

Depression Detection Through Smartphone Sensing: A Federated Learning Approach

<https://doi.org/10.3991/ijim.v17i01.35131>

Nawrin Tabassum^(✉), Mustofa Ahmed, Nushrat Jahan Shorna,
MD Mejbah Ur Rahman Sowad, H M Zabir Haque
Ahsanullah University of Science and Technology, Dhaka, Bangladesh
nawrintabassum14@gmail.com

Abstract—Depression is one of the most common mental health disorders which affects thousands of lives worldwide. The variation of depressive symptoms among individuals makes it difficult to detect and diagnose early. Moreover, the diagnosing procedure relies heavily on human intervention, making it prone to mistakes. Previous research shows that smartphone sensor data correlates to the users' mental conditions. By applying machine learning algorithms to sensor data, the mental health status of a person can be predicted. However, traditional machine learning faces privacy challenges as it involves gathering patient data for training. Newly, federated learning has emerged as an effective solution for addressing the privacy issues of classical machine learning. In this study, we apply federated learning to predict depression severity using smartphone sensing capabilities. We develop a deep neural network model and measure its performance in centralized and federated learning settings. The results are quite promising, which validates the potential of federated learning as an alternative to traditional machine learning, with the added benefit of data privacy.

Keywords—depression prediction, federated learning, mHealth, smartphone sensors, data security

1 Introduction

Depression is a mental health illness that adversely affects how people think, feel, and behave. It causes them feelings of sadness and loss of interest in their daily activities. Depression also leads to various emotional and physical difficulties. It decreases a person's ability to function well at work [1] and at home [2]. These effects of depression lead to high societal [3] and economic burdens [4]. There was a 25% increase in depression and anxiety worldwide in the first year of the COVID-19 pandemic [5]. In 2018, the economic burden of depression for adults in the United States alone was \$US 326.2 billion [6]. Even though depression is not curable [7], various treatments can minimize its adverse impacts. A recent study found that nearly 80% of people suffering from depression eventually respond well to treatment [8]. Early detection of depression followed by treatment also allows for a better prognosis [9].

A possible solution for the timely detection of depression can be achieved through smartphones. Smartphones have various built-in sensors that allow us to collect and interpret information about the users' social interactions, physical movements, and daily activities [10], [11]. A study revealed that an average user spends 145 minutes on their smartphones daily [12]. Considering the time a user spends on the smartphone combined with the ability of the smartphone to collect information, it is possible to detect depression severity in a person. As stated in [13], there had been a 79% increase in smartphone users in 2022 compared to 2016. This finding proves the scalability of using smartphones to detect depression severity in a person.

Traditional methods of diagnosing depression require filling out relevant questionnaires, face-to-face interviews, and more, which are prone to human errors. An example is when a doctor incorrectly judges a patient's mental state, as the symptoms shown by the patient are often inconsistent. Another source of error is that the patient fails to precisely recall how they have felt over a long period during self-reporting [14]. Apart from these scopes of errors, a considerable constraint that entails the traditional method is that it heavily relies on the people to initiate the treatment. However, a person undergoing depression feels hesitant to do so because they prefer being isolated during that phase. So, the traditional methods of diagnosing depression are highly unreliable.

Numerous research studies [15] – [20] successfully detected depression severity in a person using data collected through smartphones. However, apart from [17], these studies have not addressed the users' data privacy concerns. These studies used traditional machine learning approaches, which centralize users' private data. Building a good machine learning model depends on the amount and quality of data fed to the model. As our study is related to a person's mental health, it is of utmost importance to train a model using rich and diverse data that will accurately predict depression severity in a person. However, users' preference for sharing their private data for mHealth purposes varies depending on the collected data and the benefit they are receiving [21]. So, data privacy concerns hinder building a suitable model in practical cases. Hence, we need to find a strategy through which we can train a machine learning model without making the users concerned regarding their data privacy.

To overcome the data privacy gap, we used federated learning. Federated Learning (FL) is a privacy-preserving machine learning technique [22] that allows us to train a machine learning model without passing any raw data of the users to the central server. As the users' private data is never uploaded and can not be seen by the server, it resolves the data privacy concerns of the users. Another aim of our study is to compare the performance of a model that was trained using federated learning and centralized machine learning approaches. Although the authors at [17] have used federated learning on mobile data to predict depression severity, they only performed a simulation of federated learning using the TensorFlow Federated framework (TFF) [23]. The primary aim of our study is to fill this gap by showing a working demonstration of federated learning for mHealth purposes. We have developed an android application that will predict depression severity in a person using federated learning.

In our study, we determined the depression severity in a person by using 9-question Patient Health Questionnaire (PHQ-9) [24], which we considered to be the ground

truth. The PHQ-9 is a diagnostic tool used to identify the presence and severity of depression in a person. A total of 145 individuals participated in our study and were asked to complete the PHQ-9 questionnaire. The participants installed an android application through which passive sensor data were collected. Data were collected at regular intervals and stored in a real-time database. Federated Learning is an approach where we do not need to centralize the users' data. However, in our case, we centralized the users' data which allowed us to pre-train the global model before deploying it to the android smartphones. The federated learning application was developed using the Deep Learning for Java (DL4J) framework [25]. Our implementation is unlike any other existing studies, as they were limited to only performing a simulation of federated learning using the TFF.

The remaining paper is structured as follows: Section 2 contains the background study and relevant works. Section 3 gives an overview of data collection and data preprocessing. Section 4 contains the methodology of our study. Section 5 and Section 6 comprise the results and discussion respectively. Lastly, Section 7 concludes the paper.

2 Background

This section describes the core tools and approaches of our research. We further present a discussion on relevant studies and their shortcomings.

2.1 Artificial intelligence

The replication of human intelligence by machines, particularly computer systems, is referred to as artificial intelligence (AI). To replicate intelligent human behavior by automated means, researchers studying AI set out to understand the formal processes that went into games like chess, language processing, and medical diagnosis [26]. Three cognitive abilities – learning, reasoning, and self-correction are the main topics of AI programming [27]. A wide range of distinct sorts of technologies includes AI, such as machine learning (ML) – the technology of getting a computer to act without programming.

2.2 Machine learning

Machine Learning (ML) works by using data and algorithms to simulate how humans learn and gradually increasing the system's accuracy in identifying a pattern in data. ML has many fundamental strategies, and we have used supervised learning in our study. Developing algorithms for supervised machine learning involves taking samples from the outside world to create general patterns and hypotheses that can be used to anticipate how future samples will turn out [28]. Building a model for allocating class labels in respect of predictor features is the objective of supervised learning. If the predictor features are known and the value of the class label is unknown, the model can be used to predict class labels for the corresponding data [29].

2.3 Federated learning

Federated learning (FL) is a new field of study in AI for learning on decentralized data and is a privacy-preserving distributed ML paradigm. Decentralized algorithms were initially conceived to compute the mean of data collected from numerous edge devices [30]. FL provides a privacy-preserving mechanism to effectively leverage those decentralized computing resources inside end devices to train ML models. In FL, a central server connects with numerous clients such as smartphones, smart watches, IoT devices, etc. Globally, there are billions of these devices, and their combined computing power is much more than that of a sizable data center. As the availability of smartphones increases, the potential for FL increases as well. FL is a remarkably different ML approach that avoids collecting data in a centralized server. In traditional machine learning or deep learning (DL) pipelines, data are collected from numerous sources and kept in a central location, like data center. The traditional ML and DL models are trained from all the acquired data that are private and sensitive to the users. On the other hand, the users' sensitive data is safe in an FL setting. FL allows us to train an ML model using the users' data without disclosing any personal information. FL trains machine learning models using local data that are present in local nodes without explicitly exchanging the data samples.

Considering a standard FL network has three nodes and a single server. Here nodes are the smartphone devices having data and computational resources. Firstly, the server initializes a global model and sends it to the nodes. In some cases, the global model can be pre-trained using centralized data instead of random initialization. This global model is trained locally using each node's private data, and only the local update of the model is sent to the central server. Finally, the central server aggregates the updated model weights received from the nodes to generate a better model with high accuracy. These steps repeat and return better results with each iteration [31]. Depression detection requires dealing with the privacy-sensitive data of users. For this reason, we chose the federated learning approach that maintains users' privacy. This approach also cuts server and computational costs by transmitting only model parameters to the central server. A visual representation of federated learning is shown in Figure 1.

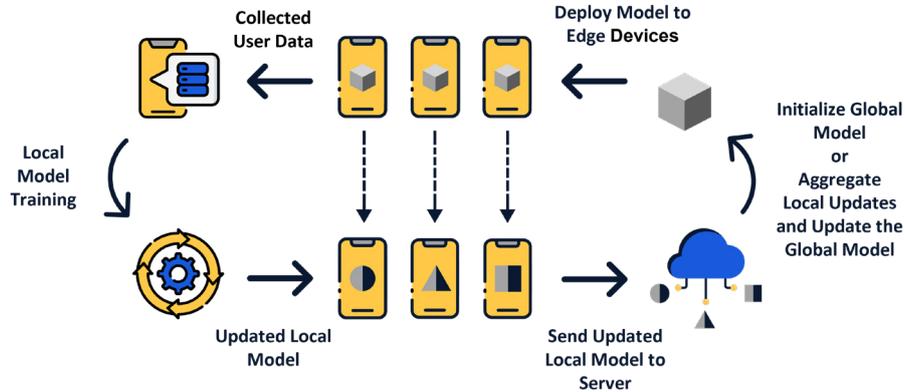


Fig. 1. Working flow of federated learning

2.4 Smartphone sensors, mHealth and PHQ-9

Android smartphones come with built-in sensors [32] that measure different useful parameters of a phone. For example, the accelerometer and the gyroscope can measure the smartphone's acceleration and angular rotational velocity, respectively. The gravity sensor determines the device's axes by interpreting the accelerometer and gyroscope data. Similarly, the magnetometer can detect the strength of the magnetic field. In our study, we used these four sensors to collect the data. Apart from these sensors, we also kept track of the battery level of the smartphones. The World Health Organization's Global Observatory for eHealth (GOe) addresses mHealth (mobile health) as a medical and public health practice assisted by mobile devices [33]. mHealth apps provide increased access to healthcare services and enhanced interaction with medical experts. The Patient Health Questionnaire (PHQ-9) is a measure used to test the presence and severity of depression and assess treatment response. The questions concern feeling down or depressed, sleeping difficulties, eating habits, self-perception, suicidal thoughts, etc. Each response group receives a score of 0, 1, 2, or 3 [34]. The overall score is computed by summing the results of the nine questions. Anyone scoring above the threshold on this scale should consult a doctor or a mental health practitioner [35].

2.5 Related works

There exist various successful studies that have used the data collected from smartphones to detect the presence of depression in a person. Using the data collected from smartphones, the authors in [15] found that students who slept less and engaged in fewer and shorter conversations were more likely to be depressed. As stated in [36], the presence of depression is associated with low performance in language lessons. The authors at [37] used a different approach and found correlations similar to [15]. Instead of complete passive sensing using smartphones, they used active input

from users. This study also reported a strong correlation between depression severity in a person and voice diary sentiment. Although most studies used various sensors and required frequent user interaction, [19] and [20] used only GPS sensor data to predict depression severity in a person. Using the GPS sensor data and time, they derived numerous features like location variance, entropy, number of places visited by the user, and more. Data analysis of these derived features revealed some distinctive habits of people with different depression severity. Due to the rising concern of data privacy, the authors at [16] used less privacy-sensitive data like total daily mobile usage duration, the number of calls received, and more to predict depression severity in a person. Furthermore, this study revealed that including privacy-sensitive data like gender and age of the users' resulted in a better-performing machine learning model. This finding shows the importance of including privacy-sensitive data for mHealth purposes.

Many studies on areas other than mHealth have used the federated learning approach. For example, the authors have used federated learning on smartphone data to perform human activity recognition in [38]. The authors at [39] have applied federated learning to image and audio data collected using webcams and microphones, respectively, and predicted human emotion. Federated learning is also widely explored in the medical field. Data collected from patients are highly confidential, and in no circumstances can they be shared. However, to develop a good machine learning model, it is essential to train the model using data from various institutions. Federated learning makes it possible without centralizing the patients' privacy-sensitive data.

Despite federated learning being used widely in other areas, it is primarily untapped in mHealth. Apart from [17] and [40], very few studies on mHealth have used this privacy-preserving approach. Therefore, in our study, we have used the federated learning approach to fill the data privacy gap in the existing studies. Also, the mentioned studies are limited to using the TensorFlow Federated framework to perform only a federated learning simulation. These studies do not show any working demonstration or implementation of federated learning. So, it is not feasible to anticipate the performance of federated learning in practical cases from the findings of these studies. So, in our study, we have developed a federated learning application for android smartphones to predict the severity of depression in a person.

3 Dataset

This section contains a detailed description of how our data was collected and pre-processed.

3.1 Data collection

We employed an android application developed by the authors at [41] to collect data. A group of students from the Computer Science and Engineering department at Ahsanullah University of Science and Technology was recruited for data collection. The interested volunteers were provided with a Google Form link that contained the

PHQ-9 questionnaire and a Google Drive link to download the app. Sensor data was collected from their smartphones for two weeks. The app was running in the device's background, and readings from the accelerometer, gyroscope, gravity, and magnetometer sensors were taken once every five minutes. Then the data was uploaded and stored in a real-time database called Firebase. A total of 205 participants filled in the PHQ-9 questionnaire, of which only 145 downloaded the app. However, we accumulated sufficient data from only 80 volunteers. Around 90% of them were aged between 21-25. Amongst them, 66.25% of the participants were male, and the remaining 33.75% were female. So, the collected data was not from a diverse group. As shown in Figure 2, the dataset was also heavily imbalanced.

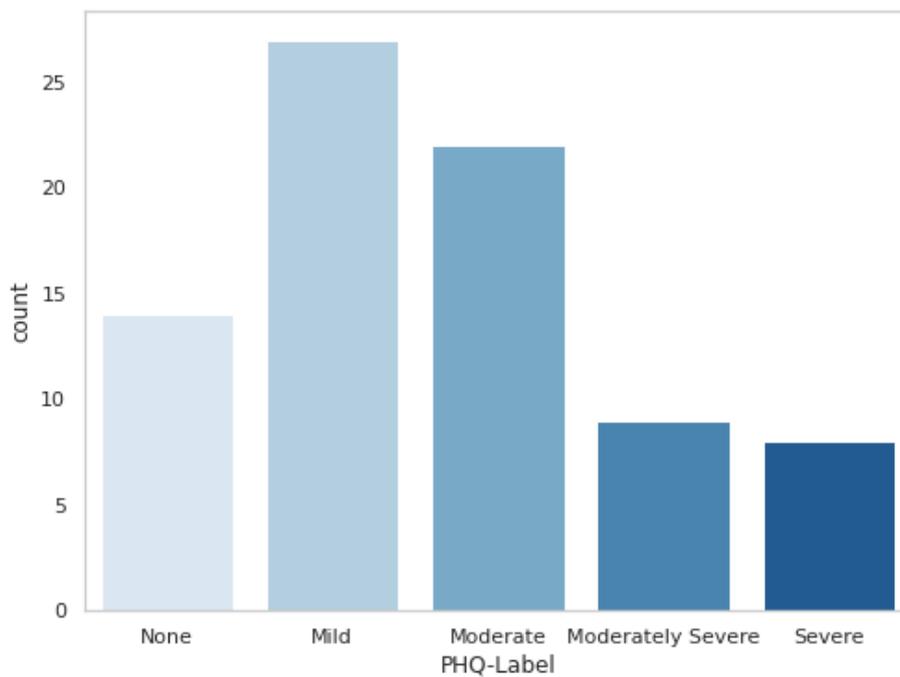


Fig. 2. Data distribution of PHQ-label

3.2 Data preprocessing

The data collected from the participants were stored in Firebase in JSON format. The data was then parsed into excel format with custom python scripts. As part of the data cleaning steps, we first removed the samples that contained invalid or too many missing values. We also considered a sample invalid if all the values from a particular sensor were 0. For example, we excluded the sample if the gravity sensor had a value of 0 on the x, y, and z-axis. In the next step, we standardized all the numeric features.

4 Methodology

In this section, we describe the methodology in detail. We further describe the working procedure of our federated learning app.

4.1 Fully connected neural network

A neural network has multiple hidden layers between input and output layers. Each layer is comprised of neurons connected to all the neurons in the next layer through a weighted link [42]. Using the input vector, each of these neurons applies a linear transformation through a weight matrix. This output then undergoes a non-linear transformation known as the activation function. The activation function determines whether the neuron should be activated or not. These steps are repeated for each neuron, and this whole process is known as forward propagation. During the training period of the neural network, the actual output is known, which helps the neural network to learn from the incorrect predictions it makes. The model's performance is evaluated based on the value calculated by the loss function. The loss function will show a higher number if the predictions are completely off. The backpropagation process then calculates the gradients for all the weights with respect to the loss function. Afterward, the optimizer uses the calculated gradient to adjust the neurons' weights. The amount of weight update is controlled by using a parameter known as the learning rate. If the value of the loss function diminishes as the neural network trains, it implies that the model's performance is improving. The neural network is trained using the same data many times. It is considered one epoch when the neural network is trained using the entire training data for one time. Within an epoch, the batch size is the number of training data samples the neural network trains on before the model weights are updated.

In our centralized neural network, we have used two hidden layers. The layers consisted of 500 and 1000 neurons, respectively. We chose *Softmax* and *Relu* activation functions for the output and hidden layers, respectively. Negative log-likelihood was used for calculating the loss. Finally, for adjusting the neural network weights, we chose the SGD optimizer, and the learning rate was set to 0.001. We trained the model for a total of 20 epochs, and the batch size we used was 8.

4.2 Transfer learning

The computation of a high number of parameters is necessary for training a neural network with several layers, which raises computational expense and energy use. Again, a significant quantity of data is required for training a large neural network. Sensor data collected in real-time for on-device training is quite limited and can not contribute significantly to training a complete deep model. The notion of transfer learning can be utilized to resolve this issue. Transfer learning uses a pre-trained model that has been trained on a big dataset, eliminating the need to train the model from scratch [43]. The pre-trained model's first few layers are fixed, and the remaining layers are fine-tuned for the specific task with the limited training data. So, we

trained a model using the data we acquired and used this model as the global model on the edge devices instead of applying random initial weights. Then local training is performed, and parameters are updated only for the last layer of the model. Transfer learning is ideal for mobile scenarios since it allows for the reduction of the training cost, an improvement in training speed, and a decrease in energy consumption.

4.3 Federated learning app

The application was built using Android Studio version 2021.2.1. To develop this app, we considered some frameworks such as TensorFlow Federated (TFF) [23], PySyft [44], and Deep Learning for Java (DL4J) [25]. Most of these frameworks lack the components required for the actual implementation of FL. In TFF and PySyft, data must be distributed from a central location rather than being directly collected from the edge devices for training. So, they mainly focus on performing a simulation and do not provide any client-server communication environment. In this study, our aim was to utilize real sensor data from smartphones and execute a client-server implementation of FL. So, we chose Deep Learning for Java (DL4J) as the framework for developing our android application (Figure 3). DL4J enables us to create and adjust a wide range of basic and complicated deep learning networks, as well as execute transfer learning in mobile devices. It comes with ND4J, a linear algebra library that simplifies mathematical and deep learning operations. In our work, we have employed version 1.0.0-beta4 of DL4J.

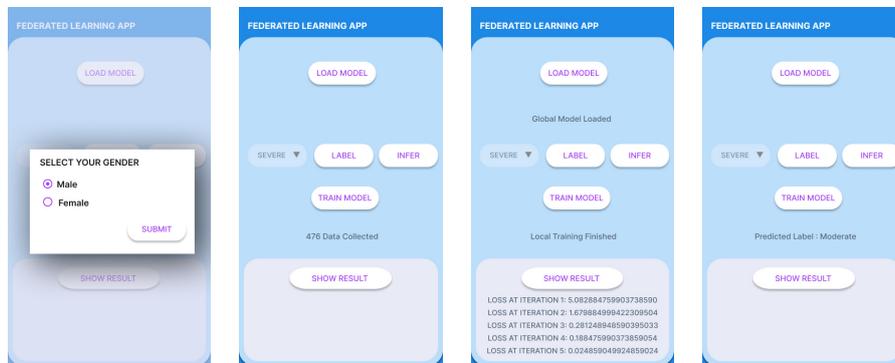


Fig. 3. Snapshot of our federated learning application

We selected five participants to train the local model on their devices. They received the app through a Google Drive link and installed it on their smartphones. To run the application, Android version 8.0 or higher is required. The app collects data from smartphone sensors and determines the depression severity of a user based on the collected data. Before the local training is carried out, the global model is used to make predictions. When the local training is completed, the prediction is performed based on the local model, providing the user with a more personalized result. Through the user's initiative to label the data, a new training dataset is stored in the device, and

then local training is performed. A batch size of 8 and an SGD optimizer with a learning rate of 0.001 are used as the hyperparameters. Local training is limited to one epoch to minimize the computational load on edge devices. To ensure data security, the local dataset is automatically deleted from the device after local training.

The communication method is implemented using Firebase. We have used Firebase cloud storage version 19.2.0 and Firebase real-time database version 19.4.0. The global weights saved in the Firebase are downloaded to the edge devices for local training. After the local training, the updated weights are uploaded back to Firebase. Then, FedAVG [45] is performed using Java to aggregate the local updates. FedAVG is an optimization technique that computes the average value of the local weights obtained from clients. Finally, the aggregated model is sent back to Firebase as the new global model. The communications are carried out for five rounds in this study, with five clients participating in each round. The complete client-server communication process is shown in Figure 4.

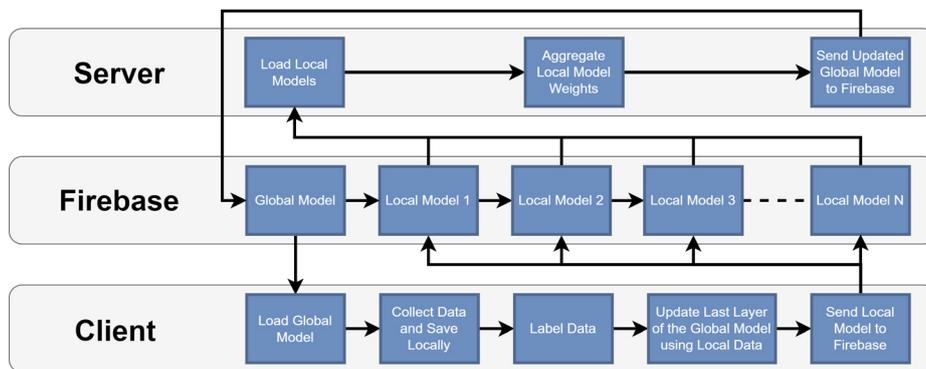


Fig. 4. Client-server communication architecture

5 Results

The findings of our experiment are described in this section. Before federated training, a centralized neural network was trained based on smartphone sensor data, and then prediction was performed on the test set. 20% of the data was utilized for testing, while 80% was used for training. To evaluate the model, we used accuracy, precision, recall, and F1-score as the performance metrics. Accuracy can be defined as how frequently the classifier makes the correct prediction. The ratio of samples that are accurately classified as positive to all the samples that are classified or misclassified as positive is known as precision. On the other hand, recall is the proportion of correctly categorized positive samples to all positive samples. F1-score is the harmonic mean of recall and precision. The model's performance was measured by accuracy, precision, recall, and F1-score using the test data after 20 epochs of training. The accuracy of the model was 0.68, and the precision, recall, and F1-score were 0.70, 0.48, and 0.52, respectively.

Local training was performed on five android devices through the app. Since smartphone sensors generate data at a relatively high frequency, gathering training data of a constant size is challenging. Hence, the training data size was different for the clients in our experiment. Each client used the same label in their respective training rounds to represent that each client had a particular level of depression severity. The performance of FL was evaluated using our test data. The accuracy, precision, recall, and F1 score of the global model were 0.65, 0.69, 0.46, and 0.47, respectively, after five training rounds. Table 1 presents the states of the global model after each round. FL showed some decline in performance due to the imbalanced and skewed dataset. However, the performance of FL was almost the same as that of centralized machine learning. The confusion matrixes of centralized machine learning and federated learning are presented in Table 2 and Table 3.

Table 1. Performance of the global model after each round

Performance Metrics	Round1	Round2	Round3	Round4	Round5
Accuracy	0.6854	0.6792	0.6661	0.6576	0.6517
Precision	0.6899	0.7146	0.7224	0.7110	0.6978
Recall	0.4844	0.4821	0.4793	0.4731	0.4665
F1-Score	0.5215	0.5139	0.5056	0.4914	0.4759

Table 2. Confusion matrix for the global model before federated training

		Predicted				
		None	Mild	Moderate	Moderately Severe	Severe
Actual	None	413	152	136	0	1
	Mild	55	1258	257	2	0
	Moderate	75	374	946	7	3
	Moderately Severe	7	58	23	33	0
	Severe	8	43	29	0	7

Table 3. Confusion matrix for the global model after federated training

		Predicted				
		None	Mild	Moderate	Moderately Severe	Severe
Actual	None	541	31	128	0	2
	Mild	218	894	460	0	0
	Moderate	192	136	1073	1	3
	Moderately Severe	43	28	32	18	0
	Severe	22	20	38	0	7

6 Discussion

6.1 Challenges of data collection

Data centralization is not required in real federated learning as model training takes place in edge devices. However, we gathered data from research participants for training a centralized deep learning model. Around 44% population dropped out of the data collection phase. They might have stopped contributing to the study out of concern for their data privacy. Again, the sensor data collector app needed to run continuously in the mobile background for data collection. Some participants might have forgotten to run the app, resulting in a high participant drop rate. We also asked the participants to fill up the PHQ-9 to acquire the ground truth of our study. Although the resulting score from this survey is self-reported, it is clinically tested and widely used for depression monitoring. It should be noted that instead of measuring the presence of depressive symptoms continuously, we assessed it once using PHQ-9. It is unlikely that the symptoms would change radically over the two weeks of the study.

6.2 Data representation

The sensor data derived from smartphones correspond to the motion and position of a person. Previous studies have shown that smartphone sensor data has a high potential for correctly predicting depression severity [15], [20]. Another key factor in determining the degree of depression is demographic data, such as age and gender [16], [18]. We collected the participants' age and gender information during the data collection phase. However, only gender was used as the demographic factor in the current study. Most of the participants were between the ages of 21 and 25, indicating that the data was not truly representative of all age groups. So, we decided to exclude the age of the participants from this study. Nonetheless, the results are positive, suggesting that combining only gender information with sensor data can assist in diagnosing depression. Only android users were the subject of our study, given that sensor data was collected from android smartphones, and an android app was developed to implement federated learning.

6.3 Advantages of pre-trained model

The centrally trained model was used as a pre-trained model for on-device training with local data. Federated learning involves refining a global model with local data to update the global model without sending data to the server. A pre-trained model eliminates the necessity of starting the training from scratch and results in fast convergence. In our experiment, we fixed the neural network's first three layers and only trained the last layer in the edge devices. Using the knowledge gained from the pre-trained model, the local training was performed in less time with reduced battery consumption.

6.4 Effectiveness of FL to detect depression

Our main goal was to test the usability of the federated learning technique to predict the severity of depression. We achieved an accuracy of 65% using this technique. We also compared FL with traditional ML and showed that we can preserve user data privacy by compromising accuracy to an acceptable extent. Another objective of our study was to demonstrate the feasibility of using this approach so that it might be used in other domains where data privacy is a concern. Several past research studies performed depression detection using mobile data, but barely any of them solved the data privacy problem. Our study resolved this issue and obtained a well-performing federated model.

6.5 Prospects of the app

We have developed the application only for demonstration purposes of federated learning. Currently, it needs manual labeling of depression severity. The app is developed as a prototype, so it requires proper updates before launching publicly i.e., the interface needs more development to make it user-friendly. This app can be integrated into different health apps that take user movement data to detect various health issues. This can also open huge opportunities to make the application commercial and apply federated learning in other health areas. Depression is a very sensitive matter, so it is suggested to take a clinical test before coming to any conclusion. Still, the result of the app can be considered for primary detection.

6.6 Limitations

A significant limitation of our work is that we did not create any server to aggregate local updates from users. Instead, we used a real-time database to simulate the communication method. Another constraint includes not being able to collect data continuously from the passive sensors, as we faced limitations in resources. So, we took readings from the sensors once every five minutes. If the collected data were continuous, we could have applied Convolutional Neural Network (CNN) or Long Short-Term Memory (LSTM) to detect mobility patterns of the participants. CNN and LSTM are excellent at detecting patterns in time series data. If the data were continuous, it would open the possibility of developing a pre-trained model that could perform better in detecting depression severity. Another limitation of our study is that we could not find a correlation between raw sensor data and depression severity. Previous studies show it is possible to find a correlation between derived sensor data [19], [20] and depression severity. Due to resource constraints, we were only able to demonstrate a working prototype using sensor data.

7 Conclusion

The societal and economic burden of depression is increasing day by day. Early detection and follow-up by treatment can drastically reduce these impacts of depression. Previous studies have shown how smartphones can be a feasible option for detecting the presence of depression in a person. However, these existing studies had to centralize users' privacy-sensitive data, which raised concern among users that their private data might get exploited. So, in practical scenarios, users might not be willing to use methods that access their private data directly. Considering this data privacy concern, we developed an android application that predicts depression severity in a person using the federated learning approach. In other words, this application allows us to train a machine learning model that detects the severity of depression in a person without centralizing users' privacy-sensitive data. A natural progression of our study would be to experiment with other smartphone data like call duration, application usage, text messages, etc., to identify the presence of depression. Our android application can also be extended for other mHealth purposes like stress or anxiety detection.

8 References

- [1] K.-O. Park, M. G. Wilson, and M. S. Lee, "Effects of social support at work on depression and organizational productivity," *American Journal of Health Behavior*, vol. 28, no. 5, pp. 444–455, 2004. <https://doi.org/10.5993/AJHB.28.5.7>
- [2] P.-W. Wang *et al.*, "Association between problematic cellular phone use and suicide: The moderating effect of family function and depression," *Comprehensive Psychiatry*, vol. 55, no. 2, pp. 342–348, 2014. <https://doi.org/10.1016/j.comppsy.2013.09.006>
- [3] K. B. Wells *et al.*, "Overcoming barriers to reducing the burden of affective disorders," *Biological Psychiatry*, vol. 52, no. 6, pp. 655–675, 2002. [https://doi.org/10.1016/S0006-3223\(02\)01403-8](https://doi.org/10.1016/S0006-3223(02)01403-8)
- [4] P. S. Wang, G. Simon, and R. C. Kessler, "The economic burden of depression and the cost-effectiveness of treatment," *International Journal of Methods in Psychiatric Research*, vol. 12, no. 1, pp. 22–33, 2003. <https://doi.org/10.1002/mpr.139>
- [5] "COVID-19 pandemic triggers 25% increase in prevalence of anxiety and depression worldwide," *Who.int*. [Online]. Available: <https://www.who.int/news/item/02-03-2022-covid-19-pandemic-triggers-25-percent-increase-in-prevalence-of-anxiety-and-depression-worldwide> [Accessed: 15-Aug-2022].
- [6] P. E. Greenberg *et al.*, "The economic burden of adults with major depressive disorder in the United States (2010 and 2018)," *Pharmacoeconomics*, vol. 39, no. 6, pp. 653–665, 2021. <https://doi.org/10.1007/s40273-021-01019-4>
- [7] S. G. Hess *et al.*, "A survey of adolescents' knowledge about depression," *Archives of Psychiatric Nursing*, vol. 18, no. 6, pp. 228–234, 2004. <https://doi.org/10.1016/j.apnu.2004.09.005>
- [8] M. Erickson, "Experimental depression treatment is nearly 80% effective in controlled study," *News Center*. [Online]. Available: <https://med.stanford.edu/news/all-news/2021/10/depression-treatment.html> [Accessed: 15-Aug-2022].
- [9] S. M. Guilfoyle, S. Monahan, C. Wesolowski, and A. C. Modi, "Depression screening in pediatric epilepsy: evidence for the benefit of a behavioral medicine service in early detec-

- tion,” *Epilepsy & Behavior*, vol. 44, pp. 5–10, 2015. <https://doi.org/10.1016/j.yebeh.2014.12.021>
- [10] G. M. Harari, S. R. Müller, M. S. H. Aung, and P. J. Rentfrow, “Smartphone sensing methods for studying behavior in everyday life,” *Current Opinion in Behavioral Sciences*, vol. 18, pp. 83–90, 2017. <https://doi.org/10.1016/j.cobeha.2017.07.018>
- [11] A. Alam, A. Das, Tasjid, and A. Al Marouf, “Leveraging Sensor Fusion and Sensor-Body Position for Activity Recognition for Wearable Mobile Technologies,” *International Journal of Interactive Mobile Technologies*, vol. 15, no. 17, pp. 141-155, 2021. <https://doi.org/10.3991/ijim.v15i17.25197>
- [12] P. Nelson, “We touch our phones 2,617 times a day, says study,” *Network World*, 07-Jul-2016. [Online]. Available: <https://www.networkworld.com/article/3092446/we-touch-our-phones-2617-times-a-day-says-study.html> [Accessed: 01-Sep-2022].
- [13] “Smartphone subscriptions worldwide 2027,” *Statista*. [Online]. Available: <https://www.statista.com/statistics/330695/number-of-smartphone-users-worldwide/> [Accessed: 23-Aug-2022].
- [14] J. E. Wells and L. J. Horwood, “How accurate is recall of key symptoms of depression? A comparison of recall and longitudinal reports,” *Psychological Medicine*, vol. 34, no. 6, pp. 1001–1011, 2004. <https://doi.org/10.1017/S0033291703001843>
- [15] R. Wang *et al.*, “StudentLife: assessing mental health, academic performance and behavioral trends of college students using smartphones,” *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pp. 3–14, 2014. <https://doi.org/10.1145/2632048.2632054>
- [16] R. Razavi, A. Gharipour, and M. Gharipour, “Depression screening using mobile phone usage metadata: a machine learning approach,” *Journal of the American Medical Informatics Association*, vol. 27, no. 4, pp. 522–530, 2020. <https://doi.org/10.1093/jamia/ocz221>
- [17] X. Xu *et al.*, “Privacy-Preserving Federated Depression Detection From Multisource Mobile Health Data,” *IEEE Transactions on Industrial Informatics*, vol. 18, no. 7, pp. 4788–4797, 2021. <https://doi.org/10.1109/TII.2021.3113708>
- [18] S. Thomée, A. Härenstam, and M. Hagberg, “Mobile phone use and stress, sleep disturbances, and symptoms of depression among young adults—a prospective cohort study,” *BMC Public Health*, vol. 11, no. 1, pp. 1–11, 2011. <https://doi.org/10.1186/1471-2458-11-66>
- [19] S. Saeb, E. G. Lattie, S. M. Schueller, K. P. Kording, and D. C. Mohr, “The relationship between mobile phone location sensor data and depressive symptom severity,” *PeerJ*, vol. 4, p. e2537, 2016. <https://doi.org/10.7717/peerj.2537>
- [20] S. Saeb *et al.*, “Mobile phone sensor correlates of depressive symptom severity in daily-life behavior: an exploratory study,” *Journal of Medical Internet Research*, vol. 17, no. 7, p. e4273, 2015. <https://doi.org/10.2196/jmir.4273>
- [21] A. A. Atienza *et al.*, “Consumer attitudes and perceptions on mHealth privacy and security: findings from a mixed-methods study,” *Journal of Health Communication*, vol. 20, no. 6, pp. 673–679, 2015. <https://doi.org/10.1080/10810730.2015.1018560>
- [22] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, “Communication-efficient learning of deep networks from decentralized data,” *Artificial Intelligence and Statistics*, pp. 1273-1282. PMLR, 2017. <https://doi.org/10.48550/arXiv.1602.05629>
- [23] “TensorFlow Federated”, *TensorFlow*. [Online]. Available: <https://www.tensorflow.org/federated> [Accessed: 26-Aug-2022].
- [24] K. Kroenke, R. L. Spitzer, and J. B. W. Williams, “The PHQ-9: validity of a brief depression severity measure,” *Journal of General Internal Medicine*, vol. 16, no. 9, pp. 606–613, 2001. <https://doi.org/10.1046/j.1525-1497.2001.016009606.x>

- [25] “DeepLearning4j Suite Overview,” *Konduit.ai*. [Online]. Available: <https://deeplearning4j.konduit.ai/> [Accessed: 26-Aug-2022].
- [26] A. Pannu, “Artificial intelligence and its application in different areas,” *Artificial Intelligence*, vol. 4, no. 10, pp. 79–84, 2015.
- [27] W. Wang and K. Siau, “Artificial intelligence, machine learning, automation, robotics, future of work and future of humanity: A review and research agenda,” *Journal of Database Management (JDM)*, vol. 30, no. 1, pp. 61–79, 2019. <http://dx.doi.org/10.4018/JDM.2019010104>
- [28] A. Singh, N. Thakur, and A. Sharma, “A review of supervised machine learning algorithms,” *2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom)*, pp. 1310-1315. IEEE, 2016.
- [29] S. B. Kotsiantis, I. Zaharakis, and P. Pintelas, “Supervised machine learning: A review of classification techniques,” *Emerging Artificial Intelligence Applications in Computer Engineering*, vol. 160, no. 1, pp. 3–24, 2007.
- [30] T. Sun, D. Li, and B. Wang, “Decentralized federated averaging,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022. <https://doi.org/10.1109/TPAMI.2022.3196503>
- [31] Q. Yang, Y. Liu, T. Chen, and Y. Tong, “Federated machine learning: Concept and applications,” *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 10, no. 2, pp. 1–19, 2019. <https://doi.org/10.1145/3298981>
- [32] “Sensors,” *Android Developers*. [Online]. Available: <https://developer.android.com/guide/topics/sensors/> [Accessed: 26-Aug-2022].
- [33] B. Martínez-Pérez, I. De La Torre-Díez, and M. López-Coronado, “Mobile health applications for the most prevalent conditions by the World Health Organization: review and analysis,” *Journal of Medical Internet Research*, vol. 15, no. 6, p. e2600, 2013. <https://doi.org/10.2196/jmir.2600>
- [34] K. Kroenke and R. L. Spitzer, “The PHQ-9: a new depression diagnostic and severity measure,” *Psychiatric Annals*, vol. 32, no. 9, pp. 509–515, 2002. <https://doi.org/10.3928/0048-5713-20020901-06>
- [35] R. L. Spitzer, J. B. W. Williams, K. Kroenke, R. Hornyak, J. McMurray, and Patient Health Questionnaire Obstetrics-Gynecology Study Group, “Validity and utility of the PRIME-MD patient health questionnaire in assessment of 3000 obstetric-gynecologic patients: the PRIME-MD Patient Health Questionnaire Obstetrics-Gynecology Study,” *American Journal of Obstetrics and Gynecology*, vol. 183, no. 3, pp. 759–769, 2000. <https://doi.org/10.1067/mob.2000.106580>
- [36] A. Stathopoulou *et al.*, “Mobile Assessment Procedures for Mental Health and Literacy Skills in Education,” *International Journal of Interactive Mobile Technologies*, vol. 12, no. 3, pp. 21–37, 2018. <https://doi.org/10.3991/ijim.v12i3.8038>
- [37] S. Nickels *et al.*, “Toward a mobile platform for real-world digital measurement of depression: User-centered design, data quality, and behavioral and clinical modeling,” *JMIR Mental Health*, vol. 8, no. 8, p. e27589, 2021. <https://doi.org/10.2196/27589>
- [38] K. Sozinov, V. Vlassov, and S. Girdzijauskas, “Human activity recognition using federated learning,” *2018 IEEE Intl. Conf. on Parallel & Distributed Processing with Applications, Ubiquitous Computing & Communications, Big Data & Cloud Computing, Social Computing & Networking, Sustainable Computing & Communications (ISPA/IUCC/BDCLOUD/SocialCom/SustainCom)*, pp. 1103-1111. IEEE, 2018. <https://doi.org/10.1109/BDCLOUD.2018.00164>
- [39] P. Chhikara, P. Singh, R. Tekchandani, N. Kumar, and M. Guizani, “Federated learning meets human emotions: A decentralized framework for human–computer interaction for

- iot applications,” *IEEE Internet of Things Journal*, vol. 8, no. 8, pp. 6949–6962, 2020. <https://doi.org/10.1109/JIOT.2020.3037207>
- [40] J. C. Liu, J. Goetz, S. Sen, and A. Tewari, “Learning from others without sacrificing privacy: Simulation comparing centralized and federated machine learning on mobile health data,” *JMIR mHealth and uHealth*, vol. 9, no. 3, p. e23728, 2021. <https://doi.org/10.2196/23728>
- [41] M. N. Chowdhury, H M Zabir Haque, K. T. Tahmid, F.-T.-Z. Salma, and N. Ahmed, “A novel approach for product recommendation using smartphone sensor data,” *International Journal of Interactive Mobile Technologies*, vol. 16, no. 16, pp. 190–204, 2022. <https://doi.org/10.3991/ijim.v16i16.31617>
- [42] B. Ramsundar and R. B. Zadeh, “Chapter 4. fully connected deep networks,” *TensorFlow for Deep Learning: From Linear Regression to Reinforcement Learning*, pp. 81–102, 2018.
- [43] C. Tan, F. Sun, T. Kong, W. Zhang, C. Yang, and C. Liu, “A survey on deep transfer learning,” *International Conference on Artificial Neural Networks*, pp. 270–279. Springer, Cham, 2018. https://doi.org/10.1007/978-3-030-01424-7_27
- [44] “OpenMined,” *Openmined.org*. [Online]. Available: <https://www.openmined.org/> [Accessed: 26-Aug-2022].
- [45] J. Konečný, H. B. McMahan, F. X. Yu, P. Richtárik, A. T. Suresh, and D. Bacon, “Federated learning: Strategies for improving communication efficiency,” *arXiv preprint arXiv:1610.05492*, 2016. <https://doi.org/10.48550/arXiv.1610.05492>

9 Authors

Nawrin Tabassum is currently doing B.Sc in Computer Science and Engineering at Ahsanullah University of Science and Technology, Dhaka, Bangladesh (Email: nawrintabassum14@gmail.com).

Mustofa Ahmed is currently doing B.Sc in Computer Science and Engineering at Ahsanullah University of Science and Technology, Dhaka, Bangladesh (Email: mustofahmed24@gmail.com).

Nushrat Jahan Shorna is currently doing B.Sc in Computer Science and Engineering at Ahsanullah University of Science and Technology, Dhaka, Bangladesh (Email: nushrat.j01@gmail.com).

MD Mejbah Ur Rahman Sowad is currently doing B.Sc in Computer Science and Engineering at Ahsanullah University of Science and Technology, Dhaka, Bangladesh (Email: mejbahurrahman13@gmail.com).

H M Zabir Haque is an Assistant Professor at Ahsanullah University of Science and Technology, Dhaka, Bangladesh. He has received his Master of Science in Computer Science from the University of Saskatchewan, Canada, and a Bachelor’s degree in Computer Science and Engineering from Ahsanullah University of Science and Technology, Dhaka, Bangladesh. His research interests include Bioinformatics, Computational Biology, and Machine Learning (Email: zabir.haque.cse@aust.edu).

Article submitted 2022-09-02. Resubmitted 2022-11-27. Final acceptance 2022-11-30. Final version published as submitted by the authors.