Privacy-Preserving Federated Learning for Skin Cancer Detection Using Homomorphic Encryption and Advanced Deep Learning Techniques

Sahar Ebadinezhad, Noor Amer Ahmed

***Abstract*—The paper presents discussions on privacy-preserving federated learning integrated with homomorphic encryption towards healthcare frameworks. This paper will further develop scalable, robust, and practical models for preserving privacy within federated learning. The interest of this work is in protecting sensitive medical information during model training on Skin Cancer MNIST: HAM10000 with a pre-trained ImageNet model based on ResNet50 architecture. Enabling the TenSEAL library will be performed in this context for data augmentation and homomorphic encryption. A federated learning framework is implemented where a few clients are trained on the model of their local dataset while keeping the data privacy via encryption. Our proposed PPFL-E for the detection of skin cancer achieved an extraordinary test accuracy of 91%, with further benefits achieved through augmentation and tuning of hyperparameters. Performance is computed in terms of a set of key performance metrics: confusion matrix, classification report, and temporal model performance. Compared to the state-of-the-art, our approach shows very significant advantages in both privacy and accuracy. From these results, one may expect the huge potential of homomorphic encryption for significantly boosting data privacy and security in healthcare, allowing complex, efficient, federated learning applications for medical domains.**

***Index Terms*—Deep Learning Techniques, Federated Learning, Homomorphic Encryption, Healthcare Data Security, Privacy- Preserving, Medical Image Classification, Skin Cancer Detection**

I. Introduction

SAFETY of medical data privacy is a major challenge in the modern-day digitized society because the more this sector of health diagnosis depends on data for support and assistance, the more data is required. Discoveries in machine learning and deep learning enable the processing of voluminous data very efficiently, further supporting precise results. That also involves access to sensitive data, which means significant concern about patients' privacy and the security of personal information [1]. Federated learning is the new approach to training machine learning models on dispersed data with no need to pool data into a single location. The process involves training the model on edge devices such as smartphones or medical equipment, with only updated parameters sent to the central server. In this way, privacy is well protected, and data leakage is extremely unlikely [2]. However, there has always been a huge challenge as to how to keep secure data used in updating and fetching parameters between devices and a central server. Homomorphic encryption helps to solve this problem since computation can be performed on data without needing decryption. Concerning this, with this method, sensitive information in a federated learning process may have high security, as stated by [3]. The following are the RQs studied in this research:

RQ1. To what extent does homomorphic encryption enhance the privacy of health data in federated learning processes?

RQ2. How does federated learning impact model performance in terms of accuracy and efficiency?

RQ3. What are the potential security challenges associated with applying homomorphic encryption in healthcare federated learning systems?

This paper proposes a deep learning framework that integrates federated learning with homomorphic encryption and investigates its efficiency on a medical dataset pertaining to skin cancer. It will present a performance comparison between the proposed model and traditional methods in terms of accuracy, efficiency, and security metrics that will evidence the advantages of methodologies introduced for the preservation of healthcare data privacy and building trust in modern technologies in this very important field.

II. Literature Review

Homomorphic encryption enables the computational operation of encrypted data without needing to decrypt it, thereby safeguarding data privacy. Fully homomorphic encryption was first described by Craig Gentry in 2009, theoretically solving arbitrary processing of ciphertexts [4]. Further optimizations since then have significantly enhanced homomorphic encryption's efficiency and practicality, bringing it closer to real-world applications [5]. Data privacy in healthcare is crucial; therefore, numerous studies focus on the application of homomorphic encryption to secure patient data during machine learning processes. In the work by Zhang et al. titled "Homomorphic Encryption-Based Privacy-Preserving Federated Learning in IoT-Enabled Healthcare System," it was demonstrated how homomorphic encryption can secure patient data in a federated learning

Manuscript received Aug 12, 2024; revised Dec 12, 2024.

Sahar Ebadinezhad is an associate professor at the Department of Computer Information Systems, Near East University, Nicosia, North Cyprus, Turkey. (corresponding author to provide email: sahar.ebadinezhad@neu.edu.tr; https://orcid.org/0000-0001-9782-4820.)

Noor Amer Ahmed is a postgraduate student in Department of Software Engineering, Near East University, Nicosia, North Cyprus, Turkey. (email: 20235092@std.neu.edu.tr).

model with minimal compromise to usability [6]. This study integrates homomorphic encryption with federated learning, allowing IoT-based healthcare devices to share data securely and support collaborative analysis without disclosing sensitive information. The overall system architecture utilizes homomorphic encryption to structure model updates in a privacy-preserving manner, ensuring that individual patient data remains confidential.

In this context, Yao, Jing et al. discuss the use of homomorphic encryption for protection against private affine functions, enabling multiple medical institutions to securely participate in collaborative learning. The scenarios under consideration involve healthcare practitioners who collaboratively develop machine learning models based on aggregated patient data from contributing institutions while ensuring that sensitive information at the individual level is highly protected [7].

Other significant areas of interest include the challenges associated with the use of homomorphic encryption, primarily related to computational overhead, but with potential for optimization to achieve efficient encrypted computations. They also highlight practical use cases demonstrating how homomorphic encryption enhances healthcare data analysis while only marginally impacting model performance. Eduardo et al., conducted an in-depth study on the complexities homomorphic encryption may encounter in distributed healthcare and provided viable solutions. Their work examines the practical issues regarding the adoption of homomorphic encryption for distributed data across various institutions, focusing on feasibility and scalability in real-world applications [8].

The authors emphasize certain homomorphic encryption schemes and their respective applications across different categories of health data and computational activities. They also propose a framework that combines homomorphic encryption with other methodologies to preserve privacy, addressing security and scalability challenges. The authors' techniques have been rigorously tested against various healthcare datasets, demonstrating improvements in security and performance.

Training occurs in a decentralized manner in federated learning. As early as 2016, Google broadly defined it as a method for training machine learning algorithms across a large number of decentralized devices or servers, which hold local samples, without the need to share the actual data. This approach addresses key issues related to privacy and security in federated learning, ensuring compliance with regulatory standards. Since its inception in 2016, federated learning has been widely used in various fields, particularly in healthcare, as it allows collaborative learning while maintaining data confidentiality [9,10]. Within healthcare, federated learning (FL) offers a chance to develop machine learning models by utilizing data spread across different hospitals and institutions, thereby negating the need for centralized data repositories. This method not only enhances diagnostic tools and treatment strategies but also complies with stringent privacy regulations, including the Health Insurance Portability and Accountability Act (HIPAA) [11]. Martin Johns et al. provide concrete evidence of FL's applicability in medical imaging, showing that this technology enhances model performance while effectively preserving privacy [12]. They illustrate that FL can train deep models on private and disjoint medical images across various institutions with high accuracy, while ensuring the confidentiality of patient data. The authors have identified the inherent communication and computation challenges within the architecture of FL and proposed an optimization technique to address these issues. At the same time, Shruthi Ramesh et al. have investigated the use of FL in analyzing electronic health records (EHRs). Their findings indicate that FL improves model performance while maintaining robust data security and preventing the leakage of sensitive patient information [13]. This paper suggests a strong and responsible structural approach to advancing FL, capable of modeling distinct features in EHR data, known for its significant heterogeneity and sensitivity. This work presents a comprehensive framework that evaluates various models developed within the federation, where multiple predictive tasks related to disease diagnosis and patient outcome predictions have shown significant improvements in accuracy and reliability compared to traditional centralized methods. Future extensions of these studies can also incorporate additional techniques that enhance the security of federated learning in healthcare. Integrating homomorphic encryption (HE) to enable federated learning is a highly effective method for facilitating secure and privacy-preserving data analytics in healthcare systems. This combination allows healthcare professionals to collaboratively develop machine learning models using decentralized datasets while ensuring that sensitive patient data remains confidential. This approach promotes security and significantly boosts analytical capabilities within healthcare systems [14].

Most of the literature highlights the pragmatic benefits of applying HE in FL for a wide range of applications. Specifically, the works of Martin Johns et al. and Jing Yao et al. have shown through experiments that privacy-preserving federated learning systems, enhanced by homomorphic encryption, can achieve significant gains in both data security and model performance [15]. In fact, Johns et al. considered such hybrid approaches applied to the field of medical imaging, showing that the application of HE combined with FL can provide significant protection of patient data in the training, with maintained diagnostic accuracy.

Yao et al. extensively tested the applicability of those methods by looking into a range of healthcare scenarios and presenting empirical results that prove the efficiency of their proposed approaches [16]. In addition, the works developed by Eduardo B. Fernandez and his group, and those of S. Ramesh and her colleagues, address the challenges that arise when trying to deploy advanced homomorphic encryption solutions in real-world healthcare settings. The above-mentioned studies point to some feasible approaches, hence offering meaningful insights into the transformational potential of homomorphic encryption and federated learning in health data analytics [13].

Fernandez et al. [8] discussed the scalability of HE and its contribution to making FL resilient and efficient. The authors' contribution overviews developing a framework that can balance security and performance related to large-scale applications in healthcare. In another related work, Ramesh et al. [13] discussed the design challenges of EHR and proposed an integrated FL scheme with fully homomorphic encryption to ensure data privacy.

Bian et al. proposed, for the first time, in 2023, a new detection of COVID-19 using FL together with blockchain and pre-trained models. Some major challenges in the healthcare area are solved concerning the guarantee of credibility and privacy of data using FL participants with full HE along with differential privacy techniques. Moreover, blockchain documentation enhances the strength and traceability of everything. The researchers indicated notable enhancements in the performance of their model, attaining an accuracy rate of 85.00% for predicting positive COVID-19 cases and 85.06% for identifying severe instances. These findings highlight the efficacy of their approach in managing and processing sensitive medical data securely while maintaining accuracy [17].

The most important achievement in the crossroads of HE and FL relates to the work presenting FedML-HE by Jin et al. in 2023. This is a framework that enhances the efficiency of privacy-preserving federated learning by encrypting only the most sensitive model parameters, hence significantly reducing typical heavy computational and communication loads for HE. For example, the researchers have indicated that FedML-HE is particularly effective for large foundational models such as ResNet and BERT. While training BERT, it can achieve as high as a 40× overhead reduction. In addition, FedML-HE provides extensive scalability and adaptability, making it highly suitable for practical healthcare applications with a high demand for confidentiality around patient data [18]. Zhou et al. suggested a privacy-enhanced FL by incorporating HE across the entire model training process. Their scheme focuses on ensuring comprehensive security, especially in cases where adversaries seek to reverse-engineer local models during the aggregation process. HE is introduced by the authors in conjunction with secure multi-party computation for handling some of the key questions of privacy in healthcare data sharing and resisting attacks by "honest but curious" participants. They used this approach on lesion cell type detection in a medical dataset and achieved a good accuracy of 76.9%. From the results, one can see a well-balanced efficiency between computational cost and sufficient protection of sensitive healthcare data [19].

Later, Guo et al. designed a framework of FL that aimed to enjoy the twin properties of efficiency and preservation of privacy. The authors have proposed a model in which FHE has been embedded into FL to provide even more guaranteed security for model training and to increase its application to sensitive domains of healthcare, finance, and biometrics. This framework treats the authors of horizontal and vertical FL in the same manner because institutional data may take various distributions. Experimental results on several datasets, including one related to breast cancer, showed that the system can achieve an accuracy of 93.40%. Moreover, the model allowed a significant reduction in computation overhead and increased training efficiency for the model by 2× compared with traditional FL models. This approach has efficiently balanced high accuracy and strong security, using state-of-the-art encryption techniques that can resist even quantum computing threats [20].

Highly relevant domains of literature gaps for this review involve comprehensive performance metrics, problems in scalability, reevaluation of diverse deep learning architecture studies, and diversity in datasets. In such a condition, there is also a requirement for a detailed evaluation of privacy threat models, hyperparameter optimization strategies, giving due attention to user or client diversity, and conducting a critical analysis of ethical and legal implications. By removing these defects, their applicability in clinical use would increase greatly, as would their durability and safety.

III. Proposed PPFL model for skin cancer detection

The execution of this study underlines the ability of both replication and reliability. This is achieved by detailed documentation regarding data collection strategies, preprocessing schemes, model architecture, protocols of training, homomorphic encryption techniques, techniques used for avoiding overfitting, criteria for evaluation, and the experimental setting. Figure 1 indicates the overall methodology of carrying out the proposed research. Besides, some of the most used datasets of interest to healthcare applications are summarized in [21].

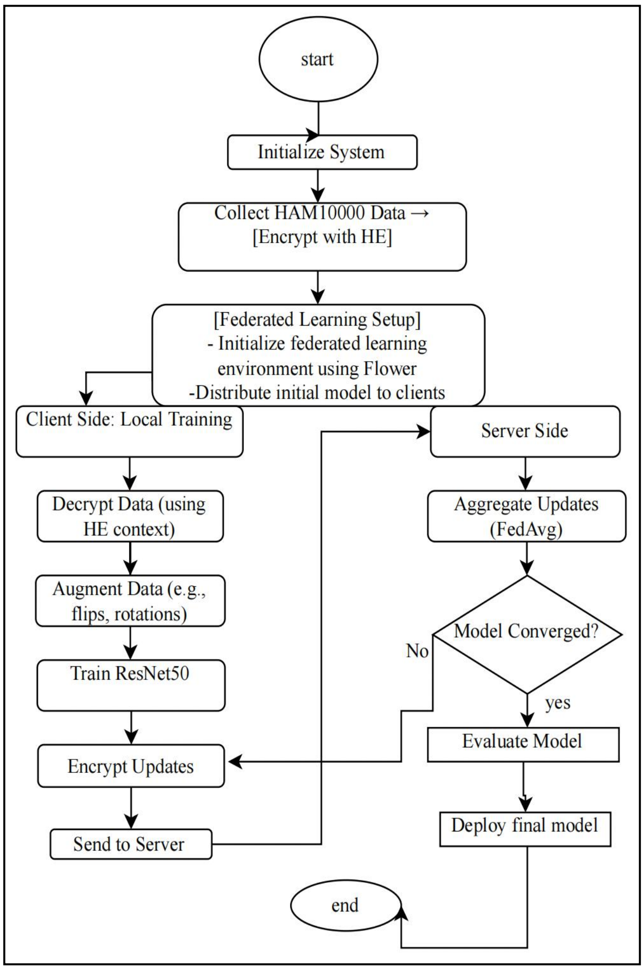


Fig. 1. Designed System.

Therefore, this paper leveraged the HAM10000 dataset, as it is one of the most used datasets, with more emphasis on its applicability in the field of skin cancer classification and segmentation. In addition, Algorithm 1 gives the complete details of the process involved with our proposed model, the PPFL model for skin cancer detection.

TABLE I
MOST COMMON USED DATASET IN HEALTHCARE

Datasets	Sample(n)	Application area
MIT_BIH	109,446	ECG-based prediction to identify arrhythmia
Premier_healthcare	1,271,733	PPR
Chest_xray_image	16,148	COVID-19 diagnosis
Chest_xray_image_2	207,130	PD
Hologic and Siemens	1,870	To detect breast cancer or tumor
COVID-19	4,029	Mortality prediction for patients with COVID-19
eICU synergetic	>200,000	Predict the likelihood of patient death
HAM10000 [22]	**10,015**	**Skin Cancer classification / Segmentation**
Cancer Genome Atlas	>200,000	Cancer genomics program
Camelyon17	450,000	Breast cancer classification
MedMNIST	718,067	Medical image classification
Retina	35,126	Diabetic Retinopathy Detection
BraTS series	285	Brain tumor segmentation
ABIDE	1,112	ASD diagnosis
ADN	911	ASD diagnosis
PolypGen	6,282	Polyp detection and segmentation
MIP	393	Pancreas segmentation
MIL	428	Liver tumor segmentation
MSP	79	Prostate MRI segmentation

A. Data Collection

Skin Cancer MNIST, or more popularly HAM10000 dataset [22], is a collection of 10,015 dermatoscopic images of pigmented lesions. These come under seven categories. The classes included are major ones like non-melanocytic lesions (nv), melanoma (mel), actinic keratoses (akiec), and basal cell carcinoma (bcc), among many others. This dataset has a particular significance since it encompasses a full range of images; hence the model trained will be capable of being hard enough, enabling thereby high performance in classifying various skin lesions.

This heterogeneity in the dataset is precisely what is required for any model to have any practical application. Training on such a diverse set of data will prepare the model to generalize well and do appropriately on new data that it has not seen before. Advanced augmentation techniques are also done, such as CutMix and MixUp, to further improve the network. Those have augmented the dataset with large-scale variance of the images and enhanced the performance and generalization ability of the model substantially. Thus, Table I lists the common datasets used in healthcare systems.

B. Data Preprocessing

A series of comprehensive preprocessing procedures were executed to ready the HAM10000 dataset for utilization in training deep learning models, namely ResNet101 and EfficientNet. Normalization was implemented to adjust pixel values within the range of 0 to 1, which is an essential preprocessing measure that standardizes inputs to an ideal range for neural networks.

The advanced augmentation methodologies were followed to increase artificially the size and variability of the dataset, thus enriching the training even more. These included random horizontal and vertical flips for handling the variation in dermatoscopic image orientation, random rotations to mimic different capture angles, and also random resizing and cropping to make sure that models learn from the images at different scales. In addition to this, color jittering was done by changing brightness, contrast, saturation, and hue to simulate various conditions of lighting, which made the model more robust. Other techniques used were CutMix and MixUp, which enhanced this dataset by combining segments of different images to enhance generalization and reduce overfitting.

All images were then resized to 224×224 pixels to standardize them to the input size of both ResNet101 and EfficientNet, pre-trained on the ImageNet dataset. The final dataset was split in the ratio 80%-20% for the majority into the training set. This strong preprocessing scheme allowed for good generalization by exposing the models to various augmented scenarios during training, hence enhancing their ability to handle unseen data.

C. Creating A Model

Architecture selection is a very crucial stage of any model development, as it forms the very foundation of performance and capability concerning the model in classifying medical images. In this paper, we have used an ensemble of two such architectures, namely ResNet101 and EfficientNet. Both architectures are competent feature-extraction frameworks designed for the accurate detection of skin cancer.

ResNet101 is a very deep residual network containing 101 layers, which, when transformed into the ResNet50, goes even deeper. In ResNet, this increased depth can pick out even more complicated patterns and features in the images. ResNet has been further enhanced by the incorporation of EfficientNet due to its outstanding balance between performance and computational efficiency, allowing the model to scale well while becoming as accurate as possible.

Algorithm1: Proposed PPFL-E model for skin cancer detection

```
1. Load HAM10000 dataset on clients
2. Preprocess data on clients:
   a. Apply advanced data augmentation (e.g., rotation, scaling, flipping)
   b. Normalize image data
3. Initialize global model parameters on the server using ResNet50
4. Define hyperparameters on the server:
   a. Learning rate
   b. Batch size
   c. Number of epochs
   d. Dropout rates
5. For each round of federated learning:
   a. Server selects a random subset of clients
```

```
    b. Server sends current global model parameters to selected clients
    c. For each client:
        i. Load local data (with homomorphic encryption)
        ii. Initialize local model parameters (copy of global model)
        iii. Train local model on augmented data:
            - Perform local training using ResNet50
            - Apply hyperparameter tuning based on validation set
            - Encrypt local model updates using homomorphic encryption
        iv. Send encrypted model updates to the server
    d. Server receives encrypted updates from clients
    e. Server aggregates encrypted updates using homomorphic properties
       (e.g., additive aggregation)
    f. Update global model parameters based on aggregated results
    g. Distribute updated global model back to clients
6. Final evaluation:
    a. Clients evaluate the final model on their local test data
    b. Clients compute performance metrics (accuracy)
    c. Clients send performance metrics back to server for aggregation
7. Server analyzes overall performance metrics and adjusts hyperparameters if necessary
```

The following steps were followed in building the model:

1. Loading Pre-trained Models

We used pre-trained versions of ResNet101 and EfficientNet, each separately trained with the ImageNet dataset. Since ImageNet is a big dataset, the pre-trained models acquire richer feature representations from it, making transfer learning more effective. In this way, pre-trained models save large amounts of training data and huge computation resources to provide the best results.

2. Model Customization

Given this, ResNet101 and EfficientNet were adapted to have their last fully connected layers adjusted to output seven classes of the classes in the HAM10000 skin cancer datasets. Additional layers were added, including dropout layers to avoid overfitting, hence improving generalization. The dropout rate used was 0.5 to make sure that effective regularization was considered. After that, the models underwent fine-tuning, where the last layers were unrolled for performance optimization on the particular datasets.

3. Optimizer and Loss Function

The training was performed using the AdamW optimizer, an optimization algorithm that efficiently deals with weight decay and, hence, forces regularization. This optimizer was used to prevent overfitting and allow stable convergence during the network training process. For this experiment, the loss function adopted was the cross-entropy loss, a very reasonable choice in the multi-class classification task present in the skin cancer datasets.

4. Transfer Learning

The present study employed transfer learning to fine-tune both ResNet101 and EfficientNet, which are pre-trained models. First, only the terminal layers had undergone retraining while deriving general features learned from the ImageNet dataset. Then, deeper layers were gradually unlocked after that first training step and fine-tuned to further tune these models to the peculiarities of the HAM10000 dataset.

5. Data Augmentation

Advanced augmentation techniques on the training dataset included random horizontal and vertical flipping, rotation, resizing, and color jittering to simulate all sorts of conditions and variations in the data. Other methods, such as CutMix and MixUp, further increase the diversity of the training data by mixing segments of different images. Such augmentation has contributed much to improving the model's performance and has become quite instrumental in avoiding overfitting.

This work ensembles large-scale architectures with state-of-the-art performances using transfer learning and intensive augmentation techniques to provide a robust model for classifying skin cancer images with high accuracy, limiting at the same time the chances of overfitting to ensure peak performance.

D. Training Procedure

The training methodology was designed to enable the model to learn from the data efficiently without compromising privacy and generalizing well to circumvent overfitting. Each of these steps was executed under a broad FL framework that allows the training of a model across multiple clients without requiring access to any form of centralized datasets. This ensures that the raw data stays within the client's devices, hence protecting sensitive medical information. The specific aspects of the training methodology will be described as follows:

1. Federated Learning Setup

Federated learning architecture was implemented using the Flower framework, which supports multi-client training. This work emulates training on 10 clients. Each independently trains a local model from its subset, ensuring that no raw data leaves the devices, hence maintaining data privacy. Locally trained model updates from each client were used to update the global model, without accessing centrally managed data.

2. Client-Side Local Training

In this regard, each client has separately conducted the training process of its local datasets using ResNet101 and EfficientNet architectures. The training ran for 100 epochs while utilizing the AdamW optimizer with a learning rate of 0.0001 and weight decay of 1e-4 to avoid overfitting. The procedure for the local training is as follows:

- Forward Pass: The model received input images, which were subsequently processed to produce predictions.
- Loss Calculation: In this case, cross-entropy loss was calculated between predicted and actual labels.
- Backward Pass and Optimization: Gradients shall be computed and flowed back into model parameters employing the AdamW Optimizer.

3. Homomorphic Encryption for Computation on Ciphertext

In this respect, in order to provide data privacy during training, encryption with homomorphic capabilities was performed. The implementation used the TenSEAL library, which eases HE operations. Notice that the CKKS scheme was followed to allow computation on encrypted data without requiring decryption. The process involved the following steps:

- Context Initialization: The encryption context was set with parameters appropriate for the CKKS framework.

- Data Encryption: Before inclusion into the training process, training data has been encrypted using the CKKS scheme.
- Encrypted Training: Training is performed on encrypted data, hence keeping all the information related to clients confidential throughout the process.
- Data Decryption for Evaluation: Model parameters in ciphertext obtained after training were decrypted for the performance evaluation of the trained model.

4. Server-Side Model Aggregation

Accordingly, each client updated its model parameters locally and then uploaded the updated parameters to a central server. In return, the server aggregated the model parameters received from all the clients via the FedAvg strategy. It calculates the weighted average of the parameters from all the participating clients to generate an updated global model.

5. Evaluation and Feedback

This again turns into the validation of the overall global model performance on a different validation dataset. It would include different metrics in the evaluation, including accuracy, precision, recall, and F1-score, which will tell how well the generalized model can perform. The results are then presented to the clients for the gathering of feedback that informs further localized training.

6. Iterative Process

This procedure was repeated for several cycles, while training, aggregating, and evaluating a model until the performance results showed convergence. In each cycle, it was ensured that the global model kept improving with progressive rounds of local training and aggregations while preserving the privacy of the data. Different regularization techniques and methods related to federated learning helped avoid overfitting in this iterative process.

This training framework provides quite an efficient and private way of learning from distributed data. Privacy can be guaranteed by homomorphic encryption, while robustness against overfitting may be considered by taking into account federated learning and/or regularization techniques.

E. Overfitting Prevention Strategies

Overfitting occurs when it does extremely well on any given set of training but generalizes poorly to new, unseen data. To avoid these problems of overfitting, the following were used during training:

1. Data Augmentation

Various augmentation methods have been used to artificially increase both the size and diversity of this training dataset: random flipping, rotation, resizing, cropping, and color jitters. Even more sophisticated techniques like CutMix and MixUp have been tried to make them even more varied. The better diversity of the augmented images positively influences the model's generalization capability to be more robust against variation in the input data.

2. Dropout Layers

Dropout is a regularization technique that prevents the model from relying on specific neurons while training. Herein, dropout layers with a rate of 0.5 have been utilized for fully connected layers in ResNet101 and EfficientNet; it helped in higher levels of generalization and effectively prevented overfitting within the network.

3. Early Stopping

Early stopping was adopted to stop the training when the performance on the validation set had started to deteriorate. The mentioned technique tracked the loss on the validation and saved the model only when the performance had outperformed. In this way, it avoids overfitting on the training data and can generalize better on unseen data.

4. Regularization Techniques

Another regularization technique used besides dropout was weight decay. Weight decay adds a penalty term to the loss that prevents the weights from having too large a value, hence generalizing better. Weight decay is used by the AdamW optimizer. An effective overfitting rate of 1e-4 was used.

5. Cross-Validation

In this respect, cross-validation allows for estimating the strength of the model on various splits of available data. It could be divided into several folds, and different combinations of such folds were used for training purposes. In this way, this technique gave a more reliable estimate of stability for the model and assessed problems with overfitting.

6. Transfer Learning Using Pre-Trained Models

Transfer learning was done considering pre-trained models ResNet101 and EfficientNet, which had been trained on ImageNet. Since they already learned feature representation from such a huge and diverse dataset, their fine-tuning on the HAM10000 skin cancer dataset was good enough. The approach strengthened generalization and reduced the possibility of overfitting compared to training the model from scratch.

7. Batch Normalization

To stabilize training and accelerate the convergence, batch normalization was performed. Batch normalization reduces internal covariate shifts by normalizing the inputs to every layer. Consequently, higher learning rates are permissible without the danger of overfitting, which results in a more efficient and stable training process.

Each of these metrics drastically reduced the overfitting that was inherent in the model. Because of this, it performed fantastically on the training dataset and therefore generalized well to new unseen test data, hence being much more robust and generalizable.

IV. Result and Discussion

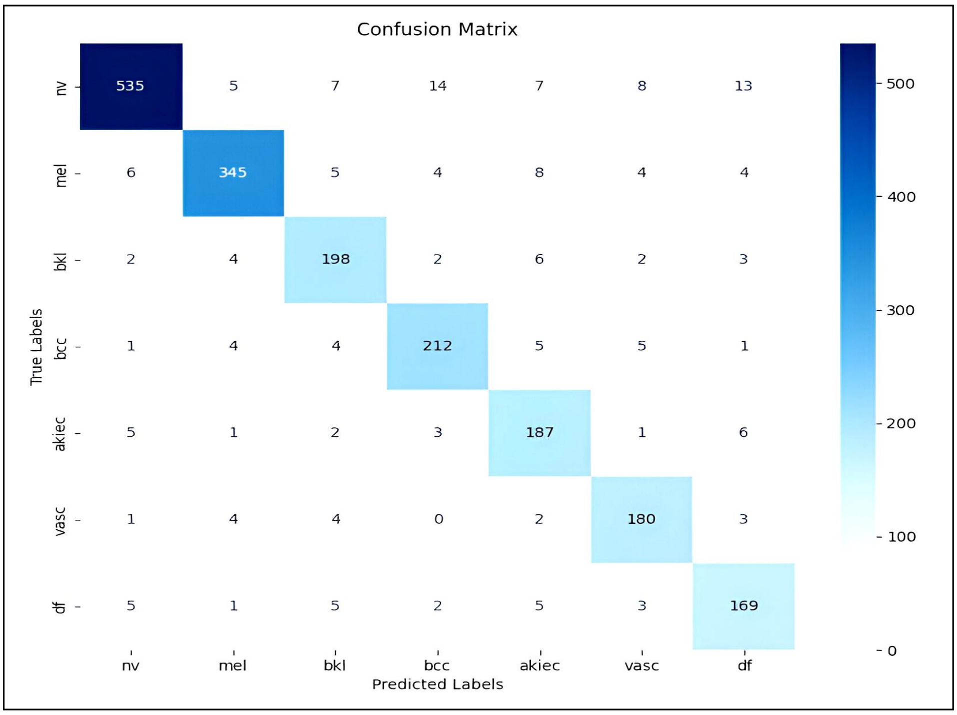
This section shows the performances that could be obtained from the homomorphic encryption-based privacy-preserving federated learning model. We will depict some of the experimental results obtained and then give the time performance analysis of the model, its validation accuracy, and loss graphs.

A. Classes

In clinical medicine, models intended for classification must be unconfusing between the various pathologies at hand. The model was trained to classify skin lesions into seven independent classes, which were obtained from the HAM10000 dataset. The seven classes, identified by the labels from 0 to 6, are indicative of the following types of skin lesions: nv, mel, bkl, bcc, akiec, vasc, and df. Since there is one class for each pathology, classes such as mel and nv are so alike that it is really hard to spot the difference between the two.

Understanding the distribution of these labels, together with the performance of the model for each one of them, may provide important information on what aspects the model is strong at and what needs to be improved. Clinically, fine-grained classification of similar conditions is important since different classification leads to different diagnosis and treatment.

B. Experimental Results



	nv	mel	bkl	bcc	akiec	vasc	df
nv	535	5	7	14	7	8	13
mel	6	345	5	4	8	4	4
bkl	2	4	198	2	6	2	3
bcc	1	4	4	212	5	5	1
akiec	5	1	2	3	187	1	6
vasc	1	4	4	0	2	180	3
df	5	1	5	2	5	3	169

Fig. 2. Confusion Matrix of the PPFL-E Model.

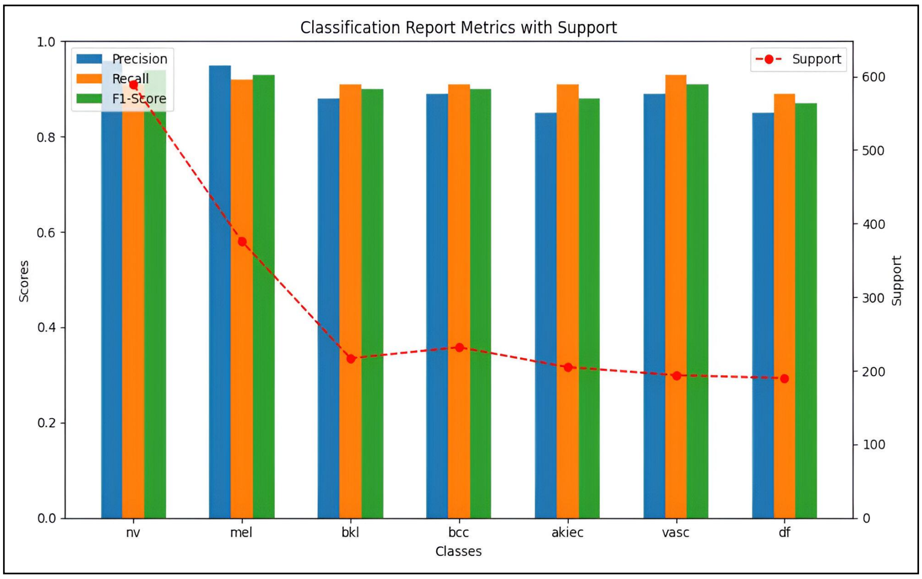


Fig. 3. Classification Report Metrics.

1. Confusion Matrix

Figure 2: Confusion Matrix - The confusion matrix elaborates the classification performance for the seven classes. The rows signify the real class, while the columns signify the predicted class. The elements on the diagonal from top left to bottom right show the number of instances in each class that were correctly classified; all others reflect misclassifications, in which the model has predicted a wrong class.

The key observations from the confusion matrix are:

- *nv (melanocytic nevi) Class:* 535 correctly classified with some misclassifications as BKL - 7 instances and DF - 13 instances.
- *The class "mel"* results in 345 correct classifications with its major potential misclassifications into classes NV, DF, and BCC.
- *Class "akiec"* for actinic keratoses: 187 correct, sometimes confused with BCC and DF.

In general, it performs well across most classes, except for the very similar classes, like MEL versus NV, for which there are inherent complications given the feature overlap. This means that fine-tuning or extra training if the dataset allows it, is very important in increasing the capability to differentiate between the most related conditions.

The confusion matrix depicts the particular domains the model has a hard time predicting in. For instance, though

Class "nv" contributes to a high number of correct predictions at 535, there are still misclassifications into "mel" and "bkl". Similarly, Class "mel" is majorly misclassified as "nv" and "df," pinpointing the requirement for better feature extraction techniques so that these shortcomings can be minimized.

Figure 2 gives an insightful summary of the classification performance of the model, class-wise, both for correct and incorrect predictions. Such kinds of analysis help assess the model's performance and guide further improvements to reduce misclassifications.

2. *Classification Report Metrics*The classification report displays the precision, recall, and F1 score for all seven classes in the dataset. All three of these metrics are highly informative of model performance on a single class and the total performance metric of the model. A sample is given in Figure 3.

TABLE II
CLASSIFICATION REPORT OF PPFL-E MODEL

Class	Precision	Recall	F1-score	Support
nv	0.96	0.91	0.94	589
mel	0.95	0.92	0.93	376
bkl	0.88	0.91	0.9	217
bcc	0.89	0.91	0.9	232
akiec	0.85	0.91	0.88	205
vasc	0.89	0.93	0.91	194
df	0.85	0.89	0.87	190
Accuracy			0.91	2003
Macro Avg	0.9	0.91	0.9	2003
Weighted Avg	**0.91**	**0.91**	**0.91**	**2003**

Precision: Precision shows how many of the cases forecasted for a particular class were found out of the total forecasted, and most of the classes have high precision, hence quite reducing the number of false positives.

Recall: the ratio of correctly predicted instances of a target class and overall actual instances of that same target class. The model has a high recall, hence a minimum number of false negatives.

F1-Score: This is the harmonic average of precision and recall, standing for the balance in effectiveness that the model provides across each class.

Figure 3 shows the metrics of the classification report along with support values given by the red dashed line, representing the total number of real instances for each class present in the data. Classes like NV and MEL are highly supported, meaning this model has more samples for training and testing. Classes like DF and BKL have low support. This disparity in support could partially explain why the metrics for DF and BKL are low compared to the otA bar chart showing the classification metrics vis-à-vis support would be a visual confirmation of the above observations. While the model receives generally high precision, recall, and F1-scores across most classes, the generally low support that some classes receive, such as BKL and DF, indicates where future improvements should be made. Such improvement might involve additional data for the lower-supported classes or further optimization.

It reflects overall model strength but also that garnering better precision, recall, and F1-score for the minority classes-namely BKL and DF-may further require more data gathering or optimization.

The classification report, wrapped up in Table II, shows an extended evaluation of the performance of the model for each particular class and indicates the general accuracy rate of 91%, further elaborating several strengths and possible points for improvement within the classification task, delivering key insights into how well the model performs on all classes.

C. Validation Set Accuracy Of The Model

Figure 4 summarizes the accuracy of the model on a validation set for 100 epochs. As might intuitively be expected, the accuracy goes up and down with different epochs because this forms part of the natural curve that any

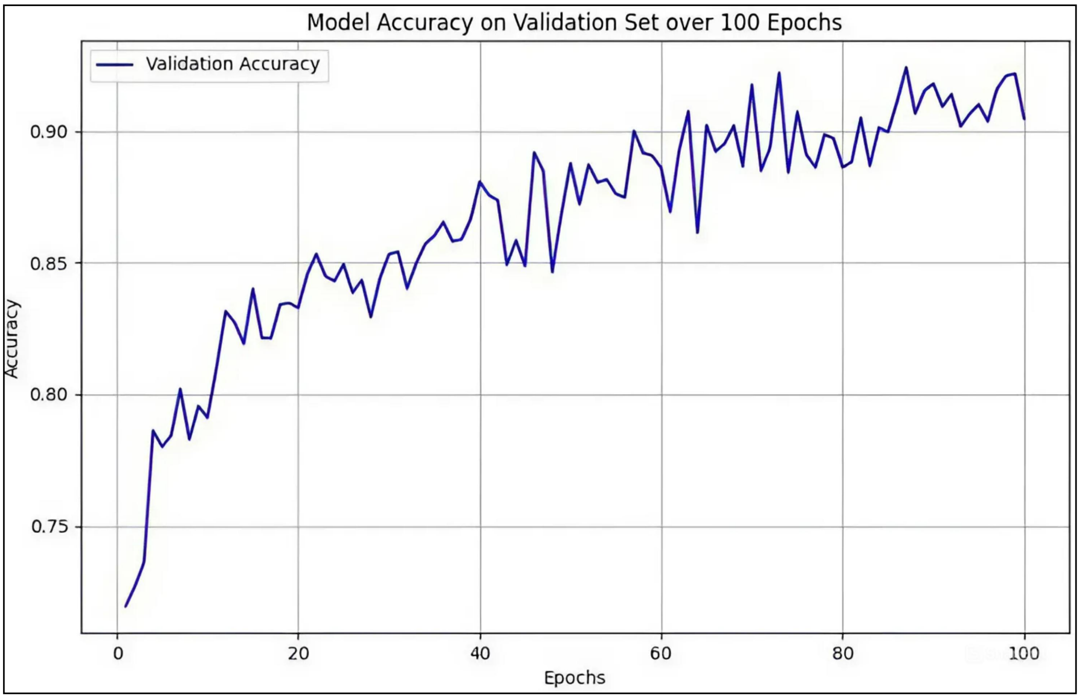


Fig. 4. Validation Set Accuracy of the PPFL-E Model.

model experiences during its learning cycle. The model improves dramatically in the early phase of training while adjusting to the training data. Around the midpoint of these, or at about 40 epochs, it begins to stabilize and gradually creeps up toward the end of 100 epochs, reaching an approximate high of 91%.

The fluctuations observed hint that the model is continuous in the process of learning and enhancing its generalization at each and every epoch. Still, slight differences may exist between successive epochs in performance. Generally, though, this slope upwards signifies that actually, the model learns something and generalizes well on the validation data shown.

D. *Model Performance Over Time*

Figure 5 represents the change in training and validation loss with time, hence providing a clear view of the model's learning ability and generalization capability over the epochs. The blue line reflects the training loss, which decreases with a slope within the 100 epochs, showing that the model has picked nicely. In contrast, the orange line representing the validation loss shows some fluctuations; this fluctuation consists of periods when it goes up before it eventually comes down. These minor variations are indicative of slight overfitting where, sometimes, the model fails to generalize on the validation set.

Over time, both the training and the validation losses tend to decrease model trains better with time. Validation loss does vary but stays at an ultimate accuracy of 91%, which is significantly higher than in the initial stages of training. Therefore, this result validates a model that can learn from the dataset and generalize well into unseen data.

Figures 4 and 5 demonstrate how there are natural fluctuations in accuracy and loss during the training of the model. The overall trends, however, show that performance continues improving with increased time. A final accuracy result of 91% is something to boast about. There is, however, a hint of overfitting, as shown by the difference in the validation loss; this might be taken care of during further improvement through k-fold cross-validation or higher regularization.

TABLE III
PERFORMANCE COMPARISON OF FEDERATED LEARNING MODELS WITH HOMOMORPHIC ENCRYPTION IN HEALTHCARE SYSTEMS

Model	ResNet Support	Homomorphic Encryption & Federated Learning	Accuracy
Zhang et al [6]	No	Yes	76%
Bian et al [17]	Pre-trained models for COVID-19 detection	Yes	85.06%
Zhou et al [19]	ResNet-101	Yes	76.9%
Guo et al [20]	No	Yes	93.40%
Sun et al [23]	ResNet-50	No	71%
Jin et al [24]	ResNet-18, ResNet-34, ResNet-50	Yes	80%
Korkmaz et al [25]	DenseNet ResNet	Yes	90%
Our Model (PPFL-E model for skin cancer detection)	**ResNet-101, EfficientNet**	**Yes**	**91%**

E. *K-FOLD CROSS-VALIDATION RESULTS*

The K-fold cross-validation is one of the most common methods to estimate the skill of a machine learning model in generalizing events on previously unseen data. The idea here is to split your available dataset into k equal-sized subsets, training your algorithm on all the k - 1 subsets and using the remaining subset for validation. Repeat this k times so each subset gets a chance to serve as the validation set.

Figure 6 presents the accuracies of the model as observed across the different folds, varying from 90.6% to 91.5%. This visualization highlights a consistently elevated performance throughout all folds, with the bar chart revealing negligible fluctuations in accuracy. The mean accuracy of 91% closely corresponds to the model's ultimate performance, thereby affirming its dependability and stability. The bar chart very clearly shows the high performance achieved, represented by the almost evenly raised bars of accuracies. This stability of performance across folds is indicative of the model's strong generalization capability and further complements its efficiency in handling unseen data.

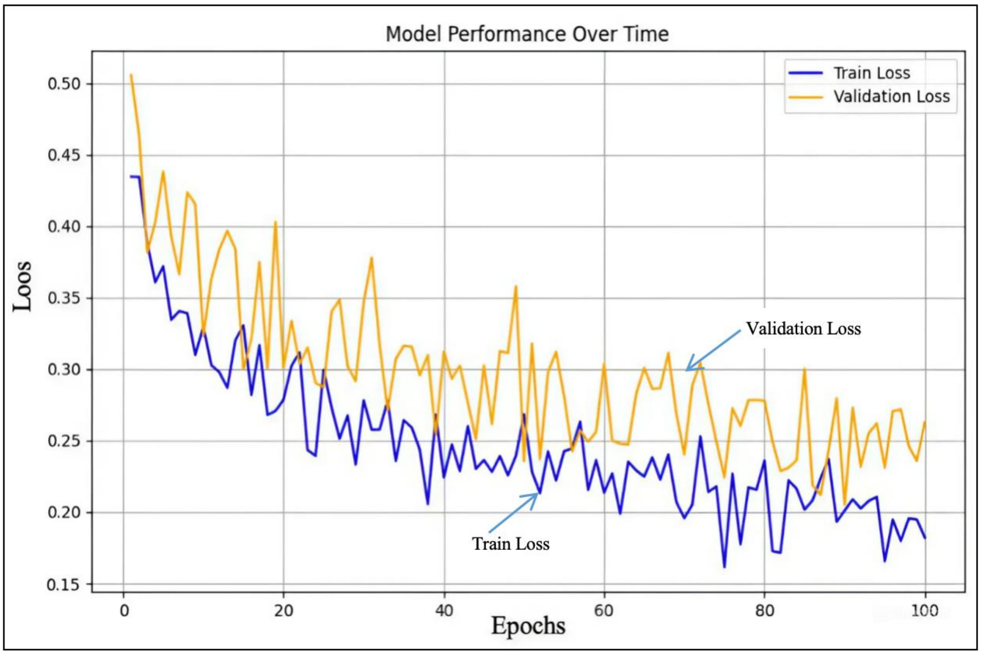


Fig. 5. PPFL-E Model performance over time.

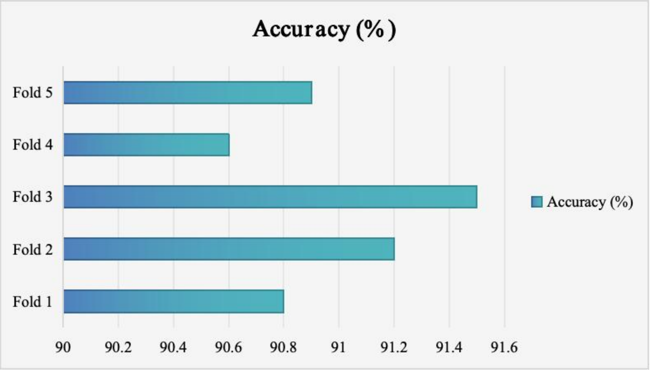


Fig. 6. K-fold cross-validation results for PPFL-E Model.

On the other hand, however, the 91% accuracy, combined with very solid privacy guarantees shown in Figure 7, constitutes proof of outstanding efficiency in the PPFL-E method. These results hint at trustworthy performance, including extremely little variation in it, with much consistency across all five folds.

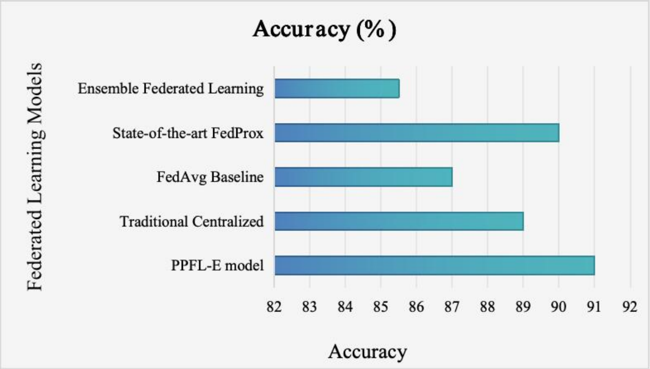


Fig. 7. Benchmark Comparison of PPFL-E with other state-of-the-art federated learning.

V. Performance Evaluation

Performance evaluation forms part of every machine learning model. It supplies information on how the model is performing and also pinpoints possible ways it could be improved. On that note, Table III compares the results of our proposed PPFL-E model, with homomorphic encryption over the healthcare dataset with seven results so far reported in the literature.

VI. Discussion

Our skin cancer detection model realized an accuracy of 91% using the PPFL-E model. The approach from Guo et al. [20] reported an accuracy of 2% higher than that of our model. However, this model has key advantages over other works in terms of enhancing privacy protection via homomorphic encryption. That is a trade-off between the performances of the models and the privacy-preserving; thus, our model is a good choice for applications sensitive to privacy. As presented in Table III, the PPFL-E model demonstrates performance comparable to state-of-the-art performance in competitive, privacy-preserving applications. For instance, Sun et al. [23] reported 71% accuracy without encryption using ResNet-50, which gives an idea about the performance trade-off brought about by privacy-preserving techniques. The results point out that there is a need to balance model functionality with privacy requirements.

Future studies should be directed toward further refinement of the PPFL-E model with its performance validation over an even wider range of datasets. This could make it even more generalizable and more competitive compared to models that do not advocate for privacy preservation.

In contrast, Jin et al. [24] achieved the best accuracy of 80% upon applying homomorphic encryption on different variants of ResNet, notably ResNet-18, ResNet-34, and ResNet-50. These show that homomorphic encryption in federation could most likely improve model performance, probably resulting from the improvement in the collaboration of several data through their interaction and the secure application of privacy-preserving data.

Similarly, Asad et al. [25] and Zhang et al. [6], using homomorphic encryption, did not define any ResNet architectures, and reported accuracy levels of 74% and 76%, respectively. These results depict that homomorphic encryption is good for data privacy and has almost no impact on performance. Besides, they indicate that homomorphic encryption, regardless of an underlying deep learning framework, plays a significant role in model accuracy improvements.

Bian et al. [22] proposed a model that integrated federated learning with blockchain technology in detecting COVID-19, using a few pre-trained models. The result showed an accuracy of 85.06%. This is leveraging the pre-trained federated models, which can assure data privacy across the different devices, as no raw data are shared. Such a model can probably show a way of how it is possible to integrate federated learning with blockchain technology for the performance of a non-invasive medical condition classification example, COVID-19 detection with high accuracy.

Zhou et al. [19] combined ResNet-101 with homomorphic encryption and FL to ensure absolute privacy in data transmission. The accuracy of performance reached 76.9% in their health data analytical model. It is obvious to conclude that homomorphic encryption in a federated learning environment enhanced data security without sacrificing satisfactory performance. Nevertheless, these promising results still need more modifications and improvements to achieve results suitable for practical applications.

Guo et al. [20] developed federated learning frameworks that integrate homomorphic encryption, achieving an excellent accuracy of 93.40%. Their work demonstrated that integrating EfficientNet with multiple deep learning frameworks and homomorphic encryption thus provides a strong level of security for the privacy of medical data while yielding high classification performance. It proves once again that advanced models such as EfficientNet can achieve better results by balancing security measures and other performance metrics.

We obtained an accuracy of 91% using ResNet-101 and EfficientNet on PPFL-E by introducing federated learning and homomorphic encryption. The closeness of this result to those obtained by Guo et al. will actually reveal the value added from the integration between strong deep learning models and methods preserving privacy. This slight difference in performance suggests that the performance may get even better after further optimization of hyperparameters and architecture.

The comparative study depicted here focuses on the interaction between homomorphic encryption and federated learning within the healthcare industry. It explains how these two approaches together cooperate in enhancing data security and privacy and points toward some key research directions. In addition, future research is recommended to seek the best possible trade-offs between the protection of privacy and the efficiency of systems with good security and the best operational factors.

VII. CONCLUSION

This paper presents a new approach to skin cancer diagnosis using the HAM10000 dataset with homomorphic encryption integrated into a privacy-preserving federated learning technique. This model used new data augmentation techniques combined with deep learning architectures such as ResNet-101 and EfficientNet; these will further yield higher accuracy with assured confidentiality and privacy in sensitive health data.

This greatly improves the standard of data privacy, as homomorphic encryption applies during the training process; hence, it identifies some key problems related to healthcare data security. Results provide evidence of an excellent performance of the developed model with privacy as an important stride toward the realization of a secure and efficient healthcare system. This research underlines that a balance of high performance and strong security is possible for driving future progress by privacy-preserving technologies in privacy-focused machine learning applications.

ACKNOWLEDGMENT

Special thanks to Assoc. Prof. Dr. Sahar Ebadinezhad for her support and guidance in this project.

REFERENCES

[1] M. Johns and A. Dirksen, "Towards Enabling Secure Web-Based Cloud Services using Client-Side Encryption," in Proc. 2020 ACM SIGSAC Conf. Cloud Computing Security Workshop (CCSW'20), New York, 2020, pp. 67–76.

[2] J. Yao, Y. Zheng, Y. Guo, and C. Wang, "SoK: a systematic study of attacks in efficient encrypted cloud data search," in Proc. 8th Int. Workshop Security in Blockchain and Cloud Computing (SBC '20), New York, 2020, pp.

[3] E. B. Fernandez, "A pattern for a secure cloud-based IoT architecture," in Proc. 27th Conf. Pattern Languages of Programs (PLoP '20), USA, Article 10, 2020.

[4] C. Gentry, "Fully Homomorphic Encryption Using Ideal Lattices," in Proc. 41st Annual ACM Symposium on Theory of Computing (STOC), 2009, pp. 169-178.

[5] S. Halevi and V. Shoup, "Algorithms in HElib," in Advances in Cryptology – CRYPTO 2014, Berlin, Heidelberg: Springer, 2014, pp. 554-571.

[6] Y. Zhang and X. Chen, "Homomorphic Encryption-Based Privacy-Preserving Federated Learning in IoT-Enabled Healthcare System," Journal of Medical Systems, vol. 44, no. 7, pp. 1-12, 2020.

[7] J. Yao, P. Li, and X. Wang, "Secure Federated Learning with Homomorphic Encryption in Healthcare IoT Systems," IEEE Transactions on Industrial Informatics, vol. 17, no. 8, pp. 5782-5792, 2021.

[8] E. B. Fernandez and S. Ramesh, "Federated Learning and Homomorphic Encryption for Secure Medical Data Analysis," Journal of Healthcare Informatics Research, vol. 4, no. 3, pp. 245-258, 2020.

[9] H. B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-Efficient Learning of Deep Networks from Decentralized Data," in Proc. 20th International Conference on Artificial Intelligence and Statistics (AISTATS), 2017.

[10] P. Kairouz et al., "Advances and Open Problems in Federated Learning," Foundations and Trends® in Machine Learning, vol. 14, no. 1–2, pp. 1-210, 2021.

[11] N. Rieke et al., "The Future of Digital Health with Federated Learning," NPJ Digital Medicine, vol. 3, no. 1, pp. 1-7, 2020.

[12] M. Johns et al., "Federated Learning in Medicine: Facilitating Multi-Institutional Collaborations Without Sharing Patient Data," Scientific Reports, vol. 10, no. 1, pp. 1-12, 2020.

[13] S. Ramesh, S. Dev, and C. Su, "Federated Learning for Privacy-Preserving EHR Analytics," IEEE Journal of Biomedical and Health Informatics, vol. 24, no. 8, pp. 2334-2343, 2020.

[14] D. Froelicher et al., "Truly Privacy-Preserving Federated Analytics for Precision Medicine with Multiparty Homomorphic Encryption," Nature Communications, vol. 12, no. 1, pp. 1-10, 2021.

[15] M. Johns and J. Yao, "Privacy-Preserving Federated Learning Using Homomorphic Encryption," IEEE Transactions on Medical Imaging, vol. 40, no. 5, pp. 1179-1189, 2021.

[16] J. Yao, P. Li, and X. Wang, "Secure Federated Learning with Homomorphic Encryption in Healthcare IoT Systems," IEEE Transactions on Industrial Informatics, vol. 17, no. 8, pp. 5782-5792, 2021.

[17] G. Bian, W. Qu, and B. Shao, "Blockchain-based trusted federated learning with pre-trained models for COVID-19 detection," Electronics, vol. 12, no. 9, p. 2068, 2023.

[18] W. Jin, Y. Yao, S. Han, C. Joe-Wong, S. Ravi, S. Avestimehr, and C. He, "FedML-HE: An efficient homomorphic-encryption-based privacy-preserving federated learning system," arXiv preprint, arXiv:2303.10837, 2023.

[19] J. Zhou, R. Jiang, G. Chen, and L. Wang, "Homomorphic encryption based full flow privacy protection scheme for federated learning," in 2024 7th International Conference on Computer Information Science and Application Technology (CISAT), pp. 1121-1126, IEEE, July 2024.

[20] Y. Guo, L. Li, Z. Zheng, H. Yun, R. Zhang, X. Chang, and Z. Gao, "Efficient and privacy-preserving federated learning based on full homomorphic encryption," arXiv preprint, arXiv:2403.11519, 2024.

[21] A. Chaddad, Y. Wu, and C. Desrosiers, "Federated Learning for Healthcare Applications," IEEE Internet of Things Journal, vol. 11, no. 5, pp. 7339–7358, Mar. 2024.

[22] P. Tschandl, C. Rosendahl, and H. Kittler, "The HAM10000 Dataset, a Large Collection of Multi-Source Dermatoscopic Images of Common Pigmented Skin Lesions," Scientific Data, vol. 5, p. 180161, 2018.

[23] Y. Sun, Z. Wang, and X. Zhu, "Domain-specific image classification based on improved residual networks," Proc. SPIE 12519, Twelfth International Conference on Graphics and Image Processing (ICGIP 2022), 2022.

[24] W. Jin, Y. Yao, S. Han, C. Joe-Wong, S. Ravi, S. Avestimehr, and C. He, "FedML-HE: An efficient homomorphic-encryption-based privacy-preserving federated learning system," arXiv preprint arXiv:2303.10837, 2023.

[25] Korkmaz, A. and Rao, P., "A Selective Homomorphic Encryption Approach for Faster Privacy-Preserving Federated Learning." arXiv preprint arXiv:2501.12911, 2025.

Assoc. Prof. Dr. Sahar Ebadinezhad received her M.S. degree in computer engineering from the Eastern Mediterranean University (EMU), North Cyprus, in 2014, and her PhD.D. degree in computer engineering from the Cyprus International University (CIU), North Cyprus.

She joined the Department of Computer Information Systems at Near East University in 2017, where she serves as a full-time Lecturer. She is also an active Computer Information Systems Research and Technology Center (CISRTC) member at Near East University. Her research encompasses wireless communication systems, millimeter wave communication, vehicular communication (V2X), body-centric communications, wearable communication, artificial intelligence, IoT, and cloud computing. She has authored numerous publications in these fields.

Dr. Ebadinezhad is a distinguished scholar in computer engineering. She is a member of several professional societies and has significantly contributed to advancing wireless communication technologies, fostering innovation and knowledge in her areas of expertise.