A Privacy-Preserving Approach to Diabetes Diagnosis Using Artificial Neural Networks and Federated Learning Integrating Homomorphic Encryption

Anik Sen1, a), Swee-Huay Heng2, b) and Shing-Chiang Tan3, c)

1Faculty of Information Science and Technology, Multimedia University, Jalan Ayer Keroh Lama, 75450 Bukit Beruang, Melaka, Malaysia.

2Centre for Intelligent Cloud Computing, CoE for Advanced Cloud, Faculty of Information Science and Technology, Multimedia University, Jalan Ayer Keroh Lama, 75450 Bukit Beruang, Melaka, Malaysia.

3Centre for Advanced Analytics, CoE for Artificial Intelligence, Faculty of Information Science and Technology, Multimedia University, Jalan Ayer Keroh Lama, 75450 Bukit Beruang, Melaka, Malaysia.

b) Corresponding author: shheng@mmu.edu.my  
b) [aniksen360@gmail.com](mailto:aniksen360@gmail.com)

c*) sctan@mmu.edu.my*

**Abstract.** The growing digitisation of healthcare data has raised significant concerns about patient privacy and data security. This study presents an innovative approach that integrates Federated Learning (FL) with Homomorphic Encryption (HE) using the CKKS (Cheon-Kim-Kim-Song) scheme, ensuring secure data sharing and privacy-preserving diabetes diagnosis. By decentralising data processing through Federated Learning (FL), sensitive patient data stays on the client device, which helps protect privacy by preventing the data from leaving its local environment. Adding in CKKS-based Homomorphic Encryption (HE) takes security a step further, allowing computations on encrypted data so that sensitive details are never exposed during the training or aggregation processes. The proposed approach achieves high accuracy with a privacy-preserving mechanism that offers a novel solution for privacy-sensitive machine learning applications in healthcare. Extensive experiments demonstrate that the federated approach, which trains the model across multiple clients, achieves 92.28% accuracy after five training rounds, comparable to centralised models while safeguarding patient data. Additionally, the proposed approach is highly adoptable, considering the performance and also the security concerns which prioritise the patient data privacy and confidentiality. This research represents a significant advancement in integrating privacy-preserving techniques into healthcare, providing an essential framework for the future of secure, data-driven healthcare innovations.

**Introduction**

Health-related information is one of the most sensitive forms of personal data, raising significant concerns around privacy and data security. As healthcare systems become increasingly digitised, the amount of sensitive patient information being collected and processed continues to grow, exposing it to potential breaches and unauthorised access [1]. Privacy issues deserve to be the highest priority for sensitive data sharing areas like diabetes prediction, where a large volume of patient data is required to train the machine learning (ML) model to get the most accurate predictions [2]. However, traditional centralised healthcare systems store patient data in a single location, making it vulnerable to cyber threats, data leakage, and breaches [3]. To mitigate these concerns, Federated Learning (FL) has become a crucial solution which offers decentralised ML model training across multiple clients without needing the raw data from the client’s end to maintain privacy and confidentiality [4]. Despite its privacy benefits, FL introduces challenges, including data consistency, client heterogeneity, and communication overhead. Besides, FL itself cannot completely tackle the risk of data leakage in the model aggregation process as well. Homomorphic encryption (HE) bridges this gap by offering the capability to perform computations on encrypted data so that the plaintext of a single user could be kept private when computing during the aggregation process. Combining FL and HE offers a promising approach to achieving high privacy and performance in ML models [5]. However, issues such as computational overhead and the balance between privacy and model effectiveness still need to be optimised. This study aims to integrate CKKS-based HE with FL, providing a secure, efficient, and privacy-preserving framework for diabetes diagnosis, offering an optimal balance between privacy and accuracy in healthcare predictions.

The objectives of this study are:

1. To develop and evaluate a privacy-preserving framework for diabetes diagnosis that ensures a secure data sharing protocol while maintaining high accuracy.

2. To assess the effectiveness of the framework by comparing its performance in a decentralised setting with a centralised model, focusing on balancing privacy, efficiency, and predictive accuracy.

**Related Works**

Recently, several studies in privacy-preserving ML for diabetes prediction have used FL and HE as well as traditional ML algorithms. ML algorithms such as logistic regression, random forest, and XGBoost, among others, have been previously used to make predictions regarding diabetes-related complications. For instance, Mora et al. [6] analysed the healthcare records of more than 600,000 patients, obtaining a model accuracy level of 60% to 75% based on the complications. Olisah et al. [7] optimised deep learning models for optimal results of up to 97.93%. However, their studies did not lay much emphasis on data privacy, particularly with consideration of healthcare data.

FL and HE have emerged as key strategies to work on data privacy and dissimilarities. Shen et al. [8] used the diagnosis dataset coupled with support vector machine and hybrid encryption, achieving an accuracy level of over 97%. Despite the success, issues such as high computational costs and leakage of model parameters remain. Huma et al. [9] combined FL and XGBoost, registering 90% accuracy levels and a balance in data privacy. Nonetheless, there are still challenges such as data imbalance and overfitting that necessitate larger as well as diverse datasets. FL is seen as a feasible option for privacy-preserving ML since it trains ML models on separate training data sources. Islam & Mosa [10] used FL to enhance performance metrics such as the F1 score and recall. Liu et al. [11] also used FL with Paillier HE to protect data privacy during aggregation, though encryption added significant computational overhead. Fully Homomorphic Encryption (FHE) has been explored by Marcel et al. [3] noted that while prediction accuracy was preserved, execution time significantly increased. This suggests the need for optimisation of HE methods to improve efficiency for real-world applications.

Overall, the integration of FL and privacy-preserving mechanisms ensures data security while maintaining high diagnostic accuracy.

**Methodolgy**

The experiment was conducted in two distinct phases: centralised and decentralised (federated). In the first phase, a centralised approach was employed, where all data was stored and processed in a single, centralised system. In the second phase, a decentralised approach utilising FL was adopted to preserve data privacy while training the model, ensuring that the data remained distributed across multiple clients without the need for central storage. The complete experimental architecture is illustrated in Figure 1.

**A diagram of a software development process

AI-generated content may be incorrect.**

**FIGURE 1.** Experimental Architecture and Workflow

## Data Processing and Inspection

The preprocessing process began with an inspection of the dataset for missing or invalid values. The public diabetes prediction dataset, sourced from Kaggle [12], contains 100,000 records, with features including age, gender, race, and medical indicators like hypertension, BMI, and blood glucose levels. Of these records, 8,500 (8.5%) are diabetes patients, and 91,500 (91.5%) are non-diabetes patients. After thorough validation, no missing values were found, removing the need for data imputation or deletion, thus ensuring a complete dataset. Some features exhibited skewed distributions, which could affect the performance of machine learning models, especially neural networks. To address this, the Box-Cox transformation was applied to the BMI and blood glucose level features, both of which had positive skewness [13]. The transformation is defined in Equation (1).

(1)

where represents the original feature value and is the transformation parameter. This transformation normalises the features, improving the model's ability to learn. Categorical features like gender, location, and smoking history were encoded using Label Encoding, expressed in Equation (2).

(2)

where represents the categorical variable, and the function maps each unique category to an integer. Feature selection was then applied to eliminate redundant variables, specifically removing the original BMI and blood glucose features after transforming them. To balance the dataset, the SMOTE-ENN algorithm was used, combining SMOTE for generating synthetic minority class samples and ENN for cleaning noisy data. The dataset was split into training and testing sets using an 80-20 ratio, with 80% utilised for training and 20% for testing. The training set included 8.5% diabetes patients and 91.5% non-diabetic individuals.

## Centralised Approach

In the centralised method, a Sequential Artificial Neural Network (ANN) model was defined with an input layer of 64 neurones (tanh activation), two hidden layers of 32 and 64 neurones (both using ReLU activation), and an output layer with sigmoid activation for binary classification. The model was built using the Adam optimiser with a learning rate of 0.001, and the binary cross-entropy loss function was employed for training. The model was trained for 15 epochs with a validation split of 0.2, utilising 20% of the data for validation. The training process involved analysing the training loss and accuracy over epochs to evaluate model performance.

## Decentralised Approach

In the decentralised approach, the dataset is distributed across clients , with each client independently performing local training on its encrypted data. The primary objective of this methodology is to preserve data privacy by employing FL in combination with HE, specifically utilising the CKKS encryption scheme during the model training process.

The CKKS (Cheon-Kim-Kim-Song) scheme, proposed by Cheon et al. [14], enables approximate computations on encrypted data, allowing secure operations on ciphertexts without requiring decryption. The process (Figure 2) begins by encoding the plaintext data into polynomials, followed by encryption into ciphertext , where and represent the encrypted components of the data. These ciphertexts are then used to compute necessary model updates, expressed as , where represents the model's function on the encrypted data. The resulting ciphertext is sent back to the client, where it is decrypted using the client’s private key to obtain the final model updates as .

In this experiment, the CKKS scheme ensures data privacy by enabling computations on encrypted data. As illustrated in Figure 2, each client encrypts its local data using CKKS before sending it to the central server. The central server computes the model updates on the encrypted data, preserving the privacy of sensitive health data throughout the FL process. This encryption scheme allows computations on encrypted data, ensuring that the data remains secure and private throughout the entire FL process.

A diagram of a function

AI-generated content may be incorrect.

**FIGURE 2.** Overview of the CKKS Scheme

## Federated Averaging (FedAvg)

Federated Averaging (FedAvg) aggregates model weights from each client by computing a weighted average. The goal is to minimise a global loss function by combining local updates from all clients [15]. The global objective function in FedAvg is defined in Equation (3).

(3)

where is the total number of clients, is the local loss function for the -th client, and are the model parameters. In each communication round , client updates the model by minimising its local loss function using an optimisation method called stochastic gradient descent (SGD), defined in Equation (4).

(4)

where is the local update, is the learning rate, and is the gradient of the local loss function. After local updates, the central server aggregates them to compute the global model, which is defined in Equation (5).

(5)

This process repeats until convergence. The steps are further detailed in Algorithm 1, which shows the full workflow, including client-side data encryption and model update aggregation.

|  |  |
| --- | --- |
| **Algorithm 1** Federated Learning with Encrypted Model Aggregation | |
| 1: | **Input:** global model, clients' data clients' labels number of epochs, encryptor (public key |
| 2: | **Output:** updated global model |
| 3: |  |
| 4: | Initialise global model with weights Galois key set for relinearisation and rotation |
| 5: | **for** each client in clients' data **do** |
| 6: | Encrypt data and labels using public key : |
| 7: | Train local model on encrypted data using homomorphic operations: |
| 8: | Encrypt local model weights using public key $pk$: |
| 9: | Send encrypted model weights to central server |
| 10: | **end for** |
| 11: | **Aggregation Step:** |
| 12: | Initialise |
| 13: | **for** each client in clients' data **do** |
| 14: | Receive encrypted model weights from client |
| 15: |  |
| 16: | **end for** |
| 17: | Perform Federated Averaging on encrypted model weights: |
| 18: | Update global model weights using private key |
| 19: | Send updated global model weights to clients |
| 20: | **for** each client in clients' data **do** |
| 21: | Decrypt the updated global model weights using client 's private key : |
| 22: | Update local model with decrypted weights |
| 23: | **end for** |
| 24: | **Return:** updated global model |

## Performance Metrics

The model's performance was assessed using typical classification matrics like accuracy, precision, recall, and F1 score. In addition to these metrics, computation time was calculated throughout each FL round, which included encryption and decryption timings for each client as well as total computation time. For each round, the total encryption time for all clients was recorded. The total calculation time for each round was the sum of encryption, federated averaging, and decryption times. The total computation time for the calculation of round is defined in Equation (6).

(6)

where is the number of clients, is the encryption time for client is the time taken for federated averaging, and is the decryption time for round . These metrics allow for both an evaluation of the model's predictive performance and an understanding of the computational efficiency in a privacy-preserving FL setting.

## Experimental Setup

The experiments were conducted using a v2-8 TPU instance for accelerated model training and privacy-preserving computations via FL and HE. The system utilised 137 GB of RAM and 21.2 GB of disc space, providing sufficient resources for data processing. The software environment included TensorFlow, TensorFlow Federated (TFF), Tenseal (for CKKS-based encryption), and Scikit-learn for data preprocessing and performance evaluation. A Keras-based ANN model was trained using the Adam optimiser and binary cross-entropy loss. The FL setup involved 8 clients, each encrypting its local data with Tenseal and training a local model. The FedAvg algorithm was used to aggregate encrypted model updates on the central server, ensuring privacy throughout the process.

# Results Analysis

This study was evaluated by comparing the performance of the centralised and decentralised (federated) training approaches, which were assessed using a range of metrics, including accuracy, precision, recall, F1-score, and computational efficiency. The analysis highlights the effectiveness of both approaches, with a particular emphasis on the trade-off between model performance and privacy-preserving capabilities in the decentralised setting.

The training performance of the centralised model over 15 epochs shows efficient convergence with minimal overfitting, as evidenced by the decrease in training loss and the stable validation accuracy. This suggests that the model learnt effectively from the entire dataset and achieved optimal generalisation without facing significant overfitting. In the decentralised (federated) approach, which ensures data privacy by distributing the training across multiple clients, the global model achieved 92.28% accuracy after five rounds of training. While slightly lower than the centralised model, this result highlights FL's effectiveness in maintaining high performance despite decentralised data. As seen in Figure 3, the global accuracy improved from 83.06% in round 1 to 92.28% in round 5, illustrating the model's learning progression through federated updates. The computational efficiency of the federated model is demonstrated in Figure 4, which shows encryption and decryption times per round. Encryption times averaged around 156 seconds per client, with the decryption time for the aggregated results averaging around 285 seconds. As shown in Figure 4, while these encryption and decryption times introduce some overhead, the federated model’s privacy benefits ensuring data confidentiality during the learning process make it a promising approach for healthcare applications. This trade-off between computational overhead and enhanced privacy ensures that the FL setup strikes a balance between performance and security, safeguarding sensitive data throughout the training process. While both models achieved high accuracy, as shown in Figure 5, the decentralised model demonstrates competitive results, with a slightly lower accuracy and precision but comparable recall and F1 score when compared to the centralised approach.

|  |  |
| --- | --- |
| A graph with numbers and lines  AI-generated content may be incorrect.  **FIGURE 3.** Total Computation Time and Global Model Accuracy per Round in FL | A graph of numbers and a number  AI-generated content may be incorrect.  **FIGURE 4.** Encryption and Decryption Times per Round in FL |
| A graph of different colored bars  AI-generated content may be incorrect.  **FIGURE 5**. Performance Comparison: Centralised vs. Decentralised Approach | |

# Conclusion

This study presented a privacy-preserving approach to diabetes diagnosis that uses FL in conjunction with HE, implementing the CKKS scheme. As FL ensures that no sensitive patient data enters the centralised environment, this approach helps tackle privacy issues that plague centralised solutions. The use of HE allows processing of encrypted data, protecting the raw information on each patient even during aggregation to provide double-layered data privacy. The results of the experimentation confirm that the federated solution’s accuracy stabilises after five communication rounds, which is not too different from the results of a centralised model. This result highlights FL's potential to maintain privacy while delivering competitive accuracy. While encryption generates overhead, this is an acceptable trade-off in systems that handle sensitive health information. The developed method is a scalable solution that can successfully be applied to a variety of healthcare issues that require processing sensitive data. By combining the frameworks of FL and HE, the proposed solution makes a significant contribution to privacy-preserving machine learning by offering a more decentralised approach to handling data. This research opens new avenues for developing ML applications that are both efficient and secure in a medical setting and emphasises the need to balance performance with privacy in sensitive data applications. Future research will explore combining this approach with other privacy solutions to integrate FL with other popular methods such as secure multi-party computation (SMPC) and differential privacy (DP).

# Acknowledgement

This research was funded by the Ministry of Higher Education of Malaysia’s Fundamental Research Grant Scheme (FRGS/1/2023/ICT07/MMU/01/1).

# References

1. N. Khairallah, A. B. C. Douss, R. Abassi, M. A. Chalouf, and O. Hamdi, “Towards secure and privacy preserving data processing in e-health,” in *Proceedings of the 2024 20th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*, IEEE, pp. 136–141 (2024).
2. C. Fang, A. Dziedzic, L. Zhang, L. Oliva, A. Verma, F. Razak, and B. Wang, “Decentralised, collaborative, and privacy-preserving machine learning for multi-hospital data,” *EBioMedicine* 101 (2024).
3. Marcel, D. Miu, and A. Rancea, “Privacy preserving neural network models based on homomorphic encryption: A case study for diabetes prediction,” in *Proceedings of the 2023 IEEE 19th International Conference on Intelligent Computer Communication and Processing (ICCP)*, IEEE, pp. 269–275 (2023).
4. S. Moon and W. H. Lee, “Privacy-preserving federated learning in healthcare,” in *Proceedings of the 2023 International Conference on Electronics, Information, and Communication (ICEIC)*, IEEE, pp. 1–4 (2023).
5. Wang, H. Li, Y. Guo, and J. Wang, “PPHLHE: A privacy-preserving federated learning scheme with homomorphic encryption for healthcare data,” *Applied Soft Computing* 146, 110677 (2023).
6. T. Mora, D. Roche, and B. Rodríguez-Sánchez, “Predicting the onset of diabetes-related complications after a diabetes diagnosis with machine learning algorithms,” *Diabetes Research and Clinical Practice* 204, 110910 (2023).
7. C. Olisah, L. Smith, and M. Smith, “Diabetes mellitus prediction and diagnosis from a data preprocessing and machine learning perspective,” *Computer Methods and Programs in Biomedicine* 230, 107673 (2022).
8. G. Shen, Z. Fu, Y. Gui, W. Susilo, and M. Zhang, “Efficient and privacy-preserving online diagnosis scheme based on federated learning in e-healthcare system,” *Information Sciences* 647, 119261 (2023).
9. E. Huma, N. Tariq, and S. Zaidi, “Predictive machine learning models for early diabetes diagnosis: Enhancing accuracy and privacy with federated learning,” *Journal of Computing & Biomedical Informatics* 8(1) (2024).
10. S. S. Aslam and A. Aiosa, “A federated mining approach on predicting diabetes-related complications: Demonstration using real-world clinical data,” in *AMIA Annual Symposium Proceedings*, Vol. 2021, p. 556 (2022).
11. J. Liu, Z. Lu, H. Yang, and L. Zhuang, “A diabetes prediction system based on federated learning,” in *Proceedings of the 2022 International Conference on Big Data, Information and Computer Network (BDICN)*, IEEE, pp. 486–491 (2022).
12. P. Chuks, “Diabetes: hypertension and stroke prediction [Data set],” *Kaggle*, https://www.kaggle.com/datasets/prosperchuks/health-dataset?select=diabetes\_data.csv, accessed February 20, 2025.
13. R. G. Fernandez, and C. Menon, “Impact of Box–Cox transformation on machine-learning algorithms,” *Frontiers in Artificial Intelligence* 5, 877569 (2022).
14. J. H. Cheon, A. Kim, M. Kim, and Y. Song, “Homomorphic encryption for arithmetic of approximate numbers,” in *Advances in Cryptology–ASIACRYPT 2017: 23rd International Conference on the Theory and Applications of Cryptology and Information Security, Hong Kong, China, December 3–7, 2017, Proceedings, Part I*, Springer International Publishing, pp. 409–437 (2017).
15. T. Sun, D. Li, and B. Wang, “Decentralized federated averaging,” *IEEE Transactions on Pattern Analysis and Machine Intelligence* 45(4), 4289–4301 (2022).