SIN-CHN SCIENTIFIC PRESS PTE.LTD.

Molecular & Cellular Biomechanics 2025, 22(5), 1699.
https://doi.org/10.62617/mcb1699

Article

Federated learning for nurse stress prediction using wearable sensors: Integrating biomechanical data

Ke Liu[1,*], Wen Xue[2], Donghan Hou[3]

[1] Department of Asset and Equipment, Xi'an Medical University, Xi'an 710021, China
[2] Teaching Department, Xi'an International Technician College, Xi'an 710077, China
[3] Faculty of Science, Xi'an Shiyou University, Xi'an 710300, China
*** Corresponding author:** Ke Liu, lke@xiyi.edu.cn

CITATION

Liu K, Xue W, Hou D. Federated learning for nurse stress prediction using wearable sensors: Integrating biomechanical data. Molecular & Cellular Biomechanics. 2025; 22(5): 1699.
https://doi.org/10.62617/mcb1699

ARTICLE INFO

Received: 26 February 2025
Accepted: 18 March 2025
Available online: 24 March 2025

Abstract: In today's fast-paced work environments, accurately predicting stress levels is essential for effective healthcare workforce management, particularly among nurses in high-pressure settings. Despite the availability of various mental health initiatives, timely detection of stress remains challenging due to concerns over sensitive personal data privacy. To address this, we propose a federated learning (FL) framework that utilizes artificial intelligence (AI) to predict nurse stress levels by integrating distributed biomechanical data from wearable sensors, thereby preventing data leakage. Biometric features from datasets at each FL client are extracted and used to train local neural network (NN) models. After several aggregation rounds, the global model converges to predict nurse stress levels. Simulations demonstrate the effectiveness of our method, achieving over 90% prediction accuracy, which enhances the feasibility of privacy-preserving stress monitoring and offers scalable solutions for occupational health management.

Keywords: artificial intelligence; neural network; federated learning; biomechanical feedback; stress prediction

1. Introduction

Nurses, as a critical component of the healthcare workforce, have long faced stressors such as high work intensity, complex doctor-patient relationships, and demanding shift systems. Research indicates that over half of nurses experience moderate to severe occupational stress, leading to both physical and mental health issues, such as anxiety and fatigue syndrome [1–3]. This stress also directly impacts the quality of nursing care, patient safety, and the operational efficiency of healthcare institutions. The increasing nurse turnover rate due to high stress exacerbates the global shortage of healthcare professionals [4,5]. Therefore, the development of precise, real-time monitoring and intervention technologies for nurse stress is crucial for improving medical service quality and ensuring the well-being of healthcare workers.

In recent years, artificial intelligence (AI) technology, coupled with wearable sensors, has provided new approaches for continuous, non-invasive stress monitoring [6–9]. However, existing methods rely on centralized data models, facing two major challenges. First, biomechanical data, such as heart rate (HR) and posture, involve user privacy concerns, and centralized storage poses a risk of data breaches. Second, data silos between medical institutions hinder the model's generalization capabilities [10,11]. Federated learning (FL) offers a distributed training framework that allows data to be processed privately on local devices, sharing only encrypted model parameters [12,13].

This framework overcomes data silos and protects nurses' biomechanical data privacy, significantly enhancing the robustness and applicability of stress prediction models.

This paper explores an FL framework for predicting nurse stress, integrating multimodal biomechanical data streams collected from wearable sensors to capture subtle physiological and behavioral patterns associated with occupational stress. By establishing a decentralized training mechanism that enables cross-institutional collaboration without the exchange of raw data, the framework addresses critical challenges in healthcare AI deployment. Specifically, the system preprocesses sensor data locally, extracting time-domain features that serve as stress indicators. Our main contributions are as follows:

1) A systematic analysis and discussion of the value of biomechanical data in nurse stress recognition, highlighting its cross-modal relationship with psychological states.
2) An FL architecture that ensures the privacy and security of data from wearable devices is constructed, while maintaining high performance over 90% in nurse stress prediction accuracy compared to traditional centralized methods.
3) The proposal of a feasible pathway for the integration of intelligent wearable devices and AI technologies, supporting the standardized use of biomechanical data in occupational health monitoring and providing guidance for the intelligent transformation of medical systems.

The remainder of this paper is organized as follows: The background on AI, FL, and biomechanics related to human stress is introduced. Subsequently, the proposed FL framework for nurse stress prediction is described, including feature extraction and neural network (NN) model aggregation for FL. Dataset partitioning and related simulation results are presented and discussed. The hyper-parameters during model training are also detailed. Finally, the paper concludes with a discussion of future work.

2. Related work

Many works have focused on health care using AI technology. Hussain et al. proposed a continuous monitoring system that not only detects falls but also identifies falling patterns and associated activities, achieving high accuracy in fall detection [14]. Ilyas et al. used a deep learning framework combining convolutional NNs and long short-term memory to detect facial expressions and assist experts in scheduling rehabilitation sessions effectively [15]. Wong et al. randomized controlled trial aims to assess the effectiveness of a remote monitoring strategy, involving wearable biosensors, in detecting subtle physiological changes in asymptomatic individuals [16]. Combining wearable seizure detection devices, including motion and electrodermal activity sensors, a machine learning (ML) algorithm was proposed to accurately detect generalized tonic-clonic seizures, offering real-time alerts and continuous monitoring [17]. A random forest model was proposed to accurately detect sleep-wake states from accelerometer data, achieving a high F1 score and providing useful estimates correlated with self-reported nap behavior [18].

Stress prediction has garnered significant attention from researchers. Delmastro et al. investigated the use of wearable technologies and proposed a mobile system architecture for online stress monitoring during motor and cognitive training for frail older adults with mild cognitive impairment [19]. To provide real-time and post-workout feedback, a system called "RecoFit" was proposed to recognize the exercise being performed. The system achieves high performance, with precision and recall exceeding 95% [20]. Dasari et al. proposed a support vector machine-based system using electroencephalographic signals to accurately detect mental stress levels [21]. Mathur et al. used the body sensors and machine learning (ML) algorithms to monitor physiological indicators for stress detection in nurses [22].

However, related works aiming to predict stress levels do not consider data privacy, particularly the sensitive data captured from human beings, which presents a significant challenge and obstacle for ML methods to achieve accurate detection.

3. Preliminary and background

3.1. AI for healthcare

As shown in **Figure 1**, the rapid evolution of AI is fundamentally reshaping healthcare paradigms by enabling data-driven insights into complex physiological and psychological datasets. Contemporary healthcare systems increasingly employ AI-driven pattern recognition architectures to process multidimensional biosignals acquired through non-invasive wearable sensors, thereby establishing novel frameworks for proactive health management. This technological advancement holds particular significance in occupational medicine, where continuous biomechanical metrics—such as autonomic nervous system indicators, postural fluctuation patterns, and electrodermal activity dynamics—provide an empirical foundation for evaluating workforce wellness. By continuously quantifying these biomarkers, healthcare administrators gain objective metrics to correlate physical strain patterns with organizational risk factors, transcending traditional subjective assessment methodologies.

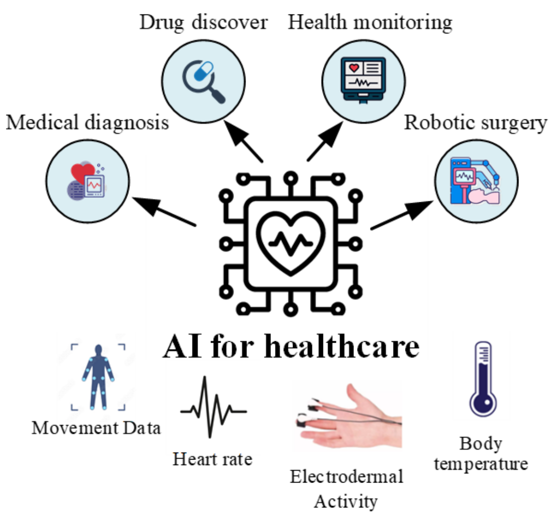

Figure 1. AI for healthcare.

AI-powered stress prediction exemplifies a paradigm shift, integrating real-time sensor data with ML models to uncover subtle correlations between biomechanical patterns and psychological states. In nursing, where high-stress environments directly affect care quality and staff retention, these systems analyze multimodal datasets, including motion kinematics, respiratory rhythms, and workload duration. By applying multi-layered feature extraction, they identify hidden associations between biomechanical disruptions and psychological distress, which is particularly critical in high-stakes nursing environments where stress undermines both patient outcomes and professional sustainability. NN architectures, especially attention-based recurrent NNs, excel at modeling spatiotemporal relationships through hierarchical abstraction, such as micro-changes in gait symmetry during long shifts or irregular muscle activation patterns preceding cognitive fatigue. These computational biomarkers enable early intervention, offering detection windows previously undetectable through traditional observational methods.

3.2. FL framework for healthcare

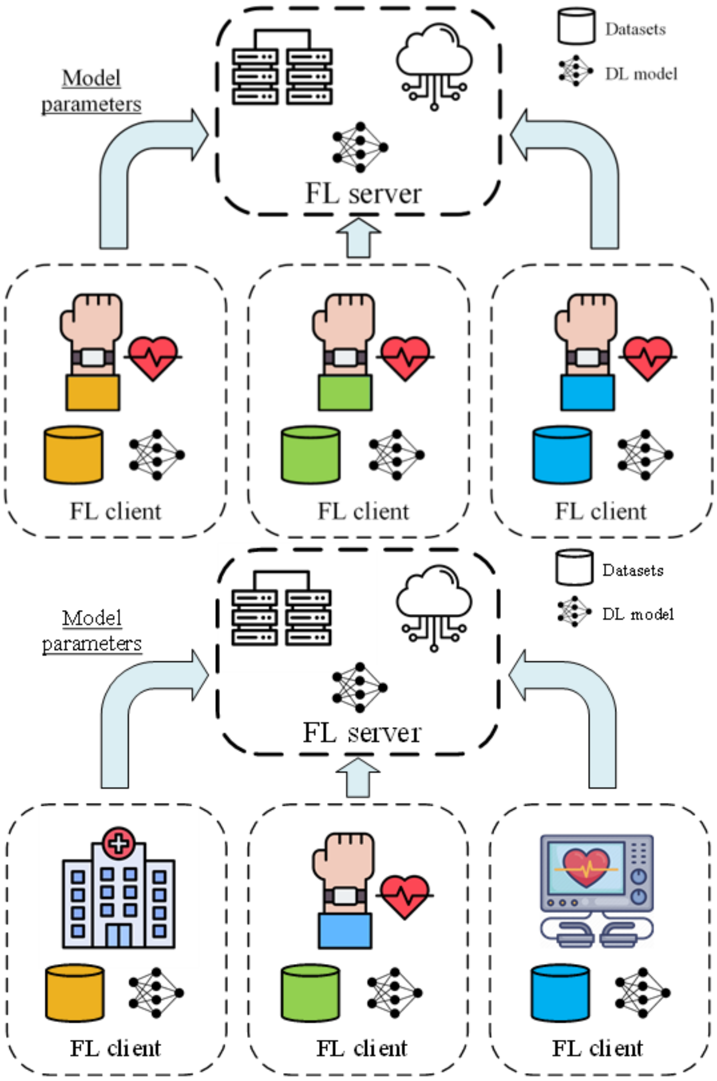

Figure 2. FL for healthcare.

FL has emerged as a transformative, privacy-preserving breakthrough in medical AI, effectively addressing the longstanding conflict between the need for multi-institutional collaboration and the paramount importance of safeguarding healthcare data confidentiality. This innovative approach enables institutions such as hospitals, research centers, and healthcare providers to collaboratively train advanced ML models without the need to directly share sensitive data. As illustrated in **Figure 2**, FL operates through a decentralized framework in which institutions participate in model training by sharing only encrypted updates to model parameters, rather than raw data. This paradigm marks a significant departure from traditional centralized methods, where data is pooled and stored in a single location, thus mitigating concerns related to data security and privacy.

The decentralized nature of FL helps resolve several ethical dilemmas that have historically complicated health data sharing. Traditionally, the sharing of medical data has raised concerns over consent, data breaches, and potential misuse of sensitive information. By ensuring that data remains within each institution's secure environment, FL maintains privacy while still facilitating the collaborative development of advanced AI models. This is particularly crucial in healthcare settings, where trust and confidentiality are foundational to healthcare.

Moreover, FL offers a unique opportunity for cross-institutional knowledge integration. By orchestrating the distributed analysis of geographically dispersed datasets, FL allows institutions to collectively benefit from a diverse range of data sources, enhancing the robustness and generalizability of predictive models. This distributed approach enables the development of more accurate and comprehensive AI models, which are trained on a broader array of data—critical in the healthcare sector, where populations, clinical practices, and disease presentations vary widely across regions and institutions.

In addition to its privacy-preserving benefits, FL accelerates medical research by enabling faster and more scalable model development. Researchers and healthcare practitioners can collaborate across institutional boundaries, reducing the time needed to develop and deploy AI-driven diagnostic tools, treatment strategies, and predictive models. The efficiency of FL also makes it an attractive option for applications requiring large datasets that are difficult to centralize due to logistical, regulatory, or privacy concerns. By ensuring that sensitive data is never directly shared, FL paves the way for more secure, collaborative, and effective advancements in the field of medical AI.

3.3. Biomechanics related to human stress

The biomechanical study of human stress has gained significant attention in recent years, particularly with the advent of wearable sensor technologies that enable real-time monitoring of physiological and kinetic responses in real-world occupational settings. These advancements have paved the way for a more comprehensive understanding of stress dynamics, as wearable systems now capture a broad range of multidimensional biomechanical parameters. Among these, postural dynamics and electrodermal activity (EDA) have emerged as key biomarkers for quantifying stress levels, especially in high-intensity professions like nursing, where the physical and emotional demands of the job can profoundly affect health and well-being.

Continuous monitoring of kinetic parameters provides valuable insights into the biomechanical patterns associated with stress. These patterns, often indicative of cumulative physical strain, can be detected through wearable sensors equipped with inertial measurement units or 3-axis accelerometers, which measure subtle changes in posture and movement over time. For example, irregular gait symmetry during extended work shifts has been shown to predict the onset of cognitive fatigue, often before workers subjectively experience symptoms. Such early indicators are crucial in preventing long-term stress-related health issues, including musculoskeletal injuries and burnout. Moreover, when analyzed in conjunction with other physiological signals, these spatiotemporal postural metrics provide objective evidence of stress—something traditional self-report tools, like questionnaires, may overlook due to their subjective nature.

EDA, which reflects sympathetic nervous system activation, has become a critical parameter in stress assessment. EDA is primarily driven by the activity of sweat glands, which are controlled by the sympathetic branch of the autonomic nervous system. In stressful situations, increased sympathetic nervous activity leads to a rise in sweat production, detectable as fluctuations in skin conductance. This physiological response provides a direct, real-time measure of emotional and psychological arousal. The temporal patterns of EDA, including both tonic and phasic variations, are crucial for understanding stress responses over time. Tonic EDA represents baseline skin conductance levels, while phasic EDA corresponds to rapid, short-term fluctuations typically triggered by acute stressors or emotional events.

In occupational settings like nursing, EDA data can reveal circadian-linked stress peaks, especially during night shifts. These peaks reflect the body's natural response to altered work-rest cycles and can indicate physiological strain due to the disruption of normal circadian rhythms. Studies have shown that EDA responses during night shifts often increase compared to daytime baselines, suggesting that prolonged work hours, high cognitive load, and disrupted sleep patterns can induce significant stress. This relationship between EDA and stress is further supported by the fact that elevated sympathetic nervous system activity, reflected in increased EDA, is associated with both physical and mental fatigue.

The integration of EDA with other biomechanical measurements, such as postural dynamics and movement patterns, offers a more holistic approach to stress assessment. For instance, during high-stress situations, such as high-pressure care environments, both EDA and biomechanical parameters like irregular gait and posture shifts often exhibit simultaneous spikes. This co-occurrence emphasizes the interconnectedness of physiological and mechanical stress responses. Together, these data points enable researchers and practitioners to develop more accurate models for predicting stress levels, potentially leading to improved strategies for mitigating stress in the workplace. The combination of continuous, objective physiological monitoring and advanced data analytics holds great promise for enhancing our understanding of the multifaceted nature of stress and its impact on human health in demanding occupational settings.

In conclusion, wearable sensors that measure both biomechanical and electrodermal parameters provide valuable tools for assessing stress in real-world environments. By capturing the dynamic interplay between postural changes, movement irregularities, and EDA responses, these technologies offer new avenues for early intervention and stress management, ultimately contributing to better health outcomes for individuals in high-stress professions like nursing.

4. FL framework for nurse stress prediction

As above-mentioned, NN has emerged as a transformative technology in healthcare, bringing significant advancements. NN models, particularly those based on artificial NNs, have demonstrated remarkable capabilities in learning complex patterns from large-scale datasets, thereby aiding healthcare professionals in making more accurate detections, particularly in stress prediction.

Supervised learning is a type of ML in which a model is trained on a labeled dataset consisting of input-output pairs. The model learns to map the input data to the correct output by minimizing a loss function, which quantifies the difference between predicted and actual values. Formally, given a training set $\{(x_1, y_1), (x_2, y_2), (x_3, y_3), \ldots\}$ where x_i represents the sample data and y_i is the corresponding class, the objective of supervised learning is to learn a mapping function $f: X \to Y$ that minimizes a loss function $\mathscr{L}(f(x_i), y_i)$ over the entire dataset. The model performance is evaluated using a suitable metric, such as accuracy for classification tasks or mean squared error (MSE). The supervised learning task can be formulated as follows.

$$f = argmin \frac{1}{n} \sum_{i=1}^{n} \mathscr{L}(f(x_i), y_i) \tag{1}$$

where $f(x_i)$ is the predicted output for the input x_i, and $\mathscr{L}$ is the close loss function. During training, the model iteratively adjusts its parameters using training algorithms to minimize this loss.

Gradient descent (GD) is a widely used optimization algorithm for training NN models, particularly in supervised learning [23]. It is an iterative method that minimizes the loss function by updating the model parameters in the direction of the negative gradient. The parameter θ updates are performed as follows:

$$\theta^{t+1} = \theta^{t} - \eta \nabla \mathscr{L}(f(x_i), y_i) \tag{2}$$

where t denotes the index of iterations and η is the learning rate during training. ∇ represents the operation for extracting the gradient.

FL is an emerging decentralized ML framework that facilitates collaborative model training across multiple institutions while preserving data privacy by avoiding the exchange of raw data. The fundamental principle of FL is to enable participants to train models independently on local devices while sharing only updated model parameters rather than raw data, thereby preserving data privacy.

Firstly, the central server generates an initial version of the global model and distributes it to all participating clients. Upon receiving the global model, each client utilizes its private data for personalized training using the GD algorithm. During this process, the client does not expose any raw data and only retains the updated local model parameters. After completing a round of local training, each client uploads the trained model parameters to the central server. Upon receiving updates from all clients, the central server aggregates them to generate a new global model. The most commonly used aggregation algorithm, FedAvg, computes a weighted average of the submitted parameters, where the weights are dynamically adjusted based on the data volume of each client to ensure fairness and representativeness [24]. Then, the aggregated model parameters ω^t are represented as follows:

$$\omega^t = \frac{1}{\mathcal{D}} \sum_{i=1}^{N} \mathcal{D}_i \theta^t \tag{3}$$

where D is the number of all samples in datasets from all FL clients. $\mathcal{D}_i$ denotes the one of them. Finally, the above steps constitute a complete FL cycle. This cycle is repeated until a stopping condition is met, such as convergence or observing no significant improvement in global model performance.

This paper aims to predict the stress levels of nurses using an FL framework based on continuous monitoring of physiological signals while preserving data privacy and providing valuable insights into the factors contributing to stress in healthcare environments. The dataset is collected from wearable sensors worn by nurses and includes both physiological and behavioral data [25]. Specifically, the dataset is captured from nurses using the Empatica E4 wearable devices. To ensure generalizability, data were collected from a total of 15 nurses, including both male and female participants. None of the participants were pregnant, had a history of heavy alcohol use, or suffered from other diseases. Specifically, the dataset captures vital signs, including orientation data from accelerometers and gyroscopes along the *X*-, *Y*-, and *Z*-axes, EDA, HR, skin temperature (ST), and timestamp information. The data are typically labeled with corresponding stress levels categorized into three levels: "High", "Normal", and "Low", based on self-reported measures. The number of samples for each class is approximately 8.5 million for "High", 0.8 million for "Normal", and 2.2 million for "Low".

Firstly, the dataset undergoes several preprocessing steps to prepare it for modeling. One critical step involves removing duplicate samples. If certain samples repeatedly appear in the dataset, the model may overfit to the features of these samples, ignoring the diversity of other instances. This reduces the model's ability to generalize and affects its performance on unseen data.

Pearson correlation analysis is used to assess the linear relationships between various physiological signals and stress levels [26]. This analysis is crucial as it helps eliminate features that are strongly correlated with each other, which would introduce redundancy in the prediction process. The Pearson correlation results ρ are obtained as follows:

$$\rho = \frac{\sum (X_i - \overline{X_i})(X_j - \overline{X_j})}{\sqrt{\sum (X_i - \overline{X_i})^2} \sum (X_j - \overline{X_j})^2} \tag{4}$$

where X_i and X_j denote two different features. $\overline{X_i}$ and $\overline{X_j}$ represent their respective mean values. After conducting the analysis, no features need to be eliminated. However, the timestamp is recorded in the format of year, month, day, hour, minute, and second, which is not suitable as input for a NN model. This feature is transformed into two new variables: "date of year" and "time of day." Specifically, the unit of "date of year" is in days, and the unit of "time of day" is in hours.

Then, potential class imbalance should be considered. Imbalanced sample sizes across different classes can cause the model to favor the majority class and neglect the minority class during training. Minority class samples are submerged during training, preventing the model from learning their features and resulting in poor performance for these classes. The undersampling is applied in this paper to balance the dataset, ensuring the model does not become biased toward the majority class. After this step, the numbers of samples in each class are all equal to about 773.6 thousand.

Another essential preprocessing step is *Z*-score normalization, which standardizes the features to ensure they contribute equally to the model. *Z*-score normalization transforms each feature to have a mean of zero and a standard deviation of one, preventing features with larger ranges from dominating the learning process. This is accomplished as follows:

$$X' = \frac{X - \overline{X}}{\sigma} \tag{4}$$

where σ denotes the standard deviation of the feature.

In the context of FL, the dataset is distributed among multiple clients in a manner that reflects both practical considerations and the non-independent and identically distributed (non-IID) nature of real-world data [27,28]. By assigning overlapping classes to each client's data, each client's data distribution is not identical, introducing a controlled form of non-IID data. This mirrors the real-world scenario where wearable device data is collected from different individuals. The dataset contains three stress-level classes, and the data is partitioned among three FL clients such that each client holds an equal number of samples from two specific classes. In detail, the distribution of data among FL clients is shown in **Table 1**.

Table 1. Data distribution among FL clients.

	"High"	**"Normal"**	**"Low"**	**Total**
Client 1	386,790	386,790	0	773,580
Client 2	0	386,790	386,790	773,580
Client 3	386,790	0	386,790	773,580

Moreover, for each client, the dataset is further split into a training and validation set, with 80% of the data allocated for training and 20% reserved for validation. This division allows each client to evaluate and fine-tune their local models while ensuring a sufficient amount of data for training. To ensure that the global model is validated across a diverse set of data, the validation data from all clients is aggregated. Half of the validation data from all clients is used to validate the global model, providing a comprehensive, cross-client dataset for evaluation. The remaining 50% of the validation data is held out and used to assess the final test set, offering an unbiased estimate of the model's generalization performance across different stress levels.

For the FL framework, an NN network is designed to process physiological data from wearable sensors and predict nurse stress levels. As shown in **Figure 3**, the model consists of an input layer, followed by two fully-connected layers with rectified linear unit (ReLU) activation functions, and a final fully-connected layer and a softmax output layer for classification. This architecture balances computational efficiency with sufficient model complexity to capture nonlinear patterns in physiological signals.

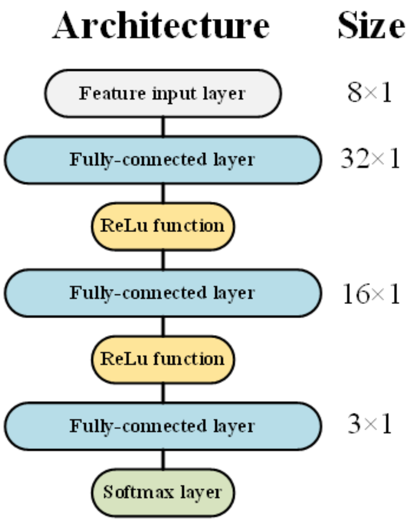

Figure 3. NN for nurse stress prediction.

5. Result and discussion

To ensure efficient and stable training of the NN model within the FL framework, the training hyperparameters are carefully configured. The training algorithm is the stochastic GD algorithm, which facilitates efficient parameter updates while accommodating decentralized, non-IID data distributions. The learning rate is set to 0.01 to balance convergence speed and stability. Each client performs local training with a mini-batch size of 1024, ensuring stable gradient estimation. Furthermore, each client updates the model using one local training epoch before transmitting updates to the central server. The cross-entropy loss is employed as the objective function for training. To enhance generalization, the samples are shuffled before training. The FL process runs for a total of 100 rounds.

Figure 4 presents the training loss of each FL client over iterations. As training progresses, all clients demonstrate a rapid initial decline in loss, eventually converging toward a stable minimum. Notably, the loss values of clients 1 and 2 converge more quickly compared to client 3. The faster convergence of clients 1 and 2 is attributed to differences in data distribution. Additionally, periodic spikes in the loss curves are observed, which can be explained by the model aggregation process at the server during each round. This aggregation introduces significant updates to the model parameters, temporarily increasing the loss before it stabilizes again.

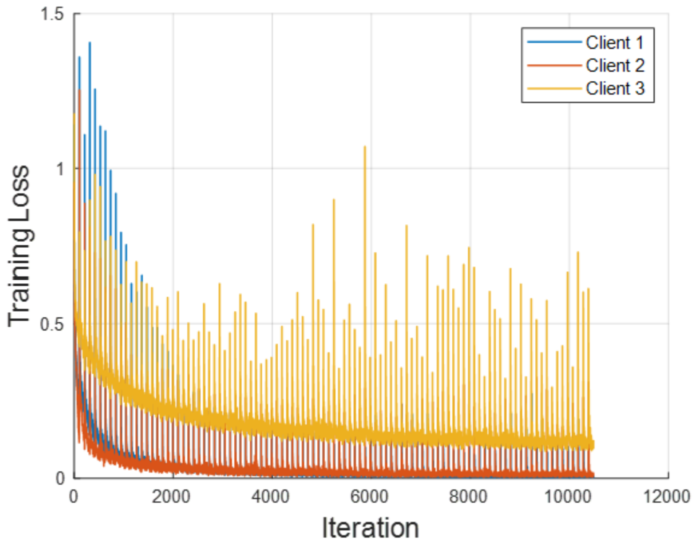

Figure 4. Training loss of FL clients.

Figure 5 illustrates the training accuracy of the three clients over iterations. All clients demonstrate a consistent upward trend in accuracy as training progresses, eventually converging to values above 95%. Client 2 achieves the fastest convergence and maintains the highest accuracy throughout the training process, closely followed by Client 1. In contrast, Client 3 converges more slowly and reaches a slightly lower final accuracy. Fluctuations in accuracy are observed across all clients during training. The observed fluctuations are also attributed to the model aggregation process at the server. Despite differences in convergence rates, all clients exhibit substantial improvements in accuracy over time, indicating the overall effectiveness of the FL training process.

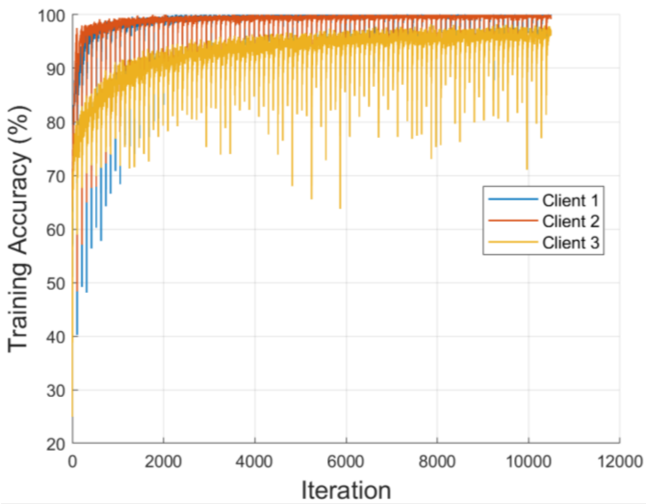

Figure 5. Training accuracy of FL clients.

Figure 6 illustrates the validation loss curves for the three clients during training. All clients exhibit a general downward trend in validation loss as training progresses, indicating continuous improvement in model performance. Compared to the training loss, the validation loss shows fewer fluctuations, particularly in the later iterations. The inset in **Figure 6** magnifies the early training phase, where periodic spikes in validation loss are observed. Despite these fluctuations resulting from aggregation, all clients demonstrate a clear convergence toward lower validation loss over time. This suggests effective generalization of the models on validation data and highlights the stability of the training process in later iterations.

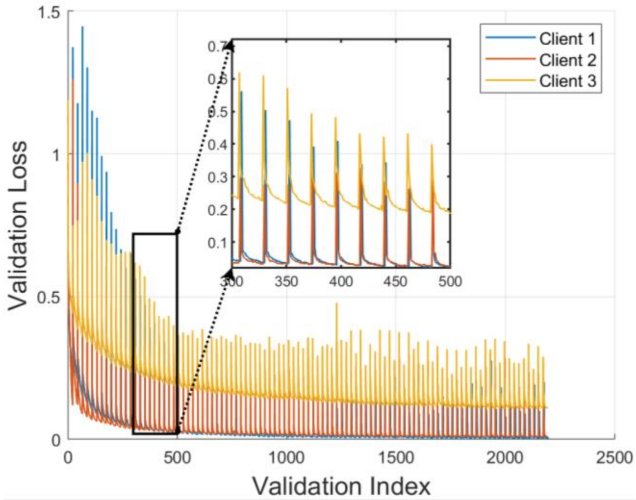

Figure 6. Validation loss of FL clients.

Figure 7 depicts the validation accuracy curves for the three clients during training. All clients demonstrate a steady increase in accuracy in the early stages, with each eventually surpassing 95% validation accuracy. Client 3 converges more slowly and attains a slightly lower final accuracy compared to Clients 1 and 2. The inset in **Figure 7** focuses on the validation index range between 400 and 700, where periodic fluctuations in accuracy are observed. The slower convergence and marginally lower final accuracy of Client 3 may be influenced by data heterogeneity. Nevertheless, the consistent upward trend and high final accuracy across all clients underscore the effectiveness of the training process in improving model generalization.

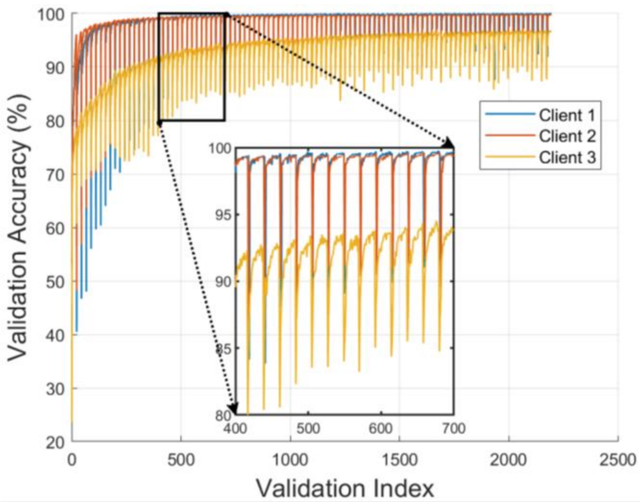

Figure 7. Validation accuracy of FL clients.

Figure 8 presents the normalized confusion matrix for the three stress classes. The diagonal elements represent the correctly classified instances, with the highest accuracy observed for the "Normal" class, followed by the "Low" and "High" stress classes. The off-diagonal elements indicate instances of misclassification between the different stress levels. The results highlight the strong overall performance of the FL model in predicting nurse stress levels. However, the slight variations in accuracy across classes suggest that the model is more effective at identifying the "Normal" stress state compared to "Low" and "High" stress levels. Despite these differences, the high accuracy across all classes demonstrates the effectiveness and reliability of the proposed FL approach.

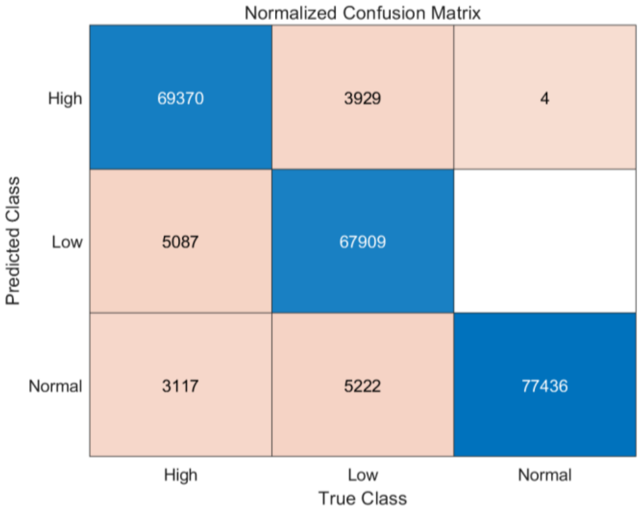

Figure 8. Confusion matrix of FL.

Figure 9 illustrates the accuracy progression of the global model within the FL framework. The global model's accuracy increases rapidly during the initial FL rounds, rising from approximately 50% to 85% within the first 20 rounds. After this initial phase, the rate of improvement slows, with accuracy gradually stabilizing around 90% by round 40. Minor fluctuations in accuracy are observed in the later rounds. The rapid accuracy improvement in the early rounds indicates efficient knowledge aggregation from the participating clients. Despite these fluctuations, the global model demonstrates effective convergence toward high accuracy, highlighting the robustness and stability of the FL training process.

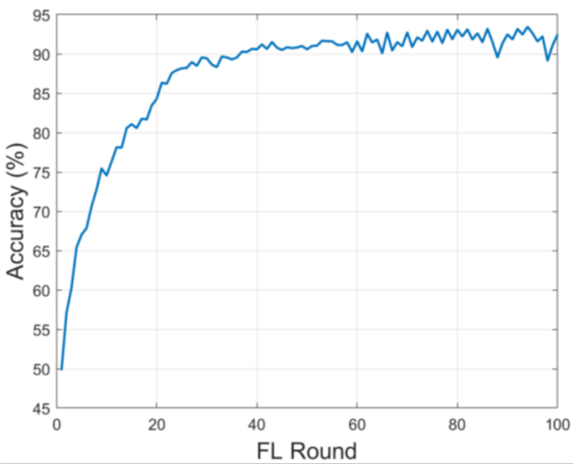

Figure 9. Accuracy versus FL round.

Figure 10 compares the performance of FL and centralized learning (CL) for predicting nurse stress levels across three stress classes. In the CL framework, all samples are aggregated into a single dataset for training, whereas in FL, data remains distributed across clients. The *y*-axis represents precision, recall, and F1-score metrics for both frameworks. For all stress classes, both FL and CL achieve high performance, with all metrics exceeding 85%. CL slightly outperforms FL in terms of recall and F1-score, with differences generally less than 0.1. The precision of FL is comparable to that of CL. The overall accuracy achieved by CL is 0.97, while FL attains 0.93. The slightly lower performance of the FL framework, particularly in recall and F1-score, may be attributed to the challenges posed by data distribution differences. Despite this gap, the FL model maintains strong performance across all metrics, demonstrating its feasibility for nurse stress prediction without centralized data aggregation. These results highlight the practicality and effectiveness of the FL approach, offering a privacy-preserving alternative to traditional CL while achieving competitive predictive accuracy.

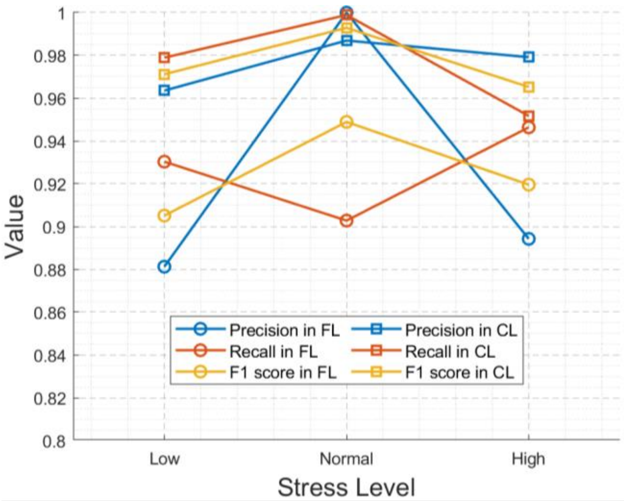

Figure 10. Performance in FL and CL.

6. Conclusion

This paper presents an FL framework for nurse stress prediction that integrates biomechanical data while preserving data privacy. Using wearable devices worn by nurses, biomechanical data such as orientation and EDA are captured, analyzed, and processed for stress prediction. An elaborate NN model is designed for FL, and client datasets are allocated considering the non-IID nature of the data. The results show that the performance, including accuracy, is comparable to that of CL, demonstrating the effectiveness of the approach, which provides an example of stress prediction based on biomechanical data while keeping sensitive information local, contributing to wearable distributed computing for timely detection of healthcare issues. Considering the limited number of stress-related features and the dynamic nature of stress, we plan to collect additional physiological data and deploy the trained model on embedded processors to enable real-time detection.

Author contributions: Conceptualization, KL and WX; methodology, KL; software, KL and DH; validation, KL and WX; formal analysis, DH; investigation, DH; resources, DH; data curation, KL; writing—original draft preparation, KL, WX and DH; writing—review and editing, KL, WX and DH; visualization, KL; supervision, KL; project administration, KL; funding acquisition, KL. All authors have read and agreed to the published version of the manuscript.

Ethical approval: Not applicable.

Conflict of interest: The authors declare no conflict of interest.

References

1. Sani MM, Jafaru Y, Ashipala DO, et al. Influence of work-related stress on patient safety culture among nurses in a tertiary hospital: a cross-sectional study. BMC nursing. 2024; 23(1): 81.
2. Lenzo V, Quattropani MC, Sardella A, et al. Depression, anxiety, and stress among healthcare workers during the COVID-19 outbreak and relationships with expressive flexibility and context sensitivity. Frontiers in Psychology. 2021; 12: 623033.
3. Hosseini S, Gottumukkala R, Katragadda S, et al. A multimodal sensor dataset for continuous stress detection of nurses in a hospital. Scientific Data. 2022; 9(1): 255.
4. Purcell SR, Kutash M, Cobb S. The relationship between nurses’ stress and nurse staffing factors in a hospital setting. Journal of Nursing Management. 2011; 19(6): 714-720.
5. Babapour AR, Gahassab-Mozaffari N, Fathnezhad-Kazemi A. Nurses’ job stress and its impact on quality of life and caring behaviors: a cross-sectional study. BMC nursing. 2022; 21(1): 75.
6. Gedam S, Paul S. A review on mental stress detection using wearable sensors and machine learning techniques. IEEE Access. 2021; 9: 84045-84066.
7. Chen J, Abbod M, Shieh JS. Pain and stress detection using wearable sensors and devices—A review. Sensors. 2021; 21(4): 1030.
8. Gil-Martin M, San-Segundo R, Mateos A, et al. Human stress detection with wearable sensors using convolutional neural networks. IEEE Aerospace and Electronic Systems Magazine. 2022; 37(1): 60-70.
9. Gonzalez-Carabarin L, Castellanos-Alvarado EA, Castro-Garcia P, Garcia-Ramirez MA. Machine Learning for personalised stress detection: Inter-individual variability of EEG-ECG markers for acute-stress response. Computer methods and programs in biomedicine. 2021; 209: 106314.
10. Hafer JF, Vitali R, Gurchiek R, et al. Challenges and advances in the use of wearable sensors for lower extremity biomechanics. Journal of biomechanics. 2023; 157: 111714.
11. Hughes GT, Camomilla V, Vanwanseele B, et al. Novel technology in sports biomechanics: Some words of caution. Sports Biomechanics. 2024; 23(4): 393-401.
12. Bonawitz K, Eichner H, Grieskamp W, et al. Towards federated learning at scale: System design. Proceedings of machine learning and systems. 2019; 1: 374-388.
13. Konečný J, McMahan HB, Ramage D, et al. Federated optimization: Distributed machine learning for on-device intelligence. arXiv; 2016.
14. Hussain F, Hussain F, Ehatisham-ul-Haq M, et al. Activity-aware fall detection and recognition based on wearable sensors. IEEE Sensors Journal. 2019; 19(12): 4528-4536.
15. Ilyas CMA, Nasrollahi K, Rehm M, et al. Rehabilitation of traumatic brain injured patients: Patient mood analysis from multimodal video. In: Proceedings of the 2018 25th IEEE International Conference on Image Processing (ICIP); 2018.
16. Wong CK, Ho DTY, Tam AR, et al. Artificial intelligence mobile health platform for early detection of COVID-19 in quarantine subjects using a wearable biosensor: protocol for a randomised controlled trial. BMJ open. 2020; 10(7): e038555.
17. Regalia G, Onorati F, Lai M, et al. Multimodal wrist-worn devices for seizure detection and advancing research: focus on the Empatica wristbands. Epilepsy research. 2019; 153: 79-82.
18. Sundararajan K, Georgievska S, Te Lindert BH, et al. Sleep classification from wrist-worn accelerometer data using random forests. Scientific reports. 2021; 11(1): 24.

19. Delmastro F, Di Martino F, Dolciotti C. Cognitive training and stress detection in mci frail older people through wearable sensors and machine learning. Ieee Access. 2020; 8: 65573-65590.
20. Morris D, Saponas TS, Guillory A, et al. RecoFit: using a wearable sensor to find, recognize, and count repetitive exercises. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems; 2014.
21. Dasari SB, Mallareddy CT, Annavarapu S, et al. Detection of Mental Stress Levels Using Electroencephalogram Signals (EEG). In: Proceedings of the 2023 2nd International Conference on Futuristic Technologies (INCOFT); 2023.
22. Mathur A, Sethia D. Body sensor-based multimodal nurse stress detection using machine learning. In: Proceedings of the 2024 16th International Conference on COMmunication Systems & NETworkS (COMSNETS); 2024.
23. Ruder S. An overview of gradient descent optimization algorithms. arXiv; 2016.
24. McMahan B, Moore E, Ramage D, et al. Communication-efficient learning of deep networks from decentralized data. In: Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS) 2017; 2017.
25. Hosseini S, Gottumukkala R. Nurse Stress Prediction Wearable Sensors. Kaggle; 2023.
26. Hall MA. Correlation-based feature selection of discrete and numeric class machine learning. In: Proceedings of the Seventeenth International Conference on Machine Learning (ICML 2000); 2000.
27. Zhao Y, Li M, Lai L, et al. Federated learning with non-iid data. arXiv; 2018.
28. Hsieh K, Phanishayee A, Mutlu O, Gibbons P. The non-iid data quagmire of decentralized machine learning. In: Proceedings of the International Conference on Machine Learning; 2020.