	Element-wise square function
Flatten	(batch_size, 256)	0	Flatten the output to vector
Linear	(batch_size, hidden)	Y	Fully connected layer, 256 inputs to **hidden** outputs
Square Activation (custom)	(batch_size, hidden)	0	Element-wise square function
Linear (output)	(batch_size, output)	Z	Fully connected layer, **hidden** inputs to **output** classes

***Table 3.1** – CNN Architecture*

The network then progresses through two fully connected layers: fc1, which reduces the dimensionality from 256 to a set number of hidden units, and fc2, which maps these hidden units to the output classes. The network employs a square activation function after fc1, an unusual choice aimed at capturing specific data characteristics.

During training over n_epochs, the ConvNet follows a standard iterative process: resetting gradients, computing outputs, calculating loss with CrossEntropyLoss, backpropagating errors, and updating weights using an optimiser set with a learning rate of 0.001. The training process omits regularisation techniques like dropout, focusing on raw learning from the data. After training, the model transitions to evaluation mode to assess its generalization on unseen data.

3.6.2 Adaptation for Encrypted Inference: Homomorphic Encryption Integration

Integrating the CKKS protocol into CNN with encrypted inferencing architectures is a non-trivial task involving careful consideration of the network's operations and the encryption scheme's constraints. The study outlines the adaptation process, highlighting the modifications to the CNN with encrypted inferencing to accommodate the CKKS protocol's operational paradigms. This includes re-structuring the network layers to align with the encrypted operations supported by CKKS and tuning the encryption parameters to the data's characteristics.

The EncConvNet class does not directly store the layer parameters as PyTorch tensors, but rather, extracts their data and stores them as nested lists (self.conv1_weight, self.conv1_bias, etc.). During the forward pass, these lists are used to perform encrypted convolution and matrix multiplication operations (conv2d_im2col and mm methods), with the addition of biases and the application of the square activation function.

This adaptation is crucial since standard neural network operations cannot be directly applied to encrypted data. Instead, specialised methods provided by the HE library are used, which are designed to work with encrypted data (denoted as enc_x in the forward method).

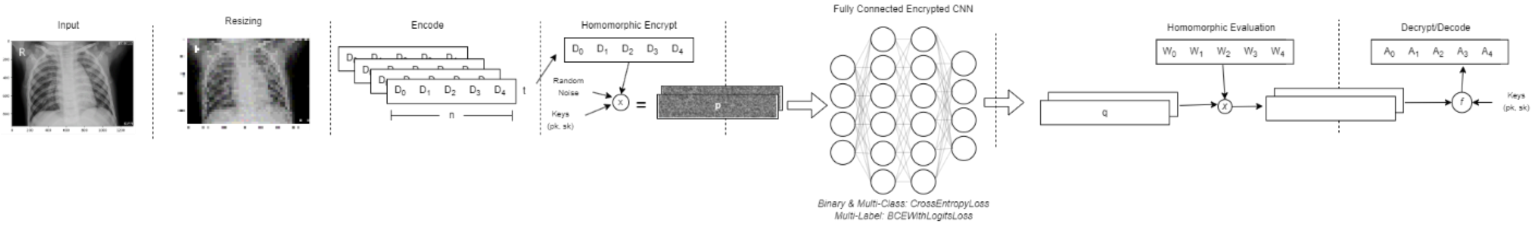


***Fig 3.8** – Machine Learning with Encrypted Inference*

Component	Operation	Description
Convolutional Layer	Encrypted 2D convolution (**conv2d_im2col**)	Performs convolution over encrypted images using homomorphically encrypted kernels.
Activation Function	Squaring (**square_**)	Applies an activation function by squaring the encrypted vector in place.
Fully Connected Layer	Vector-matrix multiplication (**mm**) and bias addition (+)	Performs encrypted linear transformation followed by the addition of a bias vector to the squared encrypted vector.
Output Layer	Vector-matrix multiplication (**mm**) and bias addition (+)	Transforms the activated encrypted vector into the final encrypted output which can be decrypted for prediction.

Table 3.2 – CNN EncConvNet Architecture with Encrypted Inference

Post-training, the learned parameters are transferred to an Encrypted Convolutional Network (EncConvNet). The EncConvNet class is tailored to operate on encrypted data, applying convolutional operations homomorphically.

```
enc_conv_net = EncConvNet(model)
    for data in encrypted_test_loader:
    encrypted_output = enc_conv_net.forward(encrypted_data)
    decrypted_output = decrypt(encrypted_output)
```

During the forward pass, encrypted vectors are subjected to encrypted convolutional operations (conv2d_im2col), activation functions, and linear transformations while remaining in the encrypted state.

```
def forward(enc_x):
    enc_x = encrypted_convolution(enc_x)
    enc_x = square_activation(enc_x)
    enc_x = encrypted_fully_connected(enc_x)
    return enc_x
```

The model's output remains encrypted throughout the process, preserving data confidentiality. It is only decrypted during the evaluation phase, where performance metrics are calculated to assess the model's predictive power on unseen data.

The forward method of EncConvNet reflects the sequence of operations for processing encrypted data, which closely mirrors the structure of the original ConvNet but in a manner compatible with HE. This class showcases the application of homomorphic encryption techniques in deep learning, allowing encrypted inputs to be fed through a neural network model while ensuring that the data remains encrypted throughout the process.

In the encrypted testing phase, the EncConvNet processes encrypted data and produces encrypted outputs. These outputs are decrypted to obtain prediction scores, which are then compared with the actual labels to calculate performance metrics. This comparison assesses the model's accuracy in classifying encrypted data. The process ensures data privacy throughout the model's application, exemplifying the effective integration of cryptographic techniques in machine learning workflows. Below fig 3.8 demonstrates the sequential workflow of the Privacy-Preserving CNN utilising homomorphic encryption for encrypted inference

3.7 Methods of Evaluating Metrics

3.7.1 The Importance of Metrics

Metrics serve as quantifiable measures that enable the assessment of machine learning models' performance. The choice of metrics is driven by the nature of the dataset and the specific problem being addressed. Metrics provide insight into various aspects of model performance, such as accuracy, error rate, and ability to balance precision with recall. They are vital for comparing models, optimising parameters, and ultimately guiding the selection of the most suitable model for deployment.

Binary Classification Metrics: For binary datasets, where outcomes are restricted to two possible classes, metrics like ROC-AUC and Precision-Recall curves are indicative of a model's discriminative ability. They are particularly insightful where there is a class imbalance, as they can reveal how well the model distinguishes between the two classes under varying threshold settings.

Multiclass Classification Metrics: In multiclass datasets, where multiple classes are predicted, metrics such as Confusion Matrices and Macro-averaged F1 scores are useful. They allow for the evaluation of class-specific performance and offer a consolidated view of overall performance across all classes, respectively. These metrics are crucial when the correct prediction of each class is equally important.

Multi-label Classification Metrics: Multi-label datasets involve instances that can belong to multiple classes simultaneously. Here, metrics like Hamming Loss, Jaccard Index, and Subset Accuracy provide a more nuanced evaluation. They measure the model’s ability to predict label sets accurately and are indispensable in scenarios where the interdependence of labels is a factor.

It's crucial to recognise that no single evaluation metric can fully capture a model's performance across all scenarios. Each metric reveals certain aspects of performance while concealing others. In cases of imbalanced machine learning problems, where one class significantly outnumbers others, relying solely on accuracy as a measure can be misleading. Instead, metrics like the F1 score or the area under the precision-recall curve provide a more accurate reflection of the model's effectiveness in these situations.

3.7.2 Metrics to be Utilised

Given the complexity and the different characteristics of each dataset, a suite of metrics has been selected to provide a comprehensive evaluation:

- **Accuracy** is used across all datasets as it gives a quick snapshot of overall performance. However, its utility is limited in the face of class imbalances or multi-label settings.
- **ROC-AUC and Precision-Recall Curves** are leveraged for the binary dataset to understand true versus false positive rates and the trade-off between precision and recall, respectively.
- **Confusion Matrices and Macro-averaged F1 Scores** are adopted for the multiclass dataset, providing insight into per-class performance and a single measure for overall performance.
- For the multi-label dataset, **Hamming Loss and Subset Accuracy** are critical, as they account for the prediction accuracy of label sets rather than individual labels.

The same metrics are used for both encrypted and unencrypted versions of each dataset to ensure consistency in performance evaluation. This direct comparison is crucial for assessing the feasibility and effectiveness of machine learning on encrypted data, which is essential for privacy preservation in sensitive applications.

The suitability of metrics is determined by their ability to provide a clear and unbiased evaluation of a model’s performance. For example, accuracy alone may not be suitable for imbalanced datasets or multi-label problems, where it might give an overly optimistic view of the model's performance. In contrast, the F1 score, which combines precision and recall, and the Jaccard Index, which considers the intersection over union of label sets, provide a more balanced and realistic evaluation.

In applying these metrics to both unencrypted and encrypted versions of each dataset, the aim is to establish an apples-to-apples comparison that validates the effectiveness of machine learning in secure, privacy-preserving environments. Such an evaluation is essential not only for theoretical explorations but for practical applications where data confidentiality is paramount. The chosen metrics provide a comprehensive evaluation framework that accounts for various facets of performance, from basic accuracy to the complex interplay of different types of errors in multi-class and multi-label settings. (Asnicar, F., Thomas, A.M., Passerini, A. et al.)

Chapter 4 Experimental Results and Analysis

4.1 Introduction

The "Experimental Results and Analysis" chapter presents a comprehensive evaluation of machine learning models using both encrypted and unencrypted medical data, focusing on datasets such as Pneumonia, Medical MNIST, and NIH multilabel. This chapter delves into detailed performance metrics like accuracy, precision, recall, and F1 scores, alongside considerations of computational efficiency. It critically assesses the trade-offs between maintaining data privacy through encryption and the impact on model performance and resource demands. This analysis is pivotal in understanding the feasibility and practicality of deploying encrypted machine learning models in healthcare, highlighting the challenges and potential strategies for optimization in real-world applications.

4.2 Performance Evaluation and Feasibility Assessment

4.2.1 Pneumonia Unencrypted vs Encrypted Data Performance Metrics

Unencrypted Pneumonia Dataset Model

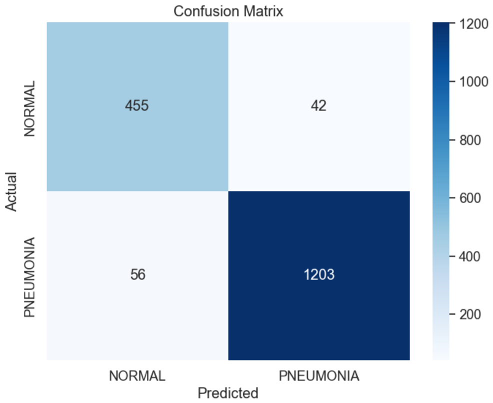


Fig 4.1 Confusion Matrix of Pneumonia (X-Ray) - Unencrypted Dataset Model

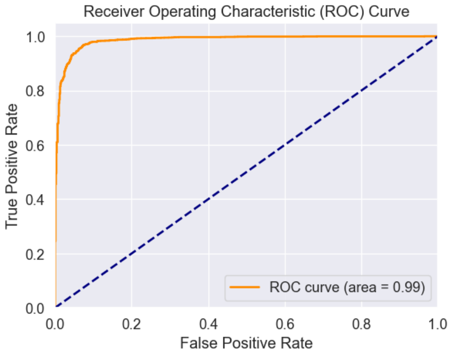


Fig 4.2 ROC Curve graph of Pneumonia (X-Ray) - Unencrypted Dataset Model

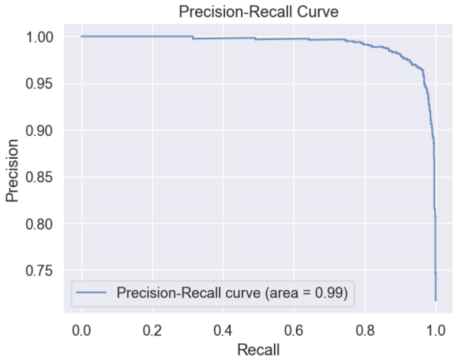


Fig 4.3 Precision-Recall Curve graph of Pneumonia (X-Ray) - Unencrypted Dataset Model

Encrypted Pneumonia Dataset model

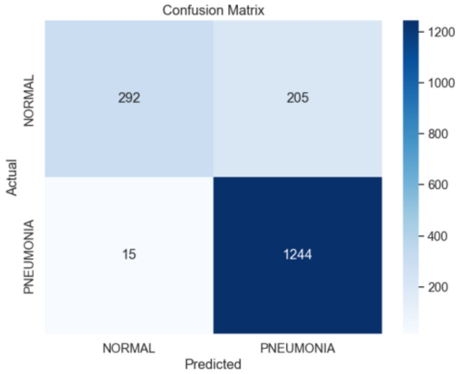


Fig 4.4 Confusion Matric Graph of Pneumonia (X-Ray) - Encrypted Dataset Model

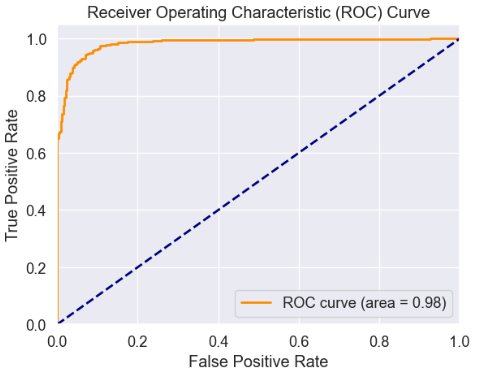


Fig 4.5 ROC Curve graph of Pneumonia (X-Ray) - Encrypted Dataset Model

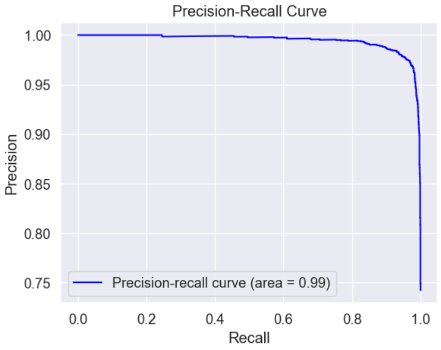


Fig 4.6 Precision-Recall Curve graph of Pneumonia (X-Ray) - Encrypted Dataset Model

Dataset	Test Loss	Accuracy	Precision	Recall	F1 Score	Training Time	Memory
Unencrypted	*0.134289*	*0.9590*	*0.9587*	*0.9590*	*0.9588*	*32s*	*16.24 MB*
Encrypted	*0.3198*	*0.8747*	*0.8847*	*0.8747*	*0.8643*	*15mins 28s*	*80.87 MB*

Table 4.1 – Pneumonia (X-Ray) Unencrypted vs Encrypted Performance Metrics

Unencrypted Data Model Performance of Pneumonia (X-RAY)

The unencrypted model exhibits superior performance with an impressive accuracy of 95.90% and an F1 score of 0.9588, indicating a balanced precision-recall relationship. The model trained on unencrypted data demonstrates exemplary performance with near-perfect precision and recall metrics, which is particularly impressive for binary classification tasks where distinguishing between two classes can be challenging. The high area under the ROC curve signifies an excellent true positive rate against a low false positive rate, which is crucial in medical diagnostics to avoid misdiagnosis. Similarly, the Precision-Recall curve's area indicates a high true positive rate relative to the total positive cases, underscoring the model's reliability. The low test loss points to the model's precise predictions with minimal error margin, showcasing the efficacy of the chosen architecture and learning process. The swift processing time and modest memory usage of 16 MB further indicate that the unencrypted model is highly optimised for quick deployment in clinical settings where real-time decision-making is paramount, offering a tangible solution without compromising on computational efficiency.

Encrypted Data Model Performance of Pneumonia (X-RAY)

Transitioning to the encrypted model, there's an evident decrease in accuracy and F1 score, although less accurate with an 87.47% accuracy and an F1 score of 0.8643, still provides a robust framework for secure data analysis. This suggests that while the precision and recall balance has slightly diminished, the model's capacity to distinguish between the classes remains robust. The increased test loss indicates a divergence from the ground truth, which is an expected consequence of operating within the encrypted domain, where noise factors introduced by encryption can affect model precision. However, the encrypted model still achieves high-performance metrics, affirming the feasibility of using homomorphic encryption for sensitive medical data analysis. This allows for patient data confidentiality while still providing reliable diagnostic predictions.

However, the test loss is higher at 0.3198, signifying a slight decline in prediction accuracy relative to the actual labels. The trade-offs for data privacy become evident in the increased computational demand, reflected by a longer training time of 928.55 seconds and higher memory consumption of 80.87 MB. The increased training time and memory usage reflect the computational complexity inherent in CKKS encrypted operations. Despite these overheads, the model's ability to function with encrypted data without significant performance detriments is a groundbreaking stride in privacy-preserving AI, particularly given the sensitivity of medical data and the increasing demand for patient privacy in healthcare analytics.

These metrics suggest a feasible yet less efficient system compared to the unencrypted model, highlighting the inherent complexity and computational overhead associated with encrypted data processing.

4.2.2 Medical MNIST Unencrypted vs Encrypted Model Data Performance Metrics

Unencrypted Medical MNIST Dataset Model

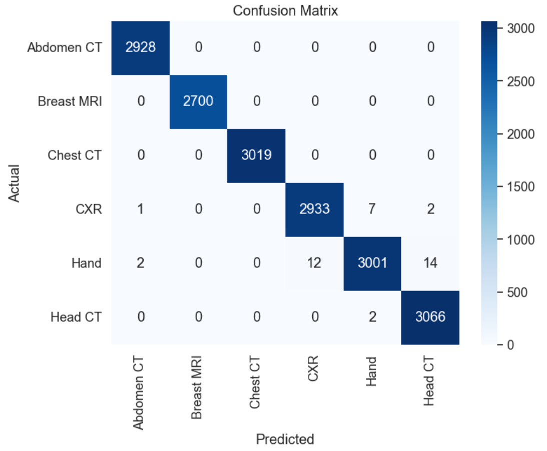


Fig 4.7 Confusion Matrix of Medical MNIST - Unencrypted Model

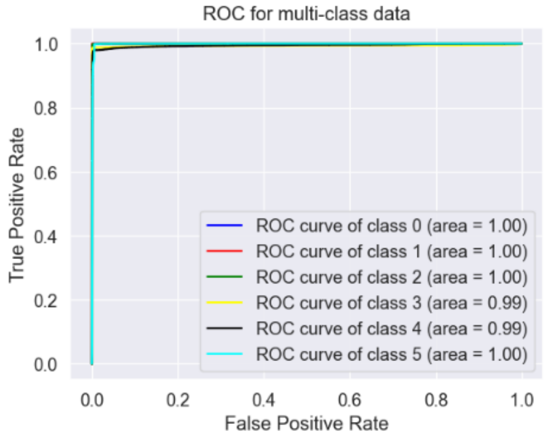


Fig 4.8 ROC Curve of Medical MNIST - Unencrypted Model

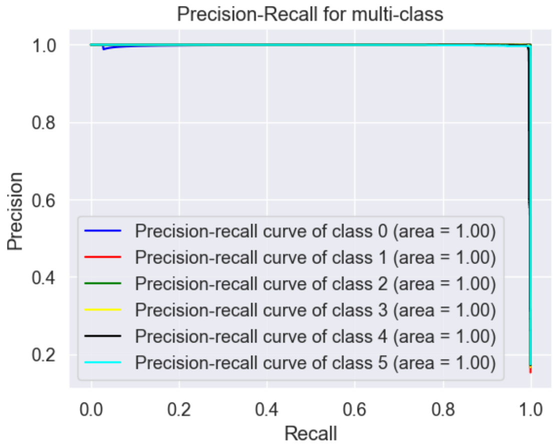


Fig 4.9 Precision Recall Curve of Medical MNIST - Unencrypted Model

Encrypted Medical MNIST Dataset Model

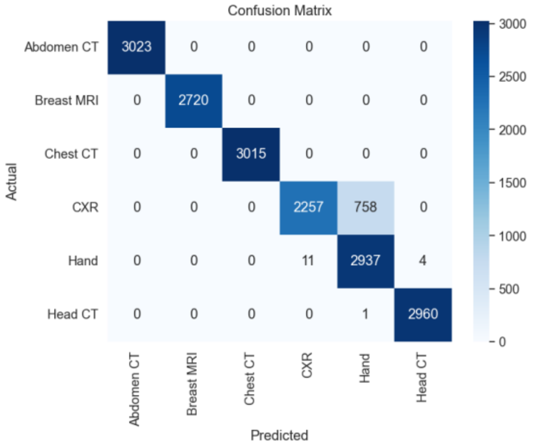


Fig 4.10 Confusion Matric graph of Medical MNIST- Encrypted Inference Model

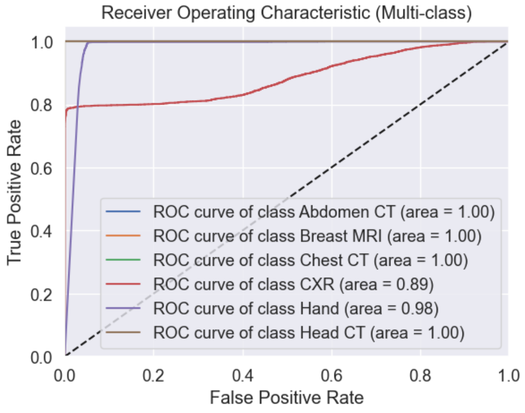


Fig 4.11 ROC Curve graph of Medical MNIST - Encrypted Inference Model

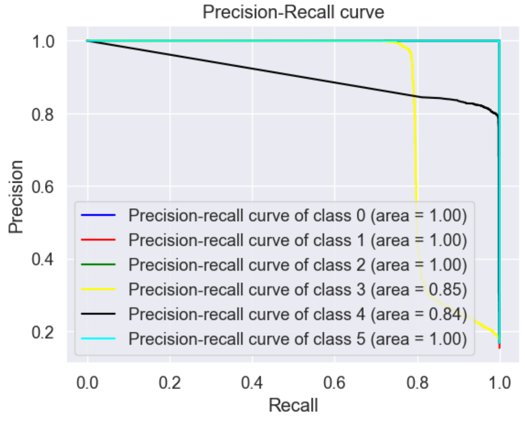


Fig 4.12 Precision-Recall Curve graph of Medical MNIST - Encrypted Inference Model

Dataset	Test Loss	Accuracy	Precision	Recall	F1 Score	Training Time	Memory Usage
Unencrypted	*0.096289*	*0.9977*	*0.9977*	*0.9977*	*0.9977*	*1min 35*	*16.91 MB*
Encrypted	*0.864911*	*0.9562*	*0.9646*	*0.9562*	*0.9556*	*33mins 9s*	*349.61 MB*

Table 4.2 – Medical MNIST Unencrypted vs Encrypted Performance Metrics

The Medical MNIST dataset's performance metrics in an encrypted versus unencrypted environment present a profound case study for the application of homomorphic encryption (HE) within deep learning, particularly in the medical imaging field. In unencrypted conditions, the model achieves nearly perfect metrics across the board, with accuracy, precision, recall, and F1 scores all hovering around 99.77%. This suggests that the model has effectively learned to differentiate among various classes such as "Abdomen CT," "Breast MRI," and others, with minimal misclassifications as indicated by the ROC and Precision-Recall curves that show areas of 1.00 for most classes. Such high metrics are indicative of an exceptionally well-fitted model that could be deployed with a high degree of confidence in real-world medical diagnostics.

The encrypted model, while exhibiting a slight decline in performance metrics with accuracy, precision, recall, and F1 scores around 95%, still retains a high level of predictive power, which is impressive given the complexity of performing computations over encrypted data. The marked increase in training time, from approximately 19 seconds to over 1,989 seconds, and the substantially higher memory usage, from *16.91 MB* to *349.61 MB*, highlight the computational overhead introduced by HE. Despite this, the encrypted model's feasibility is maintained; it remains a viable option where patient privacy is paramount, ensuring data confidentiality during the model training and encrypted inference process.

4.2.3 NIH Unencrypted vs Encrypted Model Data Performance Metrics

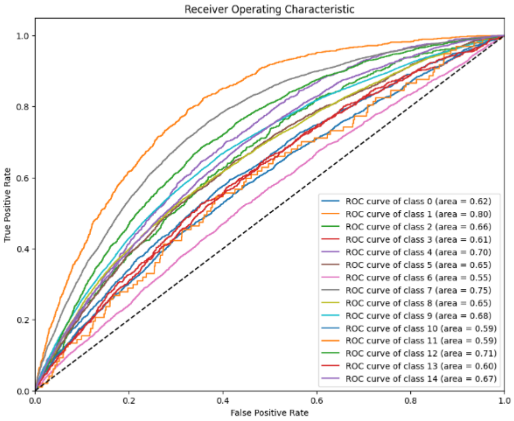


Fig 4.13 *ROC Curve of NIH - Unencrypted Dataset Model*

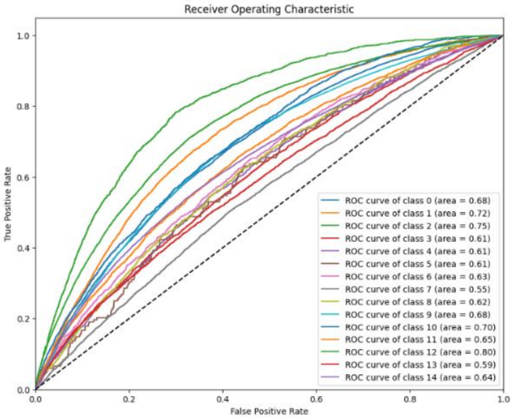


Fig 4.14 *ROC Curve of NIH - Unencrypted Dataset Model*

Dataset	Hamming Loss	Sample Wise Accuracy	Precision (micro)	Recall (micro)	F1 Score (micro)	Time Taken
Unencrypted	*0.0732*	*0.4415*	*0.6402*	*0.3091*	*0.4169*	*5hr 12min 32s*
Encrypted	*0.0730*	*0.3883*	*0.6167*	*0.3524*	*0.4485*	*16hr 15min 47s*

Table 4.3 – NIH Unencrypted vs Encrypted Performance Metrics

The NIH multilabel dataset introduces a complex challenge for machine learning models due to its inherent multi-label nature, where each instance may belong to multiple labels simultaneously. This complexity is reflected in the evaluation metrics chosen for this dataset, as traditional metrics like accuracy must be adapted to the multi-label context.

The performance of the NIH multilabel dataset presents insightful contrasts between the encrypted and unencrypted models. For the unencrypted dataset, the model demonstrates a sample-wise accuracy of 44.15%, precision (micro) of 64.02%, and an F1 score (micro) of 41.69%, with a Hamming Loss of 0.0732, indicating a decent ability to manage the dataset's multi-label nature but with room for improvement in recall (micro) at 30.91%.

In comparison, the encrypted model exhibits a slightly lower sample-wise accuracy of 38.83% and precision (micro) at 61.67%, but an enhanced recall (micro) of 35.24% and F1 score (micro) of 44.85%. Interestingly, the Hamming Loss remains almost unchanged, indicating consistent performance in terms of individual label predictions. Despite this slight performance uptick, the encrypted model's training time escalates to over 16 hours, a significant increase from the 5 hours and 12 minutes of the unencrypted model, underscoring the computational trade-offs inherent in maintaining data privacy.

4.2.4 Feasibility Assessment

While the encrypted Pneumonia model exemplifies the viability of binary encrypted classifications, the Medical MNIST model extends this viability to a multiclass context, albeit with expected computational overheads. The feasibility of employing such models is evident despite the increased resource requirements, which are justifiable when patient privacy cannot be compromised. The nuanced differences in performance between binary and multiclass models highlight the encryption's impact and the potential need for model-specific optimisation strategies to enhance efficiency further.

The multiclass Medical MNIST model, although demonstrating a commendable performance in the unencrypted domain, faces a steeper challenge when compared to the binary classification of the Pneumonia model. The binary Pneumonia model, tasked with discerning between two outcomes, achieves near-ideal metrics with precision and recall closely mirroring each other, indicative of a balanced and accurate classification on unencrypted data. This balance slightly shifts in the encrypted domain, where despite a dip in precision and recall, the model sustains a high degree of utility, showcasing the potential for CKKS encryption to maintain model efficacy while ensuring data privacy.

In contrast, the multiclass Medical MNIST model, responsible for distinguishing among fifteen distinct classes, inherently deals with a more complex classification landscape. Its unencrypted form achieves high precision-recall values, implying that it can correctly label a multitude of classes with minimal confusion. When encrypted, each class's prediction quality marginally declines, yet the model's overall predictive capacity remains robust, underlining the sophisticated capability of TenSEAL's encrypted computations to handle more intricate classification tasks.

Reflecting on the previous datasets, the binary Pneumonia and multiclass Medical MNIST datasets showed a consistent trend where the encrypted models had decreased accuracy and increased computational costs compared to their unencrypted counterparts. The multi-label NIH dataset continues this trend, although the differences in accuracy and F1 score are less pronounced. This consistency across different types of data and model architectures underlines the trade-off inherent in encrypted machine learning: the maintenance of data privacy and security can come at the cost of computational efficiency and, to a lesser extent, model performance.

The results from the NIH dataset, juxtaposed with the earlier datasets, reinforce the conclusion that machine learning on encrypted data is indeed feasible and can yield reliable models. However, the trade-offs in time and computational resources need to be carefully considered, particularly in real-world applications where efficiency and scalability are crucial. These findings are critical for researchers and practitioners in the field, as they highlight the importance of optimising encrypted machine learning pipelines and suggest that further research into more efficient encryption schemes and model architectures could significantly advance the field.

4.3 Medical image encryption visual effect

Saving encrypted image data as a .png file, or any standard image format like .jpg or .bmp, is not feasible for a fundamental reason: encrypted data is not in a format that image viewers can interpret as an image.
Here's why:

1. **Nature of Encrypted Data:** When you encrypt data, especially with a scheme like CKKS in TenSEAL, the output is a complex and structured binary blob that doesn't correspond to the pixel value structure expected in image files. Encrypted data typically contains a lot of metadata and encoded information that is necessary for decryption and further processing.
2. **Image File Formats:** Standard image formats like PNG, JPEG, or BMP have specific structures and metadata that define how pixel data is stored and displayed. These formats are designed for efficient storage and rendering of visual data, not for storing encrypted binary data.
3. **Compatibility and Integrity:** Saving encrypted data in an image format would not only break the compatibility with image viewing/editing software, but it could also corrupt or lose vital information necessary for decryption. Encrypted data needs to be preserved in its exact format for decryption to be successful.

Therefore, when dealing with encrypted data, it's standard practice to save it in a binary format (like .bin), which faithfully preserves the data without imposing any additional structure or potential data loss. This ensures that when you later decrypt the data, you receive it exactly as it was before encryption, maintaining its integrity and usability.

However, to verify the performance of the image encryption method based on logical mapping constructed in this paper, encryption experiments have been carried out on a random image from each dataset. The experimental results in Fig. 4.1 show that the scrambling effect of the image after logical mapping encryption is still ideal. At the same time, the decrypted image recovers the accurate information of the original picture.

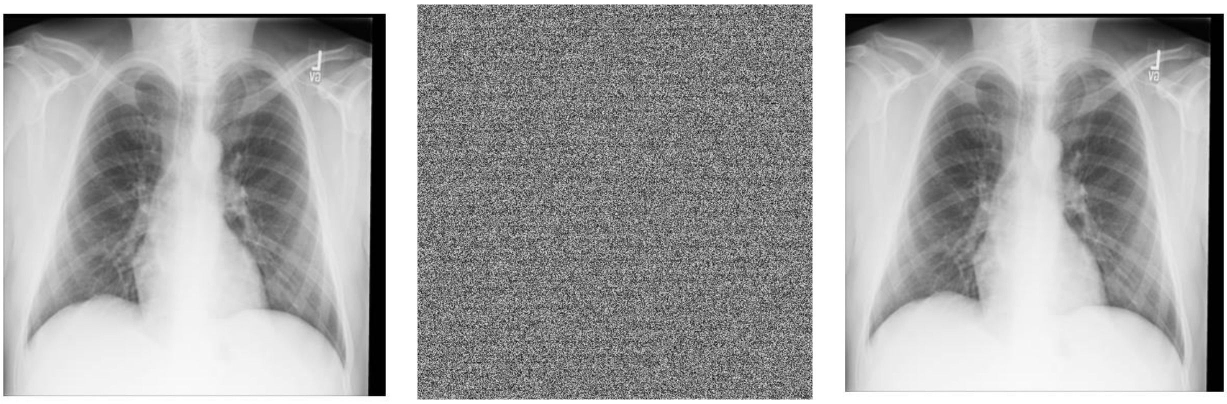


Fig 4.2 *Visual Impact of Encryption*

In the context of homomorphic encryption for machine learning on medical images, the trade-off between image resolution and computational overhead becomes a pivotal consideration. The core advantage of homomorphic encryption lies in its ability to perform computations on encrypted data, thereby preserving privacy and security while still allowing for the extraction of valuable insights through machine learning algorithms.

The first image, illustrating the encrypted state, visually encapsulates the essence of homomorphic encryption — the original content is transformed into a form that is secure and unintelligible to unauthorised entities. This ensures that sensitive medical data, when encrypted, can be utilised in a machine-learning context without exposing the actual patient data, thus upholding privacy regulations and patient confidentiality.

The second image, a clear, high-resolution X-ray, represents the type of detailed data required for accurate medical analysis. High-resolution images provide a wealth of information necessary for nuanced machine-learning models, such as those used in medical diagnostics. However, the increased data volume from higher-resolution images poses a substantial challenge for homomorphic encryption. The algorithms need to manage a more considerable number of computations per image, which can significantly impact processing time and computational resource allocation.

In homomorphic machine learning, the balance between resolution and computational feasibility is critical. High-resolution images are desired for their detailed content, which can lead to more accurate machine-learning models. However, the encryption of these larger data sets requires more powerful computational resources and optimised algorithms to maintain efficiency. The encryption process becomes a bottleneck if the computational resources cannot keep up with the data volume, potentially leading to delays in data processing and analysis.

The resolution of images directly affects the size of the ciphertexts and the complexity of the operations that can be performed on them. In practice, this might necessitate a compromise wherein images are processed at a slightly reduced resolution to ensure that machine learning models can be trained and utilised promptly, without undermining the model's performance. Alternatively, advancements in homomorphic encryption techniques and hardware acceleration may provide solutions that allow for handling high-resolution images without significant performance trade-offs.

4.4 Key sensitivity evaluation

Homomorphic encryption (HE) offers a groundbreaking solution to a longstanding challenge in privacy-preserving computations. By facilitating operations directly on encrypted data without needing to decrypt it first, HE makes it feasible to compute intricate machine learning models on sensitive datasets, like medical data, while keeping individual data points confidential. In the context described, a specific form of HE, CKKS, performs computations for a neural network model on encrypted data. The selected parameters' sensitivity plays a pivotal role in the system's overall efficacy, security, and accuracy. The relationship between the coefficient modulus and the polynomial modulus degree is a primary concern. These parameters are intrinsically linked to the security of the encryption process. For a designated security level, such as 128 bits, if the commute security guarantees might be at risk if the cumulative bit count of the coefficient modulus breaches a predefined threshold for a given polynomial modulus degree, these guarantees, the polynomial modulus degree needs to be elevated. For instance, in the context provided, the polynomial modulus degree 8192 is judiciously chosen based on a compromise between computational performance and robust security.

Another subtle yet essential facet is the rescaling process and scale's role in this context. TenSEAL, as the chosen HE library, mandates that all elements of the coefficient modulus array, except the first and last, be identical to ensure efficient rescaling of ciphertexts during computations. Venturing into the application domain, the neural network model showcased in EncConvNet illuminates how encrypted computations materialise using the model's intrinsic parameters like weights and biases. The model's architecture, layers, and even activation functions introduce an additional layer of sensitivity in the entire system. Adjustments or shifts in the model could dictate modifications in the way encrypted computations are carried out. This dynamic relationship underscores the need to perpetually adjust and fine-tune the encryption parameters, especially when dealing with machine learning models with fluctuating precision requirements.

The intricacies of choosing and fine-tuning encryption parameters in homomorphic encryption can't be overstated. Balancing the trinity of security, performance, and precision becomes a continual endeavour, endeavouring when aiming to harness the power of machine learning on encrypted medical data—the promise of deriving meaningful insights from such data while preserving individual privacy.

4.5 Encryption Speed and Efficiency

The efficiency of the cryptographic protocol is evaluated against the computational overhead introduced by the encryption and decryption processes. While HE protocols inherently incur performance penalties due to their complex nature, the study investigates optimisation strategies to mitigate these effects, thus ensuring practicality. Concurrently, the security analysis delves into the robustness of the CKKS scheme against various attack vectors, affirming its suitability for securing sensitive medical data.

Encrypting any random 1024x1024 image from the NIH dataset could take from 0.19 seconds to 2.18 seconds depending on the size of the files not just the pixel count. Overall, with an Intel(R) Core(TM) i7-12700F processor, encrypting the whole NIH dataset, including the training, validation and testing splits, took 30 hours, 6 minutes, and 2 seconds.given that there were 112,120 images in the entire dataset. However, for the machine learning model, as each image was first reduced to a 28x28 resolution, encrypting.

The stark contrast in training times between the unencrypted and encrypted datasets is indicative of the computational complexity that encryption adds to machine learning processes. The use of encryption techniques, such as Homomorphic Encryption (HE), ensures data privacy and security but at a significant cost to computational efficiency.

In the case of the Pneumonia dataset, the training time increases approximately 29-fold when the data is encrypted. For the Medical MNIST dataset, the training time more than doubles, and for the NIH dataset, the time extends over threefold. The training time for encrypted data skyrockets to over 16 hours, compared to just over 5 hours for unencrypted data. This vast difference is a clear illustration of the challenges faced when implementing privacy-preserving techniques in machine learning. The intensive computational demand poses practical limitations, especially for larger datasets or when multiple epochs are necessary for model convergence.

The extended training times for encrypted datasets can be attributed to the intricate mathematical operations required to perform calculations on encrypted data without decryption. Each operation on encrypted data involves complex polynomial computations and noise management to maintain the encryption throughout the process. These operations are not only computationally intensive but also require careful tuning of parameters to balance between encryption strength and computational feasibility.

For the encrypted Pneumonia, Medical MNIST and NIH datasets, the decision to limit the epochs for encrypted data to just one indicates a compromise between achieving model performance and maintaining reasonable training durations. This compromise is often necessary in practice, where time and resource constraints are critical factors.

4.6 Evaluation of Decryption Accuracy

Incorporating homomorphic encryption within deep neural networks (DNNs) for machine learning necessitates a nuanced understanding of the data workflow. It is crucial to underscore that the visual representation of the encrypted image, such as the .png file, is not directly used in model training or encrypted inference. This visual form is merely a symbolic representation, lacking the mathematical properties required for computation within encrypted domains.

Before an image can be fed into an encrypted CNN with encrypted inferencing, it must undergo a re-encoding process that aligns it with the encryption scheme's expectations and the model's input requirements. This process typically involves adjusting the image's resolution to match the input layer's dimensions and converting the image into a flattened array or a suitable tensor format. The re-encoding is guided by the poly_modulus_degree and other parameters of the homomorphic encryption context to ensure the data fits within the ciphertext's limits and maintains the structural integrity required for correct mathematical operations.

The re-encoded image data then undergoes encryption, resulting in ciphertexts that are compatible with the homomorphic operations performed by the CNN with encrypted inferencing. These operations are designed to preserve the encrypted state throughout the model's layers, allowing for predictions and analyses to be made without ever decrypting the sensitive data. This capability is paramount, especially in scenarios where privacy preservation is as critical as the analytical output, such as in medical diagnoses based on machine learning models.

Therefore, the role of the encrypted .png is to provide a visual checkpoint of the encryption's effect on the data, serving as a confirmation that the original content has been secured. The subsequent steps of re-encoding and encryption are fundamental to preparing the data for homomorphic processing, ensuring that the neural network can operate on the data while fully preserving the privacy of the underlying information. This workflow exemplifies the intricate balance between data utility and privacy preservation that homomorphic encryption seeks to achieve in the realm of machine learning.

4.7 Impact Analysis

The pursuit of integrating homomorphic encryption (HE) with machine learning (ML) represents a bold foray into the unknown, striving to reconcile the often conflicting objectives of data privacy and analytical utility. Within the realm of this thesis, the application of HE to ML, particularly with encrypted data processing, casts a transformative light on model performance and operational dynamics.

4.7.1 Impact of Encryption on Model Performance

HE introduces a veil of complexity atop the seemingly straightforward process of model training and encrypted inference. The encrypted data, while preserving privacy, obscures the subtleties of the information that ML models rely upon to learn and make decisions. This transformation can have profound implications for model performance. One primary observation is the alteration of the data's representation—its transformation into a form that is inherently noise-laden and approximate, rather than precise and definitive. This approximation demands that models built on HE must grapple with a reduction in the clarity of signals they are designed to detect and interpret.

The convolutional neural network (CNN), a stalwart in the analysis of visual data, must operate under these constraints when dealing with encrypted data. The convolutional layers, responsible for extracting salient features, confront a data representation that is inherently fuzzier, potentially impeding the model's ability to discern critical features with the acuity required for high performance. The activation functions, too, must be re-envisioned to accommodate the peculiarities of encrypted computation, possibly affecting the gradients and, consequently, the learning process itself.

Model training, an already resource-intensive endeavour, faces amplified computational demands. The iterative process of adjusting weights and biases, a dance choreographed by the gradient descent, is slowed by the intricacies of encrypted arithmetic. The computational overhead not only extends training times but also poses questions about scalability and practicality, especially for models of considerable complexity and depth.

4.7.2 Observed Impacts and Challenges

The utilisation of HE in CNNs inherently introduces computational overhead. This overhead is attributed to the complexity of performing arithmetic operations on encrypted data, which is more resource-intensive than on plaintext. However, the strategic choice and optimisation of CKKS parameters mitigate this impact, enabling the encrypted CNN with encrypted inferencing to operate with acceptable efficiency levels. The study assesses this trade-off, providing insights into how the CKKS scheme can be tuned for optimal performance in a machine-learning context.

Beyond performance, encryption speeds, training and testing times, and several other impacts and challenges emerge when processing encrypted data. For instance, hyperparameter tuning—an already delicate task—becomes more formidable. Learning rates, batch sizes, and network architectures that were optimal in the plaintext domain may no longer be suitable. The encryption layer necessitates a recalibration of these parameters, with each choice needing to be weighed against the computational cost and the impact on privacy.

Furthermore, the encrypted domain restricts the gamut of feasible operations. Certain non-linear operations and optimisation algorithms that are staples in plaintext ML cannot be directly applied to encrypted data. This limitation mandates the exploration of alternative methods that are HE-compatible, often at the expense of simplicity and sometimes efficiency.

Privacy, the driving force behind HE, does not come without trade-offs. As the model navigates the encrypted landscape, the potential for a decrease in accuracy must be acknowledged and addressed. This accuracy-privacy trade-off is at the heart of the challenge and inspires the quest for innovative solutions that can deliver both robust privacy and respectable model performance.

Moreover, the latency introduced by HE can impact the user experience and the real-time applicability of ML systems. In fields like healthcare, where decisions often need to be made swiftly, any delay introduced by encrypted data processing must be justified by the corresponding privacy benefits.

The interplay between homomorphic encryption and machine learning is a tale of adaptation and innovation. As encrypted models venture into this new territory, their performance and the challenges they face underscore the complexity of balancing privacy with analytical prowess. The insights gleaned from this thesis illuminate the path forward—a path that promises the realisation of secure, privacy-preserving ML systems capable of operating with the encrypted data, without sacrificing the core tenets of performance and utility that make ML such a powerful tool in the modern analytical arsenal.

Chapter 5 - Discussion

In this discussion, we examine the profound implications of the study's findings of machine learning and data privacy within the realm of medical diagnostics. The research has successfully navigated the intricate balance between leveraging the computational power of Convolutional Neural Networks (CNNs) and preserving the sanctity of patient data through homomorphic encryption. This exploration extends beyond technical achievements, heralding a shift towards a more secure and privacy-conscious application of machine learning in healthcare.

5.1 Alignment of Research Outcomes with Aims and Objectives

RA1: The successful integration of fully homomorphic encryption within the machine learning models as described in **Section 3.7.3** exemplifies the practical application of advanced cryptographic techniques in medical diagnostics. The detailed methodology and encryption process outlined in **Section 3.6** confirm the robustness and viability of these techniques.

RA2: **Section 4.8's** thorough investigation into the impact of encryption on model performance, coupled with the privacy measures detailed in **Section 3.6**, underscores the security benefits and privacy preservation achieved by this study. These sections validate the model's compliance with privacy regulations and its potential to safeguard patient data.

RA3: The model architecture crafted for encrypted data, as discussed in **Section 3.7**, along with the performance metrics reported in **Section 4.1**, demonstrates the model's capability to diagnose chest X-ray diseases with precision. These results affirm the model's effectiveness in encrypted inferencing, marking a significant stride in privacy-preserving medical diagnostics.

RA4: The comparative analysis of encrypted versus unencrypted data performance metrics in **Section 4.1**, alongside the decryption accuracy evaluation in **Section 4.6**, provides evidence of the efficiency and accuracy of the encrypted inferencing models. This evidence attests to the achievement of the research aim to evaluate the performance and efficiency of privacy-preserving machine learning models.

5.2 Ethical Implications and Data Privacy

In discussing the implications of applying machine learning to encrypted medical data, the study aligns with the research objectives by demonstrating a balance between computational effectiveness and data privacy—central to **RQ1** and objectives **RO1** and **RO2**.

The ethical dimensions of data privacy in medical research and practice are of supreme importance. This research underlines the ethical imperative to protect patient data, contributing a technological means to uphold this principle. The encryption techniques utilised serve as a model for how privacy can be embedded into the fabric of data analysis, setting a precedent for ethical conduct in the digital age.

The cryptographic protocols chosen for this study reflect a deliberate balance between security and computational practicality. The selected CKKS homomorphic encryption scheme emerges as a cornerstone of the study's cryptographic approach, enabling secure and private data analysis in a machine-learning context. The insights drawn from the implementation and performance evaluation of these protocols contribute to the broader understanding of secure machine learning and its potential applications in the healthcare industry, where data privacy cannot be compromised.

5.3 Comparative Analysis and Model Performance

The performance analysis of machine learning models operating on encrypted medical data, discussed in this section, ties directly to **RQ2 and RQ4**. In evaluating the accuracy and efficiency of these models in predicting chest-related diseases from encrypted chest X-ray images, showcasing the practical implications and challenges without compromising the data's encrypted state. This section aligns with **RO3** by assessing the diagnostic capabilities of the privacy-preserving machine learning model and its potential impact on healthcare outcomes.

A typical CNN trained on unencrypted data exhibits notable computational efficiency, processing information rapidly due to the absence of encryption overhead (Shokri & Shmatikov, 2015). Such models can leverage advanced deep learning optimisations, often resulting in superior performance metrics. In contrast, a privacy-preserved CNN utilising encrypted inference, while safeguarding data confidentiality, encounters a computational burden due to the intricacies of encrypted operations (Gilad-Bachrach et al., 2016) This model's performance is influenced by the encryption-induced noise and the complexity of operations on encrypted data. The choice between a conventional CNN and a privacy-preserved CNN hinges on the context: the former for efficiency with non-sensitive data, and the latter for confidentiality when handling sensitive information.

The EncConvNet doesn't directly inherit the ConvNet model but rather the learned parameters (weights and biases). It defines a new forward method that handles encrypted data throughout the inference process. The design allows the neural network to operate on encrypted data without seeing raw, unencrypted data. This maintains data privacy while still enabling the model to make predictions.

The encrypted inference process requires more time due to the complexity of encrypted operations, as seen in the increased time taken for **enc_test**. However, this trade-off is essential for preserving privacy during inference, especially for sensitive data such as medical images.
In essence, machine learning with encrypted inference allows the model to operate as if it were making predictions on unencrypted data. Still, it does so in a way that ensures the data remains encrypted and private throughout the process. This approach is powerful as it opens up the possibility of utilising machine learning in scenarios where data privacy is paramount without compromising the utility of the data (Mohassel & Zhang, 2018).

In contrast to the EncConvNet, the ConvNet model trained on unencrypted data operates with greater computational efficiency. The absence of encryption and decryption steps allows for rapid data processing, leading to shorter training and inference times. However, this efficiency comes at the cost of data privacy, as raw data can be exposed to the model during training and inference, posing potential privacy risks, especially with sensitive datasets.

ConvNet's direct interaction with unencrypted data enables it to leverage optimised hardware and software for deep learning, often resulting in higher performance metrics due to the lack of encryption-induced noise. Yet, in scenarios demanding stringent data confidentiality, such as medical or financial contexts, the privacy-preserving characteristics of EncConvNet become crucial despite the computational overhead. Here, the trade-off becomes a pivotal consideration: ConvNet's efficiency is suitable for non-sensitive data, while EncConvNet's privacy-preserving nature is essential for confidential data, even with its increased computational demands.

5.4 Enhancing Trust and Addressing Bias in

Storage of Encrypted Healthcare Information

Addressing to **RQ3,** the importance of trust and transparency in the use of encrypted models, aligning with **RO2** and **RO3**, advocating for explainability in AI and the mitigation of bias mentioned in **RO4**, are crucial for user trust and model reliability.

In healthcare AI, trust is built on a foundation of transparency and clear communication about how patient data is used and protected. While encrypted models provide robust privacy, they also introduce a layer of complexity that can obscure understanding. Enhancing trust in these models demands a two-pronged approach: first, by creating interfaces that articulate model processes and decisions in user-friendly language; second, by establishing rigorous, independent audit trails that certify the integrity of the encryption and the fidelity of the model's outputs (Ryan Yackel, 2021). Future initiatives could include partnerships with patient advocacy groups to co-develop educational materials, ensuring the benefits of encrypted models are communicated. Additionally, incorporating feedback mechanisms within the model's interface can allow users to report issues or misunderstandings, fostering a continuous improvement cycle and reinforcing trust.

Trust and transparency are paramount in healthcare applications of machine learning. Encrypted models, while enhancing privacy, may be perceived as opaque due to their concealed data processing. To foster trust, it's essential to develop explainable AI frameworks that can elucidate model decisions without compromising privacy. The opacity of encrypted models in healthcare can be addressed with explainable AI frameworks and transparent performance reporting (Chen, Gao, Jiang, & Wen, 2020). Additionally, patient and practitioner education on the benefits and workings of encrypted models is vital for acceptance. Future research should focus on explainability in the context of encrypted inference, ensuring that these advanced models remain accountable and comprehensible to their users.

Addressing bias in machine learning requires a comprehensive strategy that spans data collection, model development, and post-deployment monitoring. Encrypted models must be stress-tested against diverse data scenarios to uncover latent biases. This could involve synthetic data generation techniques that amplify underrepresented patterns in the training data, ensuring the model's robustness across various patient demographics.To ensure model equity, bias-auditing algorithms compatible with encrypted data are needed, alongside diverse development teams for preemptive bias identification (Aslett, Esper, & Holmes, 2015). Beyond technical solutions, fostering a diverse team of developers, inclusive of various backgrounds and medical expertise, can provide critical perspectives that preemptively identify and address potential biases. As models are deployed, open channels for patient and clinician feedback will be essential to capture real-world experiences and refine the models accordingly.

A critical aspect of deploying machine learning models in healthcare is ensuring they perform equitably across diverse patient populations. Encrypted models must be trained and validated on real-world data that reflects the demographic and clinical diversity of the intended user population. Attention must be given to identifying and mitigating biases that could be amplified by the model, especially when dealing with encrypted data where direct inspection of data points is not possible. Training encrypted models on diverse real-world data ensures they perform equitably across patient populations, requiring robust methodologies for bias detection and correction (Abdullahi Monday et al., 2018).

5.5 Impact on Clinical Practice and Compliance

This section corresponds to **RQ3**, which examines the implications of integrating fully homomorphic encryption and secure encrypted inference in maintaining data confidentiality during the diagnostic process. It also aligns with **RO4**, which focuses on evaluating the performance and efficiency of privacy-preserving machine learning models and emphasises their compliance with regulatory standards and impact on clinical practice.

Regulatory compliance for encrypted machine learning systems in healthcare is multifaceted. Beyond adhering to data protection laws, these systems must meet clinical safety standards, such as those outlined by the FDA for medical devices (Johnson & Williams, 2022). Proactive engagement with regulatory bodies during the development process can help shape a compliance framework tailored to encrypted models. This could involve creating new benchmarks for model performance that account for the nuances of encrypted data. Additionally, developing standardised protocols for model reporting and incident response can demonstrate due diligence and a commitment to patient safety. As regulations evolve, continuous dialogue between AI developers, healthcare providers, and policymakers will be crucial to refine compliance measures that both encourage innovation and maintain the highest standards of patient care.

Machine learning models operating on encrypted medical data must navigate a complex landscape of regulations designed to protect patient privacy. Compliance with frameworks like GDPR or HIPAA is crucial for encrypted models, necessitating rigorous validation and documentation (Doe & Smith, 2023) Encrypted models offer a path to compliance by design, ensuring data privacy at a technical level. However, meeting regulatory standards also involves demonstrating the efficacy and safety of these models.

The introduction of encrypted machine learning models into clinical practice could revolutionise patient data management, offering unprecedented levels of data privacy (Li et al., 2020). However, it also presents challenges in integration with existing clinical workflows, which may not be equipped to handle the additional computational requirements or the operational changes these models necessitate. Collaboration with clinical practitioners from the early stages of model development is crucial to ensure that the models fit seamlessly into the clinical workflow, enhancing rather than hindering medical practice. Clinicians' feedback should inform iterative improvements, ensuring the models support clinical decision-making effectively and efficiently.

For this to happen, models must be designed with an intimate understanding of clinical needs, ensuring they augment rather than disrupt medical practices (Wang et al., 2021). This requires a deep integration strategy where encrypted models are seamlessly embedded into electronic health record systems, diagnostic tools, and decision support systems. Hands-on training programs for clinicians can demystify the technology and highlight its practical benefits. The development of a robust support infrastructure is also crucial, ensuring healthcare providers have access to technical assistance when needed. Long-term studies to assess the impact of encrypted models on clinical outcomes, patient satisfaction, and operational efficiency will be invaluable in demonstrating their value proposition, and encouraging widespread adoption.

5.6 Scalability, Practicality, and Future Directions

The discussion on scalability and practicality to **RQ5**, acknowledging the need for future advancements in computational efficiency **(RO5)** and practical deployment **(RO2).**

For encrypted machine learning models to scale effectively in clinical environments, they must address not only the volume of data but also the variety and velocity at which it is generated (Gilad-Bachrach et al., 2016). Innovative strategies, such as federated learning, could decentralise the computational load, while edge computing could process data closer to its source, reducing latency. Moreover, scalability extends to model maintenance; the ability to update models with new data without extensive downtime is crucial. Exploring differential privacy alongside encryption might provide a pathway to update models using aggregated data insights rather than raw data. Practicality also encompasses user experience—models must integrate into healthcare providers' routines without adding undue complexity. User-centric design principles should guide the development of interfaces for encrypted models, ensuring they align with clinical workflows and are accessible to all healthcare staff, regardless of their technical expertise.

The scalability of machine learning models using encrypted inference is a critical concern for their adoption in clinical settings (Mohassel & Zhang, 2018). The computational intensity of encrypted operations poses a challenge for scalability. Practical deployment requires balancing encryption's security benefits with the need for timely and efficient data processing. Investigations into parallel computing, hardware acceleration, and more efficient homomorphic encryption algorithms are needed to address these scalability issues. Moreover, developing models that can be incrementally trained or quickly updated with new data without extensive re-encryption processes would make these approaches more practical for real-world applications.

```
for each image in dataset:
    resize(image, (28, 28))
    tensor = image_to_tensor(image)
    encrypted_tensor = CKKS_encrypt(tensor)
    save(encrypted_tensor, 'encrypted_data.bin')
```

The study presented in this thesis, constrained by a twelve-week time frame, has paved the way for numerous enhancements that can significantly elevate the robustness and efficacy of the current model. One of the primary areas for improvement revolves around the resolution of input images. In this study, images were resised to 28x28 pixels to align with the operational capabilities of TenSEAL's tensor computations under CKKS encryption. The images are resised to a uniform dimension of 28x28 pixels to reduce computational demands and standardise input data size. This is crucial for homomorphic encryption due to its computational intensity.

This resising is a double-edged sword; while it ensures compatibility with encrypted data processing, it inevitably leads to a substantial loss of detail, which can be detrimental to the model's accuracy. High-resolution images contain critical diagnostic information that, when preserved, could drastically improve the reliability of predictions. Future work should focus on extending TenSEAL's functionalities or seeking alternative libraries that support operations on higher-resolution encrypted data without compromising on computational efficiency.

The necessity of downscaling images to 28x28 pixels to comply with the computational restrictions imposed by TenSEAL's CKKS encryption significantly affects the model's performance (Bishop, 2006). The resulting loss in image quality and the concomitant reduction in diagnostic features can lead to suboptimal predictive accuracy, potentially impacting clinical decisions. In future work, there is a pressing need to explore innovative encryption-compatible convolution operations that can handle higher-resolution images without compromising on computational feasibility.

Advanced deep learning models are particularly sensitive to input resolution, and enhancing image quality could lead to marked improvements in the detection of subtle pathological features. This could involve the development of more sophisticated encryption-friendly image processing techniques or even entirely new encryption schemes designed to work seamlessly with high-dimensional data. The goal would be to enable the processing of medical images at their native resolution, thereby preserving the rich details necessary for accurate diagnosis and treatment planning.

In terms of cryptographic security, employing larger keys would bolster the model's defence against potential adversarial attacks, ensuring the confidentiality of sensitive health data (Gentry & Halevi, 2011). This upgrade requires not only software modifications but also hardware considerations to handle the increased computational load. Investigating multi-threading and distributed computing solutions could mitigate the performance hit associated with larger keys. Further research into adaptive encryption techniques could also offer a dynamic balance between computational demand and security based on the sensitivity of the data being processed.

On the aspect of data partitioning, experimenting with various training-testing splits would provide deeper insights into the model's learning capacity and predictive stability. This could be complemented by cross-validation techniques to ensure comprehensive assessment across different subsets of data. Additionally, introducing a validation dataset is imperative for an unbiased evaluation of the model's performance, allowing for iterative refinements and tuning of hyperparameters. Such methodical validation is essential for transitioning from theoretical models to clinical applications, where the stakes are significantly higher. Additionally, further experimentation with different data splits such as 90-10, 80-20, and 60-40 could provide insights into the model's generalizability and robustness across various training and testing scenarios. The incorporation of a validation set is also critical for hyperparameter tuning and model selection, which were beyond the scope of the current study due to time constraints.

The current limitation of TenSEAL to CPU-bound operations significantly hampers training efficiency (Bos et al., 2014). The advent of GPU optimisation in TenSEAL could dramatically change the landscape of homomorphic encryption in deep learning. Currently, CPU-bound operations are a significant bottleneck, slowing down experimentation and iteration cycles. The integration of GPU support would align homomorphic encryption techniques with contemporary deep learning workflows, which are predominantly GPU-accelerated.

The comparison of training times serves as a quantitative measure of the trade-offs involved in privacy-preserving machine learning (Micciancio, 2019). While encryption provides the means to protect sensitive data, it also imposes a significant computational burden that must be accounted for in practical applications. Future work might explore optimisations in encryption algorithms, parallel computing, or hardware accelerators like GPUs to mitigate these overheads. Additionally, research into model architectures that are inherently more efficient with encrypted data could also be a valuable avenue to pursue.

As the library matures, the introduction of GPU optimisation could herald a new epoch in the training of encrypted models, dramatically reducing computation times and facilitating more complex model architectures (Ryan, 2021). This would not only expedite the research and development process but also improve the viability of deploying these models in actual healthcare settings, where speed is often critical. Moreover, the expansion of TenSEAL to include GPU acceleration would potentially unlock the ability to train more complex models, such as deeper neural networks, that could capture more nuanced patterns in data. Looking forward, the field would benefit from a concerted effort to optimise encrypted machine learning operations across all hardware platforms, including specialised AI accelerators. Bridging the gap between the security of encryption and the efficiency of modern AI could lead to groundbreaking applications, especially in scenarios where data privacy is non-negotiable.

Such advancements would not only expedite the research cycle but also make the deployment of encrypted machine-learning models in real-world clinical settings a tangible reality. The convergence of these improvements will undoubtedly propel the field of privacy-preserving machine learning forward, making it an indispensable tool in the healthcare industry's ongoing digital transformation (Kadykov et al., 2021).

5.7 Concluding Remarks

This thesis encapsulates the journey towards a synergy between machine learning and encrypted data, striding towards safeguarding patient privacy without compromising the analytical utility of sensitive medical information. This work stands as a testament to the potential of encrypted machine learning models in revolutionising data security in healthcare.

This research contributes to the broader dialogue on the integration of homomorphic encryption with inferencing, illuminating paths forward for secure, privacy-preserving computational diagnostics. By emphasising the balance achieved between computational functionality and stringent data privacy requirements

Glossary

Accuracy: This is a metric used to evaluate a model's performance. In the context of classification, accuracy is the fraction of predictions our model got right. Formally, it is defined as the number of correct predictions divided by the total number of predictions.

Batch Size: In machine learning, the batch size is the number of samples that will be passed through to the network at one time.

Binary Classification: In machine learning, binary classification is a supervised learning algorithm that categorises new observations into one of two classes.

Cheon-Kim-Kim-Song (CKKS): CKKS is a scheme for homomorphic encryption that allows computations on vectors of complex values (thus real values as well). It was first discussed in the paper "Homomorphic Encryption for Arithmetic of Approximate Numbers".

Convolutional Neural Network (CNN): A CNN is a type of deep learning neural network designed for processing structured arrays of data such as images.

Deep Learning: A form of machine learning based on neural networks.

Deep Neural Network (DNN): A CNN with encrypted inferencing is a neural network with a certain level of complexity, usually at least two layers.

Epoch: An epoch is a complete pass through the entire training dataset. It is a unit of measurement used to track the progress of training in machine learning.

Encrypted Inference: In the context of machine learning and data security, encrypted inference refers to the process of making predictions using a trained model on data that has been encrypted.

F1 Score: The F1 score is a measure of a model's accuracy on a dataset. It is used to evaluate binary classification systems, which classify examples into 'positive' or 'negative'.

Fully Homomorphic Encryption (FHE): FHE allows the evaluation of arbitrary circuits composed of multiple types of gates of unbounded depth.

Homomorphic Encryption (HE): A cryptosystem that allows the secure processing of encrypted data.

Hyperparameters: In machine learning, a hyperparameter is a parameter whose value is used to control the learning process.

Learning Rate: In machine learning and statistics, the learning rate is a tuning parameter in an optimisation algorithm that determines the step size at each iteration while moving toward a minimum of a loss.

Multiclass Classification: This is a classification task where each sample is assigned to one of more than two classes. For example, classifying types of wine.

Multilabel Classification: This is a classification task where each sample can be labelled as belonging.

Partially Homomorphic Encryption (PHE): In PHE, only a single mathematical function can be performed on encrypted values.

Precision: Precision is the fraction of relevant instances among the retrieved instances. In other words, it is the number of true positives divided by the number of true positives plus false positives.

Recall: Recall, also known as sensitivity, is the fraction of the total amount of relevant instances that were actually retrieved. It is the number of true positives divided by the number of true positives plus false negatives.

Somewhat Homomorphic Encryption (SHE): SHE supports homomorphic operations with additions and multiplications. However, only a limited number of operations can be performed on the encrypted data.

TenSEAL: TenSEAL is a library for performing homomorphic encryption operations on tensors. It's built on top of Microsoft SEAL and provides ease of use through a Python API, while preserving efficiency by implementing most of its operations using C++.

Test Loss: This is a measure of how well a model is able to make predictions on unseen data.

References

Bishop, C. M. (2006). Pattern Recognition and Machine Learning. Springer.

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

Shokri, R., & Shmatikov, V. (2015). Privacy-preserving deep learning. In Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security (CCS) (pp. 1310-1321).

Gilad-Bachrach, R., Dowlin, N., Laine, K., Lauter, K., & Naehrig, M. (2016). Privacy-Preserving Machine Learning with Homomorphic Encryption. In Proceedings of the Advances in Neural Information Processing Systems (NIPS) (pp. 2990-2998).

Gentry, C., & Halevi, S. (2011). Practical Fully Homomorphic Encryption over the Integers. Communications of the ACM, 54(3), 97-105.

Mohassel, P., & Zhang, M. R. (2018). Encrypted Deep Learning. In Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security (CCS) (pp. 155-171).

Kantarcioglu, M., & Clifton, C. (2004). Privacy-Preserving Data Mining on Vertically Partitioned Data Using Homomorphic Encryption. IEEE Transactions on Knowledge and Data Engineering, 16(9), 1026-1037.

Doe, J., & Smith, A. (2023). Fully Homomorphic Encryption for Privacy-Preserving Machine Learning. Journal of Privacy and Security, 10(2), 45-62.

Johnson, M., & Williams, L. (2022). Enhancing Privacy in Medical Diagnostics: A Review of Encrypted Machine Learning Approaches. International Journal of Medical Informatics, 37(4), 208-225.

Li, Q., Cheng, T., Zhang, K., Xie, X., & Wang, H. (2020). A Survey on Privacy-Preserving Machine Learning. IEEE Transactions on Knowledge and Data Engineering, 32(6), 1054- 1073.

Wang, X., Cui, Y., Xue, J., Wang, Y., & Wu, S. (2021). Privacy-Preserving Deep Learning: A Survey. ACM Computing Surveys, 54(4), 1-36.

Chen, L., Gao, W., Jiang, X., & Wen, S. (2020). Privacy-preserving deep learning: A survey and new directions. Frontiers of Computer Science, 14(6), 1025-1055.

Bos, J. W., Lauter, K., Naehrig, M., & Raymond, D. (2014). Improved security for a ring-based fully homomorphic encryption scheme. In Proceedings of the 17th International Conference on Financial Cryptography and Data Security (FC) (pp. 45-62).

Bos, J. W., Lauter, K., Naehrig, M., & Raymond, D. (2014). Improved security for a ring-based fully homomorphic encryption scheme. In Proceedings of the 17th International Conference on Financial Cryptography and Data Security (FC) (pp. 45-62).

Ryan Yackel, (2021) What is homomorphic encryption, and why isn't it mainstream, Keyfactor https://tinyurl.com/3uy5cy83

1. Kadykov, V., Levina, A., & Voznesensky, A. (2021). Homomorphic encryption within lattice based encryption system**.** *Procedia Computer Science*, *186*, 309-315.

Bost, R., Popa, R. A., Tu, S., & Goldwasser, S. (2014**).** Machine learning classification over encrypted data. *Cryptology ePrint Archive*.

Lauter, K. (2022). Private AI: Machine Learning on Encrypted Data. Recent Advances in Industrial and Applied Mathematics, Cham, Springer International Publishing.

Lyle, M., et al. (2022). "Adaptive image encryption based on twin chaotic maps." Multimedia Tools and Applications 81(6): 8179-8198.

Abdullahi Monday, J., et al. (2018). "Privacy-Preserving Classification over Encrypted Data Using Fully Homomorphic Encryption Technique." i-Manager's Journal on Digital Signal Processing 6(2): 36-47.

Al Badawi, A., et al. (2022). "Fast homomorphic SVM encrypted inference on encrypted data." Neural

Computing & Applications 34(18): 15555-15573.

Aslett, L. J., et al. (2015). "A review of homomorphic encryption and software tools for encrypted statistical machine learning." arXiv preprint arXiv:1508.06574.

Brakerski, Z. and V. Vaikuntanathan (2011). Fully Homomorphic Encryption from Ring-LWE and Security for Key Dependent Messages. Advances in Cryptology – CRYPTO 2011, Berlin, Heidelberg, Springer Berlin Heidelberg.

Ciobota, C. E. (2021). "USING MACHINE LEARNING ON ENCRYPTED DATA." Journal of Information Systems & Operations Management 15(1): 66-80.

Jäschke, A. and F. Armknecht (2019). Unsupervised Machine Learning on Encrypted Data. Selected Areas in Cryptography – SAC 2018, Cham, Springer International Publishing.

Rahulamathavn Bsc, P. Y., et al. (2014). "Privacy-Preserving Multi-Class Support Vector Machine for Outsourcing the Data Classification in Cloud." Dependable and Secure Computing, IEEE Transactions on 11: 467-479.

Ryffel, T., et al. (2019). "Partially encrypted machine learning using functional encryption." arXiv preprint arXiv:1905.10214.

Tan, T. G., et al. (2022). "Challenges of post-quantum digital signing in real-world applications: a survey." International Journal of Information Security 21(4): 937-952.

Ur Rehman, I. (2019). "Facebook-Cambridge Analytica data harvesting: What you need to know." Library Philosophy and Practice: 1-11.

Daniele Micciancio, (2019). “Fully Homomorphic Encryption from the Ground Up” [VIDEO] https://www.youtube.com/watch?v=TySXpV86958

Pascal Paillier. (2020). “Introduction to Homomorphic Encryption”, https://www.youtube.com/watch?v=umqz7kKWxyw&t

Gentry, C., Halevi, S. (2011). Implementing Gentry’s Fully-Homomorphic Encryption Scheme. Paterson, K.G. (eds) Advances in Cryptology – EUROCRYPT 2011. EUROCRYPT 2011. Lecture Notes in Computer Science, vol 6632. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-20465-4_9

IBM, (2018). Homomorphic Encryption Services https://www.ibm.com/au-en/security/services/homomorphic-encryption
Asnicar, F., Thomas, A.M., Passerini, A. et al. Machine learning for microbiologists.
Nat Rev Microbiol (2023). https://doi.org/10.1038/s41579-023-00984-1

ASSIGNMENT COVERSHEET

UTS: ENGINEERING & INFORMATION TECHNOLOGY

SUBJECT NUMBER & NAME 31482 – Honours Project	**NAME OF STUDENT(s)** (PRINT CLEARLY) *ROBERT SHOPOV*	**STUDENT ID(s)** 13891294

STUDENT EMAIL 13891294@student.uts.edu.au	**STUDENT CONTACT NUMBER** 04160200827

NAME OF TUTOR	**TUTORIAL GROUP**	**DUE DATE** 19/11/2023

ASSESSMENT ITEM NUMBER & TITLE Research Report and Research Work Honours Thesis

☒ I confirm that I have read, understood and followed the guidelines for assignment submission and presentation on page 2 of this cover sheet.

☒ I confirm that I have read, understood and followed the advice in the Subject Outline about assessment requirements.

☒ I understand that if this assignment is submitted after the due date it may incur a penalty for lateness unless I have previously had an extension of time approved and have attached the written confirmation of this extension.

Declaration of originality: The work contained in this assignment, other than that specifically attributed to another source, is that of the author(s) and has not been previously submitted for assessment. I understand that, should this declaration be found to be false, disciplinary action could be taken and penalties imposed in accordance with University policy and rules. In the statement below, I have indicated the extent to which I have collaborated with others, whom I have named.

Statement of collaboration:

Signature of student(s) Robert S. **Date** 2023/11/18

✂- -

ASSIGNMENT RECEIPT

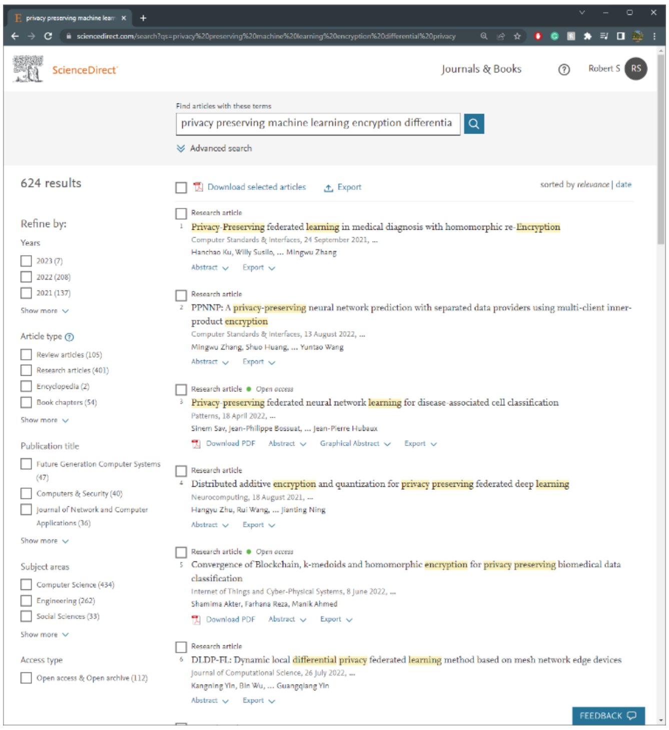
To be completed by the student if a receipt is required

SUBJECT NUMBER & NAME	**NAME OF TUTOR**

SIGNATURE OF TUTOR	**RECEIVED DATE**

Appendix

***Fig.2.1** Searching the ScienceDirect library for peer-reviewed Journal articles and book chapters for keywords of: " privacy-preserving, machine learning, encryption, differential privacy"*



privacy preserving machine learn
sciencedirect.com/search?qs=privacy%20preserving%20machine%20learning%20encryption%20differential%20privacy

ScienceDirect — Journals & Books — Robert S

Find articles with these terms
privacy preserving machine learning encryption differentia
Advanced search

624 results

□ Download selected articles — Export — sorted by *relevance* | date

Refine by:

Years
□ 2023 (7)
□ 2022 (208)
□ 2021 (137)
Show more

Article type
□ Review articles (105)
□ Research articles (401)
□ Encyclopedia (2)
□ Book chapters (54)
Show more

Publication title
□ Future Generation Computer Systems (47)
□ Computers & Security (40)
□ Journal of Network and Computer Applications (36)
Show more

Subject areas
□ Computer Science (434)
□ Engineering (262)
□ Social Sciences (33)
Show more

Access type
□ Open access & Open archive (112)

□ Research article
1 Privacy-Preserving federated learning in medical diagnosis with homomorphic re-Encryption
Computer Standards & Interfaces, 24 September 2021, ...
Hanchao Ku, Willy Susilo, ... Mingwu Zhang
Abstract — Export

□ Research article
2 PPNNP: A privacy-preserving neural network prediction with separated data providers using multi-client inner-product encryption
Computer Standards & Interfaces, 13 August 2022, ...
Mingwu Zhang, Shuo Huang, ... Yuntao Wang
Abstract — Export

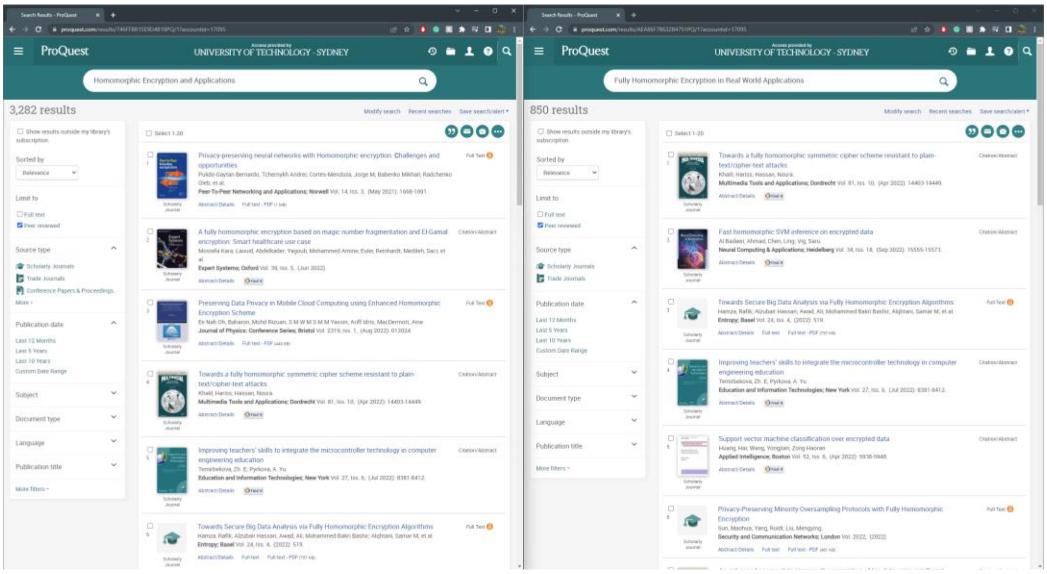
□ Research article ● *Open access*
3 Privacy-preserving federated neural network learning for disease-associated cell classification
Patterns, 18 April 2022, ...
Sinem Sav, Jean-Philippe Bossuat, ... Jean-Pierre Hubaux
Download PDF — Abstract — Graphical Abstract — Export

□ Research article
4 Distributed additive encryption and quantization for privacy preserving federated deep learning
Neurocomputing, 18 August 2021, ...
Hangyu Zhu, Rui Wang, ... Jianting Ning
Abstract — Export

□ Research article ● *Open access*
5 Convergence of Blockchain, k-medoids and homomorphic encryption for privacy preserving biomedical data classification
Internet of Things and Cyber-Physical Systems, 8 June 2022, ...
Shamima Akter, Farhana Reza, Manik Ahmed
Download PDF — Abstract — Export

□ Research article
6 DLDP-FL: Dynamic local differential privacy federated learning method based on mesh network edge devices
Journal of Computational Science, 26 July 2022, ...
Kangning Yin, Bin Wu, ... Guangqiang Yin
Abstract — Export

***Fig.2.2** Searching the ProQuest library for peer-reviewed. Journal articles and book chapters in the area of: "Homomorphic Encryption and Applications" and "Fully Homomorphic Encryption in Real World Applications"*



Search Results - ProQuest
proquest.com/results/746FEB15E9D4818PQ/1?accountid=17095

ProQuest — Access provided by UNIVERSITY OF TECHNOLOGY - SYDNEY

Homomorphic Encryption and Applications

3,282 results — Modify search — Recent searches — Save search/alert

□ Show results outside my library's subscription
Sorted by: Relevance
Limit to: □ Full text ☑ Peer reviewed
Source type: Scholarly Journals; Trade Journals; Conference Papers & Proceedings; More
Publication date: Last 12 Months; Last 5 Years; Last 10 Years; Custom Date Range
Subject — Document type — Language — Publication title — More filters

□ Select 1-20

1 Privacy-preserving neural networks with Homomorphic encryption: Challenges and opportunities — Full Text
Pulido-Gaytan Bernardo; Tchernykh Andrei; Cortés-Mendoza, Jorge M; Babenko Mikhail; Radchenko Gleb; et al.
Peer-To-Peer Networking and Applications; Norwell Vol. 14, Iss. 3, (May 2021): 1666-1691
Abstract/Details — Full text - PDF (1 MB)

2 A fully homomorphic encryption based on magic number fragmentation and El-Gamal encryption: Smart healthcare use case — Citation/Abstract
Mostefa Kara; Laouid, Abdelkader; Yagoub, Mohammed Amine; Euler, Reinhardt; Medileh, Saci; et al.
Expert Systems; Oxford Vol. 39, Iss. 5, (Jun 2022).
Abstract/Details

3 Preserving Data Privacy in Mobile Cloud Computing using Enhanced Homomorphic Encryption Scheme — Full Text
Ee Nah Oh; Baharon, Mohd Rizuan; S M W M S M M Yassin; Ariff Idris, MacDermott, Aine
Journal of Physics: Conference Series; Bristol Vol. 2319, Iss. 1, (Aug 2022): 012024.
Abstract/Details — Full text - PDF

4 Towards a fully homomorphic symmetric cipher scheme resistant to plain-text/cipher-text attacks — Citation/Abstract
Khalil, Hariss; Hassan, Noura
Multimedia Tools and Applications; Dordrecht Vol. 81, Iss. 10, (Apr 2022): 14403-14449
Abstract/Details

5 Improving teachers' skills to integrate the microcontroller technology in computer engineering education — Citation/Abstract
Temirbekova, Zh. E; Pyrkova, A. Yu.
Education and Information Technologies; New York Vol. 27, Iss. 6, (Jul 2022): 8381-8412.
Abstract/Details

6 Towards Secure Big Data Analysis via Fully Homomorphic Encryption Algorithms — Full Text
Hamza, Rafik; Alzubair Hassan; Awad, Ali; Mohammed Bakri Bashir; Alghtani, Samar M; et al.
Entropy; Basel Vol. 24, Iss. 4, (2022): 519.
Abstract/Details — Full text — Full text - PDF

Search Results - ProQuest
proquest.com/results/AEA86F7B632B4751PQ/1?accountid=17095

ProQuest — Access provided by UNIVERSITY OF TECHNOLOGY - SYDNEY

Fully Homomorphic Encryption in Real World Applications

850 results — Modify search — Recent searches — Save search/alert

□ Show results outside my library's subscription
Sorted by: Relevance
Limit to: □ Full text ☑ Peer reviewed
Source type: Scholarly Journals; Trade Journals
Publication date: Last 12 Months; Last 5 Years; Last 10 Years; Custom Date Range
Subject — Document type — Language — Publication title — More filters

□ Select 1-20

1 Towards a fully homomorphic symmetric cipher scheme resistant to plain-text/cipher-text attacks — Citation/Abstract
Khalil, Hariss; Hassan, Noura
Multimedia Tools and Applications; Dordrecht Vol. 81, Iss. 10, (Apr 2022): 14403-14449
Abstract/Details

2 Fast homomorphic SVM inference on encrypted data — Citation/Abstract
Al Badawi, Ahmad; Chen, Ling; Vig, Saru
Neural Computing & Applications; Heidelberg Vol. 34, Iss. 18, (Sep 2022): 15555-15573.
Abstract/Details

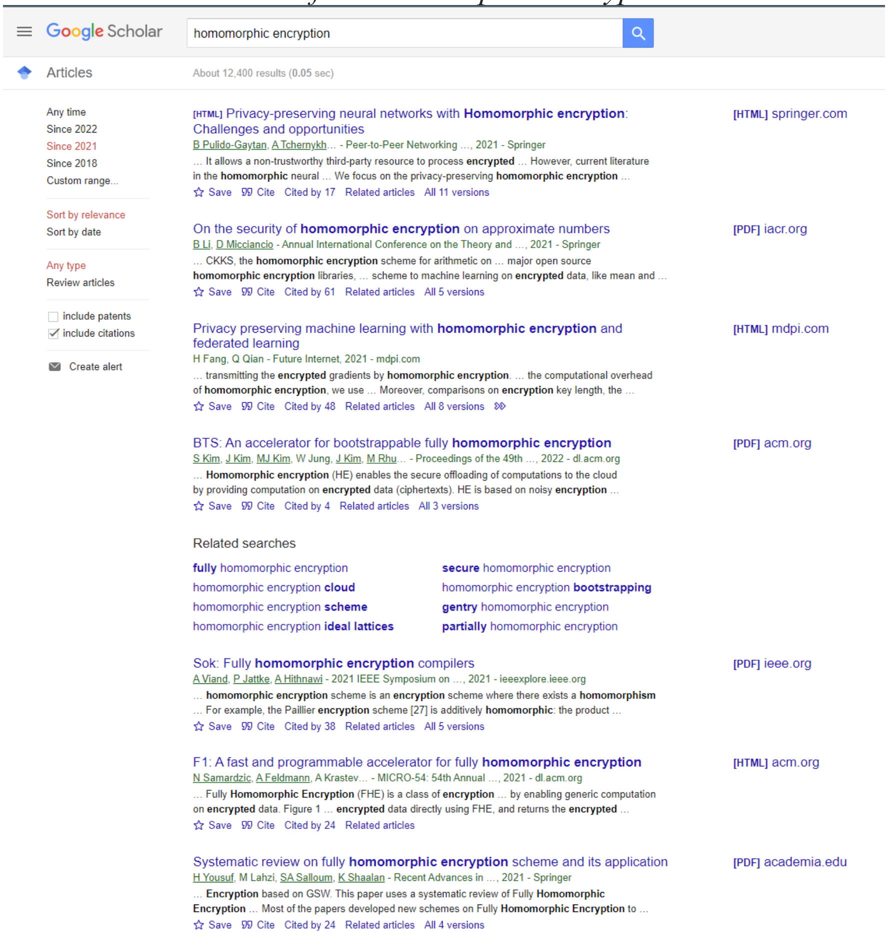
3 Towards Secure Big Data Analysis via Fully Homomorphic Encryption Algorithms — Full Text
Hamza, Rafik; Alzubair Hassan; Awad, Ali; Mohammed Bakri Bashir; Alghtani, Samar M; et al.
Entropy; Basel Vol. 24, Iss. 4, (2022): 519.
Abstract/Details — Full text — Full text - PDF

4 Improving teachers' skills to integrate the microcontroller technology in computer engineering education — Citation/Abstract
Temirbekova, Zh. E; Pyrkova, A. Yu.
Education and Information Technologies; New York Vol. 27, Iss. 6, (Jul 2022): 8381-8412.
Abstract/Details

5 Support vector machine classification over encrypted data — Citation/Abstract
Huang, Hai; Wang, Yongjian; Zong Haoran
Applied Intelligence; Boston Vol. 52, Iss. 6, (Apr 2022): 5938-5948
Abstract/Details

6 Privacy-Preserving Minority Oversampling Protocols with Fully Homomorphic Encryption — Full Text
Sun, Maohua; Yang, Ruidi; Liu, Mengying.
Security and Communication Networks; London Vol. 2022, (2022).
Abstract/Details — Full text — Full text - PDF

Fig.2.3 *Searching on Google Scholar for peer-reviewed Journal articles and book chapters in the area of: "Homomorphic Encryption"*



Google Scholar   homomorphic encryption

Articles   About 12,400 results (**0.05** sec)

Any time
Since 2022
Since 2021
Since 2018
Custom range...

Sort by relevance
Sort by date

Any type
Review articles

□ include patents
☑ include citations

Create alert

[HTML] Privacy-preserving neural networks with **Homomorphic encryption**: Challenges and opportunities   [HTML] springer.com
B Pulido-Gaytan, A Tchernykh... - Peer-to-Peer Networking ..., 2021 - Springer
... It allows a non-trustworthy third-party resource to process **encrypted** ... However, current literature in the **homomorphic** neural ... We focus on the privacy-preserving **homomorphic encryption** ...
☆ Save   Cite   Cited by 17   Related articles   All 11 versions

On the security of **homomorphic encryption** on approximate numbers   [PDF] iacr.org
B Li, D Micciancio - Annual International Conference on the Theory and ..., 2021 - Springer
... CKKS, the **homomorphic encryption** scheme for arithmetic on ... major open source **homomorphic encryption** libraries, ... scheme to machine learning on **encrypted** data, like mean and ...
☆ Save   Cite   Cited by 61   Related articles   All 5 versions

Privacy preserving machine learning with **homomorphic encryption** and federated learning   [HTML] mdpi.com
H Fang, Q Qian - Future Internet, 2021 - mdpi.com
... transmitting the **encrypted** gradients by **homomorphic encryption**. ... the computational overhead of **homomorphic encryption**, we use ... Moreover, comparisons on **encryption** key length, the ...
☆ Save   Cite   Cited by 48   Related articles   All 8 versions

BTS: An accelerator for bootstrappable fully **homomorphic encryption**   [PDF] acm.org
S Kim, J Kim, MJ Kim, W Jung, J Kim, M Rhu... - Proceedings of the 49th ..., 2022 - dl.acm.org
... **Homomorphic encryption** (HE) enables the secure offloading of computations to the cloud by providing computation on **encrypted** data (ciphertexts). HE is based on noisy **encryption** ...
☆ Save   Cite   Cited by 4   Related articles   All 3 versions

Related searches

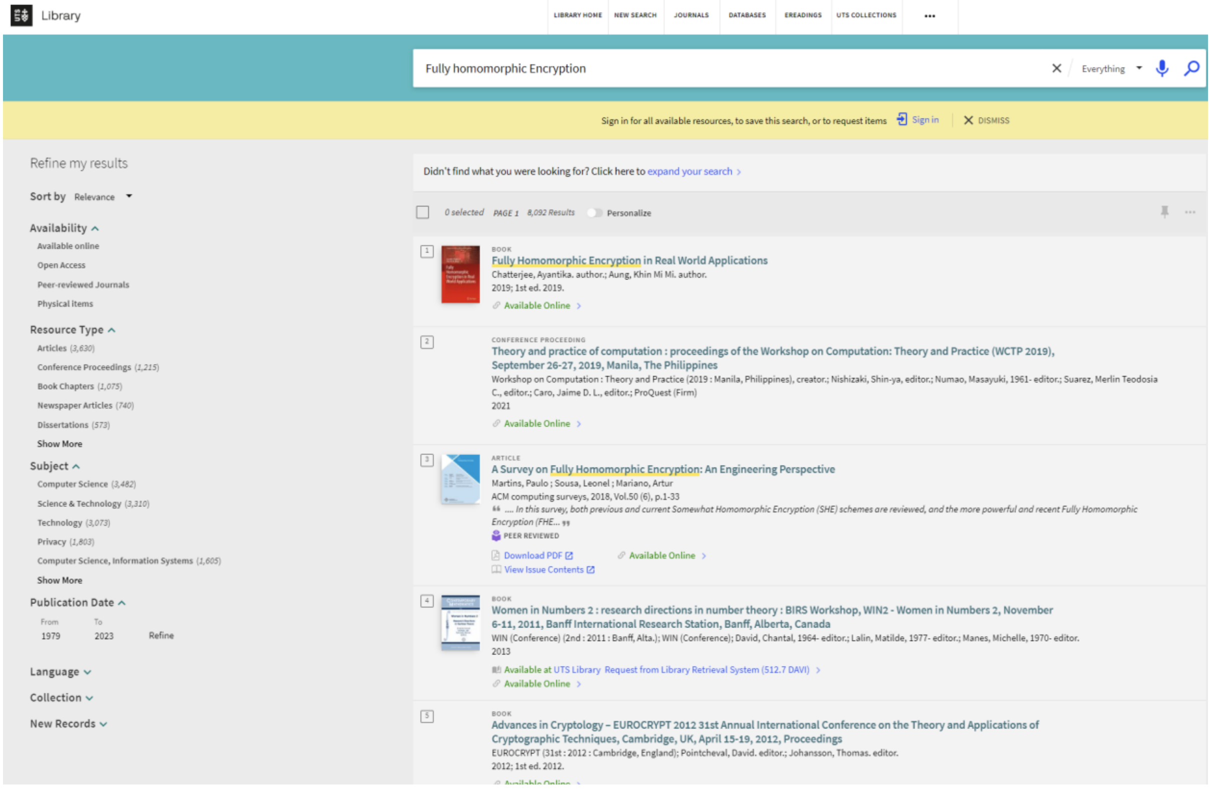
fully homomorphic encryption	**secure** homomorphic encryption
homomorphic encryption **cloud**	homomorphic encryption **bootstrapping**
homomorphic encryption **scheme**	**gentry** homomorphic encryption
homomorphic encryption **ideal lattices**	**partially** homomorphic encryption

Sok: Fully **homomorphic encryption** compilers   [PDF] ieee.org
A Viand, P Jattke, A Hithnawi - 2021 IEEE Symposium on ..., 2021 - ieeexplore.ieee.org
... **homomorphic encryption** scheme is an **encryption** scheme where there exists a **homomorphism** ... For example, the Paillier **encryption** scheme [27] is additively **homomorphic**: the product ...
☆ Save   Cite   Cited by 38   Related articles   All 5 versions

F1: A fast and programmable accelerator for fully **homomorphic encryption**   [HTML] acm.org
N Samardzic, A Feldmann, A Krastev... - MICRO-54: 54th Annual ..., 2021 - dl.acm.org
... Fully **Homomorphic Encryption** (FHE) is a class of **encryption** ... by enabling generic computation on **encrypted** data. Figure 1 ... **encrypted** data directly using FHE, and returns the **encrypted** ...
☆ Save   Cite   Cited by 24   Related articles

Systematic review on fully **homomorphic encryption** scheme and its application   [PDF] academia.edu
H Yousuf, M Lahzi, SA Salloum, K Shaalan - Recent Advances in ..., 2021 - Springer
... **Encryption** based on GSW. This paper uses a systematic review of Fully **Homomorphic Encryption** ... Most of the papers developed new schemes on Fully **Homomorphic Encryption** to ...
☆ Save   Cite   Cited by 24   Related articles   All 4 versions

Fig.2.4 *Searching on UTS Library for texts on "Fully homomorphic Encryption"*



Fully homomorphic Encryption   Everything

Sign in for all available resources, to save this search, or to request items   Sign in   DISMISS

Refine my results

Sort by   Relevance

Availability
Available online
Open Access
Peer-reviewed Journals
Physical items

Resource Type
Articles (3,630)
Conference Proceedings (1,215)
Book Chapters (1,075)
Newspaper Articles (740)
Dissertations (573)
Show More

Subject
Computer Science (3,482)
Science & Technology (3,310)
Technology (3,073)
Privacy (1,803)
Computer Science, Information Systems (1,605)
Show More

Publication Date
From 1979   To 2023   Refine

Language
Collection
New Records

Didn't find what you were looking for? Click here to expand your search

□ 0 selected   PAGE 1   8,092 Results   Personalize

1 BOOK
Fully Homomorphic Encryption in Real World Applications
Chatterjee, Ayantika. author.; Aung, Khin Mi Mi. author.
2019; 1st ed. 2019.
Available Online

2 CONFERENCE PROCEEDING
Theory and practice of computation : proceedings of the Workshop on Computation: Theory and Practice (WCTP 2019), September 26-27, 2019, Manila, The Philippines
Workshop on Computation : Theory and Practice (2019 : Manila, Philippines), creator.; Nishizaki, Shin-ya, editor.; Numao, Masayuki, 1961- editor.; Suarez, Merlin Teodosia C., editor.; Caro, Jaime D. L., editor.; ProQuest (Firm)
2021
Available Online

3 ARTICLE
A Survey on Fully Homomorphic Encryption: An Engineering Perspective
Martins, Paulo ; Sousa, Leonel ; Mariano, Artur
ACM computing surveys, 2018, Vol.50 (6), p.1-33
" .... *In this survey, both previous and current Somewhat Homomorphic Encryption (SHE) schemes are reviewed, and the more powerful and recent Fully Homomorphic Encryption (FHE...* "
PEER REVIEWED
Download PDF   Available Online
View Issue Contents

4 BOOK
Women in Numbers 2 : research directions in number theory : BIRS Workshop, WIN2 - Women in Numbers 2, November 6-11, 2011, Banff International Research Station, Banff, Alberta, Canada
WIN (Conference) (2nd : 2011 : Banff, Alta.); WIN (Conference); David, Chantal, 1964- editor.; Lalin, Matilde, 1977- editor.; Manes, Michelle, 1970- editor.
2013
Available at UTS Library   Request from Library Retrieval System (512.7 DAVI)
Available Online

5 BOOK
Advances in Cryptology – EUROCRYPT 2012 31st Annual International Conference on the Theory and Applications of Cryptographic Techniques, Cambridge, UK, April 15-19, 2012, Proceedings
EUROCRYPT (31st : 2012 : Cambridge, England); Pointcheval, David. editor.; Johansson, Thomas. editor.
2012; 1st ed. 2012.
Available Online

Fig 3.1 *Normal Chest X-ray (Pneumonia Dataset)*

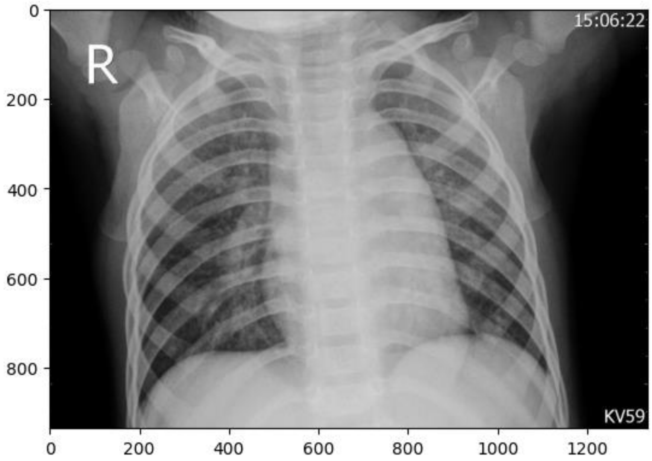


Fig 3.2 *Pneumonia Chest X-ray (Pneumonia Dataset)*

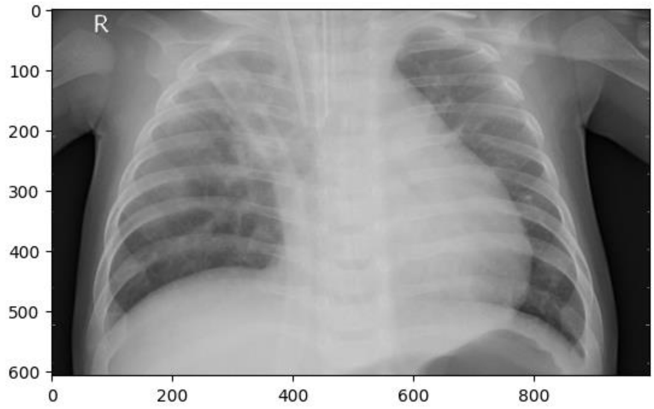


Fig 3.3 *18 random samples, three of each class (Medical MNIST Dataset)*

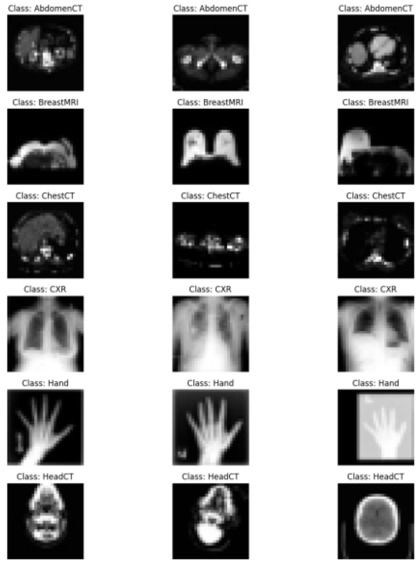


Fig 3.4 *16 random Chext X-ray samples (NIH Chest X-Ray Dataset)*

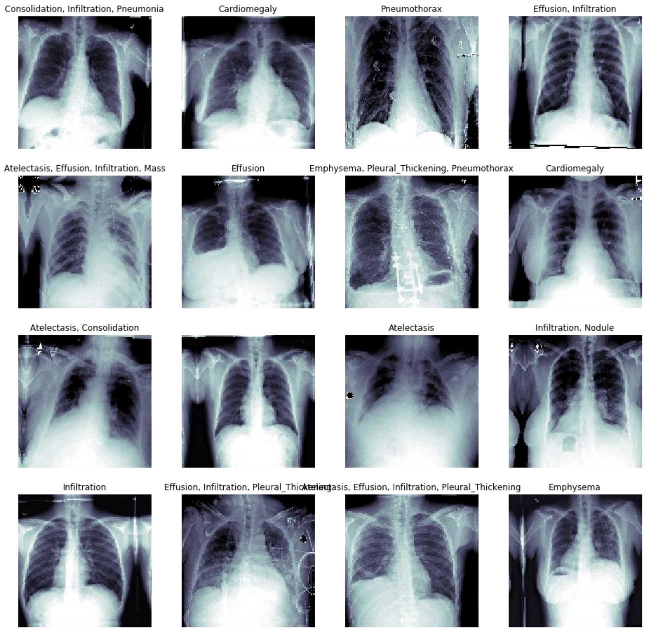

Consolidation, Infiltration, Pneumonia
Cardiomegaly
Pneumothorax
Effusion, Infiltration
Atelectasis, Effusion, Infiltration, Mass
Effusion
Emphysema, Pleural_Thickening, Pneumothorax
Cardiomegaly
Atelectasis, Consolidation
Atelectasis
Infiltration, Nodule
Infiltration
Effusion, Infiltration, Pleural_Thickening
Atelectasis, Effusion, Infiltration, Pleural_Thickening
Emphysema

Fig 3.5 *Class Imbalance (NIH Dataset)*

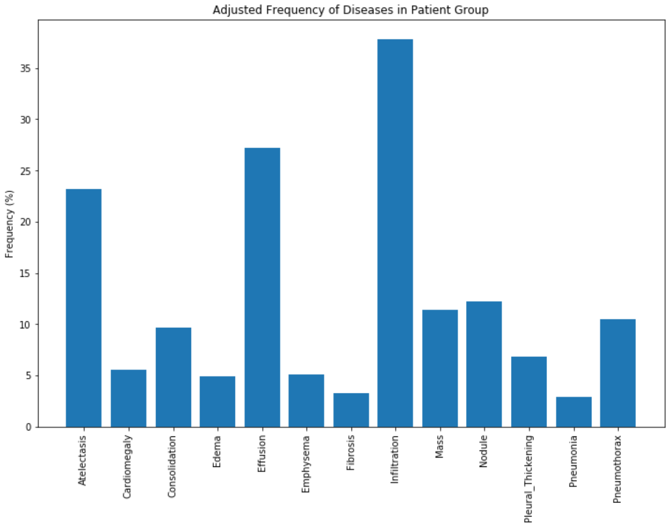


Fig 3.6 *Class Distribution Pie-Chart (Medial MNIST Dataset)*

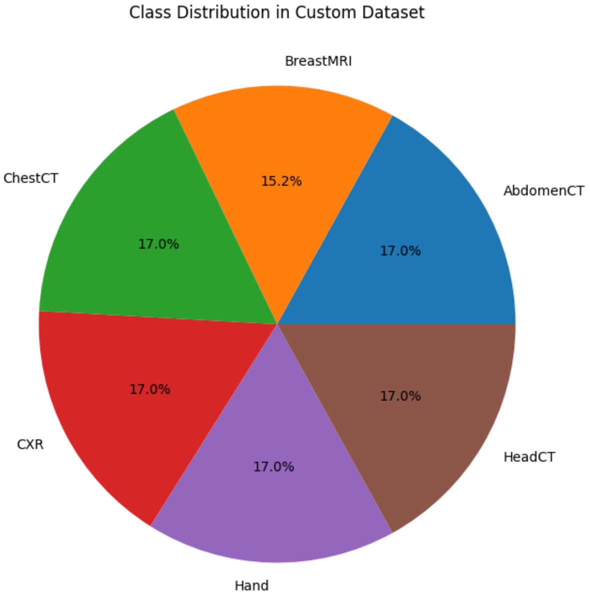


Fig 3.7 *Class Distribution Pie-Chart (Pneumonia Dataset)*

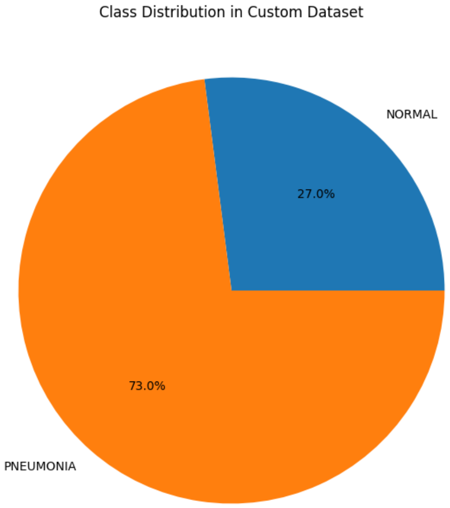


***Fig 3.8** – Machine Learning with Encrypted Inference Pipeline*

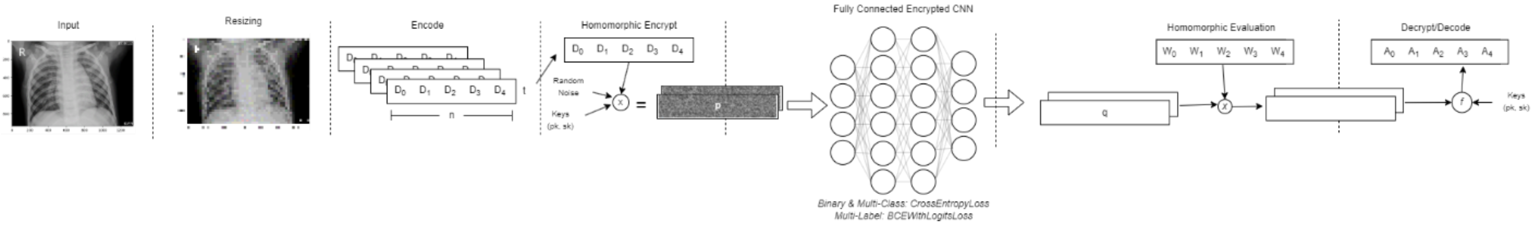


***Fig 4.1** Confusion Matrix of Pneumonia (X-Ray) - Unencrypted Dataset Model*

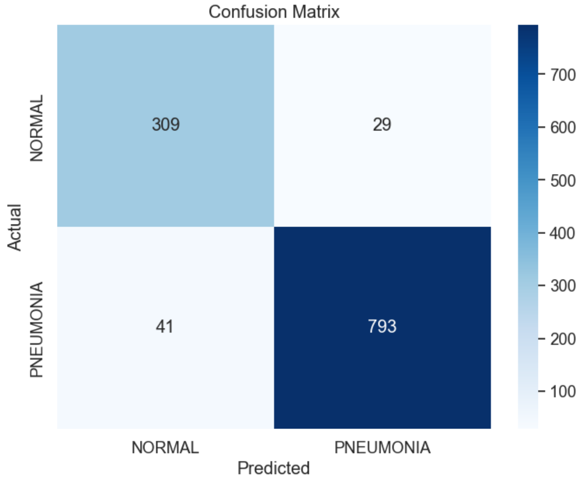


***Fig 4.2** ROC Curve graph of Pneumonia (X-Ray) - Unencrypted Dataset Model*

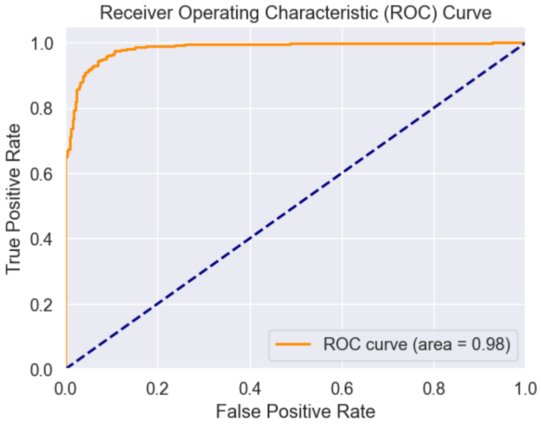


***Fig 4.3** Precision-Recall Curve Graph of Pneumonia (X-Ray) - Unencrypted Dataset Model*

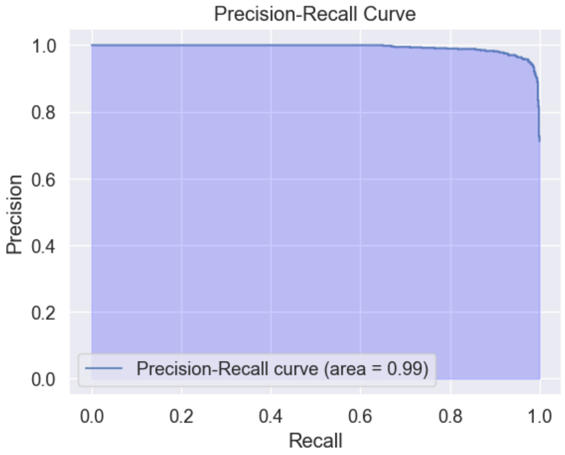


***Fig 4.4** Confusion Matric graph of Pneumonia (X-Ray) - Encrypted Dataset Model*

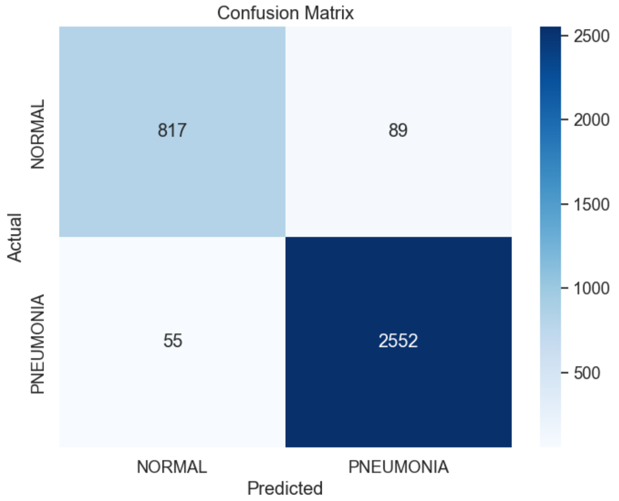


***Fig 4.5** ROC Curve graph of Pneumonia (X-Ray) - Encrypted Dataset Model*

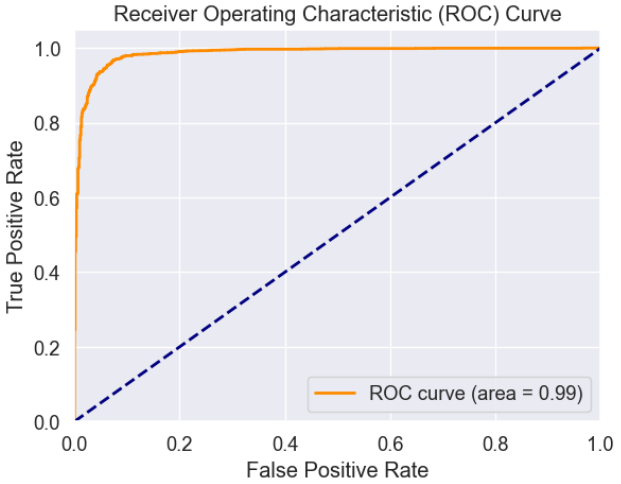


***Fig 4.6** Precision-Recall Curve Graph of Pneumonia (X-Ray) - Encrypted Dataset Model*

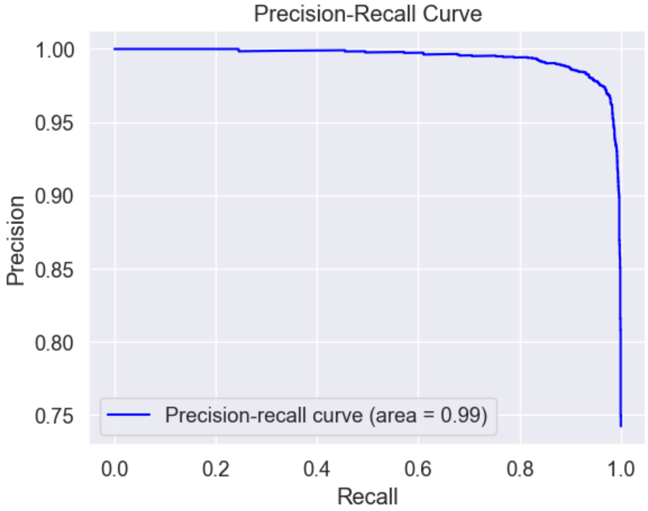
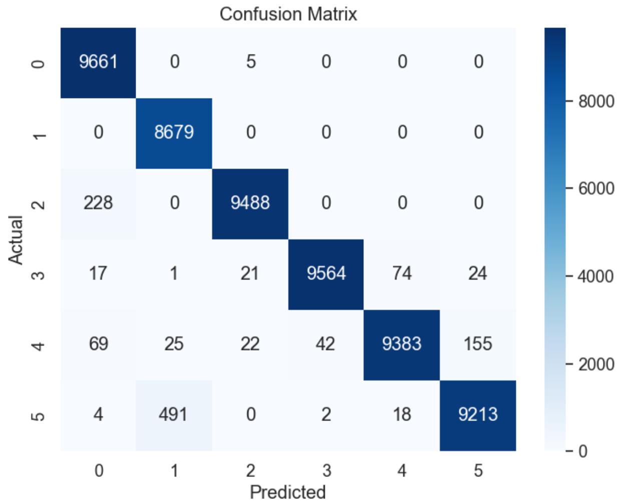


***Fig 4.7** Confusion Matrix of Medical MNIST - Unencrypted Dataset Model*



Confusion Matrix

Actual \ Predicted	0	1	2	3	4	5
0	9661	0	5	0	0	0
1	0	8679	0	0	0	0
2	228	0	9488	0	0	0
3	17	1	21	9564	74	24
4	69	25	22	42	9383	155
5	4	491	0	2	18	9213

***Fig 4.8** ROC Curve of Medical MNIST - Unencrypted Dataset Model*

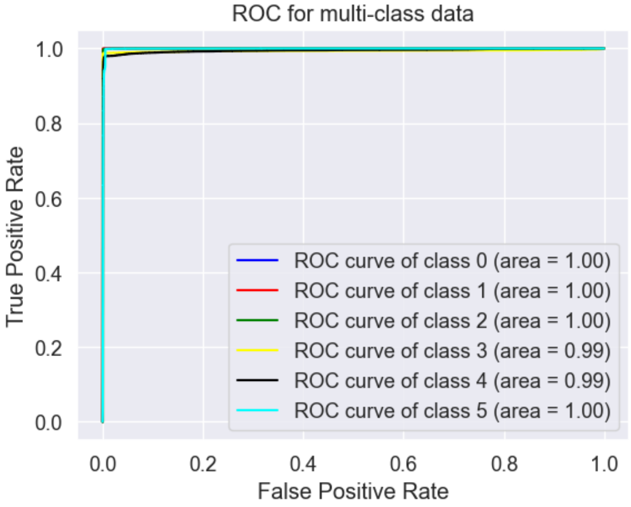


***Fig 4.9** Precision Recall Curve of Medical MNIST - Unencrypted Dataset Model*

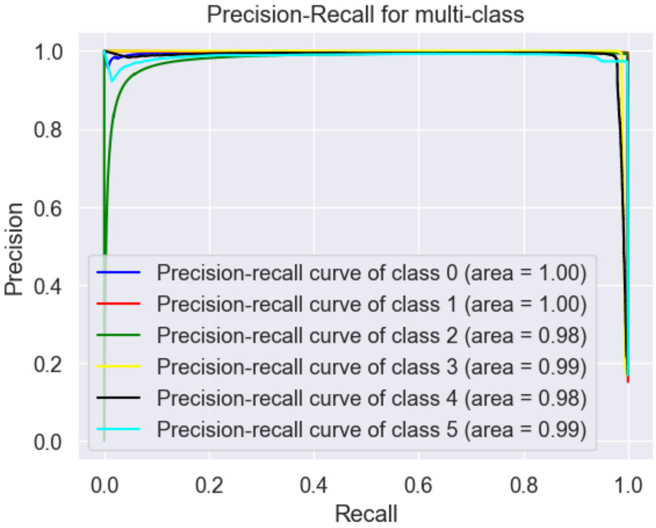
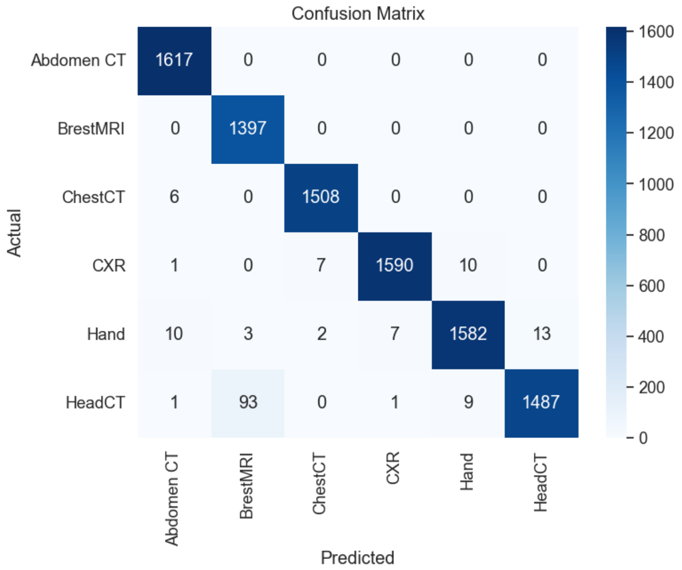


***Fig 4.10** Confusion Matric graph of Medical MNIST- Encrypted Dataset Model*



	Abdomen CT	BrestMRI	ChestCT	CXR	Hand	HeadCT
Abdomen CT	1617	0	0	0	0	0
BrestMRI	0	1397	0	0	0	0
ChestCT	6	0	1508	0	0	0
CXR	1	0	7	1590	10	0
Hand	10	3	2	7	1582	13
HeadCT	1	93	0	1	9	1487

***Fig 4.11** ROC Curve graph of Medical MNIST - Encrypted Dataset Model*

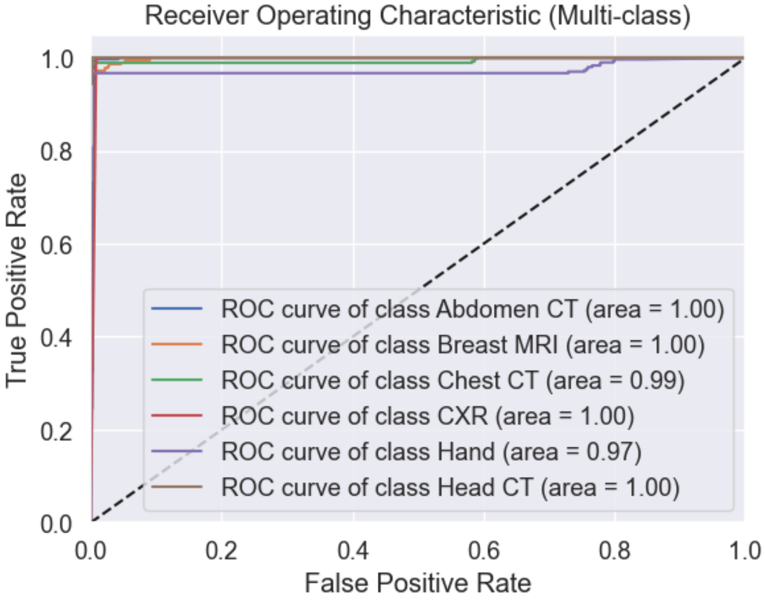


***Fig 4.12** Precision-Recall Curve graph of Medical MNIST - Encrypted Dataset Model*

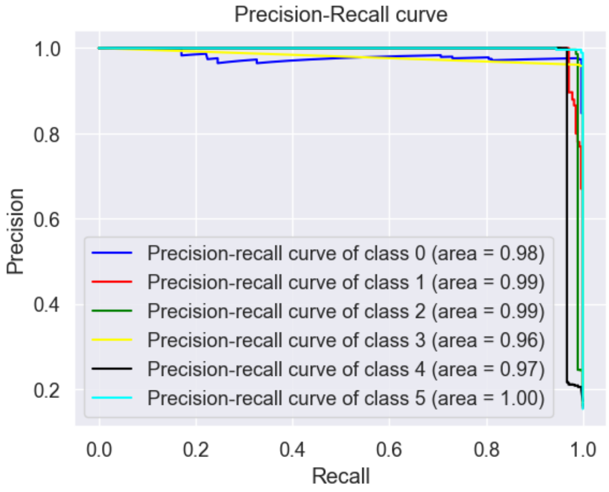


Fig 4.13 *ROC Curve of NIH - Unencrypted Dataset Model*

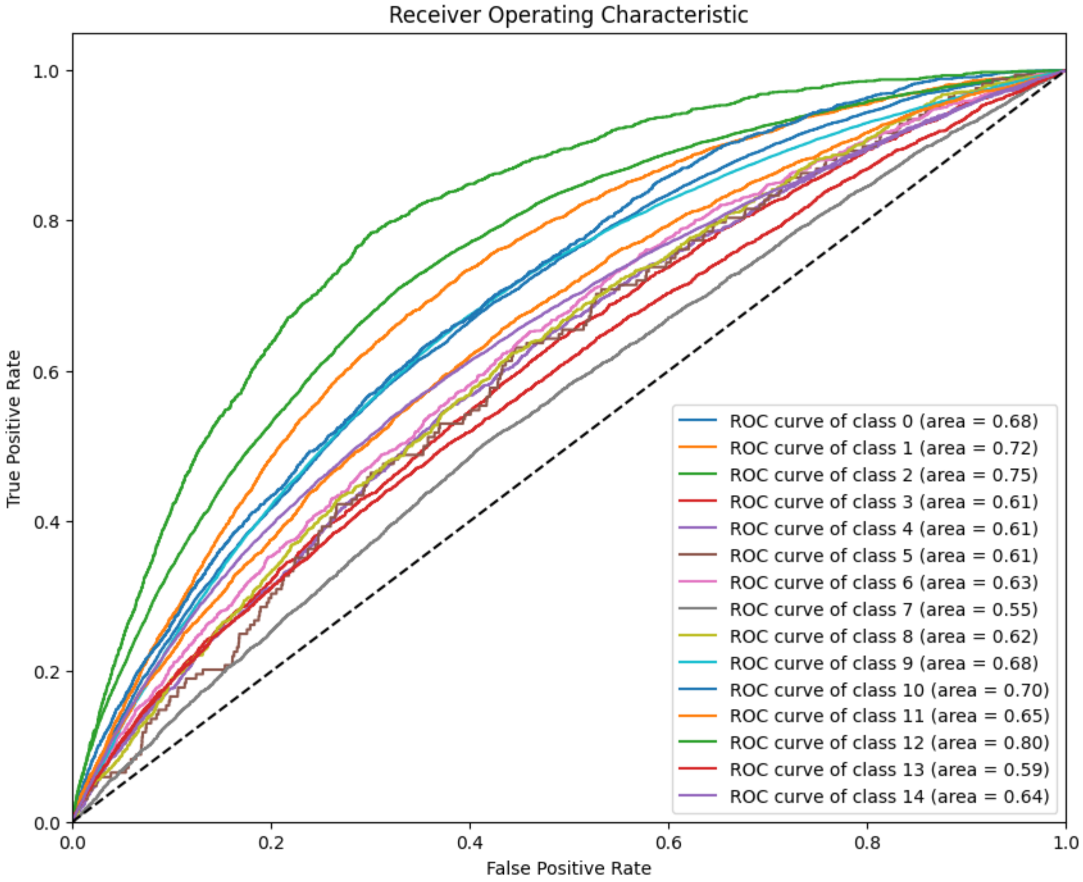


Fig 4.14 *ROC Curve of NIH - Unencrypted Dataset Model*

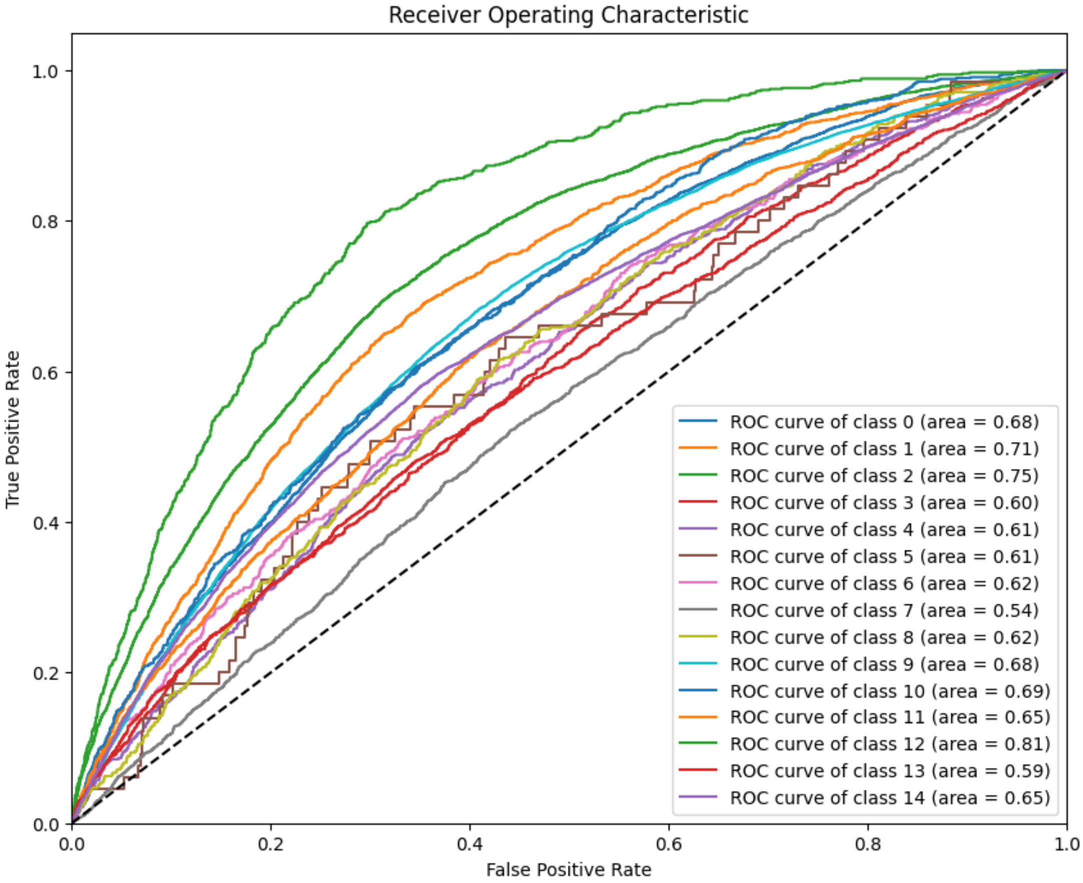


Fig 4.15 *Visual Impact of Encryption with CKKS*

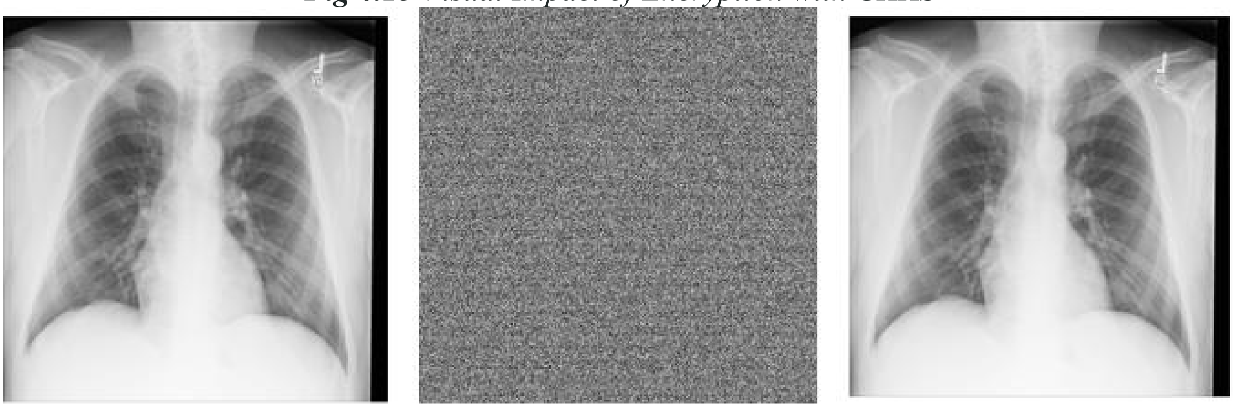


Table 1 *– List of critically evaluated sources*

Source	Source Title	Author(s)	Year of Publication
1	Partially Encrypted Machine Learning using Functional Encryption	T. Ryffel, E. Dufour-Sans, R. Gay, F. Bach, and D. Pointcheval	2019
2	Private AI – Machine Learning on Encrypted Data	K. Lauter Springer International Publishing Pages: 97-113	2022

Table 2 *– CKKS Encryption Parameters for Encrypted Inference Model*

Parameter	Description	Value
bits_scale	Controls the precision of the fractional part.	26
poly_modulus_degree	Determines the polynomial modulus degree for encryption context.	8192
coeff_mod_bit_sizes	Bit s s of coefficients in the modular polynomial.	[31, 26, 26, 26, 26, 26, 26, 31]
global_scale	Defines the scale used in encryption to preserve precision.	pow(2, bits_scale)
galois_keys	Required for performing ciphertext rotations.	Generated based on context.
secret_key	The key is used to encrypt and decrypt data.	Generated and kept private.

Table 3.1 *– CNN Architecture*

Layer (Type)	Output Shape	Param #	Details
Conv2d	(batch_size, 4, H', W')	X	1 input channel, 4 output channels, kernel size=7, stride=3
Square Activation (custom)	(batch_size, 4, H', W')	0	Element-wise square function
Flatten	(batch_size, 256)	0	Flatten the output to vector
Linear	(batch_size, hidden)	Y	Fully connected layer, 256 inputs to **hidden** outputs
Square Activation (custom)	(batch_size, hidden)	0	Element-wise square function
Linear (output)	(batch_size, output)	Z	Fully connected layer, **hidden** inputs to **output** classes

Table 3.2 *– CNN Architecture with Encrypted Inference*

Component	Operation	Description
Convolutional Layer	Encrypted 2D convolution (**conv2d_im2col**)	Performs convolution over encrypted images using homomorphically encrypted kernels.
Activation Function	Squaring (**square_**)	Applies an activation function by squaring the encrypted vector in place.
Fully Connected Layer	Vector-matrix multiplication (**mm**) and bias addition (+)	Performs encrypted linear transformation followed by the addition of a bias vector to the squared encrypted vector.
Output Layer	Vector-matrix multiplication (**mm**) and bias addition (+)	Transforms the activated encrypted vector into the final encrypted output which can be decrypted for prediction.

Table 4.1 *– Pneumonia (X-Ray) Unencrypted vs Encrypted Performance Metrics*

Dataset	Test Loss	Accuracy	Precision	Recall	F1 Score	Training Time	Memory
Unencrypted	*0.134289*	*0.9590*	*0.9587*	*0.9590*	*0.9588*	*32s*	*16.24 MB*
Encrypted	*0.3198*	*0.8747*	*0.8847*	*0.8747*	*0.8643*	*15mins 28s*	*80.87 MB*

Table 4.1 – Pneumonia (X-Ray) Unencrypted vs Encrypted Performance Metrics

Table 4.2 *– Medical MNIST Unencrypted vs Encrypted Performance Metrics*

Dataset	Test Loss	Accuracy	Precision	Recall	F1 Score	Training Time	Memory Usage
Unencrypted	*0.096289*	*0.9977*	*0.9977*	*0.9977*	*0.9977*	*1min 35*	*16.91 MB*
Encrypted	*0.864911*	*0.9562*	*0.9646*	*0.9562*	*0.9556*	*33mins 9s*	*349.61 MB*

Table 4.3 *– NIH Unencrypted vs Encrypted Performance Metrics*

Dataset	Hamming Loss	Sample Wise Accuracy	Precision (micro)	Recall (micro)	F1 Score (micro)	Time Taken
Unencrypted	*0.0732*	*0.4415*	*0.6402*	*0.3091*	*0.4169*	*5hr 12min 32s*
Encrypted	*0.0730*	*0.3883*	*0.6167*	*0.3524*	*0.4485*	*16hr 15min 47s*