IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 8 May 2024, accepted 12 August 2024, date of publication 19 September 2024, date of current version 14 October 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3464240

APPLIED RESEARCH

Privacy-Preserving Hybrid Federated Learning Framework for Mental Healthcare Applications: Clustered and Quantum Approaches

ARTI GUPTA, MANISH KUMAR MAURYA, KHYATI DHERE, AND VIJAY KUMAR CHAURASIYA, (Senior Member, IEEE)
Department of Information Technology, Indian Institute of Information Technology at Allahabad, Allahabad, Uttar Pradesh 211015, India
Corresponding author: Arti Gupta (arti2910gupta@gmail.com)

ABSTRACT Privacy-preserving approaches are essential in health- care applications where sensitive data is involved. Federated learning (FL) has emerged as a widely adopted approach for collaboratively training decentralized models without sharing individual health records. However, ensuring privacy in health- care data, both during training and when clients exchange their models with a central server, remains a challenge. Bias, fairness, clients heterogeneity, and constrained computation are also challenging factors. To address this challenge, in this paper, a communication-efficient and privacy-preserving hybrid Federated learning (HFL) framework is specifically designed for mental healthcare applications. Two HFL approaches, namely Clustered Federated Learning (CFL) and Quantum Federated Learning (QFL), have been proposed. CFL focuses on leveraging the learning behaviour of clients and Conversely, QFL introduces a new phase to FL by incorporating a variational quantum classifier (VQC) for classification tasks. Angle encoding is used for a quantum state preparation to enhance data encoding and learning the quantum model. Experiments were conducted on independent and identically distributed (iid) and non-independent and identically distributed (non-iid) data to evaluate the performance of the proposed methods with state-of-the-art results available in the literature. The results demonstrate exceptional performance in the case of QFL, achieving an accuracy of 84.00%. CFL also exhibits promising results with an accuracy of 78.396%. Additionally, QFL achieves 18.75% better recall and CFL has 6.24% better precision than traditional FL. Nevertheless, it's crucial to remember that every model has advantages and disadvantages of its own.

INDEX TERMS Hybrid federated learning, mental healthcare, clustered federated learning, quantum federated learning, variational quantum classifier, angle encoding.

I. INTRODUCTION

In recent years, mental healthcare has gained significant attention due to the rising prevalence of mental health disorders and the need for effective interventions. Since people do not want to share their personalized mental health information on social networks. The lack of effective machine learning (ML)/deep learning (DL) models to maintain patient data privacy often leads to fragmented data silos. With technological advancements, researchers have explored innovative approaches to address the challenges of mental healthcare data while preserving privacy and ensuring efficient data analysis. One such approach proposed by Google in 2016 is federated learning (FL) to break the problem of medical data silos caused by patient data privacy [1]. FL enables distributed data analysis across multiple sources while keeping the data decentralized, thereby preserving privacy [2]. Additionally, in 2018, the European Union (EU) passed the general data protection regulation (GDPR) data privacy law aims to enhance the protection of individual's data and give them more control over how their data is collected, processed, and stored by organizations [3], [4].

The associate editor coordinating the review of this manuscript and approving it for publication was Wentao Fan.

TABLE 1. Nomenclature.

Symbols	Acronyms
Δw	Weight update
w_t	Weight at communication round t.
n	Number of client
η_t	Learning rate
q_j	Importance parameter of node j
r_j	Weighted number of samples reaching node j
G_j	Impurity value of node j
F_i	Importance parameter of feature i
RfF_i	Importance parameter of feature i from Rf model
$normF_i$	Normalized importance of feature i
x	Data value
S_j	Cluster j
C_i	Client i
m	Number of cluster
$g_t^{S_j}$	Model parameter of cluster S_j at round t
ψ	Quantum state
β	Adjustable parameter of VQC
J	Cost function
L	Layer number
$\mathcal{H}$	Hilbert space
Lf	loss function

In this context, FL has emerged as a promising approach to address privacy concerns in data analysis while leveraging the power of decentralized data sources. In FL, instead of centralizing data on a single server, the learning process is distributed among multiple devices or entities that hold local data [5], [6]. This distributed learning paradigm offers advantages such as preserving data privacy, reducing the need for data transfer, and enabling collaboration among different organizations [7], [8]. Privacy is affected by data shared or hijacking the artificial intelligence (AI) model, no matter what training method is used. In FL, the weakest link occurs when the client trades their working model with the central server. Therefore, a model that preserves privacy during training and exchanging the model parameters is required when dealing with highly sensitive and regulated data. Instead of this, traditional FL approaches often face challenges in handling complex and high-dimensional mental health data. Also, the data heterogeneity, variability, and scale can limit the effectiveness of FL models.

There are limitations of traditional FL techniques that necessitate the development of new strategies that address challenges related to communication efficiency, data heterogeneity, privacy, security, model bias, and fairness.

To overcome these limitations, hybrid federated learning (HFL) has been proposed as an extension of FL that integrates other techniques to enhance its performance and efficiency. This paper develops clustered federated learning (CFL) and quantum federated learning (QFL) by integrating innovative techniques such as clustering and quantum computing. CFL incorporates clustering algorithms by classifying clients into clusters based on learning rates and characteristics. It offers improved control over the training process, enhanced data privacy, reduced communication costs, and superior model

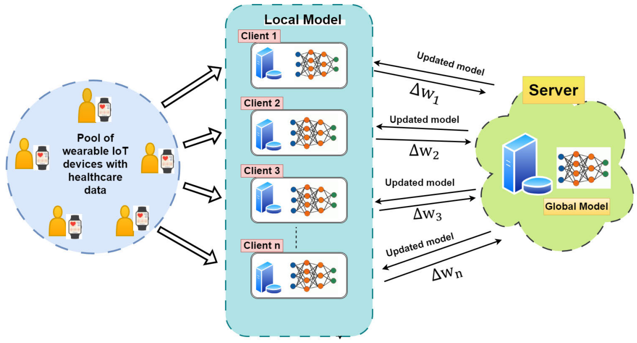


FIGURE 1. FL process having a pool of n clients with mental health data stored in their wearables which participate in FL process to get optimized global model by sharing weight update Δw_i^t.

performance. It provides a flexible and efficient framework for detection analysis. Another emerging approach within HFL is QFL. With the rapid advancements in quantum computing, researchers are exploring its potential applications in various domains, including ML [9], [10], [11]. QFL utilizes the principles and computational power of quantum computing to enhance the learning process. Quantum computing offers advantages such as parallelism, optimization capabilities, and more efficient processing of complex mathematical operations. By incorporating quantum computing into the HFL framework, QFL aims to overcome classical computing limitations in FL. It can potentially optimize the training process, handle large-scale datasets, and improve the accuracy and efficiency of mental healthcare models. QFL presents a novel approach that offers significant data analysis and model development advancements. The need for HFL approaches, such as CFL and QFL, arises from the requirements of privacy preservation, improved model performance, and efficient data analysis in the domain of mental health monitoring.

The major contribution are summarized as:

1) The paper introduces and explores the concept of two HFL as an extension of traditional FL named as: CFL (FL+clustering) and QFL (FL+quantum computing). By integrating additional techniques, such as clustering and quantum computing, into the FL framework, HFL aims to enhance its performance, scalability, communication efficiency, and privacy-preserving capabilities. This framework provides a novel approach to address the challenges associated with complex and high-dimensional datasets.
2) The paper focuses on applying the edge-based framework on wearable devices for stress detection analysis in mental healthcare. Wearable Internet of Things (IoT) devices, equipped with sensors to capture physiological signals, provide a rich data source for monitoring stress levels. The research demonstrates how the proposed framework can effectively utilize wearable devices to detect real-time stress, offering insights into

individuals' stress patterns and potential interventions for improved mental well-being.

3) The paper delves into the application of CFL that incorporates clustering algorithms into FL, enabling the grouping of clients for more efficient analysis and model training.
4) QFL framework is also proposed with quantum-classical classifiers VQC. Here is an approach for pre-processing to map data into quantum states for further quantum classification. Angle encoding is used for the state preparation approach for data encoding.
5) This paper ensured privacy in healthcare data during training and when clients exchange their models with a central server.
6) A comparative analysis between traditional FL, CFL, and QFL methodologies and classification performance is conducted. The paper provides insights into their strengths, limitations, and suitability by evaluating them. This comparison contributes to selecting and optimizing appropriate techniques based on different healthcare scenarios' specific requirements and constraints.

The subsequent sections are organized in the following manner: Section II describes the literature review and process of FL. A detailed description of the problem definition is present in section III. Section IV describe the proposed methodology. Section V covers the dataset description, analysis of experimental work, and the result discussion. Section VI describes the limitation and future work that could be done. Finally, Section VII represents the conclusion.

II. BACKGROUND

A. RELATED WORK

In the literature survey, several existing research works on privacy-preserving HFL models for healthcare analysis are discussed. The focus is on ensuring data privacy while leveraging the benefits of FL in mental health monitoring.

Relevant approaches from the literature are summarized in Table 2, highlighting the models, research gaps, and experimented datasets. Wu et al. [23] addressed privacy concerns in in-home health monitoring using FedHome. However, mental stress detection remained unexplored until Firouzi et al. [16] presented a study on privacy-preserving stress monitoring using FL. [24], [25], [26] provide privacy concerns and an efficient approach for IoT-based healthcare systems. Further contributions include Bn and Abdullah [15] comparing FL with centralized models for decentralized speech analysis and Sáez-de-Cámara et al. [22] addressing FL challenges in heterogeneous environments. However, there was still an issue of privacy with FL while sharing learning with server and computational overhead then Yoo et al. [13] propose a personalized federated cluster model (PFCM) for Major Depressive Disorder (MDD) severity evaluation which share learning within a cluster and focused on reducing computational overhead. Advancements in CFL are demonstrated by Yan et al. [27] with an iterative clustered FL (ICFL) framework, Sattler et al. [28] with federated Multitask Learning (FMTL), and Briggs et al. [12] with FL and hierarchical clustering ($FL + HC$). The main difference between these existing CFL and proposed model is, they use limited epochs, fixed clusters, and communication rounds are also limited. They are not specifically designed for stress detection using physiological parameters.

The emergence of quantum machine learning (QML) further improves privacy in FL is explored by Sierra-Sosa et al. [29], [30] using TensorFlow Quantum. Gupta et al. [31] analyzing VQC models for binary classification on the PIMA diabetes and EDA datasets. Huang et al. [19] focused on decentralized feature extraction in quantum datasets, and Chen et al. [18] discussed a QFL architecture for dog vs cat classification. Furthermore, Chehimi and Saad [9] introduced the first QFL framework exclusively for quantum data, and Maheshwari et al. [10] enhanced VQC prediction rates through feature selection and state preparation techniques.

All these research works contribute to understanding privacy-preserving HFL models, addressing various challenges and proposing novel approaches that help apply HFL in mental health monitoring and healthcare analysis. Table 1 represents the Notation and symbol used in this paper.

B. FL: REVEALING THE OPERATIONAL WORKFLOW

FL is a decentralized ML approach that enables models to learn from distributed data sources without sharing the actual data. It preserves data privacy by allowing personal data to remain on local servers. In wearable IoT devices, clients collect data from these devices and perform local model training parameter updates using FL techniques [32]. The local models are then aggregated on a central server to create a global model that benefits from diverse data sources. This approach enables real-time data utilization and personalized applications while ensuring data security [33]. FL (Fig. 1) involves operational steps that facilitate the collaborative training of ML models using distributed data sources.

1) **Client Data Collection:** Each client possesses local data, which may comprise user behaviour data, sensor data, or other pertinent information. The decentralized approach to data collection safeguards data privacy and security, as the raw data remains on the client's devices.
2) **Model Initialization:** In this step, an initial ML/DL model is initialized by a central server or by distributing pre-trained models to the client devices. This model serves as a starting point for the FL process, and the global model w_0 is initialized.
3) **Distribute the global model to clients:** The global model w_0 is distributed to n clients, where n is the total number of clients.
4) **Local Model Training:** Each client i ($i \in n$) independently trains a local copy of the global model w_t on their local data, producing a new set of model weights w_i^t. This training occurs locally on the client's device, ensuring data privacy as the raw data does not leave the client's environment. The client uses its local

TABLE 2. A comparative analysis of the existing approach.

Ref.	Year	Description	FL	Hybrid FL CFL	Hybrid FL QFL	Research finding and Gap	Dataset
[12]	2020	FL with hierarchical clustering .		✓		They use limited epochs, fixed clusters, and communication rounds are also limited; as a result, doesn't inform about the effect on varying those parameters.	MNIST
[13]	2021	PFCM:Predict major depressive disorder severity from heart rate variability (HRV).		✓		It used quite a small dataset, so results are not generalised.	self generated HRV
[14]	2021	FedMood:Analysis and diagnosis of depression.	✓			For non-iid data, accuracy decreased by 13% which can be improve by incorporating other techniques with FL.	self generated typing strokes.
[15]	2022	Privacy sensitive speech analysis :FL for speech analysis to assess depression.	✓			Speech is a well-known way to detect depression, and they got an accuracy of 74%.It help in understanding FL.	DIAC-WOZ
[16]	2022	Case study of FL on wesad data.	✓			—-	
[17]	2023	Wrist-Based Electrodermal Activity Stress Monitoring and Detection using Federated Learning:FL-based DNN.	✓			They use only two clients for their study ,which is far away from the practical scenario.	WESAD
[10]	2021	Variational Quantum classifier for binary Classification:			✓	It helps in tuning VQC but doesn't focus on privacy.	Sonar and the Dia-betes .
[18]	2021	hybrid quantum-classical ML models based on federated training			✓	They use pre-train model as a base model and give a hybrid quantum approach ,and its application in image classification.	Dog vs cat and CIF-AR
[19]	2022	find approximate optima develop an extension of the conventional VQA.			✓	They help in explaining VQA and explain qubit concept. Experimented on image dataset	MNIST
[9]	2022	Quantum federated learning with quantum data.			✓	They consider purely quantum data, which is not real-life application-based data.	Excitations of quan-tum cluster states.
[20]	2022	Article investigated the blockchain -empowered cluster-based imbalanced FL model BLADE for the blade icing detection of the wind turbine.	✓	✓		Achieved precision is quiet low.	self generated wind farm data.
[21]	2023	An iterative CFL (ICFL) framework.		✓		Explain the concept of CFL although Overfitted.	MNIST and CIFAR 10
[22]	2023	An architecture for unsupervised models training for network intrusion detection		✓		Application of unsupervised FL is explained.	Gotham testbed cap-turing of network traf-fic traces for dataset generation.
Our	2023	A comparative study of hybrid FL for stress detection	✓	✓	✓	Comparison of hybrid FL is given.Accuracy could be impr-ove by further hypertuning.	WESAD

optimization algorithm,i.e. Adam optimizer, to update the model parameters based on its local data.

5) **Send weight differences to the server:** Each client calculates the weight difference between their local model w_i^t and the global model w_t as $\Delta w_i^t = w_i^t - w_t$, and sends this weight difference to the server.
6) **Model Aggregation:** The server aggregates the weight differences from all clients by averaging them:

$$\Delta w_{t+1} = \frac{1}{n}\sum_{i=1}^{n} \Delta w_t^i \tag{1}$$

This aggregation process ensures that the collective knowledge from different clients is captured in the global model while preserving data privacy.
7) **Model Update and Iterations:** The server updates the global model w_t by adding the aggregated weight differences Δw_{t+1}.

$$w_{t+1} = w_t - \eta_t \cdot \Delta w_{t+1} \tag{2}$$

where η_t is the learning rate. Clients receive model parameters, and the process is repeated iteratively to improve the global model by incorporating collective intelligence from client devices.
8) **Convergence and Final Model:** Steps 3 to 7 are repeated for a predetermined number of cycles. The FL process converges as the iterations go on, and the global model reaches a point where additional iterations are no longer significantly better for performance. The final trained model is obtained at this stage, representing the collaborative knowledge learned from the decentralized data sources.

III. HYBRID FL PROBLEM DESCRIPTION

A. PROBLEM DEFINITION

When applying federated training in healthcare applications, it is crucial to consider multiple factors carefully. These factors encompass ensuring the secure aggregation of model parameters, effectively managing computational overhead, addressing concerns related to bias and fairness in data, and accommodating the heterogeneity of clients. Notably, only a few researchers have focused on decentralized model training for mental health monitoring while ensuring data privacy [16]. It is important to mention that the existing work in this area does not specifically encompass the aspects mentioned earlier of federated training. While federated training offers a degree of data privacy by keeping the

TABLE 3. A comparative analysis of the existing approach.

Ref.	Year	Description	FL	Hybrid FL CFL	QFL	Research finding and Gap	Dataset
[12]	2020	FL with hierarchical clustering .		✓		They use limited epochs,fixed clusters, and communication rounds are also limited; as a result,doesn't inform about the effect on varying those parameters.	MNIST
[13]	2021	PFCM:Predict major depressive disorder severity from heart rate variability (HRV).		✓		It used quite a small dataset, so results are not generalised.	self generated HRV
[14]	2021	FedMood:Analysis and diagnosis of depression.	✓			For non-iid data, accuracy decreased by 13% which can be improve by incorporating other techniques with FL.	self generated typing strokes.
[15]	2022	Privacy sensitive speech analysis :FL for speech analysis to assess depression.	✓			Speech is a well-known way to detect depression, and they got an accuracy of 74%.It help in understanding FL.	DIAC-WOZ
[16]	2022	Case study of FL on wesad data.	✓			—-	
[17]	2023	Wrist-Based Electrodermal Activity Stress Monitoring and Detection using Federated Learning:FL-based DNN.	✓			They use only two clients for their study ,which is far away from the practical scenario.	WESAD
[10]	2021	Variational Quantum classifier for binary Classification:			✓	It helps in tuning VQC but doesn't focus on privacy.	Sonar and the Diabetes .
[18]	2021	hybrid quantum-classical ML models based on federated training			✓	They use pre-train model as a base model and give a hybrid quantum approach ,and helps in understanding its application with FL.	Dog vs cat and CIFAR
[19]	2022	find approximate optima develop an extension of the conventional VQA.			✓	They help in explaining VQA and explain qubit concept .	MNIST
[9]	2022	Quantum federated learning with quantum data.			✓	They consider purely quantum data ,which is not real-life application based data.	Excitations of quantum cluster states.
[20]	2022	FL based on blockchain empowered clustered.	✓	✓		Achieved precision is quiet low.	self generated wind farm data.
[21]	2023	An iterative CFL (ICFL) framework.		✓		Explain the concept of CFL although Overfitted.	MNIST and CIFAR 10
[22]	2023	An architecture for unsupervised models training for network intrusion detection		✓		Application of unsupervised FL is explained.	Gotham testbed capturing of network traffic traces for dataset generation.
Our	2023	A comparative study of hybrid FL for stress detection	✓	✓	✓	Comparison of hybrid FL is given.Accuracy could be improve by further hypertuning.	WESAD

data at its source, it alone is insufficient. Somebody can compromise the privacy of sensitive user data if the model parameters are compromised during the model aggregation process.

Safeguarding user data throughout the model training and transmission processes presents substantial challenges. It is of utmost importance to ensure secure aggregation for model updates and prevent unauthorized access to client data to maintain data privacy effectively.

Data heterogeneity also poses a significant challenge in federated training. The models used in FL rely on local data, which can introduce bias and may not accurately represent the entire population. This can lead to fairness concerns when deploying the approach on a larger scale. One additional challenge to consider in FL is the communication overhead involved. Since FL involves significant communication between clientss, it can lead to increased computational burdens and longer training times. This is particularly true when a larger number of communication rounds are required to exchange the model updates.

To tackle these challenges and bridge the existing research gap, this paper introduces the Hybrid FL approach, which aims to address the aforementioned issues associated with FL.

IV. PROPOSED METHODOLOGY

In this section, the methodology and workflow of CFL and QFL is presented. Each method has its own benefit and advantage with respect to improving classification etc.

A. DATA PREPOSSESSING

We utilized feature selection techniques to address the dataset's high dimensionality and limited computational resources. By selecting the most relevant features, we improved model performance and accuracy, overcoming the potential impact of excessive features that can hinder ML models.

Two ML classifiers are used: 1) Random Forest (entropy-based) 2) Extra tree classifier-based feature importance. Random Forest consists of 4 to 12 hundred decision trees, where each tree is created by randomly selecting observations and attributes. This randomness ensures decorrelation among trees and is less prone to overfitting. Each tree consists of nodes with affirmative or negative questions based on attributes, dividing observations into distinct buckets. The purity of each bucket determines attribute significance.

Nodes importance is calculated by Scikit-learn using each decision tree's Gini Importance [34], assuming only two child

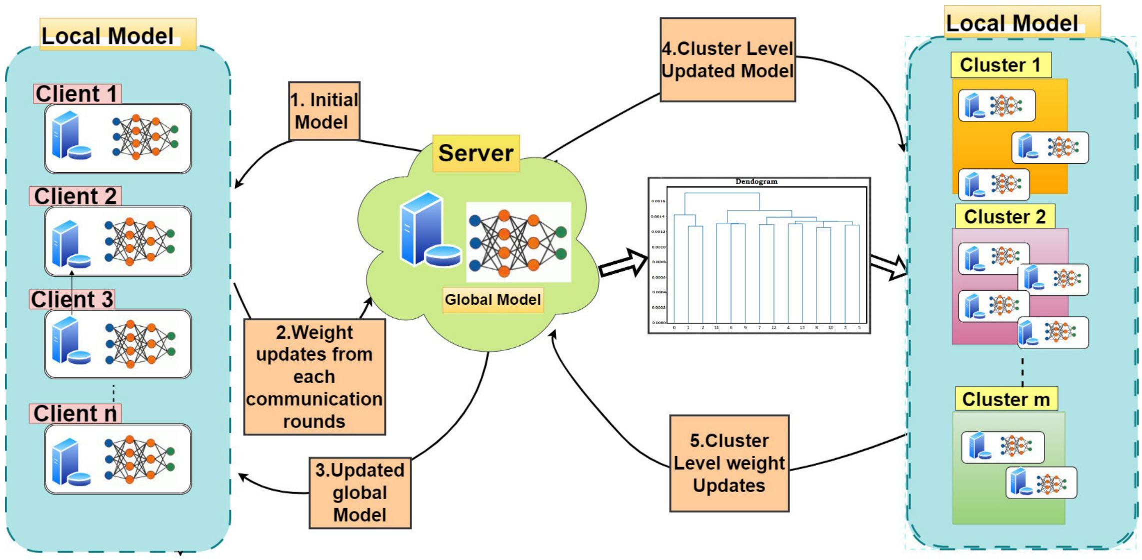


FIGURE 2. CFL process consisting of local model weight updates based clustering with FL on cluster level.

nodes (binary tree):

$$q_j = r_j G_j - r_{left(j)} G_{left(j)} - r_{\text{right}(j)} G_{right(j)} \tag{3}$$

where q_j is the importance of node j, r_j is the weighted number of samples reaching node j, G_j is node j impurity value, $left(j)$ is child node from left split on node j, $right(j)$ is child node from right split on node j.

The following formula is used to determine the significance of each feature on a decision tree:

$$F_i = \frac{\sum_{j\text{: node } j \text{ splits on feature } i} q_j}{\sum_{k \in \text{ all nodes}} q_k} \tag{4}$$

where Fi_i is the importance of feature i, q_j^i is the importance of node j for feature i. These can then be divided by the total amount of feature importance to be normalized to a value between 0 and 1:

$$\text{norm } F_i = \frac{F_i}{\sum_{j \in \text{ all features}} F_j} \tag{5}$$

The final feature's weight, at the level of the Random Forest, is the average across all the trees. The total number of trees is divided by the sum of the important values of the features on each tree:

$$RfF_i = \frac{\sum_{j \in \text{ all trees}} \text{ norm } F_{ij}}{T} \tag{6}$$

where RfF_i is the feature i importance calculated from all trees in the Rf model, $normF_{ij}$ the normalized feature importance for i in tree j and T total number of trees.

The feature importance attribute of the model is used to determine the importance of each feature in the dataset. Using the Extra Tree Classifier, the top 10 features are extracted. The intersection technique is applied to combine the standard features from each classifier, resulting from two classifiers. These classifiers identify the ten crucial features from the dataset.

The data is standardized using the normalization method to scale it between 0 and 1, reducing training time. Each attribute's maximum and minimum values are represented as 0 and 1, respectively.

$$Xnorm = \frac{X - Xmin}{Xmax - Xmin} \tag{7}$$

B. CLUSTERED FEDERATED LEARNING

CFL is a variation of FL where clientss are partitioned into clusters based on weight similarity during FL training using hierarchical clustering [35]. A separate model is trained on each cluster, which is later combined to form a global model. CFL involves two steps: first, clustering using agglomerative clustering based on weight similarity from the FL process and then CFL to produce a global model.

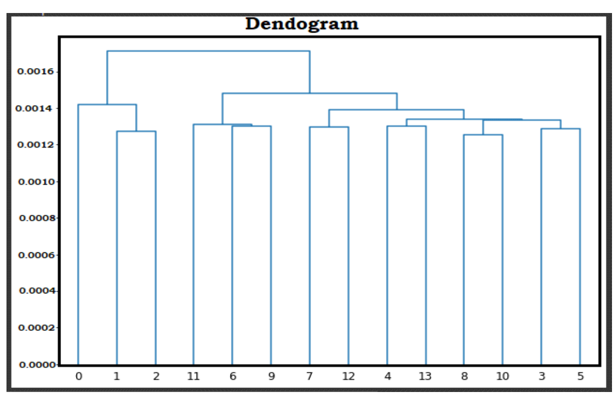


FIGURE 3. Dendogram representing the hierarchical relationship between clusters.

The proposed approach combines hierarchical clustering and the traditional FL method. This method consists of four major steps given in Fig. 2, i.e. 1) Basic traditional FL using all clients and recording their weight updates (also explained in Section II-B). 2) Hierarchical clustering on the basis of those weight updates and dividing them into clusters. 3) Applying FL on a cluster level, get a focused model. 4) Cluster-based aggregation for weight updates to get a final generalized model.

1) AGGLOMERATIVE HIERARCHICAL CLUSTERING

Agglomerative hierarchical clustering is a technique that follows a bottom-up approach to group similar data points into clusters. Each data point is first treated as a separate cluster in the approach, which then iteratively joins the nearest clusters until all data points are in a single cluster.

The process involves the following steps:

1) Distance Computation: Distance metric such as Euclidean distance or Manhattan distance is used to compute distance or similarity between each pair of data points to quantify how dissimilar or similar the data points are.
2) Initialising Clusters: Initially, assign each data point to its own cluster, which means the number of clusters is equal to the number of data points.
3) Compute Cluster Similarity: Calculate the similarity or distance between each pair of clusters. The distance between two clusters can be calculated using techniques such as single-linkage, complete-linkage, or average-linkage, based on the separations between the clusters' individual data points.
4) Merging Closest Clusters: Merge these two clusters into a single cluster by identifying the two closest clusters based on the computed similarity or distance measure. The choice of merging strategy depends on the linkage method used.
5) Update Cluster Similarity: To determine the next pair of clusters to merge, recalculate the similarity or distance between the newly formed cluster and the remaining clusters.
6) Iterative Merging: Repeat the process of merging until all data points are part of a single cluster.
7) Dendrogram Construction: As clusters are merged, a dendrogram is constructed as shown in Fig.3 to represent the hierarchical relationship between the clusters. The height of each fusion in the dendrogram corresponds to the similarity or distance at which the clusters were merged.
8) Cluster Selection: The desired number of clusters can be chosen by cutting the dendrogram at a specific height. The clusters below the chosen height will be considered the final clusters.

Agglomerative Hierarchical Clustering is a versatile clustering method that can accommodate different distance metrics and linkage strategies. It provides a hierarchical representation, allowing for various levels of granularity in clustering results.

2) OPERATIONAL STEPS OF CFL

Mathematically, it can be explained as: Let set of clients and clusters are $\mathcal{C} = \{\mathcal{C}_1, \mathcal{C}_2, \ldots, \mathcal{C}_n\}$ and $\mathcal{S} = \{\mathcal{S}_1, \mathcal{S}_2, \cdots, \mathcal{S}_m\}$ respectively. All clusters are formed based on weight similarity among clients using agglomerative hierarchical clustering. Initial cluster parameter set $g_0^{\mathcal{S}} = \{w_0^{\mathcal{S}_1}, w_0^{\mathcal{S}_2}, \ldots, w_0^{\mathcal{S}_i}\}$ which is actually the copy of global model parameters w_t from FL and $k \in n$ where k is any variable number of clients. $\mathcal{S}_j.clients = \{C_k | C_k \text{ is in cluster} \mathcal{S}_j, \forall C_k \in \mathcal{C}\}$. The overall global model parameter is w_0 each communication round T, for each group S_j in $\mathcal{S}$ clients $C_{k,j}$ that belong to S_j. Perform the training on its data and update parameters using current cluster model parameters $w_t^{S_j}$ parallelly.

$$w_{t+1}^{C_k} \leftarrow \text{update}\{C_k, w_t^{C_k}\} \tag{8}$$

After each communication t, all Cluster parameter updates parallelly as

$$g_{t+1}^{S_j} = \frac{1}{k} \sum w_t^{C_k} \tag{9}$$

When each cluster got trained after T communication rounds. Global model parameter update as

$$w_{t+1} = \frac{1}{m} \sum g_{t+1}^{S_j} \tag{10}$$

Above steps are repeated for multiple iterations to improve the performance and convergence of the global model.

C. QUANTUM FEDERATED LEARNING

QFL comprises key steps for collaborative training using quantum data sources. This approach enables collaborative learning, ensuring data privacy and leveraging quantum computing for enhanced model training and detection. The pictorial representation of QFL is given in Fig. 4.

1) FEDERATED QUANTUM DATA PREPARATION

For the QML, it is required to prepare the classical data as federated quantum data called state preparation. For example, a common classification of the function in supervised learning determines $y = f(x)$ by mapping the input data x and the output labels y. The primary objective of categorization is to raise the predictive model's accuracy. The training data for the binary classification $B = \{c1, c2, \ldots, cn\}$, where B is the target label, is a set of data that can be represented in the traditional ML domain as [36].

$$D = \{(x_1, y_1), \ldots (x_i y_i), \ldots (x_l, y_l)\} \tag{11}$$

where, x_i denotes some of the features (l) on the order's attributes of data point i, and y_i designates the matching data point. In the instance of binary classification, $y_i \in \{c1, c2\}$, where $x_i \in R^d$ and d is a set of features with real values. In the realm of QML, the initial step involves converting conventional data into quantum data, which is

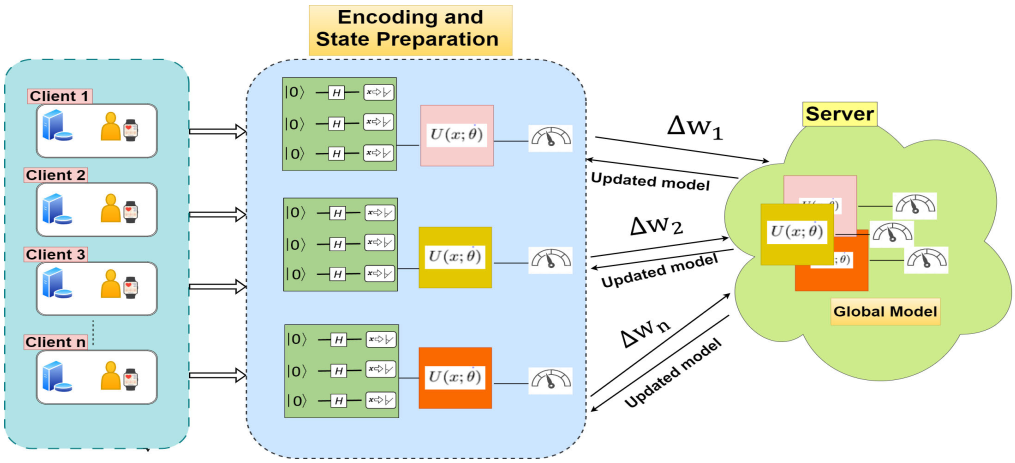


FIGURE 4. QFL model consisting of variational quantum classifier along with FL process.

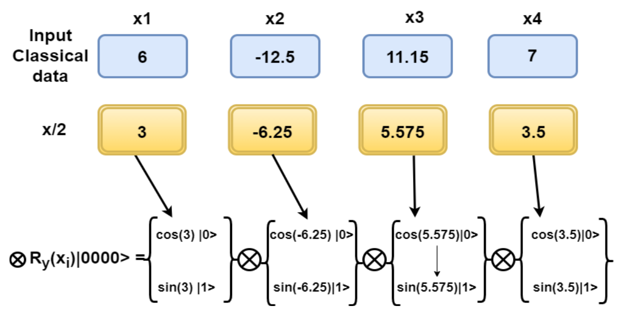


FIGURE 5. Angle encoding of classical data points demonstrating using four qubits.

then represented within the training data. This transformation is crucial as it serves to elucidate the framework used for analysis within the QML domain.

$$F_l = \{(|\psi_1\rangle, y_1), \ldots (|\psi_i\rangle, y_i), \ldots, (|\psi_n\rangle, y_n)\} \tag{12}$$

where $|\psi_i\rangle$ denotes the quantum state of F_l, $|\psi_i\rangle \in C^{2^d}$, and in the case of data stratification, $y_i \in \{c1, c2\}$. The classical data can be embedded in high-dimensional quantum data using a variety of approaches. There are many methods for doing this, i.e. basis encoding, amplitude encoding, and angle encoding. In this work, angle encoding is used as it is most suitable for floating data points.

Angle encoding [37] is a straightforward and efficient way for embedding data. The values that need to be encoded are applied rotations on the $x-$ or $y-$axis using quantum gates to produce the angle embedding. The rotations required to apply angle embedding to a dataset, as shown in Fig. 5, will be equal to the number of features in the dataset. It would require n qubits to create the collection of quantum states in the n-dimensional sample. Angle embedding [38] has several benefits, one of which is that it may be carried out with parallelism in a constant amount of time because each qubit will pass through a rotation gate in parallel. Each qubit's encoding requires just one gate, which lowers noises that are crucial in today's NISQ computers.

Angle embedding converts a single floating-point value $x \in R$ into a quantum state with the mapping.

$$x \mapsto R_\xi(x)|0\rangle = e^{-ix\sigma_\xi/2}|0\rangle \tag{13}$$

where $\xi \in \{x, y, z\}$ is the axis of rotation in the Bloch sphere.

2) LOCAL MODEL: VQC

In this research paper, QFL leverages the power of quantum computing by integrating VQC as the local model at each client, ensuring privacy preservation. The VQC is a quantum algorithm that uses quantum circuits for classification tasks. The VQC's ability to capture complex patterns and correlations in data makes it a promising choice for FL. Each client trains a VQC on its local data using classical optimization techniques in this approach. A variational circuit [39] is a type of quantum circuit that contains adjustable parameters, represented by $\boldsymbol{\beta}$, which can be optimized to minimize a cost function. The circuit consists of layers of unitary gates applied sequentially on qubits, with each gate acting on a subset of the qubits, which has two 4-qubit heavily entangled layers with rotations R and CNOTs. The circuit's output state is represented by $|\Psi(\boldsymbol{\beta})\rangle$. The goal is to optimize the parameters such that the output state closely approximates the target output state $|\psi_{\text{target}}\rangle$, corresponding to the desired classification result.

The optimization of the parameters $\boldsymbol{\beta}$ can be performed using classical optimization algorithms, such as gradient descent. In this procedure, the parameters are iteratively updated, and the gradient of the cost function concerning the parameters is determined until the cost function is minimized. In the case of binary classification, the output of the circuit can be mapped to a probability of the input belonging to one of two classes. The cost function is typically chosen as a function of the difference between the predicted probability and the true label for each input in the training set.

The variational quantum circuit can be represented mathematically as follows:

Let $\boldsymbol{\beta}$ be a vector of parameters for the VQC. The input data $\boldsymbol{x}$ is encoded into the quantum state $|\psi(\boldsymbol{x}, \boldsymbol{\beta})\rangle$ using quantum gates. The state is then measured on a computational basis, yielding a classical output y.

$$|\psi(\boldsymbol{x}, \boldsymbol{\theta})\rangle = U(\boldsymbol{\theta})|\phi(\boldsymbol{x})\rangle \tag{14}$$

Here, $|\psi(\boldsymbol{x}, \boldsymbol{\theta})\rangle$ represents the quantum state resulting from the VQC with input $\boldsymbol{x}$ and parameter vector $\boldsymbol{\theta}$. $U(\boldsymbol{\theta})$ denotes the unitary operator corresponding to the variational quantum circuit, which acts on the initial state $|\phi(\boldsymbol{x})\rangle$ that encodes the input $\boldsymbol{x}$. The probability of the input $\boldsymbol{x}$ belonging to class c is given by:

$$p_c(\boldsymbol{x}, \boldsymbol{\beta}) = |\langle c \mid \psi(\boldsymbol{x}, \boldsymbol{\beta})\rangle|^2 \tag{15}$$

where $|c\rangle$ is the state representing class c.The cost function, denoted as $J(\boldsymbol{\beta})$, calculates the sum of the differences between the predicted probability and the true label for all training examples.:

$$J(\boldsymbol{\beta}) = \sum_{i=1}^{N} \left(p_{y_i}(\boldsymbol{x}_i, \boldsymbol{\beta}) - y_i\right)^2 \tag{16}$$

where y_i is the true label of the i-th training example.

Single qubit rotation and entangler layers using circuit-centric classifier design inspiration. The weights for each layer are contained in the parameter weights. Thus, the first dimension of weights determines the layer count, L. The imprimitive parameter specifies the type of the 2-qubit gates, which then sequentially act on the M wires, where $i = 1, \ldots, M$. The equation $(i + r) mod M$ determines each gate's second qubit, where r is a hyperparameter known as the range and $0 < r < M$.

3) OPERATIONAL STEP OF QFL

Let $\mathcal{H}$ be a Hilbert space representing the quantum states, and $\mathcal{M}(\mathcal{H})$ denote the space of linear operators on $\mathcal{H}$. The initial quantum model is represented as a parameterized quantum circuit $\mathcal{C}_\beta : \mathcal{H} \to \mathcal{H}$, where β is a vector of real-valued parameters.

The steps of QFL can be summarised as follows:

1) QFL starts by dividing the training data among multiple clients, each associated with a quantum model. Let $D_i = (x_1^{(i)}, y_1^{(i)}), \ldots, (x_{l_i}^{(i)}, y_{l_i}^{(i)})$ be the local dataset of the i-th client, where $x_j^{(i)}$ is the input feature vector and $y_j^{(i)}$ is the corresponding binary label.
2) Each client optimizes the shared quantum model by minimizing the empirical risk on their local dataset:

$$\min_{\beta} \frac{1}{n_i} \sum_{j=1}^{n_i} Lf\left(y_j^{(i)}, \left\langle \psi\left(\mathcal{C}_\beta, x_j^{(i)}\right)\right\rangle\right) \tag{17}$$

where $\psi(\mathcal{C}_\beta, x_j^{(i)})$ is the output state of the quantum circuit on the input $x_j^{(i)}$, and $Lf(y, \hat{y})$ is a loss function that measures the discrepancy between the true label y and the predicted label $\hat{y}$.
3) Each client's quantum model is trained using local data from that client, and the updated parameters are shared with a central server. The quantum global model is updated using a central server to aggregate the parameters.
4) To update the global model, the server collects the modified parameter vectors from all clients.:

$$\theta_{k+1} = \frac{1}{N} \sum_{i=1}^{N} \beta_{i,k+1} \tag{18}$$

where N is the total number of clients, $\beta_{i,k+1}$ is the updated parameter vector from the i-th client at iteration $k + 1$.

Each client repeats the process of optimizing the model on their local data after receiving the server's broadcast of the new global model. When the process has reached convergence, further data classifications can be made using the final model parameters.

V. EXPERIMENTAL ANALYSIS

This section presents a demonstration and analysis of the experiment result for stress detection in mental healthcare applications. We analyze the result of the FL, CFL, and QFL approach in both iid and non-iid scenarios.

A. DATASET DESCRIPTION

The wearable stress and affect detection (WESAD) dataset [40] is a valuable resource for detecting mental states, capturing physiological signals from wearable sensors like blood volume pulse, electrocardiogram, electrodermal activity, electromyogram, respiration, body temperature, and three-axis acceleration. It includes data from wrist and chest-worn devices of 15 subjects, providing multimodal physiological and motion data. The dataset features multiple sensor modalities and three affective states (neutral, stress, amusement), facilitating ML model training and evaluation. With 64 features and around 100,000 samples, it has implications for mental health monitoring, stress management, and affective computing.

B. EXPERIMENTAL SETUP

Our model is implemented using TensorFlow, Keras, and Pennylane as a backend. All experiment are conducted on ubuntu 20.04.4 Intel® Xeon(R) CPU E5-2630 v3 @ 2.40GHz × 16.Adam optimizer as our training optimizer is used.90% data for training and the rest for testing is used.

Parameters for communication rounds, local epochs, batch size, and learning rate are selected based on experimental comparisons as shown in table 4. Zero (0) is considered a non-stress stage, and one (1) is considered a stress stage.

Data is distributed among 14 clients using two scenarios: iid and non-iid scenarios. Iid data is prepared by data partitioning, which involves dividing the entire dataset into

TABLE 4. Hyper parameters.

Parameter	FL	CFL	QFL
Batch size	16	16	16
Learning Rate	0.00025	0.00025	0.00025
Qubit	-	-	10
Optimizer	Adam	Adam	Adam
Layers	7	7	5
Loss function	binary cross entropy	binary cross entropy	Square loss
Epochs	5	5	5

separate subsets that are disjoint and represent different clients or data silos and then randomly sampling data from the overall dataset to create client-specific datasets. For non-iid data distribution, it is prepared in two manners. In the first manner, a real-world scenario where data is gathered from different clients inherently introduces non-iid as data is quantity-wise and class imbalanced. The second manner is distributing labels unbalanced across clients and evenly distributing the number of samples. This can be done by sorting the entire dataset based on labels and then dividing the samples among the clients; in this way, each client will have an unbalanced class. FL is performed for both manner of non-iid in which the first technique gives better results, is closer to real-world scenarios, and is reliable. Therefore, all the experiments are performed using the first manner of non-iid for the HFL model.

C. MODEL ARCHITECTURE

A sequential deep neural network (DNN) model is employed with the specific configuration for FL and CFL. The input layer consists of 8 units and utilizes the ReLU activation function. Following the input layer, there are five dense hidden layers. These hidden layers have 256, 128, 64, 54, and 50 units, respectively. The output layer has 1 unit and employs the sigmoid activation function. To address the binary classification task, each dense layer incorporates ReLU, sigmoid activation functions, and a fully connected layer. Adam optimizer is used for calculating and minimizing the loss, with binary cross-entropy as the loss function.

In the case of QFL model architecture, a VQC model is used with 10 qubits, 5 layers, square loss as loss function, the learning rate of 0.00025, angle encoding is used for encoding data and for state preparation and classification strongly entangle layer circuit is used which is an inbuilt library of Pennylane.

D. RESULT ANALYSIS AND DISCUSSION

In this section, a thorough study of FL, CFL and QFL with the help of results analysis and discussion is presented. To investigate the performance of FL, the most widely used and effective algorithm, FedAvg [41], was implemented, utilizing all 15 clients with their associated data. Accuracy, precision, F1-score, and recall metrics are used to evaluate the stress detection model performance, and descriptions of these parameters are given in [42].

TABLE 5. Classification comparison of all proposed model.

Method	Accuracy(%)	Precision(%)	Recall(%)	F1 score(%)
ML	98.00	98.00	98.75	98.45
FL	87.339	80.180	87.3	87.903
CFL	78.396	95.00	75.45	79.7
QFL	84.25	78.50	92.75	84.52

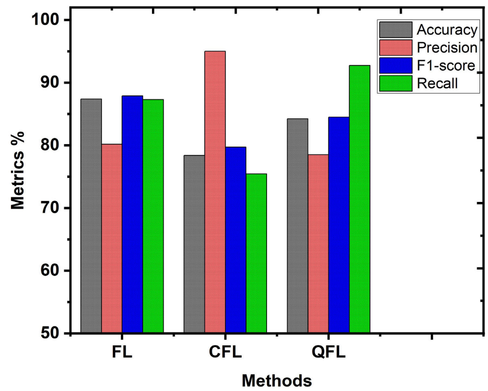


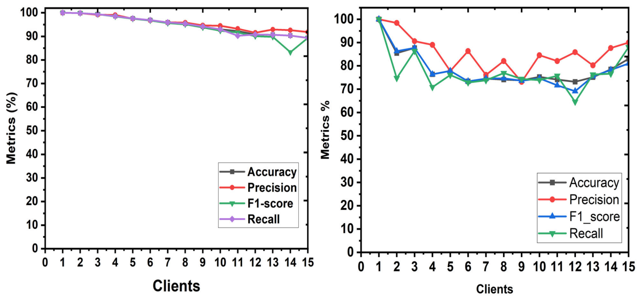
FIGURE 6. Clasification performace of HFL models.

1) CLASSIFICATION COMPARISON

As shown in table 5, comparison of the model's detection is made, ML model using a neural network is giving the best result in all parameters while sacrificing privacy. FL is providing good accuracy with the benefit of privacy but with high communication rounds. It is also observed as shown in fig.6, that the CFL worked on focusing on the client and shows 95% precision, i.e., 6.24% better over traditional FL with privacy. On the contrary, QFL achieves almost the same accuracy with 92.75% recall, i.e., 18.75% better performance during shorter communication rounds and with improved privacy. Concurrently, dealing with non-iid, which is real-world scenario then also, HFL gives exceptionally good results.

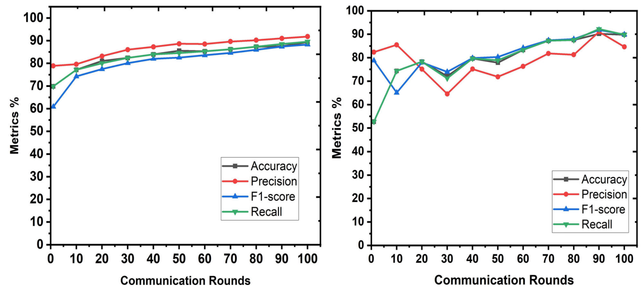
2) COMMUNICATION EFFICIENCY

The performance of FL algorithms such as FedAvg is influenced by the number of communication rounds, with varying trends observed in both iid and non-iid data settings. While FL models generally see increased accuracy with higher communication rounds in iid data scenarios, non-iid data settings initially experience fluctuations before eventually improving accuracy as shown in Fig. 9a and 9b. CFL also demonstrates varying accuracy trends across different clusters as the communication rounds progress, as shown in Fig.10. Some clusters exhibit an increasing accuracy trend, while others experience a decrease, as shown in Fig.11. This heterogeneity within clusters arises from variations in data samples or the quality of collected data. QFL exhibits fluctuating accuracy with increasing communication rounds for both iid and non-iid data, as depicted in Figs. 12. Fluctuations can be attributed to the quantum operations and



(a) iid (b) Non-iid

FIGURE 7. Effect of data partition Vs different numbers of clients.



(a) iid (b) Non-iid

FIGURE 9. Effect of FL in communication rounds.

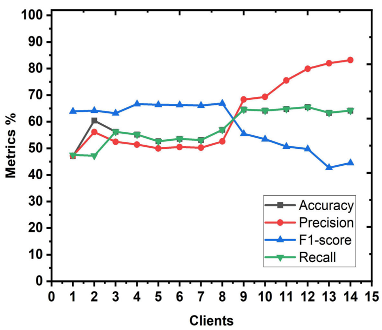


FIGURE 8. Effect of client participation on local training in FL (Non-iid).

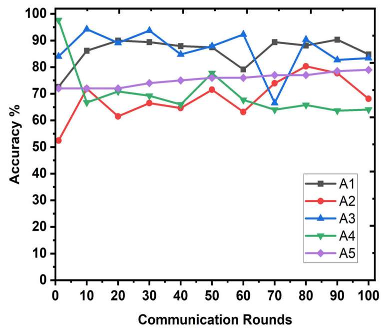


FIGURE 10. Effect of CFL in communication round.

measurements performed on the simulator accumulate noise and errors. Moreover, as the number of rounds increases, the circuit depth (the number of gates or operations in the circuit) also increases. This increased circuit depth can amplify the effects of noise and errors, leading to fluctuations in accuracy.

3) COMMUNICATION OVERHEAD

Regarding communication overhead, FL typically incurs higher overhead compared to CFL and QFL due to direct communication between all clients and the central server. CFL mitigates this overhead through local aggregation within clusters, while QFL introduces quantum communication overhead, influenced by the complexity of quantum states and measurements. These insights underscore the importance of understanding communication dynamics and data characteristics in optimizing FL performance.

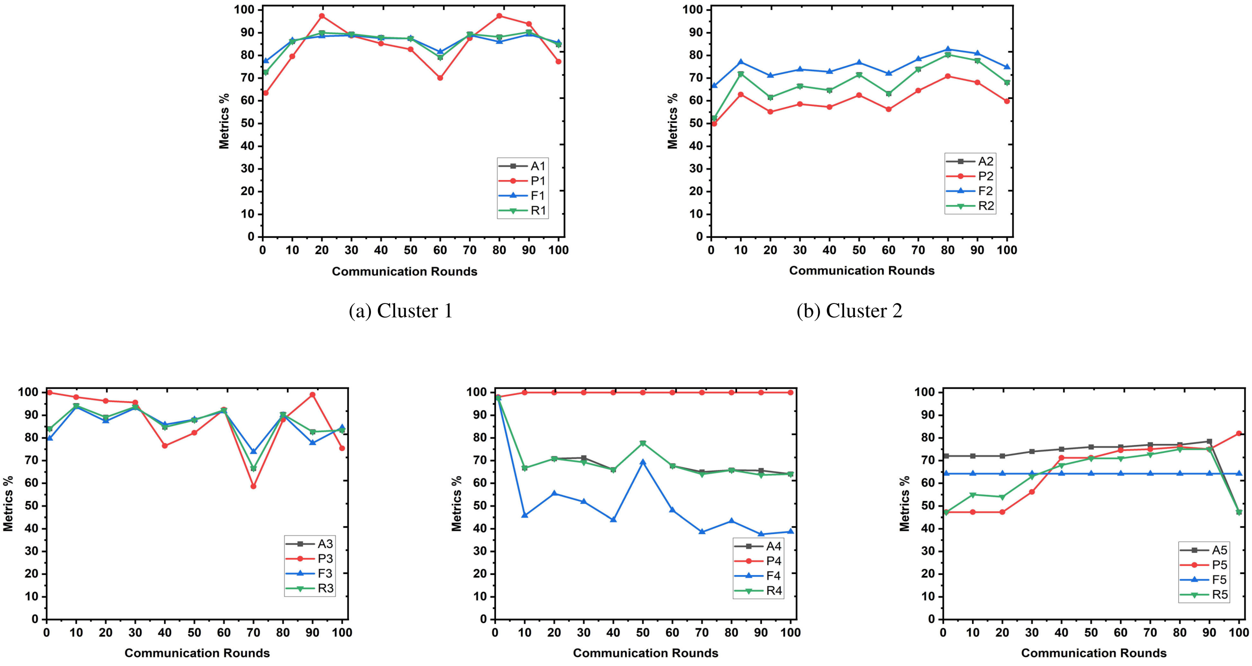
4) IMPACT OF DATA HETEROGENEITY

The impact of data heterogeneity is a crucial aspect addressed in the experiments conducted within the context of the FL framework for specifically, the experiments focus on both iid data and non-iid data to understand how different data distributions affect the performance of the proposed framework.

When iid data is utilized in FL, notable improvements are observed across various performance metrics including accuracy, recall, F1-score, and precision. This improvement is illustrated through the evolution of these metrics over communication rounds, as depicted in Fig. 9a. Furthermore, when QFL is applied to iid data, there are initial fluctuations in performance metrics before eventually achieving improved accuracy, as illustrated in Fig. 12a. However, the scenario shifts when dealing with non-iid data. In FL, the performance exhibits simultaneous but fluctuating increases in recall and accuracy, as depicted in Fig. 9b. Conversely, QFL experiences a improved in accuracy around 10-15 rounds, as evidenced in Fig. 12b. Additionally, when considering CFL, the trends observed are diverse and influenced by factors such as data diversity, model initialization, and convergence behaviors during rounds (Figs. 10 and 11).

a: IMPACT OF DATA PARTITION TO A DIFFERENT NUMBER OF CLIENTS

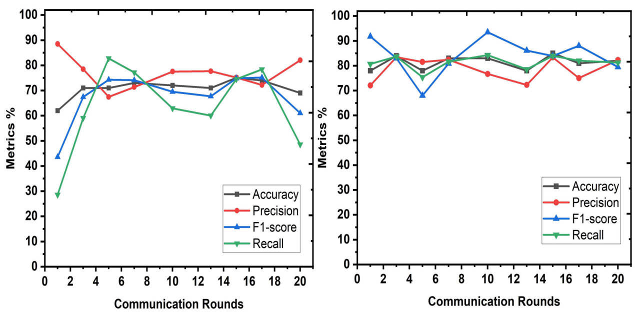
Conventional training (ML) can lead to a high-performance generalised model. However, the FedAvg model trained on individuals caused degradation in task performance due to data heterogeneity shown in Table 9. To address this, subjects were grouped to achieve customization with improved performance. The performance of FL using data partitioning across varying numbers of clients was evaluated and shown in Fig. 7. Each client holds data from multiple subjects such that when *client* $= 1$, then the client holds data from all 15 subject, and *client* $= 3$, then each client hold data from 5 subjects, and data is partitioned from all 15 clients to n clients where $n \in [1, 14]$. Fig. 7 represents the Accuracy, precision, F1-score, and recall also slightly decrease as the number of clients increases. As data heterogeneity degrades



(a) Cluster 1 (b) Cluster 2

(c) Cluster 3 (d) Cluster 4 (e) Cluster 5

FIGURE 11. Performance of CFL with the effect of communication rounds on each cluster in local training.



(a) iid (b) Non-iid

FIGURE 12. Effect of QFL in communication rounds.

classification performance. When the number of clients is one, the metrics parameter is highest because when all the data is given to one client, this process behaves like a centralized model, resulting in the highest accuracy value. It is observed that in the case of iid data Fig. 8a degrading accuracy with increasing clients. In contrast, non-iid (sorted class manner) Fig. 7b data give the best performance initially and then fluctuate as the number of clients increases. This pattern is due to the different partitioning of the data with the unbalanced class. Beyond the tenth client, fluctuation is also minimized.

b: EFFECT OF THE NUMBER OF PARTICIPATING CLIENTS

The Impact of different client participation rates (1-14 clients) on classification performance was investigated. All client hold their own data such that when $client = 1$, the client holds data from one patient (id1); when $client = 3$, each client holds data from one patient (id1, id2, id3 respectively). Fig. 8 shows that the accuracy, precision, and recall improve as the number of clients increases. Additional data from new clients contribute to this improvement. However, a degradation in F1-score performance was observed when increasing eight clients to nine. This degradation can be attributed to the disparity between the new client's data distribution and the existing data. Nevertheless, increased data samples benefit overall performance.

5) PRIVACY PRESERVATION COMPARISON

FL, CFL, and QFL are three prominent approaches for privacy-preserving ML in distributed settings. FL ensures privacy by keeping data on client and aggregating model updates instead of sharing raw data. CFL enhances privacy by clustering data and performing model updates within each cluster, reducing exposure of individual client data. CFL enables localized data processing within clusters, reducing the need for data transmission between clients and the central server. By performing computations locally, CFL minimizes the amount of shared sensitive data, thereby reducing the exposure to potential security threats. QFL leverages quantum computing principles for secure computations and employs quantum encryption techniques to protect data privacy. QFL also considers the future threat of quantum computers capable of breaking classical encryption algorithms. By adopting quantum encryption techniques now, QFL prepares for the post-quantum era, where traditional encryption methods may become vulnerable. The choice of privacy-preserving approach depends on specific requirements and trade-offs. FL is suitable for scenarios where data privacy and minimal data transfer are important. CFL is preferred when stronger privacy guarantees are desired at the cost of additional complexity. QFL is promising but requires further advancements in hardware and algorithms.

6) TRAINING TIME AND CONVERGENCE SPEED

QFL exhibits longer training time than FL and CFL, taking approximately 8-10 hours versus 2-4 hours. However, QFL achieves convergence with fewer communication rounds. CFL demonstrates rapid convergence with low training loss, while FL initially has a higher loss but eventually achieves similar convergence. QFL exhibits high initial loss but gradually converges after the 10th round. CFL shows efficient utilization of distributed data, FL catches up to CFL's performance, and QFL requires more iterations. These findings emphasize each approach's convergence speed and stability, paving the way for further research on optimizing convergence dynamics in FL.

TABLE 6. Comparison with existing State-of-the-art.

Research	Method	Accuracy(%)
[16]	FL	83.9
[14]	FL	76.95
[17]	FL	86.82
Proposed	FL	**87.339**
	CFL	**78.396**
	QFL	**84.25**

7) COMPUTATIONAL OVERHEAD

Each client trains a local model using its data and sends the model updates to a central server for aggregation. The computational overhead mainly involves training the local models on each client and the communication overhead between clients and the server. The computational overhead in FL is generally manageable as it scales linearly with the number of clients. In contrast, computational overhead in CFL involves the clustering process, which can be computationally intensive depending on the clustering algorithm and dataset size. The computational overhead in QFL includes the execution of quantum circuits on quantum simulators for each communication round. Quantum circuits can have significant computational requirements; additionally, the optimization process in QFL may require additional computational resources due to the need for quantum gradient calculations or other quantum-specific operations.

8) COMPARISON WITH STATE-OF-THE-ART

On comparing the proposed work with state-of-the-art as shown in Table 6, it is observed that as per the literature survey, HFL for mental healthcare application is a novel work; no such research work has ever presented with recent techniques. Also, it can be observed in Fig.13, the proposed FL 4.09%, 13.50%, and 0.59% performed better than the state-of-the-art approach [14], [16], and [17], respectively. Besides, [16] takes more than 13 hours for total runtime while our proposed method takes 1.76, 4, and 10.30 hours for FL, CFL, and QFL respectively. Consequently, the proposed method is 6.38 times faster than previous work.

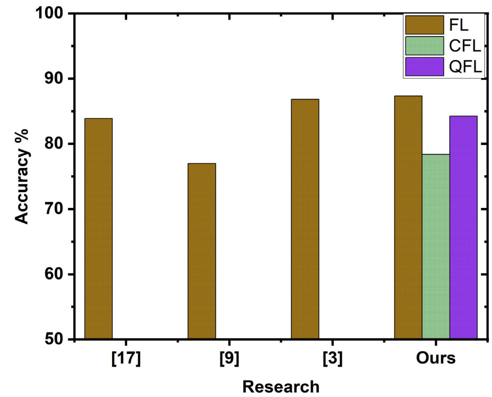


FIGURE 13. Comparison of our research approach with State-of-the-art.

VI. LIMITATION AND FUTURE WORK

The proposed scheme of this paper has some drawbacks that in CFL prediction score can be further improved by incorporating adaptive clustering algorithms, and in QFL, careful circuit designing can improve predictions. Future research should focus on privacy-preserving techniques, robust security measures, adaptive clustering algorithms, and efficient solutions for resource-constrained environments. As a result, SplitNN, blockchain, and even encryption can be included in future work for more reliable and future-ready deployment. Since blockchain enabled FL by removing single points of failure, blockchain technology strengthens the robustness of the training procedure. Secondly, a harmful global model might be mostly rejected because of the blockchain's consensus mechanism. Another improvement can be in communication overhead by sending the compressed parameter to a server, reducing this overhead.

VII. CONCLUSION

This paper proposed two privacy-preserving HFL frameworks for mental healthcare, CFL and QFL. CFL integrates FL with clustering techniques, while QFL combines FL with quantum computing for enhanced privacy and efficiency. These frameworks address various challenges outlined in the problem statement, such as bias and fairness in data, communication efficiency, classification performance, and privacy-preserving capability. This paper demonstrated the proposed framework effectively utilizes W-IoT data to perform real-time stress detection and potential interventions for improved mental well-being. CFL improved the individual's model training efficiency, while QFL ensures secure model transmission and data privacy. Results show CFL achieving 6.24% better precision and QFL shows 18.75% better recall compared to traditional FL method. Additionally, our frameworks outperform state-of-the-art approaches by 4.09%, 13.50%, and 0.59%, respectively, and are 6.38 times faster than previous methods. In summary, our proposed HFL frameworks offer a promising solution to mental healthcare challenges, prioritizing data privacy and model performance.

REFERENCES

[1] M. Asad, A. Moustafa, and T. Ito, "FedOpt: Towards communication efficiency and privacy preservation in federated learning," *Appl. Sci.*, vol. 10, no. 8, p. 2864, Apr. 2020.

[2] T. Li, A. K. Sahu, A. Talwalkar, and V. Smith, "Federated learning: Challenges, methods, and future directions," *IEEE Signal Process. Mag.*, vol. 37, no. 3, pp. 50–60, May 2020.

[3] N. Truong, K. Sun, S. Wang, F. Guitton, and Y. Guo, "Privacy preservation in federated learning: An insightful survey from the GDPR perspective," *Comput. Secur.*, vol. 110, Nov. 2021, Art. no. 102402.

[4] G. A. Reina, A. Gruzdev, P. Foley, O. Perepelkina, M. Sharma, I. Davidyuk, I. Trushkin, M. Radionov, A. Mokrov, D. Agapov, J. Martin, B. Edwards, M. J. Sheller, S. Pati, P. Narayana Moorthy, S.-H. Wang, P. Shah, and S. Bakas, "OpenFL: An open-source framework for federated learning," 2021, *arXiv:2105.06413*.

[5] C. Zhang, Y. Xie, H. Bai, B. Yu, W. Li, and Y. Gao, "A survey on federated learning," *Knowl.-Based Syst.*, vol. 216, Sep. 2021, Art. no. 106775.

[6] J. Du, B. Jiang, C. Jiang, Y. Shi, and Z. Han, "Gradient and channel aware dynamic scheduling for over-the-air computation in federated edge learning systems," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 4, pp. 1035–1050, Apr. 2023.

[7] G. A. Kaissis, M. R. Makowski, D. Ruckert, and R. F. Braren, "Secure, privacy-preserving and federated machine learning in medical imaging," *Nature Mach. Intell.*, vol. 2, no. 6, pp. 305–311, Jun. 2020.

[8] M. K. Maurya, M. Kumar, and M. Kumar, "Sound event detection using federated learning," in *Proc. IEEE 9th Uttar Pradesh Sect. Int. Conf. Electr., Electron. Comput. Eng. (UPCON)*, Dec. 2022, pp. 1–6.

[9] M. Chehimi and W. Saad, "Quantum federated learning with quantum data," in *Proc. ICASSP - IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2022, pp. 8617–8621.

[10] D. Maheshwari, D. Sierra-Sosa, and B. Garcia-Zapirain, "Variational quantum classifier for binary classification: Real vs synthetic dataset," *IEEE Access*, vol. 10, pp. 3705–3715, 2022.

[11] M. K. Maurya, V. K. Singh, S. K. Shaw, and M. Kumar, "FFT-ASVR: An adaptive approach for accurate prediction of IoT data streams," *J. Supercomput.*, vol. 80, no. 10, pp. 13976–13999, Jul. 2024.

[12] C. Briggs, Z. Fan, and P. Andras, "Federated learning with hierarchical clustering of local updates to improve training on non-IID data," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2020, pp. 1–9.

[13] J. H. Yoo, H. M. Son, H. Jeong, E.-H. Jang, A. Y. Kim, H. Y. Yu, H. J. Jeon, and T.-M. Chung, "Personalized federated learning with clustering: Non-IID heart rate variability data application," in *Proc. Int. Conf. Inf. Commun. Technol. Converg. (ICTC)*, Oct. 2021, pp. 1046–1051.

[14] B. Cao, L. Zheng, C. Zhang, P. S. Yu, A. Piscitello, J. Zulueta, O. Ajilore, K. Ryan, and A. D. Leow, "DeepMood: Modeling mobile phone typing dynamics for mood detection," in *Proc. 23rd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2017, pp. 747–755.

[15] S. Bn and S. Abdullah, "Privacy sensitive speech analysis using federated learning to assess depression," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2022, pp. 6272–6276.

[16] F. Firouzi, S. Jiang, K. Chakrabarty, B. Farahani, M. Daneshmand, J. Song, and K. Mankodiya, "Fusion of IoT, AI, edge–fog–cloud, and blockchain: Challenges, solutions, and a case study in healthcare and medicine," *IEEE Internet Things J.*, vol. 10, no. 5, pp. 3686–3705, Mar. 2023.

[17] A. Almadhor, G. A. Sampedro, M. Abisado, S. Abbas, Y.-J. Kim, M. A. Khan, J. Baili, and J.-H. Cha, "Wrist-based electrodermal activity monitoring for stress detection using federated learning," *Sensors*, vol. 23, no. 8, p. 3984, Apr. 2023.

[18] S. Y.-C. Chen and S. Yoo, "Federated quantum machine learning," *Entropy*, vol. 23, no. 4, p. 460, Apr. 2021.

[19] R. Huang, X. Tan, and Q. Xu, "Quantum federated learning with decentralized data," *IEEE J. Sel. Topics Quantum Electron.*, vol. 28, no. 4, pp. 1–10, Jul. 2022.

[20] X. Cheng, W. Tian, F. Shi, M. Zhao, S. Chen, and H. Wang, "A blockchain-empowered cluster-based federated learning model for blade icing estimation on IoT-enabled wind turbine," *IEEE Trans. Ind. Informat.*, vol. 18, no. 12, pp. 9184–9195, Dec. 2022.

[21] A. Ghosh, J. Chung, D. Yin, and K. Ramchandran, "An efficient framework for clustered federated learning," in *Proc. Int. Conf. Neural Inf. Proc. Sys. (NeurIPS)*, 2020, pp. 19586–19597.

[22] X. Sáez-de-Cámara, J. L. Flores, C. Arellano, A. Urbieta, and U. Zurutuza, "Clustered federated learning architecture for network anomaly detection in large scale heterogeneous IoT networks," *Comput. Secur.*, vol. 131, Aug. 2023, Art. no. 103299.

[23] Q. Wu, X. Chen, Z. Zhou, and J. Zhang, "FedHome: Cloud-edge based personalized federated learning for in-home health monitoring," *IEEE Trans. Mobile Comput.*, vol. 21, no. 8, pp. 2818–2832, Aug. 2022.

[24] A. Gupta, A. Raut, R. Yadav, M. Kumar, and V. K. Chaurasiya, "A hybrid approach based stress monitoring system for office environment using IoT," in *Proc. IEEE 19th India Council Int. Conf.*, Nov. 2022, pp. 1–6.

[25] C. Maurya, A. Gupta, and V. K. Chaurasiya, "Efficient mutual authentication scheme for IoT-based healthcare system," in *Proc. IEEE 6th Conf. Inf. Commun. Technol. (CICT)*, Nov. 2022, pp. 1–5.

[26] A. Gupta and V. K. Chaurasiya, "Efficient task-offloading in IoT-fog based health monitoring system," in *Proc. OITS Int. Conf. Inf. Technol. (OCIT)*, Dec. 2022, pp. 495–500.

[27] Y. Yan, X. Tong, and S. Wang, "Clustered federated learning in heterogeneous environment," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 1, no. 1, pp. 1–14, May 2023.

[28] F. Sattler, K.-R. Müller, and W. Samek, "Clustered federated learning: Model-agnostic distributed multitask optimization under privacy constraints," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 8, pp. 3710–3722, Aug. 2021.

[29] D. Sierra-Sosa, M. Telahun, and A. Elmaghraby, "TensorFlow quantum: Impacts of quantum state preparation on quantum machine learning performance," *IEEE Access*, vol. 8, pp. 215246–215255, 2020.

[30] D. Sierra-Sosa, J. D. Arcila-Moreno, B. Garcia-Zapirain, and A. Elmaghraby, "Diabetes type 2: Poincare data preprocessing for quantum machine learning," *Comput., Mater. Continua*, vol. 67, no. 2, pp. 1849–1861, 2021.

[31] H. Gupta, H. Varshney, T. K. Sharma, N. Pachauri, and O. P. Verma, "Comparative performance analysis of quantum machine learning with deep learning for diabetes prediction," *Complex Intell. Syst.*, vol. 8, no. 4, pp. 3073–3087, Aug. 2022.

[32] M. Kumar, P. K. Singh, M. K. Maurya, and A. Shivhare, "A survey on event detection approaches for sensor based IoT," *Internet Things*, vol. 22, Jul. 2023, Art. no. 100720.

[33] A. Gupta and V. Kumar Chaurasiya, "Reinforcement learning based energy management in wireless body area network: A survey," in *Proc. IEEE Conf. Inf. Commun. Technol.*, Dec. 2019, pp. 1–6.

[34] M. I. Prasetiyowati, N. U. Maulidevi, and K. Surendro, "Determining threshold value on information gain feature selection to increase speed and prediction accuracy of random forest," *J. Big Data*, vol. 8, no. 1, p. 84, Dec. 2021.

[35] A. Gupta, C. Maurya, K. Dhere, and V. K. Chaurasiya, "Wellness detection using clustered federated learning," in *Proc. IEEE 6th Conf. Inf. Commun. Technol. (CICT)*, Nov. 2022, pp. 1–5.

[36] Z. Abohashima, M. Elhosen, E. H. Houssein, and W. M. Mohamed, "Classification with quantum machine learning: A survey," 2020, *arXiv:2006.12270*.

[37] B. Roy, "All about data encoding for quantum machine learning," DataDrivenInvestor, 2021. Accessed: Sep. 29, 2024. [Online]. Available: https://medium.datadriveninvestor.com/all-about-data-encoding-for-quantum-machine-learning-2a7344b1dfef

[38] PennyLane. (Aug. 2022). *How to Embed Data Into a Quantum State*. Accessed: Sep. 29, 2024. [Online]. Available: https://pennylane.ai/blog/2022/08/how-to-embed-data-into-a-quantum-state/

[39] S. Thanasilp, S. Wang, N. A. Nghiem, P. Coles, and M. Cerezo, "Subtleties in the trainability of quantum machine learning models," *Quantum Mach. Intell.*, vol. 5, no. 1, p. 21, Jun. 2023.

[40] P. Schmidt, A. Reiss, R. Duerichen, C. Marberger, and K. Van Laerhoven, ''Introducing WESAD, a multimodal dataset for wearable stress and affect detection,'' in *Proc. 20th ACM Int. Conf. Multimodal Interact.*, Oct. 2018, pp. 400–408.

[41] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. Y. Arcas, ''Communication-efficient learning of deep networks from decentralized data,'' *Proc. Artif. Intell. Statist.*, vol. 1, pp. 1273–1282, Jun. 2017.

[42] S. A. Alvarez, ''An exact analytical relation among recall, precision, and classification accuracy in information retrieval,'' Boston College, Boston, MA, USA, Tech. Rep. BCCS-02-01, 2002, pp. 1–22.

ARTI GUPTA received the M.Tech. degree in information technology (wireless communication engineering) from Indian Institute of Information Technology, Allahabad, where she is currently pursuing the Ph.D. degree with the Department of Information Technology. Her research interests include healthcare, machine learning, quantum machine learning, WBAN, the Internet of Things, and sensor networks.

MANISH KUMAR MAURYA received the M.Tech. degree in IT (wireless communication engineering) from IIIT-Allahabad, India. He is currently working as a Research Scholar (Ph.D.) with the Department of IT, IIIT-Allahabad. His research interests include the IoT, sensor networks, federated learning, and data aggregation.

KHYATI DHERE received the Bachelor of Technology degree in electronics and communication engineering from Bhilai Institute of Technology, Durg. She is currently pursuing the M.Tech. degree with the Department of Information Technology (Cyber Law and Information Security), Indian Institute of Information Technology, Allahabad. Her research interests include machine learning, image processing, quantum machine learning, Raspberry Pi and Arduino-based the IoT circuit designing, deep neural networks, and natural language processing.

VIJAY KUMAR CHAURASIYA (Senior Member, IEEE) received the master's degree in wireless communication and computing and the Ph.D. degree in information technology from Indian Institute of Information Technology, Allahabad, Prayagraj, in 2004 and 2010, respectively. He is currently working as an Associate Professor with the Department of Information Technology, Indian Institute of Information Technology, Allahabad. His current research interests include resource utilization in the IoT networks, wireless sensor networks, routing protocols, network security, load balancing, data backup, and reliability and information processing and security.