# Evaluating Homomorphic Encryption with TenSEAL For Secure Healthcare Predictive Modeling

Logan Choi

A Capstone Thesis Proposal
submitted in partial fulfillment of the
requirements of the degree of

Master of Science in Computer Science & Software Engineering

**University of Washington**
February 17th, 2025

**Thesis Committee:**
Dr. Wooyoung Kim, Committee Chair
Dr. Brent Lagesse, Committee Member
Dr. Yang Peng, Committee Member

## I. Introduction

The healthcare industry relies heavily on data and machine learning to improve diagnostic accuracy and patient outcomes [4]. However, the data is often highly sensitive, creating significant challenges when it comes to privacy and security. With increasing reliance on artificial intelligence (AI) and machine learning (ML), finding ways to protect patient information while maintaining model performance has become more important than ever [5].

Homomorphic encryption (HE) offers a unique solution to this problem. It allows computations to be performed directly on encrypted data, ensuring that sensitive information remains secure throughout the analysis process. TenSEAL, a Python-based library, provides a framework for combining HE with popular machine learning models [6]. By using TenSEAL, it is possible to securely analyze encrypted medical data without compromising privacy or accuracy.

This research evaluates the effectiveness of machine learning models using TenSEAL across multiple encrypted healthcare datasets. The study will compare model performance on both encrypted and unencrypted data to assess the impact of homomorphic encryption on predictive accuracy. Additionally, it will analyze the suitability of the Brakerski/Fan-Vercauteren (BFV) and Cheon-Kim-Kim-Song (CKKS) encryption schemes across different datasets, with a primary focus on accuracy as the key metric for evaluating each scheme's performance in secure healthcare applications [7] [8].

HE offers a way to prepare for emerging challenges, such as the potential threats posed by quantum computing to traditional encryption methods [9] [10]. By exploring the practical use of TenSEAL in predictive healthcare modeling, this study aims to provide a foundation for using HE in safeguarding patient data while enabling secure and accurate medical diagnoses such as cancer assessments, diabetes prediction, etc.

## II. Thesis Goals/Vision

### A. Goals of Completed Work

The primary goal of this thesis is to evaluate the feasibility and effectiveness of homomorphic encryption in accurately predicting medical outcomes while preserving patient data privacy, with TenSEAL serving as the core framework for encrypting data and enabling secure machine learning computations.This involves comparing traditional ML models on unencrypted dataset with ML models on encrypted datasets using TenSEAL, focusing on both predictive accuracy and computational performance. By analyzing the BFV and CKKS encryption schemes, the study further aims to identify which scheme is more suitable for secure healthcare applications through similar metrics of accuracy and computational efficiency.

The completed work will provide insights into the trade-offs between encryption and plaintext, different encryption schemes, and model performance, helping to establish benchmarks for using encrypted machine learning in healthcare. Additionally, this research will lay the foundation for secure third-party platforms that enable private computations in healthcare diagnostics, ensuring sensitive medical data remains protected throughout the analysis pipeline.

### B. Identification of Problem

The increasing use of AI and machine learning in healthcare brings substantial benefits in terms of predictive diagnostics and personalized treatment. However, the sensitive nature of medical data poses significant challenges related to privacy and security. Traditional encryption methods can safeguard data in transit but fail to protect it during computation, leaving it vulnerable to breaches.

Homomorphic encryption offers a solution by allowing computations on encrypted data without requiring decryption. While promising, its practical implementation in real-world scenarios like healthcare diagnostics remains underexplored. Questions persist about whether homomorphic encryption can provide accurate predictions without excessive computational overhead. Furthermore, the lack of comparative studies on different encryption schemes, such as BFV and CKKS, makes it difficult to determine their suitability for specific use cases.

This research addresses the critical gap in evaluating the viability of homomorphic encryption, specifically through TenSEAL, in healthcare settings. By focusing on real-world medical diagnostic scenarios, the study aims to demonstrate whether this approach can achieve the dual goals of preserving data privacy and delivering accurate, actionable insights.

### C. Stakeholders and Beneficiaries of Research

The findings of this research will benefit multiple groups. Healthcare providers and institutions, such as hospitals, clinics, and medical research centers, will gain valuable insights into secure diagnostic solutions that prioritize patient privacy without compromising predictive accuracy. This work can guide these organizations in adopting encrypted machine learning models with confidence, addressing the growing need for secure data processing in healthcare.

Patients are another key beneficiary of this research. By enabling predictive analytics on encrypted data, the study ensures that sensitive medical information remains private and protected, even when shared across institutions or analyzed by third-party platforms. This level of security fosters greater trust in healthcare systems and the use of AI-driven diagnostic tools.

In addition, the research offers significant value to AI and machine learning researchers exploring privacy-preserving computation. By showcasing the capabilities and limitations of TenSEAL, this study provides a foundation for future innovations in homomorphic encryption and its practical applications in healthcare. It serves as a resource for advancing encrypted computation techniques while ensuring data privacy in sensitive environments.

Finally, this work lays the groundwork for future researchers who wish to explore comparative analyses of other libraries, such as Microsoft SEAL or IBM HELib, or investigate further optimizations in encrypted computation. The study highlights areas for improvement, such as leveraging parallel programming or addressing challenges posed by emerging technologies like quantum computing.

## III. Criteria

### A. Levels of Success

1. ***Minimum:*** At the minimum level, this thesis will compare the performance of machine learning models on unencrypted versus encrypted data using TenSEAL.

The focus will be on evaluating predictive accuracy and identifying any significant differences between encrypted and unencrypted workflows. This baseline analysis will demonstrate whether TenSEAL can effectively process encrypted medical data while maintaining acceptable levels of accuracy.

2. ***Expected:*** At the expected level, the thesis will expand upon the baseline by conducting a more comprehensive analysis. This includes comparing encrypted and plain data, evaluating the performance of the BFV and CKKS encryption schemes, and analyzing key metrics such as memory usage and computational efficiency. This level of research will provide a deeper understanding of the trade-offs involved when using homomorphic encryption for medical diagnostics.

3. ***Aspirational:*** At the aspirational level, the research will aim to develop and test parallel programming methodologies to optimize the performance of encrypted machine learning models. Additionally, the study will compare TenSEAL to other leading homomorphic encryption libraries, such as Microsoft SEAL and IBM HELib, to offer a broader perspective on the state-of-the-art in encrypted computation. This level of success would push the boundaries of existing research and establish a foundation for future advancements in privacy-preserving machine learning.

### B. Quality and Associated Measurement Metrics

The quality of this research will be assessed based on several key performance indicators, including predictive accuracy, computational efficiency, and resource utilization when applying machine learning models to encrypted data. Predictive accuracy will be measured using metrics such as F1-score, precision, recall, and overall accuracy. The target is to achieve results comparable to unencrypted models, with no more than a 5% reduction in accuracy when using encrypted inputs. Computational efficiency will be evaluated by measuring the time required for encryption, training, and inference, aiming for execution times practical for real-world healthcare applications. Additionally, the BFV and CKKS schemes will be compared based on their performance with real-valued and integer-valued data, accuracy, and overall computational demands. The objective is to identify the most suitable scheme for specific diagnostics in healthcare.

### C. Targets

This research sets clear and progressive targets to ensure meaningful outcomes at every stage. The baseline target is to demonstrate that TenSEAL can process encrypted data effectively, maintaining predictive accuracy with minimal loss compared to unencrypted workflows. By achieving this, the study will validate the feasibility of using encrypted machine learning models for healthcare diagnostics.

The expected target is to conduct a detailed evaluation of the BFV and CKKS encryption schemes, analyzing their trade-offs in terms of accuracy, computational time, and memory usage. This comparison will establish benchmarks for encrypted and unencrypted workflows in medical diagnostic applications, providing a practical understanding of how homomorphic encryption impacts performance.

## IV. Positioning of Thesis

Research on TenSEAL and its applications remains relatively limited. One notable study provides a comparative analysis of the BFV and CKKS schemes on IoT data, demonstrating the effectiveness of homomorphic encryption in preserving IoT data privacy [1]. Another paper explores the use of homomorphic encryption for securing patient data in Thailand's healthcare system, focusing on encryption and decryption processes [2]. A third study discusses a secure healthcare predictive modeling system implemented using TenSEAL and homomorphic encryption. However, this work lacks a detailed analysis of the system's viability for accurately diagnosing patients, leaving a critical gap in understanding its practical applications [3].

This thesis aims to address this gap by conducting a comprehensive evaluation of TenSEAL's capabilities in healthcare diagnostics. By focusing on encrypted and unencrypted workflows, this study seeks to provide a foundational analysis that future researchers can build upon to further explore and enhance the use of homomorphic encryption in predictive modeling and healthcare diagnostics.

## V. Thesis Plan

### A. General Weekly Plan (20 weeks)

Over the first two weeks, I'll focus on setting up the foundation for my research by conducting a literature review on TenSEAL, homomorphic encryption, and the BFV and CKKS schemes. During this time, I'll identify research gaps, finalize my research questions, explore potential datasets, and install the necessary tools. Once that's complete, I'll move on to understanding and preparing the data, addressing inconsistencies, handling missing values, and splitting it into training, validation, and testing sets. By week three, I'll finalize my experimental design, deciding on key performance metrics like accuracy, computational efficiency, and memory usage while drafting a structured workflow for both encrypted and unencrypted models.

Weeks four and five will be spent implementing basic models on unencrypted data, starting with logistic regression and random forest, followed by initial evaluations. Then, I'll shift my focus to familiarizing myself with TenSEAL by encrypting datasets using the BFV and CKKS schemes, ensuring encryption and decryption work as expected. In weeks six through nine, I'll train and evaluate logistic regression models on encrypted data, first using the BFV scheme and then the CKKS scheme, measuring their performance along the way.

By week ten, I'll compare results across both encryption schemes, analyzing accuracy, efficiency, and memory usage. Over the following two weeks, I'll work on optimizing the encrypted models, fine-tuning hyperparameters, and exploring lightweight strategies to improve performance while beginning my thesis draft. In weeks thirteen and fourteen, I'll conduct a comparative analysis of encrypted vs. unencrypted models, identifying trends and drawing conclusions from my findings. The final phase, spanning weeks fifteen to twenty, will be dedicated to writing, refining, and submitting my thesis. I'll also prepare visual aids, incorporate feedback, and finalize my presentation before presenting to my committee.

### B. Testing Plan

The datasets I plan to use come from Kaggle and focus on cancer diagnoses, heart disease, and diabetes prediction [11][12][13]. Each dataset will have critical health information, such as patient attributes, diagnostic results, and relevant biomarkers, which will be preprocessed to ensure consistency and quality before model training.These datasets will then be used to train and evaluate various ML models, including logistic regression. The goal is to compare the accuracy of these models when trained on plaintext data versus when trained on encrypted data using TenSEAL's BFV and CKKS encryption schemes. My hypothesis is that the encrypted models will maintain a similar level of accuracy to their plaintext counterparts while operating with acceptable computational efficiency on all three datasets. The testing process will involve evaluating accuracy, training time, and memory usage to determine the trade-offs between security and performance.

## VI. Constraints, Risks, and Resources

### A. Key Constraints

Homomorphic encryption, particularly when using the BFV and CKKS schemes, is computationally demanding. The process of encrypting and processing data can significantly increase model training times and computational overhead, making the research more resource-intensive. Additionally, access to high-quality, anonymized healthcare datasets may present a challenge. Due to privacy regulations and limited availability, obtaining suitable datasets for testing encrypted machine learning models could hinder the progress of the research.

### B. Resources Needed For Success

To ensure the success of this research, access to high-performance computing facilities and cloud resources at the University of Washington will be essential. These resources will help handle the computational demands of homomorphic encryption and large-scale data processing. Collaboration with faculty members who specialize in cryptography, machine learning, and healthcare data will provide valuable expertise and guidance. Additionally, regular feedback and support from advisors and peers will be critical for troubleshooting and refining the research approach. Acquiring suitable healthcare datasets, either publicly available or anonymized, will also be crucial for model training and validation.

### C. Anticipated Risks

There are several risks associated with this research. One significant challenge is the technical complexity of implementing homomorphic encryption with TenSEAL, especially when working with large datasets. The process may involve troubleshooting and adjustments, potentially delaying progress. Another risk is the difficulty in acquiring appropriate healthcare datasets, which may limit the study's real-world applicability. In the absence of suitable data, synthetic datasets might need to be used, potentially affecting the validity of the results. Finally, resource constraints, such as limited computing power, could impact the ability to conduct extensive experiments with large datasets or complex models, further slowing down the research.

## VII. Research Methods & Design

The research will use publicly available or anonymized healthcare datasets for model training and testing. TenSEAL will be utilized to implement homomorphic encryption, allowing machine learning models to process encrypted data. Python libraries like scikit-learn, pandas, and NumPy will handle machine learning tasks, data manipulation, and numerical computations. Jupyter Notebooks will facilitate the execution of experiments and visualization of results. The performance of the models will be evaluated based on accuracy, computational efficiency, and memory usage, measuring the time taken for training, testing, and inference, as well as peak memory consumption during these processes.

Data preprocessing will involve cleaning and normalizing the datasets, handling missing values, and splitting the data into training, validation, and testing sets. Baseline models, such as logistic regression and random forest, will first be trained on unencrypted data. Encrypted models will then be trained using the BFV and CKKS schemes via TenSEAL, with performance compared across both approaches. A consistent experimental workflow will be defined to ensure reliable comparisons between encrypted and unencrypted models. Statistical analysis will be conducted to assess differences in performance, using tests to compare accuracy and efficiency metrics. To improve model performance, hyperparameter tuning will be applied, alongside optimization strategies such as batching encrypted data operations. Finally, the study will compare the BFV and CKKS encryption schemes, focusing on trade-offs in terms of accuracy, memory usage, and computational time.

## References

[1] Aimé, F. L., Wiryen, Y. B., Vigny, N. A., & Ngono, M. J. (2024). Leveraging TenSEAL: A Comparative Study of BFV and CKKS Schemes for Training ML Models on Encrypted IoT Data. Int. J. Inf. Sec. Priv., 18(1), 1–17. doi:10.4018/IJISP.356402

[2] Lohlah, Y., & Boonyopakorn, P. (2024). Application of Homomorphic Encryption for Encrypting and Decrypting Patient Data in Thailand's Healthcare System. 2024 Research, Invention, and Innovation Congress: Innovative Electricals and Electronics (RI2C), 231–237. doi:10.1109/RI2C64012.2024.10784418

[3] Shekhu, P., Sethi, S., & Chaudhary, A. (2024). Secure Healthcare Predictive Modeling with Homomorphic Encryption. 2024 11th International Conference on Computing for Sustainable Global Development (INDIACom), 1628–1631. doi:10.23919/INDIACom61295.2024.10498904

[4] S. Flam, "Machine Learning in Healthcare - Benefits & Use Cases," ForeSee Medical, Mar. 23, 2020. https://www.foreseemed.com/blog/machine-learning-in-healthcare

[5] N. Yadav, S. Pandey, A. Gupta, P. Dudani, S. Gupta, and K. Rangarajan, "Data privacy in healthcare: In the era of artificial intelligence," Indian Dermatology Online Journal, vol. 14, no. 6, pp. 788–792, Nov. 2023, doi: https://doi.org/10.4103/idoj.idoj_543_23.

[6] A. Benaissa, B. Retiat, B. Cebere, and A. E. Belfedhal, "TenSEAL: A Library for Encrypted Tensor Operations Using Homomorphic Encryption," arXiv.org, Apr. 28, 2021. https://arxiv.org/abs/2104.03152

[7] J. Cheon, A. Kim, M. Kim, and Y. Song, "Homomorphic Encryption for Arithmetic of Approximate Numbers." Accessed: Feb. 05, 2025. [Online]. Available: https://eprint.iacr.org/2016/421.pdf?

[8] J. Fan and F. Vercauteren, "Somewhat Practical Fully Homomorphic Encryption," Cryptology ePrint Archive, 2012, Available: https://eprint.iacr.org/2012/144

[9] Abel, "Homomorphic Encryption Based on Lattice Post-Quantum Cryptography," arXiv.org, 2024. https://arxiv.org/abs/2501.03249 (accessed Feb. 05, 2025).

[10] "Quantum Computing - How it Changes Encryption as We Know It | Division of Information Technology," Division of Information Technology, Oct. 18, 2024. https://it.umd.edu/security-privacy-audit-risk-and-compliance-services-sparcs/topic-week/quantum-computing-how-it-changes-encryption-we-know-it

[11] erdemtaha, "Cancer Prediction %96.5 With Logistic Regression," Kaggle.com, Oct. 22, 2023. https://www.kaggle.com/code/erdemtaha/cancer-prediction-96-5-with-logistic-regression/input (accessed Feb. 05, 2025).

[12] “Logistic regression To predict heart disease,” www.kaggle.com. https://www.kaggle.com/datasets/dileep070/heart-disease-prediction-using-logistic-regression

[13] “Diabetics prediction using logistic regression,” www.kaggle.com. https://www.kaggle.com/datasets/kandij/diabetes-dataset