2024 46th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) | 979-8-3503-7149-9/24/$31.00 ©2024 IEEE | DOI: 10.1109/EMBC53108.2024.10782516

Differential Private Federated Transfer Learning for Mental Health Monitoring in Everyday Settings: A Case Study on Stress Detection

[1]Ziyu Wang*, [1]Zhongqi Yang*, [1,3]Iman Azimi, and [1,2,3]Amir M. Rahmani
[1]*Department of Computer Science, University of California, Irvine*
[2]*School of Nursing, University of California, Irvine*
[3]*Institute for Future Health, University of California, Irvine*
{ziyuw31, zhongqy4, azimii, a.rahmani}@uci.edu

***Abstract*—Mental health conditions, prevalent across various demographics, necessitate efficient monitoring to mitigate their adverse impacts on life quality. The surge in data-driven methodologies for mental health monitoring has underscored the importance of privacy-preserving techniques in handling sensitive health data. Despite strides in federated learning for mental health monitoring, existing approaches struggle with vulnerabilities to certain cyber-attacks and data insufficiency in real-world applications. In this paper, we introduce a differential private federated transfer learning framework for mental health monitoring to enhance data privacy and enrich data sufficiency. To accomplish this, we integrate federated learning with two pivotal elements: (1) differential privacy, achieved by introducing noise into the updates, and (2) transfer learning, employing a pre-trained universal model to adeptly address issues of data imbalance and insufficiency. We evaluate the framework by a case study on stress detection, employing a dataset of physiological and contextual data from a longitudinal study. Our finding show that the proposed approach can attain a 10% boost in accuracy and a 21% enhancement in recall, while ensuring privacy protection.**

***Index Terms*—Stress Monitoring, Differential Privacy, Federated Learning, Transfer Learning, Personalized Machine Learning, Health Informatics**

I. Introduction

Mental health concerns impact a diverse demographic cross-section of the population [1]. The prolonged effects of these mental health conditions can have detrimental impacts on physical health [2], [3], psychological states, and overall quality of life [4]–[6]. In response, the field of mental health monitoring has increasingly turned to data-driven methodologies [7], [8]. These methods, utilizing advanced analytics and machine learning algorithms, provide precise, real-time assessments of various mental health states [9]. However, the ascendancy of data-driven techniques in mental health monitoring brings forth pressing concerns regarding data privacy [10]. The sensitive nature of the data involved demands methodologies that are not just accurate and efficient in monitoring mental health but also stringently protective of individual privacy [11]–[13].

*Authors equally contributed.

Recent studies in mental health monitoring (e.g., mood, loneliness, depression, stress) increasingly adopt federated learning to enhance privacy [14]–[20], showing promise in using local data for model training. For example, [17] developed a framework for depression detection using smartphone data (location, accelerometer, call logs) for training on devices, although vulnerable to privacy breaches like membership inference attacks. [18] and [19], [21] similarly leveraged federated learning for loneliness and stress detection, utilizing sensor and PPG signal data from smartphones and wearables, respectively, to update models locally, thus preserving personal data privacy.

Although existing approaches have shown success in enhancing privacy in certain domains, they still exhibit vulnerabilities to a range of cyber threats, notably backdoor and inference attacks. The absence of more sophisticated privacy-preserving mechanisms, such as those offered by differential privacy, represents a gap in the current approaches for mental health monitoring. Furthermore, a significant hurdle in the practical deployment of federated learning in mental health studies stems from the reliance on limited data sources [22], [23]. When models are trained locally in such studies, there's often a marked inconsistency in the diversity and volume of data, as noted by Dixit et al. [24]. This variability can lead to inadequate training of decentralized models, especially when real-world data lacks the comprehensiveness or balance needed for effective training [25]. Therefore, resolving this issue is imperative for the effective application of federated learning in practical scenarios.

In this study, we propose a differential private federated transfer learning framework for mental health monitoring. We incorporate a differential privacy mechanism into federated learning by introducing noise into the combined updates prior to their transmission back to the central server. Furthermore, we utilize transfer learning to tackle data insufficiency and imbalance in stress detection as a case study. Specifically, we commence by pre-training a universal model on an extensive, non-sensitive dataset, subsequently refining it with individual user data on-device. We evaluate the proposed framework via a case study on stress detection on a dataset from a longitudinal study. The dataset includes physiological and

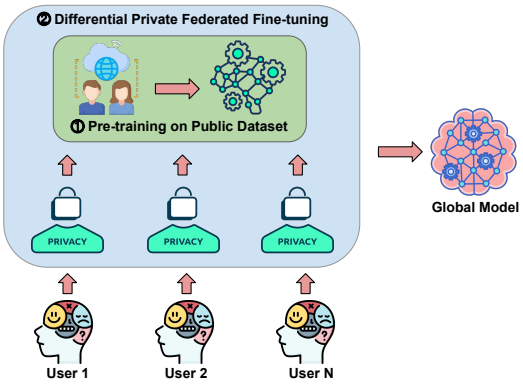

Fig. 1. Overview of the proposed framework.

contextual data gathered from Samsung Galaxy Active 2 Watches and smartphones from 54 individuals. The label stress levels of these individuals were collected through multiple daily ecological momentary assessments (EMAs) conducted via their smartphones.

II. Methodology

Our methodology addresses data scarcity and privacy challenges in mental health monitoring by integrating transfer learning with federated learning, enhanced by differential privacy techniques. We initially pre-train a foundational model on a comprehensive public dataset to capture broad health trends. This model is then fine-tuned on sparse, user-specific data in a distributed manner, improving accuracy with limited data. Differential privacy is applied during weight updates to protect individual data privacy. This combined approach effectively balances privacy concerns with the need for precise health monitoring. Our methodology is depicted in Fig. 1.

A. Differential Private Federated Transfer Learning Framework for Mental Health Monitoring

The learning process is detailed in Algorithm 1, and we provide an in-depth explanation of each module in the subsequent sections.

1) Pre-training: In real-world scenarios, gathering extensive, high-quality data for mental health studies is often fraught with difficulties. Participants in such studies may not consistently engage, leading to gaps or irregularities in the data. Additionally, the collected data often contains artifacts and noise, further complicating its utility and quality. These factors combine to create a scenario where the available data is scarce, inconsistent, and noisy.

We initiate our approach by pre-training our model $\mathcal{M}$ on a comprehensive publicly available dataset $\mathcal{D}$pt. This dataset, rich in both quantity and quality, comprises N pairs of well-curated features and labels:

$$D_{pre-train} = (x_1, y_1), (x_2, y_2), ..., (x_N, y_N). \tag{1}$$

The pre-training phase utilizes the binary cross-entropy loss, and is aimed at learning generalized mental health patterns:

$$\mathcal{L} = -\frac{1}{N}\sum_{i=1}^{N}\left[y_i \cdot \log(\mathcal{M}(x_i)) + (1-y_i)\cdot \log(1-\mathcal{M}(x_i))\right]. \tag{2}$$

Algorithm 1 Differential Private Federated Transfer Learning for Mental Health Monitoring

Pre-training Phase:
$\mathcal{D}_{pre-train} \leftarrow$ Load publicly available dataset
Initialize model $\mathcal{M}$
for each batch (x_i, y_i) in $\mathcal{D}_{pre-train}$ **do**
  Update $\mathcal{M}$ to minimize $\mathcal{L}_{pre-train}(\mathcal{M})$
end for
Federated Fine-Tuning Phase:
for each client k in the federated network **do**
  Distribute pre-trained model $\mathcal{M}$ to client k
  $\mathcal{D}_{\text{user},k} \leftarrow$ Load user-specific dataset for client k
  for each batch (x'_j, y'_j) in $\mathcal{D}_{\text{user},k}$ **do**
    Fine-tune $\mathcal{M}$ on (x'_j, y'_j)
  end for
  Compute gradients and apply gradient clipping
  Add differential privacy noise Lap $\left(\frac{\Delta f}{\epsilon}\right)$ to gradients
  Send updated model $\mathcal{M}_k$ to server
end for
Global Model Aggregation:
Initialize global model $\mathcal{M}_{\text{global}}$
$\mathcal{M}_{\text{global}} \leftarrow$ Aggregate updates $\mathcal{M}_k$ from all clients
return $\mathcal{M}_{\text{global}}$

2) Differential Private Federated Fine-tuning: To tackle the critical privacy concerns in mental health monitoring, our approach incorporates differential privacy (DP) within the federated learning paradigm, specifically through the Federated Averaging (FedAvg) algorithm [26] enhanced with Laplacian noise. This implementation is key to defending against sophisticated privacy attacks against traditional federated learning frameworks, including model inversion [27], membership inference [28], and backdoor [29], which pose risks to user data confidentiality.

a) Federated Averaging: FedAvg is a widely-used algorithm in federated learning that involves averaging model updates (weights) from multiple clients. In our implementation, each client represents a participant in the mental health monitoring study, first trains a local model using their private data. The local models are then aggregated to update the global model. This process is formalized as:

$$\mathcal{M}_{\text{global}} = \frac{1}{K}\sum_{k=1}^{K}\mathcal{M}_k, \tag{3}$$

where $\mathcal{M}_{\text{global}}$ represents the global model, $\mathcal{M}_k$ is the model trained on the k-th client, and K is the total number of clients.

b) Laplacian Noise for Differential Privacy: To ensure differential privacy, we add Laplacian noise to the model updates during aggregation. The noise is calibrated to the sensitivity of the model's gradients and the privacy budget ϵ. For a model parameter θ, the update with Laplacian noise is:

$$\theta_{\text{updated}} = \theta + \text{Lap}\left(\frac{\Delta f}{\epsilon}\right), \tag{4}$$

where Δf denotes the sensitivity of the function (gradient norm), and $\text{Lap}(\cdot)$ represents the Laplacian noise.

c) Fine-tuning: Following the pre-training phase, the model $\mathcal{M}$ utilize private, user-specific datasets, represented as $D_{fine-tune} = \{(x'_j, y'_j)\}_{j=1}^{M}$ to fine-tune. This process occurs within a federated learning framework and is essential for adapting the generalized model to the specific mental health profiles and solve the data scarcity issue of the private dataset. During fine-tuning, the model parameters are adjusted according to each user's data. The fine-tuning employs the same loss function, as Eq. 2, used in the pre-training phase, ensuring consistency in the optimization approach across both stages of model development.

B. Case Study on Stress Detection

The efficacy of our framework is showcased through a case study in stress detection, leveraging data from an extensive longitudinal study [30]. This study engaged distinct groups of participants across its two phases, with 30 individuals in the initial phase and 24 in the subsequent phase, including both undergraduate and graduate students. Data collection occurred over two periods: from June 2020 to June 2021 and from March 2022 to May 2023, resulting in 109,586 and 23,012 refined samples for each period, respectively. The dataset from the first phase, $D_{pre-train}$, was critical for the pre-training process, and the dataset from the second phase, $D_{fine-tune}$, was essential for fine-tuning. Conducted using our ZotCare health monitoring system [31], this strategy leveraged the comprehensive dataset from the first phase for initial model training, while the more constrained dataset from the second phase was used for precise model enhancement, effectively tackling the challenges posed by data scarcity.

1) Dataset: The collected dataset comprises raw PPG and motion data, along with partial annotations for stress levels, emotions, and physical activity through EMA at semi-random intervals. To comprehensively process the PPG signals, we employ the HeartPy library [32], extracting 12 specific features from both electrical activations and pressure waveforms in the dataset. These features[1] are heart rate (HR) and heart rate variability (HRV) measures. Stress levels equal to or less than 2 are categorized as 'unstressed,' while higher values indicate 'stressed' individuals.

2) Stress Detection Model: To develop our stress detection model, we began with a pre-processing stage to refine bio-signals from wearable devices. Initial data cleaning eliminated erroneous readings by using motion data to filter out noise and artifacts, utilizing a bandpass Butterworth filter specific to PPG signal frequencies. This was followed by a moving average technique to smooth the data, minimizing motion-related distortions. We then applied min-max normalization, scaling feature values uniformly between 0 and 1, to reduce individual data variations and enhance bio-signal interpretability.

To model stress levels, we opted for a Multilayer Perceptron (MLP). This MLP is structured with a three-layer design, including hidden layers that consist of 128 and 32 neurons, respectively. It processes an input comprising the 12 extracted features related to heart rate (HR) and heart rate variability (HRV). The output of the MLP is a binary classification, differentiating between states of stress and no stress.

[1]BPM, IBI, SDNN, SDSD, RMSSD, pNN20, pNN50, MAD, SD1, SD2, S, SD1/SD2, and BR1.

III. Evaluation

This section is dedicated to the evaluation of our proposed framework. We establish a baseline using the pre-trained model to highlight the benefits of incorporating transfer learning. Additionally, we explore the delicate balance between privacy preservation and model accuracy by adjusting the privacy budget parameter, ϵ.

1) Implementation Details: We train the proposed stress detection model by the Adam optimizer (learning rate=0.001) and cross-entropy loss function. A dropout strategy (rate=0.5) is implemented after the initial hidden layer to reduce overfitting. Differential privacy via Laplacian noise addition to the gradients was introduced, with an ϵ value of 1, aiming for a balance between privacy protection and model accuracy.

We pre-train on $D_{pre-train}$ with 3271 labeled samples for 50 epochs to capture a wide array of stress signals. Subsequently, we applied FedAvg to fine-tune the model on $D_{fine-tune}$, consisting of 1220 labeled samples, over 30 epochs for personalized user data adaptation.

The dataset $D_{fine-tune}$ was divided in a manner that allocated the last 30% of each user's chronologically ordered data to the test set, ensuring the model was trained on past data and evaluated on future data. Client partitioning for federated learning was conducted based on individual users, reflecting a real-world scenario where fine-tuning occurs on each user's client device in a personalized manner.

2) Results: The results of our proposed framework for stress detection are shown in Table I. The Plain model, trained solely on the original dataset $D_{fine-tune}$ without incorporating transfer learning techniques, yielded an accuracy of 0.43, an F1 Score of 0.39, a recall of 0.54, and a precision of 0.31. Conversely, training the model on the public, more extensive dataset $D_{pre-train}$ resulted in improved metrics: accuracy rose to 0.51, F1 Score to 0.44, recall to 0.25, and precision to 0.36. Notably, the integration of transfer learning within our differential privacy federated learning framework enhances performance, achieving an accuracy of 0.53, an F1 Score of 0.52, a recall of 0.75, and a precision of 0.40.

TABLE I
STRESS DETECTION PERFORMANCE OF DIFFERENT MODELS

Model	Accuracy	F1 Score	Recall	Precision
Plain	0.43	0.39	0.54	0.31
Pre-trained	0.51	0.44	0.58	0.36
Fine-tuned	**0.53**	**0.52**	**0.75**	**0.40**

3) Privacy Budget Analysis: In the final stage of our analysis, we conducted comparisons of our outcomes with two reference benchmarks: one involving the identical neural network framework without the implementation of federated

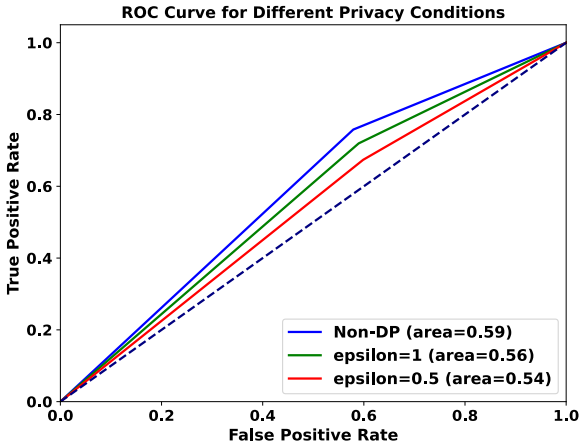

Fig. 2. Overview of the proposed approach.

learning or differential privacy, and the other utilizing the same neural network structure with federated learning but absent differential privacy. This comparative evaluation was performed across two scenarios lacking differential privacy, with ϵ values set at 0.5 and 1, respectively. It's important to note that a lower ϵ value signifies enhanced privacy safeguards. As depicted in Fig. 2, integrating differential privacy within a federated learning context did not significantly alter the metrics of the receiver operating curve (ROC). This observation suggests that it is feasible to strike an effective equilibrium between preserving model efficacy and upholding privacy standards.

IV. Discussion

In this study, we introduce a differential privacy federated transfer learning framework and evaluate it on stress detection. By applying transfer learning, our approach effectively leverages the strengths of both public and private data sources, enhancing the model's performance in real-world scenarios. Especially noteworthy is the improvement in recall indicates that our framework is effective in reducing potential overlooking true stress cases. This advance is crucial, minimizing missed detections of stress and thereby mitigating their potential adverse effects on individual well-being and public health. Additionally, the implementation of differential privacy within the framework ensures the protection of sensitive health data, which is a critical concern in the field of mental health. By infusing noise into the model updates, our framework mitigates risks associated with several cyber attacks [27]–[29].

This study's limitation is its focus on stress detection, which may not encompass other mental health conditions or capture long-term trends effectively due to its design for short-term analysis. Future efforts could aim to broaden the framework's applicability to various mental health states and for different cohorts. This could improve its predictive power for both immediate and extended periods that were shown feasible in prior research [33].

V. Conclusion

This study proposed a differential private federated transfer learning framework addressing the challenges of data scarcity and privacy protection in mental health monitoring. The proposed framework exhibited strong performance in stress detection while effectively preserving data privacy. Specifically, our framework achieved a substantial 10% improvement in accuracy and a noteworthy 21% enhancement in recall. These results highlighted the potential of our approach to provide more effective and privacy-conscious mental health monitoring, addressing the pressing need for efficient solutions in this domain.

References

[1] M. Prince *et al.*, "No health without mental health," *The lancet*, vol. 370, no. 9590, pp. 859–877, 2007.
[2] M. Cheng, X. Diao, Z. Zhou, Y. Cui, W. Liu, and S. Cheng, "Toward short-term glucose prediction solely based on cgm time series," 2024.
[3] Z. Yang *et al.*, "Chatdiet: Empowering personalized nutrition-oriented food recommender chatbots through an llm-augmented framework," *Smart Health*, p. 100465, 2024.
[4] Z. Zhou *et al.*, "Glumarker: A novel predictive modeling of glycemic control through digital biomarkers," 2024.
[5] M. Cheng *et al.*, "Saic: Integration of speech anonymization and identity classification," 2023.
[6] Z. Zhou *et al.*, "Crossgp: Cross-day glucose prediction excluding physiological information," *arXiv preprint arXiv:2404.10901*, 2024.
[7] S. Bucci *et al.*, "The digital revolution and its impact on mental health care," *Psychology and Psychotherapy: Theory, Research and Practice*, vol. 92, no. 2, pp. 277–297, 2019.
[8] H. H. Koshkak *et al.*, "Seal: Sensing efficient active learning on wearables through context-awareness," in *Proceedings of the IEEE/ACM Design, Automation and Test in Europe Conference (DATE'24)*, (Spain), 2024.
[9] E. Garcia-Ceja *et al.*, "Mental health monitoring with multimodal sensing and machine learning: A survey," *Pervasive and Mobile Computing*, vol. 51, pp. 1–26, 2018.
[10] A. Kanduri, S. Shahhosseini, E. K. Naeini, H. Alikhani, P. Liljeberg, N. Dutt, and A. M. Rahmani, *Edge-Centric Optimization of Multi-modal ML-Driven eHealth Applications*, pp. 95–125. Cham: Springer Nature Switzerland, 2024.
[11] Z. Wang *et al.*, "Guardhealth: Blockchain empowered secure data management and graph convolutional network enabled anomaly detection in smart healthcare," *Journal of Parallel and Distributed Computing*, 2020.
[12] X. Yang *et al.*, "Zebra: Deeply integrating system-level provenance search and tracking for efficient attack investigation," *arXiv preprint arXiv:2211.05403*, 2022.
[13] H. Alikhani, A. Kanduri, P. Liljeberg, A. M. Rahmani, and N. Dutt, "Dynafuse: Dynamic fusion for resource efficient multi-modal machine learning inference," *IEEE Embedded Systems Letters*, 2023.
[14] X. Xu *et al.*, "Fedmood: Federated learning on mobile health data for mood detection," *arXiv preprint arXiv:2102.09342*, 2021.
[15] A. Alahmadi *et al.*, "A privacy-preserved iomt-based mental stress detection framework with federated learning," *The Journal of Supercomputing*, pp. 1–20, 2023.
[16] Y. Chen *et al.*, "Fedhealth: A federated transfer learning framework for wearable healthcare," *IEEE Intelligent Systems*, vol. 35, no. 4, pp. 83–93, 2020.
[17] B. Suruliraj *et al.*, "Federated learning framework for mobile sensing apps in mental health," in *2022 IEEE 10th International Conference on Serious Games and Applications for Health (SeGAH)*, pp. 1–7, IEEE, 2022.
[18] M. M. Qirtas *et al.*, "Privacy preserving loneliness detection: A federated learning approach," in *2022 IEEE International Conference on Digital Health (ICDH)*, pp. 157–162, IEEE, 2022.
[19] Y. S. Can *et al.*, "Privacy-preserving federated deep learning for wearable iot-based biomedical monitoring," *ACM Transactions on Internet Technology (TOIT)*, vol. 21, no. 1, pp. 1–17, 2021.
[20] Y. Yao *et al.*, "Privacy-preserving and energy efficient task offloading for collaborative mobile computing in iot: An admm approach," *Computers & Security*, vol. 96, p. 101886, 2020.
[21] Z. Sharifi-Heris *et al.*, "Phenotyping the autonomic nervous system in pregnancy using remote sensors: potential for complication prediction," *Frontiers in Physiology*, vol. 14, p. 1293946, 2023.

[22] H. Hu, Z. Qiao, M. Cheng, Z. Liu, and H. Wang, "Dasgil: Domain adaptation for semantic and geometric-aware image-based localization," *IEEE Transactions on Image Processing*, vol. 30, pp. 1342–1353, 2020.
[23] M. Cheng, Z. Zhou, B. Zhang, Z. Wang, J. Gan, Z. Ren, W. Feng, Y. Lyu, H. Zhang, and X. Diao, "Efflex: Efficient and flexible pipeline for spatio-temporal trajectory graph modeling and representation learning," *arXiv preprint arXiv:2404.12400*, 2024.
[24] N. K. Dixit *et al.*, "Managing the scarcity of monitoring data through machine learning for human behavior in mental health care," in *Applications of Deep Learning and Big IoT on Personalized Healthcare Services*, pp. 147–175, IGI Global, 2020.
[25] M. Cheng *et al.*, "Vetrass: Vehicle trajectory similarity search through graph modeling and representation learning," *arXiv preprint arXiv:2404.08021*, 2024.
[26] B. McMahan *et al.*, "Communication-efficient learning of deep networks from decentralized data," in *Artificial intelligence and statistics*, pp. 1273–1282, PMLR, 2017.
[27] M. Fredrikson *et al.*, "Model inversion attacks that exploit confidence information and basic countermeasures," in *Proceedings of the 22nd ACM SIGSAC conference on computer and communications security*, pp. 1322–1333, 2015.
[28] M. Nasr *et al.*, "Comprehensive privacy analysis of deep learning," in *Proceedings of the 2019 IEEE Symposium on Security and Privacy (SP)*, pp. 1–15, 2018.
[29] Bagdasaryan *et al.*, "How to backdoor federated learning," in *International conference on artificial intelligence and statistics*, pp. 2938–2948, PMLR, 2020.
[30] A. Tazarv *et al.*, "Active reinforcement learning for personalized stress monitoring in everyday settings," in *2023 IEEE/ACM Conference on Connected Health (CHASE)*, (Los Alamitos, CA, USA), pp. 44–55, IEEE Computer Society, jun 2023.
[31] L. Sina *et al.*, "Zotcare: a flexible, personalizable, and affordable mhealth service provider," *Front. Digit. Health*, 2023.
[32] P. Van Gent *et al.*, "Heartpy: A novel heart rate algorithm for the analysis of noisy signals," *Transportation research part F: traffic psychology and behaviour*, vol. 66, pp. 368–378, 2019.
[33] Z. Yang *et al.*, "Loneliness forecasting using multi-modal wearable and mobile sensing in everyday settings," in *2023 IEEE 19th International Conference on Body Sensor Networks (BSN)*, pp. 1–4, IEEE, 2023.