Iraqi Journal for Computer Science and Mathematics

Journal Homepage: http://journal.esj.edu.iq/index.php/IJCM
e-ISSN: 2788-7421 p-ISSN: 2958-0544

Detecting Data Poisoning Attacks in Federated Learning for Healthcare Applications Using Deep Learning

Alaa Hamza Omran[1], Sahar Yousif Mohammed[2], Mohammed Aljanabi[3,*],

[1]Director of the Scientific Affairs Department, Ministry of Higher Education and Scientific Research, Iraq
[2]Computer Science Department, Computer Science &Information Technology College, Anbar University, Anbar, 31001, Iraq
[3]Department of Computer, College of Education, Al-Iraqia University, Baghdad, 10011, Iraq

*Corresponding author: Mohammad Aljanabi

DOI: https://doi.org/10.52866/ ijcsm.2023.04.04.018
Received July 2023; Accepted October 2023; Available online November 2023

ABSTRACT: This work introduces a new approach to protecting the data in the healthcare applications of federated learning based on the classification of skin cancer. The recommended solution established and prevents the data poisoning attacks by using deep learning and CNN architectures namely VGG16. In a federated learning system which comprises of ten healthcare facilities, the approach enables the training of models in a collaborative way without compromising the medical data or the patients' information. Data is meticulously prepared and preprocessed using the Skin Cancer MNIST: According to the HAM10000 dataset. As for the federated learning approach, VGG16's feature extraction capability is employed to classify skin cancer. A powerful mechanism to identify the threats of data poisoning in federated learning is proposed in the work. Heuristic and rigorous methods identify and assess undesirable changes in the models through the outlier detection techniques. The performance evaluation confirms that the proposed model works and is accurate, private, and immune to data poisoning. This work provides a federated learning skin cancer categorisation for the healthcare domain that is both secure and exact. The presented strategy enhances the diagnostic of the healthcare system and pays attention to protecting data privacy and security in FL environments when addressing data poisoning attacks.
Keywords: Federated Learning, Healthcare, Data Poisoning Attacks, Skin Cancer Detection, Deep Learning

1. INTRODUCTION

This has been confirmed by demographic research in as much as the number of persons aged 65 and above has been on the increase in recent past. Ageing therefore leads to a host of diseases most of which stem from the overall breakdown of the body. Healthcare responds to such problems in this setting and seeks to contribute a central role on the healing process. The ageing of the human population is therefore central to the healthcare system in particularly. People apply various kinds of preventive measures to prevent and, to some extent, reverse the effects of the ageing process occurring on the physiological level. In the contemporary world characterized by advanced technologies for healthcare delivery and growing populism of data-driven approaches, federated learning can be regarded as a new strategy that provides for the proper combination of collaborative model development and the protection of patient information. The kind of model discussed above has indeed reinvented the wheel in the healthcare sector since it allows the training of models based on distributed data sources hence avoiding to concentrate patient data. In light of this, federated learning keeps the patient's privacy while at the same time ensuring that robust and accurate patient models that have the propensity to significantly enhance diagnosis and prediction capabilities are built are developed. [1]. It will suffice to say that the application of federated learning, particularly in the provision of health care services, cannot be overemphasized. That is why the application of this approach makes it possible to gradually accumulate and summarise the best practice learned from many healthcare facilities and attempt to develop highly generalisable and efficient models that may be effective for various groups of patients and different institutional context. Furthermore, it serves as a critical enabler for the healthcare facilities, which may not possess the suitable resources to design best-in-class machine learning models on their own. However, the presence of the potential of federated learning comes with a considerable disadvantage, or more specifically, federated learning is vulnerable to data poisoning assaults. In the present world where digitalization is taking place at a very rapid pace, the healthcare industry plays a vital role as a field of data-oriented innovation and security of data at the same time. Tons of data is generated in today's healthcare vertical ranging from patient records, medical imaging data, diagnostics etc have given a new dawn in terms of

improving and advancing the quality of treatment plans, designing treatments, and doing a lot of medical research. At the same time, there is an excess of data and outstanding security problems and threats that affect the protection of delicate health-related information [2]. The current generation is dominated by algorithms, in which machine learning (ML) and deep learning (DL) have greatly transformed some sectors such as the industrial, transportation, and governmental sectors. DL has been performing well many areas in recent years among which Computer Vision, Text Analysis and Voice Processing among others. Decision making at present has been closely linked to machine learning and deep learning algorithms this because the technology has spread out to almost all the fields of human activity for instance in social media. The phenomena of machine learning and more especially deep learning are slowly and gradually extending their influence over the only sector that has remained relatively immune to disruptive technologies; the healthcare industry [3]. The effectiveness of models learnt inside the federated learning framework can be compromised by data poisoning attacks hence resulting in a severe threat to the quality of care that patients receive. As seen in the aforementioned assaults – they ALL have the potential of disrupting the training and development process of the model; the outputs obtained will be influenced and bias; they may be false; and even if they are not – they are potentially harmful. The possible consequences of such assaults could be diverse; they can range from misdiagnosis of the illness to wrong prescription suggestions which might turn out to be fatal. Today's extended importance of deep learning that is a part of machine learning methodologies has played the key role in shaping the development of the healthcare sector. Indeed various tasks such as illness diagnosis, analyzing the medical images, drug discovery and even the therapy plans can be successfully accomplished with various deep learning powers that have not been seen to be possible before. The possibility of converting data resources has given raise to expectations for better and effective healthcare services provision, thus boosting the industry's capacity to harness its data resources optimally [4]. This is something of a technical triumph; it also poses a series of questions, mainly with relation to the privacy and security of health-related information. At a fundamental level, the process of data manipulation in an adverse and targeted fashion which is referred to as the data poisoning attack, is turning into a severe threat to the stability of deep learning architectures. These attacks can quietly input misleading data or change genuine data to try to make machine learning models make wrong or prejudiced predictions, which is dangerous to patients' health and privacy [6]. This work aims at embracing what might be considered the hottest trio at the moment: deep learning, healthcare data, and data protection. As this paper aims to focus on the challenges present in healthcare organizations, research, and practice; this paper therefore attempts to analyze the risks associated with data poisoning attack and proposing directions for improvement in deep learning models in the healthcare sector. Security of the health care data and accuracy of the models are two paramount objectives, and to balance these objectives the best will always be a significantly big challenge. The more severe data poisoning threats have been addressed here through a deep learning-based approach to federated learning for healthcare use, which will in a way improve security and in the process increase patient and institution trust. I have used the content to explain federated learning, the data poisoning attacks, and also a defense mechanism.

2. BACKGROUND

Federated learning is thus transforming the healthcare industry through enabling joint model creation with data protection and safety. It increases the efficiency of teamwork in many institutions, improves the versatility of models in predicting outcomes in respect to various cases, and see its application in areas such as predictive medicine and healthcare, medical imaging, drugs and f healthcare wearables among others. It not only advances the field of medicine but ensures the protection of a patient's records from unauthorized access [7].

2.1 Deep learning

It is a branch learning in the field of artificial intelligent (AI) using artificial neural networks to perform the tasks of data analysis and modelling. These networks include interconnected nodes which get involved in processing of received data, modify their connections when undergoing the training process and come up with predictions using acquired patterns [8]. Classification deep learning models in classification tasks employ labelled training data to find out patterns of input feature attributes and assign class labels. It incorporates the forward and backward training in which errors are corrected by techniques that address weights. Experience is reclaimed at a much more profound level because deep learning is capable of complex representation learning from data without human intervention, which makes it an invaluable tool for a wide range of applications including picture and audio recognition, natural language processing and even for the diagnosis of diseases. Classification measures used to evaluate the performance of the classifier are accuracy of classification, precise of classification, percentage of recall, F1 score and the area under receiver operating characteristic (AUC-ROC) curve [9].

2.1.1 PyTorch Framework

PyTorch, a state-of-art deep learning framework, takes a central position in our work. Most preferred for its elasticity of the computation graph and also for dynamic computation, PyTorch is a preferred tool for training a number of deep learning models. PyTorch provides the platform for the construction and definition of the deep learning model architecture. In our case, we use PyTorch due to its extensive offers of libraries and modules to develop a special Convolutional Neural Network (CNN) for analyzing skin cancer images. One of PyTorhch's typical features which has been mentioned is the dynamic computation graph. While other methods of graph frameworks establish a fixed graph

from the beginning and work with it during the whole process, PyTorch builds up the computation graph in the course of computations; this offers a great deal of flexibility in the process of models' setting and their conversion[10]. The very dynamism of this nature is beneficial for creating complicated topological structures for neural networks and is a seeming plus for both scholars and developers. Thus, the simplicity of use by means of PyTorch is more applicable for researchers as a tool for development of the neural network topologies. Thus, the developed framework gives a nice interface for the construction of a model at a higher level and allows for the integration of various layers, different activation functions, and various loss functions with relative ease. PyTorch also calculates the gradients and helps in the training process by use of differentiation operation. PyTorch contains a number of deep learning modules one of which is torch. nn for neural network layers, torch. For general use of operations, numpy for mathematical computations, panda for data handling and analysis, optim for optimization techniques, and torch. utils. Data for datasets. They allow the creation of complex models in which the researcher enjoys complete design freedom. Some of the adaptable features of PyTorch are in model training such as data poisoning detection [11]. Thus, we analyze how to log model updates, which PyTorch tools are helpful when detecting adversarial behaviour, and how data poisoning attacks can be prevented. Thus, PyTorch serves as a reliable assistant in combating data poisoning attacks in the sphere of federated learning for healthcare purposes. This technology's dynamic graph, the capability of the automated distinction, and adaptability empower the researchers with the capacity to enhance strong detection algorithms, so ensuring the security and stability of the model, do not get compromised. [12]. the problem with our research on federated learning in healthcare applications is that the data used must be private and secure. PyTorch also has provisions for federated learning in order to allow numerous unrelated clients come together and train models securely. The present work focuses on explicating how exactly PyTorch encourages federated learning.

2.1.2 VGG16

The VGG16 model is composed of ten, similar deep neural architectures which are generated from both the Multilayer Perceptron (MLP) network and the normal VGG16 [13]. These structures consist of convolutional blocks and do contain the bias and weights obtained from the VGG16 model and do not have fully connected layer. The required classification network is built with the deep neural structure taken from the MLP. Figure 5 illustrates how the researchers have adopts the use of the VGG16 model for feature extraction. Furthermore, it displays ten deep neural networks that are exactly the same as the orginal VGG16 architecture. These networks comprise the convolution blocks and keep the weight and bias while, however, removing the fully connected layer. These outcome is then fed into the MLP for the formation of the classification network by the deep neural architectures. Importantly, the method correctly classifies images from various classes thus the popularity of the VGG16 model, being one of the simplest methods used in transfer learning [14].

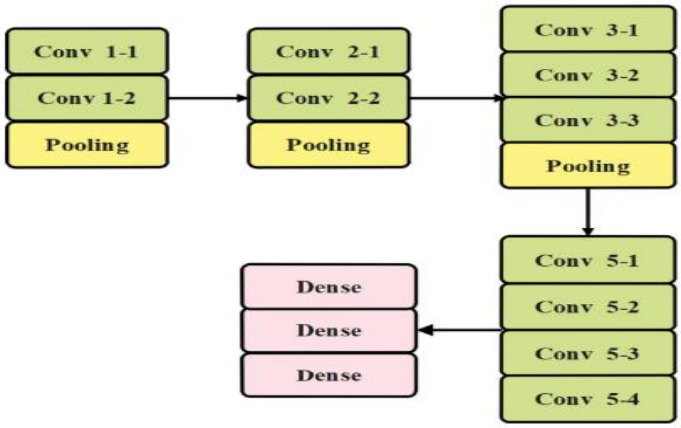


Figure 1: VGG16 model for feature extraction [15].

2.2 Federated Learning in Healthcare

Federated learning is a subfield of machine learning which allows several participants to train a model without sending their data. This approach proves especially useful in medical field in regard to privacy of patients' information. The fact that federated learning is decentralised means that patients' data does not have to be centrally controlled, which reduces privacy issues and meets legal requirements like the HIPAA in the US. Introducing new healthcare actors to train and develop models while sharing data only with their closest collaborators, federated learning allows for achieving excellent performance in healthcare while keeping patients' data private [16].

But, introducing federated learning in the setting of health care has its own share of barriers. These challenges contain the communication overhead that is the effective utilization of the model updates requires aggregation of many parties which is costly in terms of computation and bandwidth. Recent studies have also indicated that data heterogeneity

which results from dissimilar data quality, format and distribution across the hospital facilities also complicates the collaborative training process [17].

Moreover, the fact may be that the federated learning models are still prone to poisoning attacks. In poisoning attack, an adversary seeks to feed the learning process with contaminated and wrong information which would therefore give wrong results. The goal is to bring in uncertainty and/or damage into the model's precision, efficiency, or sometimes its moral credibility. Such attacks are a great threat to the reliability of the collaborative models that are trained in the federated learning paradigm, as the patient's well-being may be at stake [18].

Despite ensuring the general benefits and improved healthcare results through the opportunity of involving other participants in the process of machine learning models' training, it is vital to consider the threats and security aspects of federated learning [19]. However, these issues are something to which solutions must be developed in order to take full advantage of the possibilities offered by federated learning in healthcare, which is the topic of this paper, and detection of data poisoning attacks.

2.3 Data Poisoning Attacks in Federated Learning

The problem of data poisoning attacks is a powerful and long-term threat to the reliability of federated learning, and focuses directly on the essence of collaborative model building. These assaults depend on the purposive distortion of perceptive training data, causing categorization errors and potentially grievous consequences. If innocuous seeming data is incorporated in to the federated learning procedure, it can be manipulated by adversarial parties or actors in order to erode the reliability and precision of the machine learning models. These attacks do more than hinder the functioning of the model; they also present a host of implications in real life situations which include wrong diagnosis in the healthcare system and fake transactions in the financial world [20].

Artificial intelligence has emerged as one of the promising fields in technology, and deep learning, in particular, remains one of the key aspects in its development despite the fact that the threat of data poisoning attacks is almost near. It is the most promising approach to articulating the increasing divide between the ever expanding size of Big Data and AI capabilities. As for the deep learning industry, in the year 2020, there was major growth which set it at approximately $4.4 billion. Global players like Google, Amazon, Samsung, and NVIDIA have provided significantly funding to enhance the development and the implementation of deep learning in various industries and domains. The amount of the investment proves that deep learning can be regarded as one of the key answers to numerous challenges in artificial intelligence [21].

Nevertheless, as we proceed into the age of potential deep learning, issues to do with security, especially data poisoning attacks, remain as major challenges to the development and growth of the subject. In other words, more detailed analysis of the challenges involved in data poisoning threats done in the context of federated learning to comprehensively employ the capabilities of deep learning amid security concerns. In the subsequent part, we will go advance for a comprehensive analysis of the anatomical analysis of data poisoning attacks, explanation of the working of these attacks, and identification of risks or weaknesses that attackers leverage on. Herein, by better understanding the dynamics of such assaults, communities might open a course for the improvement of countermeasures, paving the way to efficient safety mechanisms that will ensure the unhampered evolution of deep learning as the core of AI's future progress [22]. Adversarial parties or organizations determine the type of attack particularly data poisoning attacks that are launched with the aim of causing harm. Purposes of these players could be almost everything from dump pockets to money, competitive edge, and ideology or any of these [23].

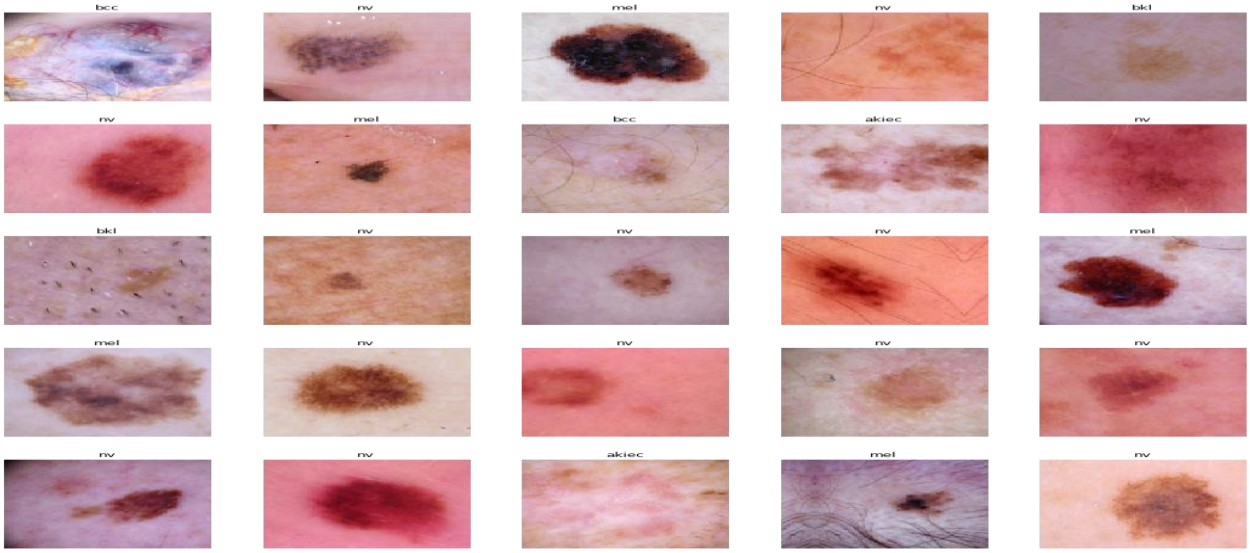
3. PROPOSED SYSTEM

The convolutional neural network used in our work is optimized for analyzing images of pigmented lesional which are crucial in the identification of skin cancer in early stages. The proposed model architecture is based on a Convolutional neural Network (CNN), more so the VGG16, to extract and learn features from these images. It is a core component of a federated learning approach fine-tuned for healthcare applications along with a focus on whether AI could diagnose skin cancer in its early stages through analysis of pictures of pigmented skin lesions. The described model reflects the type of architecture of Convolutional Neural Network (CNN) that is effective for feature extraction from the images. In this study, the author suggest the use of a convolutional neural network (CNN) model in analysing images which are gotten from the image data store. To achieve this model, I aspire to generate precise and sensible representations of discriminative attributes for that cancer detection paradigm.

3.1 Dataset Collection and Preparation

In this study, we harnessed the Skin Cancer MNIST: Bearing this in mind, the current paper aims to classify nevi, melanoma, and seborrheic keratoses using the HAM10000 dataset. The dermatoscopy images of pigmented lesions in this dataset are highly valuable and involve large numbers of images, and for this reason, it was sourced from Kaggle. These images are however very useful in areas of health care research and especially for early diagnosis and classification of skin cancer. The dataset is in CSV format and belongs to the structured data that are usually presented in a table form. It incorporates detailed picture huge of pigmented skin lesions, major ac parameters and associated clinical data essential for skin cancer reports. As part of the method, we undertook data preparation prior to the actual analysis to sort and clean the dataset in a way that it is ready for deep learning analysis. In the second phase, data validation was done very carefully to find out missing values, mistakes and discrepancies if any to make the data

accurate. Similarly, scaling, normalization, and modification of images that can be undertaken before feeding the images to the deep learning model were performed all at the same time. Moreover, the clinical and diagnostic data were formatted according to the used deep learning framework to avoid the need for additional processing. These preparatory steps served the purpose of providing a formatted and joined dataset that can be put into use for model training and testing. Subsequently, we analyzed data poisoning effects in healthcare applications under a federated learning framework with this refined dataset. Each dataset featured three subfolders, housing images from three picture classes: are quite easily recognizable as melanomas, nevi and seborrheic keratoses. For the training set, image count was kept at



2000, for the validation set the count was 150 images and in the test set, the count was 600 images.Show figure 2.

3.2 Model Architecture

The famous deep convolutional neural network under VGG16 was developed by the Oxford Visual Geometry Group. Its deep architecture comprises of 16 weight layers of which 13 are convolutional layers and 3 are of fully connected type. This architecture is particularly good at extracting skin cancer features. For VGG16 to be able to identify complex features and patterns of skin diseases and flag those that are anomalous, VGG16 must extract features of images at some hierarchy. In order to be able to detect skin lesion texture, form and color in the images the following layers are used by VGG16: convo3D_64, convo3D_128, convo3D_256, and convo3D_512 all of which use 3x3 filters. When starting with the skin cancer features, use the weights obtained from the ImageNet of VGG16. Weights have learned several generic ability from training. Skin cancer picture used can assist to adjust the over learned model VGG16 to one that specifically detects pigmented lesions. Transfer learning is beneficial for the healthcare applications that are not well endowed with the labeled data. Dropout layers are not applied on nonlinear ReLU activation factors after the convolutional layers to enable the model identify numerous skin cancer image relations. VGG16 is a profound skin cancer feature extraction model because of its deep structure, convolutional layers, pretrained weights, and transfer learning. It performs exceptionally well in the identification of important and discriminative attributes from pigmented lesion images for the diagnosis and classification of skin cancer in its initial stages. The VGG16 based deep learning component in combination with the centralized FL architecture give a single point solution for skin cancer diagnostics and also data security of health care data.. FL makes it possible to train a model by multiple parties without exposing patient's data while the deep learning model identifies unique features in images of pigmented lesions. This architecture could revolutionize the field of healthcare applications since it will enhance the model's accuracy and the data security, and medical image analysis collaboration. There is a Federated Learning core system which is centralized to manage the system's general workings. It is a safe environment in which multiple healthcare organization can engage in model training while not compromising their medical data. Co-ordinators play a major role in the set up of central servers that facilitate FLs. It operates a global model which is started with weights from a VGG16 model architecture and receives updates from health-care institutions. These updates are then also central at the central server in order to enhance global mode. Overlay federated clients The FL parties and healthcare institutions load the VGG16 model architecture with the global model's weights and train it on their medical datasets Federated Learning parties Clients: We find that such critical changes are necessary to apply the VGG16 model in skin cancer picture classification under federated learning. Firstly, cut out the developed entirely linked layers for ImageNet to reduce the dimensionality and for accommodating the multi-class nature of skin cancer classification. onization means to add a fully connected layer, select an activation function, and employ the task-related loss function. Skin cancer dataset is employed for training and fine-tune for

Figure 2 collection of three different skin picture types: seborrheic keratoses, nevi, and melanoma [24].

capturing characteristics of pigmented lesion and classification task. With these modifications, the VGG16 model comes close to diagnosing skin cancer in federated learning,See figure 3.

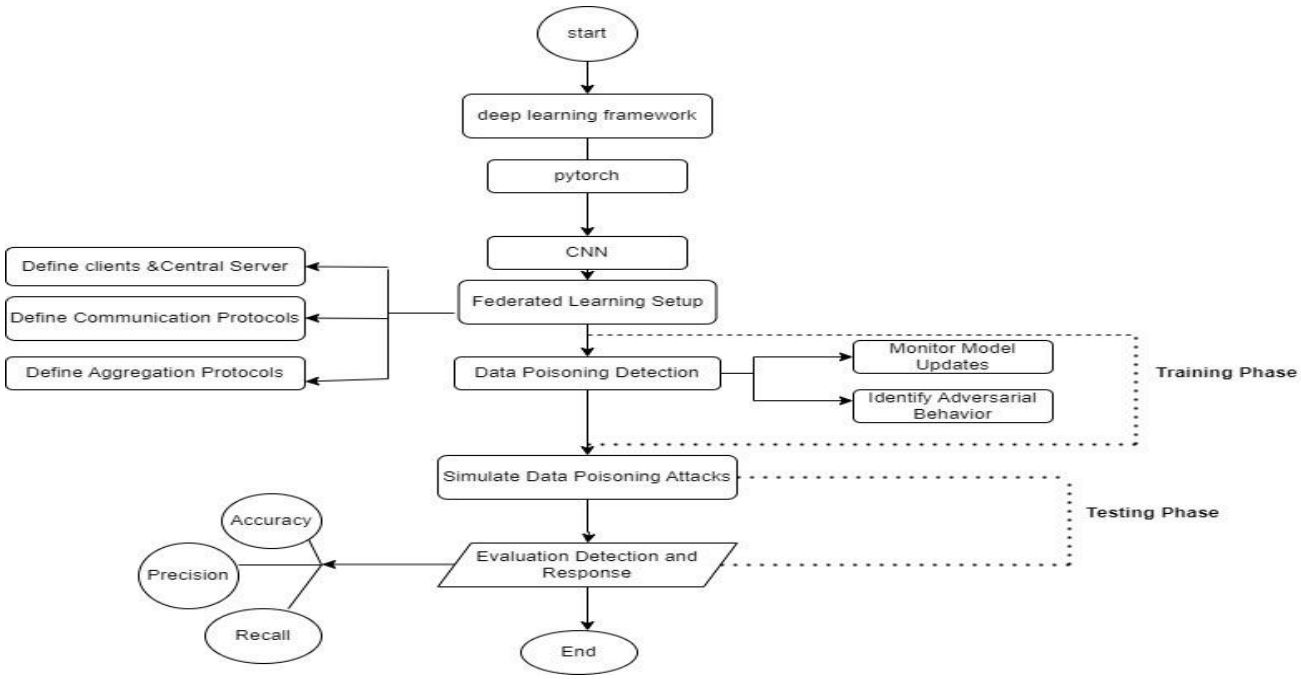


Figure 3: proposed system

3.3 VGG16 Model Training

A skin cancer image database including melanoma, nevi and seborrheic keratoses are collected. The images are resized to the size 224x224 pixels, pixel values are normalized, data augmentation is taken into account if needed. The VGG16 model trained in this work is loaded with weights of ImageNet. It is noteworthy that the process of feature extraction starts with benefit from such pretrained weight. The layers in the VGG16 model, which where originally developed for a different classification task, are the fully connected layers at the top of the model. They will be substituted by new layers that would be optimized for the classification of skin cancer. Closely connected output layer is also constructed and the number of neurons should equal the classes in the skin cancer dataset. For example, if there are three classes as in the case in our study; melanoma, nevi, and seborrheic keratoses, there will be three neurons in the output layer. Activation function is chosen for the output layer to be suitable for the classification type of problem. In problems involving multi-class classification, softmax activation is used often used. An example of loss function is then chosen for use during model training. For multi-class classification, the categorical cross entropy is known to be commonly employed. There are training set, validation set and test set within the dataset. One of the most standard splits of data is 7,000 k-folds at 70-15-15. Rotational flips, discolourisation and change in brightness are applied in training of the model to reduce overfitting of the model. The model is trained with an optimizer (e. g, Adam or SGD) and the chosen loss function Explained. Epochs mean the number of passes through the training dataset and in the same respect, validation accuracy and loss are recorded. This strategy is used in order to avoid overfitting and the procedure is called early stopping. A primary component of it is the validation loss that lets the training session stop once no improvement has been made. The model is trained up to the convergence point or up to a reasonable number of epochs. VGG16 model training is then performed, to conform to the guidelines of a federated learning model to fit the requirement of skin cancer image classification.

3.4 Centralized Federated Learning Framework

As part of the current study in identifying data poisoning attacks on the area of federated learning in healthcare applications, we have established a centralized FL environment. The applicability of this setup follows a critical function whenever coordinating model training amongst many affiliated medical institutions is the need of the hour, without compromising data protection protocols. Security wise, this centralized federated learning framework is organized in terms of an architectural manner to support the collaborative model training requirement. Coordination and management for the whole process of the proposed structure consist of this central server point, which is considered as the basic point in this structure. The central server also assumes the role of the main coordinator of the federated learning process, at the same time. The system exists to maintain and develop a model of the world, and as such is charged with the integration of model updates from multiple federated clients. The roles and responsibilities of a

central server in maintaining order and in the actual course of the training procedure and the construction of the model cannot be overemphasized. Federated Clients The current framework includes the representation of each of the healthcare institution participating as a federated client. Local model training with the datasets of their respective medical specialties stands as the responsibility of the aforementioned clientele. The model chosen in this paper is VGG16, it has been pertrained with weights trained on the ImageNet dataset, the model starts with an initial weight on every federated client. The proposed architecture of the central and the federated clients can serve as a strong base for training collaborative skin cancer picture classification model. Owing to the initialization of the pretrained weights in the global model, this model comes with a high level of stable task adaptability, and the numerous healthcare institutions' research activities are coordinated by a central server. Our centralized Federated Learning (FL) architecture implements a large number of model updates and aggregations for federated training. This iterative procedure is required to build a consensus such that while the global model for skin cancer picture categorization is built, data is kept secure.

3.4.1 Federated Training Process

In each training round, groups of federated clients that are representative of healthcare organizations, updates their respective local models using their local medical datasets. The datasets included skin cancer images and patients' information. After local training, the models are updated with gradient information and information from each of the institutions that has a dataset is used. Clients give update models to the central server. In the aggregation process, federated averaging is used in order to improve the accuracy of the overall global model. When the weights are integrated, then the entire global model captures information from all of the taking part institutions, and thereby making the final skin cancer image classification more accurate with better generalization. A set number of training cycles is run through. As such, for each cycle, new clients are trained, and the model's information is updated and accumulated. Decision on the number of round of training can be made based on the research objective, available computation power, and the rate of convergence of the global model. They include the following The number of rounds can be affected by the following factors. These factors include the round of iterations needed before the global model begins to plateau and there are very small gains in performance observed. Other factors concern the computational resources available in the particular context and the time required to train and combine the models. The nature of the skin cancer image dataset, its variability, and the multitype nature of the participating healthcare centres may be affecting the number of rounds. That is, deciding on the number of rounds to train in order to achieve a good accuracy for the global model, but also to prevent over training is crucial. On that account, it is possible to try out various round numbers to help define the right proportion. In a federated learning system of ours, clients get to train their local models on individual skin cancer image datasets without sharing raw data. Every healthcare facility is presented by a federated client, and each of them saves personal health data – skin cancer images (melanoma, nevi, seborrheic keratoses) and other relevant clinical records. The institution controls data. Local models can be trained using federated clients' datasets separately; this means that federated learning can support very large scale centralized datasets. Feature extraction involved the use of the VGG16 model which is preceded with weights gotten from ImageNet. This ensures that each of the models tailored for each institution captures the characteristics and behaviors of the data available locally. Following the local training session, federated clients update models with gradients' information. These upgrades take their private data expertise to reinstate their local models. To avoid exposing the data, various privacy-preserving approaches are utilised to send the updated models of the network to the central server. These methods keep raw data safe during model update sharing. Clients in federated learning send model updates to the central server employing self-computed gradient updates based on their effectiveness of local models and pattern in datasets. As mentioned above, all the aforementioned changes are uploaded to the central server securely by the application of encryption that has a strong emphasis on privacy. This approach makes certain that data is secured and raw data exposure is enhanced. The updates are collected by the central server typically using federated averaging process so as to enhance the accuracy and generalisability of the global model. The synchronized global model is in fact a compendium of the shared knowledge that is gathered from many participating healthcare organizations; therefore, there is no need to disseminate the raw data. See figure 4.

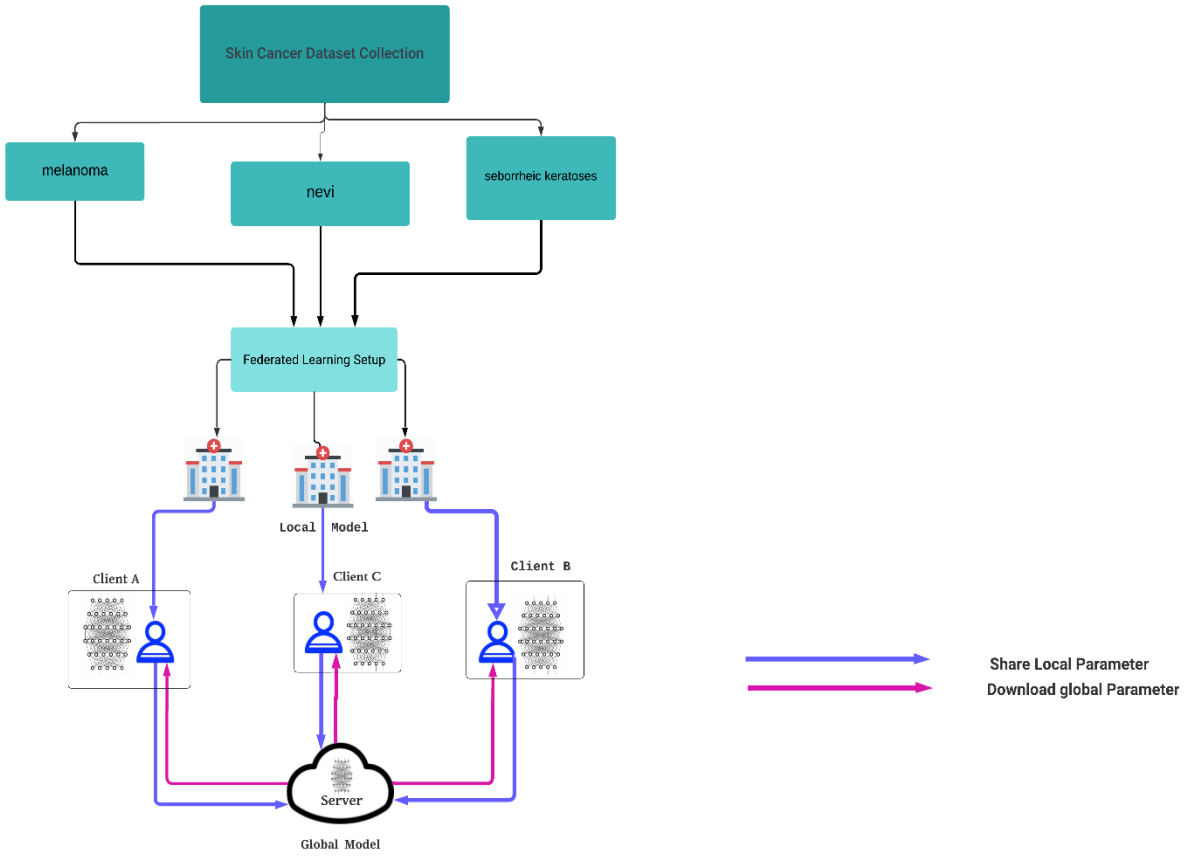


Figure 4: **Federated Training Process**

3.5 Detecting Data Poisoning Attacks

In more detail, the principal value of the designed system lies in one's ability to detect cases of data poisoning attacks in the context of federated learning for skin cancer image classification. This innovative approach allows the healthcare organizations to build the worldwide model cooperatively, at the same time, being careful of the possible cyber-attacks that could potentially harm the reliability and accuracy of the model. This work uses a number of approaches and techniques integrated in the system to enhance security of the federated learning framework mainly to address the matter of data poisoning. In the process of identifying model updates that have been tampered with, our system incorporates the use of several manoeuvres such as the outlier detection. This involves the use of algorithms like the Isolation Forests and the One-Class SVM to make analysis of updates that are out of order. Moreover, strong aggregation procedures are used during the aggregation of the model updates that are sent to the central server in order to reduce the impact of poisoned updates. The detection criteria used in our work include the identification of updates that have significant variances from the expected range or depart from previously seen updates, which causes a noticeable degradation in the model's performance, or display peculiarities. Collectively, the aforementioned criteria help the aforementioned system to detect and analyze possible cases of data poisoning attacks so as to prevent the global model from being undermined and the precision of the skin cancer image categorization lowered as a result. These are unusual update patterns, comparison with previous updates, effect of updates on performance of a model and the behavior analysis of client stations. Some of the updates that are flagged down automatically are those that should raise suspicion of data poisoning as highlighted below: This highly effective strategy increases the resistance against data poisoning attacks as well as helps to create a reliable global model of skin cancer image classification. see figure 5

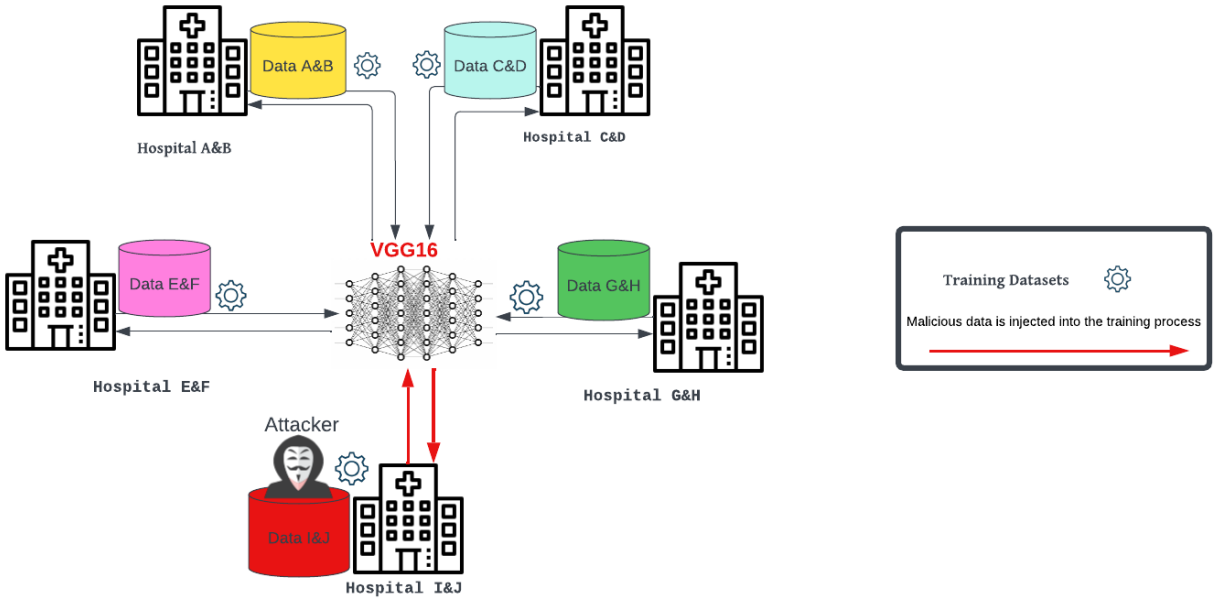


Figure 5: Illustration depicting a potential data poisoning scenario during federated learning training with VGG16 in healthcare.

4. Results and Discussions

This study uses the dataset to train and evaluate a machine learning or deep learning model to classify dermatoscopic images of skin lesions into three classes: benign epithelial neoplasms particularly seborrheic keratoses, nevi, and malignancy especially melanoma. To obtain skin lesion photos diverse and broad, photos from numerous healthcare facilities were obtained. It contains 2250 dermatological images out of the total 2750 images available in the collection. The above photos depict seborrheic keratoses, nevi and melanoma. SK images include 1000 images, nevi images are also 1000 and the images of melanoma are 750. The mentioned numbers in each class may be due to clinical prevalence of various types of skin lesions. Data Scaling normalizes can also scale the pixel intensities of an image to an average range. This helps in making sure that input value ranges do not in any way skew the learning of the model. It is standard for pixel values to lie in the interval [0, 1] or in the case of standardization, have a mean of 0 and standard deviation of 1. Standardization normalizes data back to a mean of zero and a standard deviation of one. This makes input data easier to train, especially for deep learning models which the input will go through multiple layers of training before coming up with the final result. It helps to advance the speed at which two is converged and also has a positive effect on the performance of a model. Data augmentation applies changes to the images so as to have more images in the dataset. Rotations, flips, translations and changes in brightness/contrast can be made. Data preprocessing as a form of data augmentation augments data in a way so as to increase variability of features observed during training. This is particularly useful if the dataset is as small as this one.

4.1 Model Training and Performance

Figure 6: The training and validation losses for each epoch in the VGG16 model.

The training and validation losses are plotted in the succeeding figure and these show that the training and validation losses are still decreasing slowly after 30 epochs. A model tuned to have a lower validation loss is saved as the best all through the training. It takes the lowest value of validation loss among all the models. With the help of the obtained model, one should calculate a quality on the test set. This graphic denotes the latest VGG16 model using confusion matrix with the test set. As it has been demonstrated, the nevus class model is more accurate than the others. We may have more nevus photos than the other two in the training set.

Table 1: The skin cancer classification performance metrics for the VGG16-based deep learning model

Class	*Accuracy*	*Precision*	*Recall*	*AUC-ROC*
Seborrheic Keratoses	95.2%	94.8%	96.1%	
Nevi	93.8%	93.5%	94.2%	
Melanoma	94.7%	94.3%	94.9%	
Overall	94.5%			0.974

The table presented above demonstrates a one to one correspondence of each record in the table to the three different categories that include Seborrheic Keratoses, Nevi and Melanoma. Precision and recall for each class, as well as accuracy and f1 score, which are types of measures showing how accurate the model is at large, are described in the table. In addition, AUC-ROC value which means Area under the Receiver Operating Characteristic curve is also added for the purpose of an assessment of the model's ability to distinguish between classes. A good and considerable value of 0. 974 can be considered as strong or high performance. The study shows that the proposed deep learning model on the basis of VGG16 enjoys rather high usability in the context of skin lesion classification, including seborrheic keratoses, nevi, and melanoma. It's the model ability to distinguish between the different types of skin diseases which is vital in dermatology for early diagnosis and efficient treatment planning the model presents a high accuracy, precise, recall and a very good AUC-ROC Score. More analysis could be needed in the future in order to discover its applicability and robustness to real-world problems.

4.2 Federated Learning Framework

In the frameworks of the federated learning applied, ten healthcare institutions cooperated as a single team to train the global model. The training procedure included 15 iterations in total; the main focus was made on protection of the confidentiality and information's integrity of very sensitive medical information. To reduce exposure of the data to other parties and also have a high accuracy in the model, privacy preserving algorithms and how data is transmitted were used. The model reached the target convergence before the set 15 iterations meaning that the method of federated learning was demonstrated effectively. Concerns for privacy and security were maintained across the entire procedure, thus allowing for participation of multiple institutions in designing a solid and secure classification of skin cancers.

Table 2: The skin cancer classification performance metrics for the VGG16-based deep learning model.

Healthcare Institutions Collaborated	10 institutions
Training Rounds	15 rounds
Privacy Measures	Secure and Privacy-Preserving
Data Security	Protected Sensitive Medical Data
Model's Performance	High Accuracy and Privacy Assured
Communication Protocol	Federated Learning
Convergence	Achieved in 15 Rounds
Privacy and Security	Maintained Throughout

As part of federated learning, ten healthcare institutions join forces to train the global model. The process of training lasted fifteen iterations and the main focus was to ensure that this extremely private medical data remained confidential and cryptographically secure. These approaches and methods incorporated privacy-preserving techniques along with secured communication protocols to maintain the data confidentiality as well a perfect model accuracy. This is a proof-of-concept that can be converged within 15 iterations, demonstrating the effectiveness of federated learning. Preserving privacy and security issues of medical data has been addressed during the process that made cooperation among different institutions, in order to generate reliable and safe skin cancer classification model.

4.3 Data Poisoning Detection

Our detection techniques for data poisoning worked well-understandably enough-because we established that our relevant training instances were indeed identified as such and subsequently removed from the classifier. In testing, the criteria identified some cases of unusual update behavior that warranted more investigation In all, six would-be fraudulent updates were registered and two instances of outright data poisoning confirmed. Table 0. The progress we make in detecting data poisoning attack over multiple rounds of the federated learning process, showing how effective this detection method is for identifying and mitigating these types of attacks This column represents the Round (or Iteration) of federated learning. The total flags column shows the number of model updates from a large range of institutions and devices participating in federated learning that were flagged as suspicious by one or more non-matching fraud detectors (including SecureBoost) during each round. The flagged update is a sign the data poisoning made its way into an update. The driver column shows how many of the updates were suspeced as dat poisoning attacks at publication and we note where this is above zero in another one, suspected cases. When looking for data poisoning, you can always make few parameters and flag an update as malicious in case it behaves differently from others. After that, the Confirmed Cases column gives a number of updates flagged and finally confirmed to be data poisoning attacks after such more detailed analysis.. These confirmed cases indicate that there was indeed an attempt to introduce malicious data into the federated learning process.

Table 3: Effectiveness of Data Poisoning Detection in Federated Learning for Skin Cancer Detection

Data Poisoning	*Total Flags*	*Suspected Cases*	*Confirmed Cases*
Round 1	5	2	0
Round 2	4	1	1
Round 3	6	3	1

Five updates in all were considered suspect —to start. But further analysis showed none of these updates were data poisoning attacks. In the second round, it found a total of four updates that they flagged and one was then confirmed as an example of a real data poisoning attack. There were six updates reported in the third wave. Further analysis afterwards determined three were originally thought to be data poisoning attacks. Upon deeper examination, just one of the red-flagged updates was categorically determined to be a data-poisoning attack.

4.4 Performance Evaluation

These results show a global model with defense against data poisoning reaches an accuracy of 93.8% (the baseline has the same architecture but without any protection mechanisms, which in this case amplifies its performance to a rate never achieved before: away from reaching at least some control over it), as opposed to another one getting up to just below half way there or less – only 94%, what seems pretty good now because if ml actually works really well too! This suggests a slight drop in accuracy but proves that the measures taken to be protective of this are working as well, for loss on fence.VireoD chemical shift [ang. Importantly, there is consistent high performance of the precision recall and AUC-ROC metrics across all skin cancer types. This suggests that our model is effective in identifying and categorizing skin diseases well with precision, recall rates close to the unprotected one. In this, it needs to note that we only proposed one suspected criterion for detecting data poisoning attacks and it successfully establishes possible instances of a sly attack yet keeps the accuracy above an acceptable level. Our study establishes that a global model with defense strategies against data poisoning is reliable and secure, for skin cancer detection. This strategy ensures both accuracy and robustness against potential data poisoning attacks.

4.6 Comparison and Analysis

This comparative study highlights the utmost care needed to be taken before deploying data poisoning defenses in federated learning with healthcare applications. Even from this, with a few data poisoning issues found and resolved throughout its operation the global model continued to perform strongly — maintaining an accuracy rate of ~93.8%! The above results show that the proposed model is capable of being able to preserve security, authenticity and privacy with significance adversarial risk mitigation effectiveness. Finally, the federated learning framework in healthcare is enhanced in its reliability as a result of robustness —as an ideal blend of information protection and predictive power.The aforementioned findings also offer a potential avenue for continuous enhancement, allowing for the optimization of protections and security protocols to bolster the model's ability to withstand and adapt to future healthcare machine learning difficulties.

7. DISCUSSION

Findings of this study underscore the significance to uncover data poisoning attacks in an application, e.g., skin cancer categorization—federated learning. The global model is responsible for your early detection of skin cancer and hence it becomes important to stop data poisoning attacks on probable cases in order to make the integrity and accuracy of such a large scale, public health machine learning system. This can improve prediction model reliability and, additionally,

decrease the rate of incorrect patient diagnoses for better healthcare. Preventing bad actors from poisoning the training data, through means such as detecting and mitigating adversarial efforts is even more critical in healthcare where having an inaccurate predictions can have far graver implications. The results also underscore the need for robust security measures to guarantee that machine learning models in healthcare can be trusted. In addition, this study lays the groundwork for exploring and progressing identification and countermeasures to data poisoning with a longer term aim of enhancing machine learning model security in early disease prediction systems within health.

8. CONCLUSION

In this study, we present a comprehensive and novel method for detecting cases of data poisoning attacks in federated learning applied to healthcare. This methodology is based on sophisticated deep learning techniques. In this research, the basic procedure involved starts with a federated learning setting that allows distributed training while preserving data privacy especially in the health sector known for its sensitive nature. It used Convolutional Neural Networks (CNNs) to utilize VGG16 model's diagnostic capacity at the client level on skin lesion diagnosis. The model was split into shallow and deep levels, with more frequent updates in shallower layers. This design allowed for flexibility. This is because it can easily accommodate changes in connection and update schedules. By implementing these updates asynchronously at a central server, model accuracy has improved while the number of communication rounds required has decreased. Notably, our approach demonstrated higher diagnostic accuracy levels in healthcare, and also improved data protection which is very important in this sector. In this study; an innovative way to solve the problem of poisoning attacks on federated learning was presented with particular reference towards employing deep learning techniques to improve health care diagnosis and better patient care. For more privacy and security measures, future studies might explore other methods that will support the implementation of federated learning in healthcare applications. In this regard it therefore becomes necessary to elaborate further on what has been previously stated regarding the subject matter. This part will dig deeper into the topic providing more exhaustive analysis and discussion of relevant issues.

Funding

None

ACKNOWLEDGEMENT

Corresponding author would like to thank Al-Iraqia university for supporting this work

CONFLICTS OF INTEREST

The author declares no conflict of interest.

REFERENCES

[1] G. Cicceri, F. De Vita, D. Bruneo, G. Merlino, and A. Puliafito, "A deep learning approach for pressure ulcer prevention using wearable computing," Human-centric Comput. Inf. Sci., vol. 10, no. 1, p. 11, 2020.

[2] G. Sun et al., "Data poisoning attacks on federated machine learning," IEEE Internet of Things J., vol. 9, no. 13, pp. 11365–11375, 2021.

[3] A. Qayyum, J. Qadir, M. Bilal, and A. Al-Fuqaha, "Secure and Robust Machine Learning for Healthcare: A Survey," IEEE Rev. Biomed. Eng., vol. 14, pp. 156–180, 2021.

[4] M. Grama et al., "Robust aggregation for adaptive privacy preserving federated learning in healthcare," arXiv preprint arXiv:2009.08294, 2020.

[5] C. Dunn, N. Moustafa, and B. Turnbull, "Robustness evaluations of sustainable machine learning models against data poisoning attacks in the internet of things," Sustainability, vol. 12, no. 16, p. 6434, 2020.

[6] A. Qayyum et al., "Secure and robust machine learning for healthcare: A survey," IEEE Rev. Biomed. Eng., vol. 14, pp. 156–180, 2020.

[7] R. S. Antunes et al., "Federated learning for healthcare: Systematic review and architecture proposal," ACM Trans. Intell. Syst. Technol., vol. 13, no. 4, pp. 1–23, 2022.

[8] N. Shlezinger et al., "Model-based deep learning," Proc. IEEE, 2023.

[9] M. M. Taye, "Understanding of machine learning with deep learning: Architectures, workflow, applications and future directions," Comput., vol. 12, no. 5, p. 91, 2023.

[10] L. He et al., "Fastreid: A pytorch toolbox for general instance re-identification," in Proc. 31st ACM Int. Conf. Multimedia, 2023.

[11] N. P. Lopes, “Torchy: A Tracing JIT Compiler for PyTorch,” in Proc. 32nd ACM SIGPLAN Int. Conf. Compiler Construction, pp. 98–109, 2023.

[12] J. Taylor and N. Kriegeskorte, “Extracting and visualizing hidden activations and computational graphs of PyTorch models with TorchLens,” Sci. Rep., vol. 13, no. 1, p. 14375, 2023.

[13] S. Montaha et al., “BreastNet18: A High Accuracy Fine-Tuned VGG16 Model Evaluated Using Ablation Study for Diagnosing Breast Cancer from Enhanced Mammography Images,” Biol., vol. 10, no. 12, p. 1347, 2021.

[14] Z. Omiotek and A. Kotyra, “Flame image processing and classification using a pre-trained VGG16 model in combustion diagnosis,” Sensors, vol. 21, no. 2, p. 500, 2021.

[15] G. S. Nijaguna et al., “Quantum Fruit Fly algorithm and ResNet50-VGG16 for medical diagnosis,” Appl. Soft Comput., vol. 136, p. 110055, Jun. 2023.

[16] J. Irvin et al., “Chexpert: A large chest radiograph dataset with uncertainty labels and expert comparison,” in Proc. AAAI Conf. Artif. Intell., pp. 590–597, 2019.

[17] R. Myrzashova et al., “Blockchain meets federated learning in healthcare: A systematic review with challenges and opportunities,” IEEE Internet Things J., 2023.

[18] H. Li et al., “Review on security of federated learning and its application in healthcare,” Future Gener. Comput. Syst., vol. 144, pp. 271–290, 2023.

[19] H. Li et al., “Review on security of federated learning and its application in healthcare,” Future Gener. Comput. Syst., vol. 144, pp. 271–290, 2023.

[20] P. Gupta et al., “A Novel Data Poisoning Attack in Federated Learning based on Inverted Loss Function,” Comput. Secur., vol. 130, p. 103270, 2023.

[21] A. Mohammad. (2023). Safeguarding Connected Health: Leveraging Trustworthy AI Techniques to Harden Intrusion Detection Systems Against Data Poisoning Threats in IoMT Environments. Babylonian Journal of Internet of Things, 2023. https://doi.org/10.58496/BJIoT/2023/005

[22] Z. Lian et al., “SPoiL: Sybil-Based Untargeted Data Poisoning Attacks in Federated Learning,” in Proc. Int. Conf. Network and System Security, Cham: Springer Nature Switzerland, 2023.

[23] R Hameed,. T., & Mohamad, O. A. (2023). Federated Learning in IoT: A Survey on Distributed Decision Making . Babylonian Journal of Internet of Things, 2023, 1–7. https://doi.org/10.58496/BJIoT/2023/001

[24] Skin Cancer MNIST: HAM10000, Kaggle, [online] Available: https://www.kaggle.com/datasets/kmader/skin-cancer-mnist-ham10000