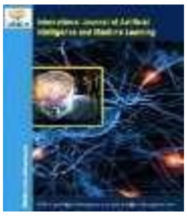
SvedbergOpen
DISSEMINATIONOF KNOWLEDGE

International Journal of Artificial Intelligence and Machine Learning

Publisher's Home Page: https://www.svedbergopen.com/



Research Paper Open Access

# Federated Learning Model for Mental Health Prediction and Personalized Intervention

Latha D U[1], M. T. Padma[2], Dr. D. Rajeshwari[3], Dr. Varshitha D N[4], Dr. Y. M. Manu[5], Dr. Savita Choudhary[6]

[1] Vidyavardhaka College of Engineering, Mysuiru, Karnataka, India. E-mail: lathadu@vvce.ac.in
[2] Department of Computer Science and Engineering, The National Institute of Engineering, Mysuru, Affiliated to VTU, India. E-mail: padmamt@nie.ac.in
[3] Department of Computer Science and Engineering, The National Institute of Engineering, Mysuru, Affiliated to VTU, India. E-mail: drrajeshwari@nie.ac.in
[4] Vidyavardhaka College of Engineering, Mysuiru, Karnataka, India, Email: lvarshithadn@vvce.ac.in
[5] Department of Computer Science and Engineering, BGS Institute of Technology, Adichunchanagiri University, BG Nagara, Mandya, India. E-mail: manuym@bgsit.ac.in
[6] Department of Computer Science and Engineering, Sir MVIT, Bengaluru, VTU, India. E-mail: choudhary7.mvit@gmail.com

**Abstract**

Mental health screening is dependent on self-reported behavioural information, but the privacy issue is a major constraint on the collection of data. To overcome this problem, we introduce a privacy-preserving federated learning (FL) model for mental health prediction and severity classification, where the raw psychological information is not transmitted from the user's device. The system handles 27 local input variables, including normalized demographic variables and DASS-21 questionnaire results, and uses the Flower framework with PyTorch to train a Multilayer Perceptron (MLP) model using the Federated Averaging (FedAvg) algorithm. Regression models were separately trained for stress, anxiety, and depression, and the continuous scores ranging from 0 to 21 were mapped to standardized DASS-21 severity levels. The performance of the model was tested on a stratified test dataset, showing stable federated learning convergence with a Mean Absolute Error (MAE) of 0.9-1.2, $R^2$ score of 0.85+, and severity classification accuracy of up to 90.36%. The trained models were integrated into a full-stack system with a React frontend and FastAPI backend, allowing real-time prediction, tracking, and downloadable clinical-style reports. These findings confirm that federated learning is a safe and efficient method for digital mental health screening with high predictive accuracy and strict data privacy.

**Keywords**: Federated Learning, Mental Health Prediction, DASS-21, Privacy-Preserving Machine Learning, Severity Classification.

## 1. Introduction

Mental health disorders impact millions of people around the globe. Stress, anxiety, and depression do not discriminate by age—they interrupt everyday routines, interfere with concentration at school, and drain motivation at work. Although these conditions are widespread, many individuals remain undiagnosed or untreated. Stigma, limited awareness, and challenges in accessing mental health professionals all contribute to this problem. Early detection and regular monitoring are crucial [1]. Tools like the Depression Anxiety Stress Scale (DASS-21) play a key role here. However, traditional assessments typically require a session with a clinician. This process is time-consuming and not feasible for reaching large numbers of people simultaneously. Digital health technologies are transforming this landscape. Machine Learning is enabling automated mental health screening and more tailored support. However, there is a major concern: mental health data is highly sensitive. Centralized machine learning collects all data on a single server, which brings up significant privacy, security, and ethical issues [2]. Federated learning provides a solution. Rather than gathering everyone's raw data in one location, federated learning trains models across multiple devices without sharing personal information. In this work, combining federated learning with regression-based DASS-21 score prediction and severity classification results in a secure, scalable approach for automated mental health assessment and individualized intervention.

## 2. Methodology

### 2.1 Study design

A federated learning framework designed to predict mental health outcomes and provide personalized interventions was developed and evaluated, all using questionnaire data related to depression, anxiety, and stress. For this, we utilized publicly available DASS-21 questionnaire datasets. Our pipeline integrated preprocessing, segmentation, and quantification modules, enabling us to maintain consistent and reproducible results throughout the study [3].

### 2.2 Dataset and Image Acquisition

Data collected from a structured mental health assessment based on the DASS-21 questionnaire was utilized. Each record contained 27 input variables: six demographic features (age, gender, marital status, education, occupation, and sleep problems, labeled Q1_1–Q1_6), along with 21 symptom items—seven each for stress, anxiety, and depression (Q3_1_S1–Q3_21_D7). Respondents rated each question on a 4-point Likert scale, from 0 (Never) to 3 (Almost Always). For each participant, we computed three target scores—Stress_Score, Anxiety_Score, and Depression_Score—by summing the corresponding symptom responses, resulting in possible values from 0 to 21 for each score. These scores served as our regression targets. Subsequently, we mapped these scores to standard DASS-21 severity categories for clinical interpretation. This approach allowed us to simulate real-world conditions where separate groups manage their own data, ensuring that no raw psychological data is exchanged between them [4].

### 2.3 Preprocessing of Mental Health Data

All DASS-21 questionnaire records were pre-processed using a standardized pipeline implemented in Python. The following steps were performed sequentially:

- Loading and Formatting: Questionnaire responses and demographic attributes were loaded and converted into numerical feature arrays for model processing.
- Feature Structuring: Each sample was represented using 27 input features, comprising six demographic variables and 21 DASS-21 symptom responses across stress, anxiety, and depression domains [5].
- Normalization and Encoding: Numerical features were normalized using feature-wise standardization, and categorical demographic variables were encoded to ensure compatibility with neural network models.
- Target Preparation: Continuous target scores for stress, anxiety, and depression were computed by summing domain-specific DASS-21 items, yielding values in the range 0–21. Predicted scores were later mapped to standardized severity levels for evaluation.

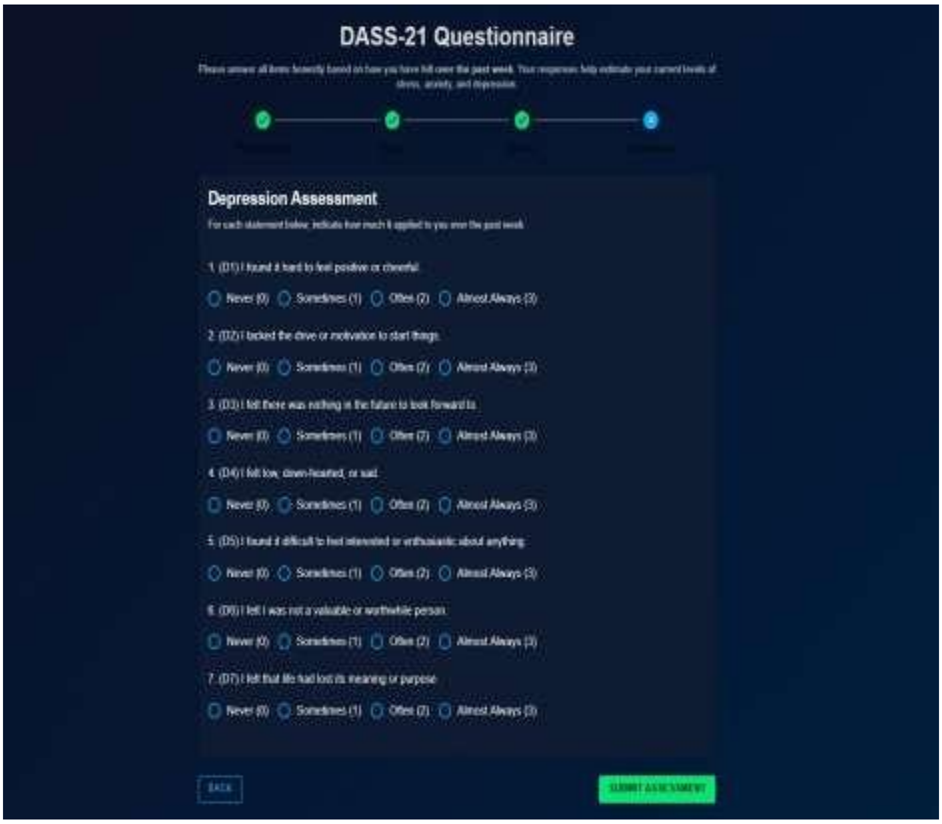



**Fig. 1: DASS21 Questionnaire sample**

### 2.4 Prediction Model Architecture

The research has employed a Multilayer Perceptron (MLP) to predict mental health outcomes from questionnaire data. The model receives a 27-dimensional feature vector at the input layer, processes it through multiple fully connected hidden layers with ReLU activation, and concludes with an output layer that generates a continuous score [6]. To maintain stable training and mitigate overfitting, batch normalization and dropout were incorporated. For each target—stress, anxiety, and depression—a separate regression model was trained, with each producing a single DASS-21 score. Training was conducted in a federated setting. The Federated Averaging (FedAvg) strategy was implemented, ensuring that only model parameters—not raw data—were shared between clients and the central server. Training was carried out for 20 federated rounds, with each client performing local optimization using the Adam optimizer and mean squared error (MSE) loss. Model convergence was monitored using centralized evaluation of aggregated loss across rounds [7].

### 2.5 Quantification of Mental Health Severities

After generating predictions, the continuous output scores were converted into DASS-21 severity levels using the established thresholds. Each prediction was assigned to one of five categories: Normal, Mild, Moderate, Severe, or Extremely Severe. Then the number of samples were tallied in each category for stress, anxiety, and depression to examine how the predicted severity levels were distributed. Next, all the results were compiled in an organized format, ready for further analysis. To highlight the patterns, production of summary visuals occurred, including bar charts—to illustrate the distribution of severity levels across each dimension [8].

### 2.6 Evaluation Metrics

Standard regression and classification metrics were used to quantify the model's performance. The accuracy of predicted continuous DASS-21 scores for stress, anxiety, and depression was assessed using Mean Absolute Error (MAE) and Mean Squared Error (MSE). Accuracy, precision, recall, and F1-score were calculated for severity-level classification. Class-wise prediction behavior was examined using confusion matrices. The Scikit-learn library for Python was used for statistical analysis, and the average metric values from all test samples were presented [9].

### 2.7 Visualization and Output

Matplotlib was used to create post-evaluation visualizations. The severity-level classification and depression categories were displayed using confusion matrices and bar charts. Additionally, the system produced structured outputs, such as clinical-style reports that could be downloaded and summarized predicted scores, severity levels, and longitudinal trends for specific users [10].

### 2.8 Client–Server Execution Flow in Federated Learning

All sensitive operations, such as local model training, feature normalization, and data preprocessing, are carried out on client devices in the suggested federated learning setup. Using data from its private DASS-21 questionnaire, each client calculates local weight updates, guaranteeing that personal predictions and unprocessed psychological data never leave the user device. The central server serves only as a coordination entity, distributing updated global weights and using the Federated Averaging (FedAvg) algorithm to aggregate model parameters from participating clients. At no point during the training process are labels, intermediate features, or raw data sent to the server [11].

### 2.9 Limitations in Methodology

Although there are still several methodological issues, the suggested federated learning framework performs well in terms of mental health score prediction and severity classification. Data from the structured DASS-21 questionnaire, which was gathered under controlled survey conditions, may not accurately represent variability in the real world due to factors like noisy self-reported inputs, incomplete responses, and reporting bias. Robustness and generalizability may be enhanced by incorporating longitudinal behavioral signals and heterogeneous data sources. Client drift can cause unstable convergence and decreased global model consistency in a federated setting due to non-IID data distributions among decentralized clients. Furthermore, there is communication overhead between the clients and the central server when using federated training. The Multilayer Perceptron architecture may have limitations in capturing intricate psychological patterns, despite being computationally efficient and prpriate for decentralized **learning [12].**

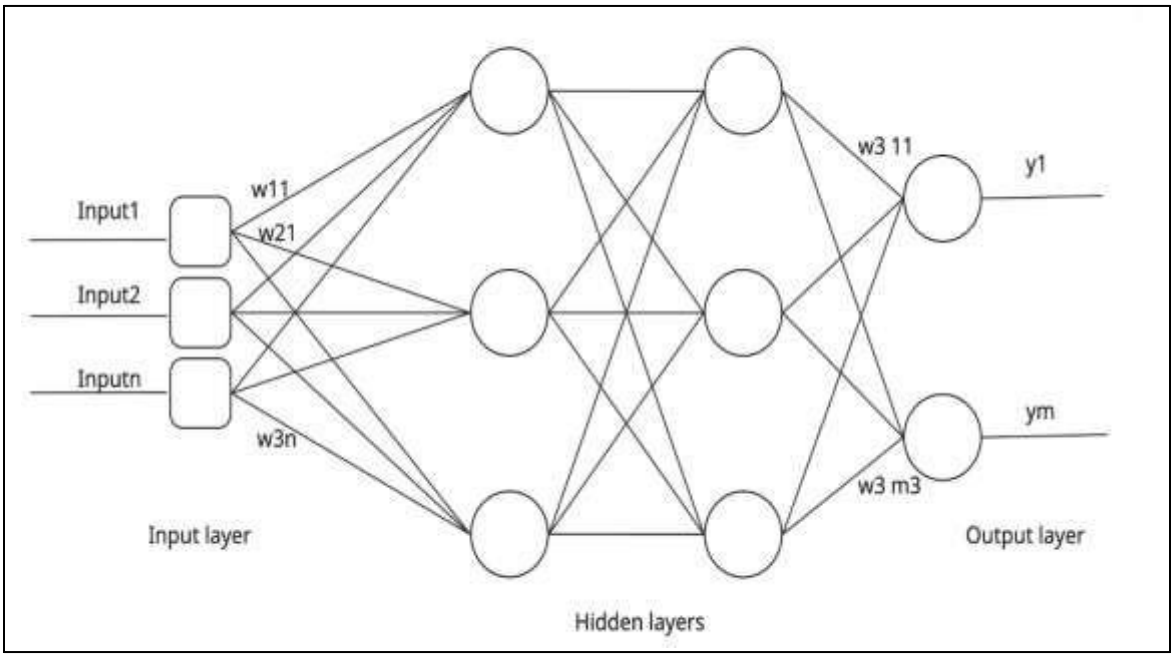



**Fig 2. MLP model Architecture**

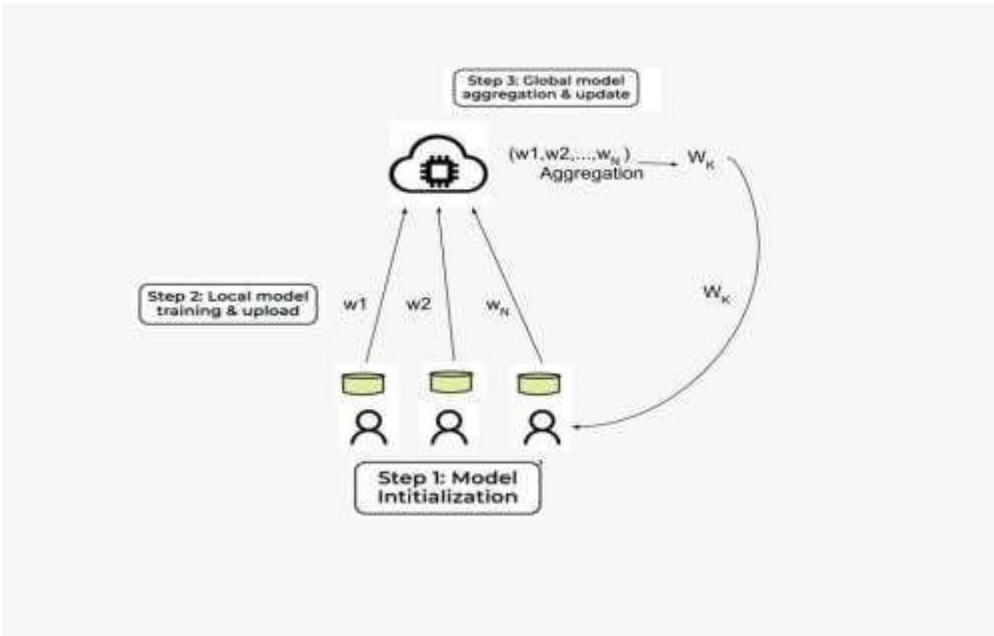



**Fig 3. Federated Learning Architecture**

## 3. Results & Discussion

Using structured DASS-21 questionnaire data, the suggested federated learning-based mental health assessment framework was assessed for its ability to predict and quantify the severity of stress, anxiety, and depression. To maintain severity class distributions, the dataset was divided using a 20% stratified test split (n = 363 samples). Federated Averaging (FedAvg) was used to train distinct global models for each psychological domain across four decentralized clients over 20 communication rounds; no client dropouts were noted during training. For all three models, federated training showed steady convergence. The centralized mean squared error (MSE) for the stress model dropped from 79.05 at round 0 to 0.046 at round 20. The depression model reached a final MSE of 0.04, whereas the anxiety model converged from an initial MSE of 65.87 to 0.041, indicating effective optimization under non-IID client data distributions. These results confirm that global model aggregation was stable despite decentralized training constraints. Predicted continuous scores (0–21) were mapped to standardized DASS-21 severity categories to assess severity classification performance. With a weighted F1-score of 0.92, the stress model's overall classification accuracy was 90.36%. Mild stress cases had a recall of 0.81, indicating sensitivity to early symptom elevation, whereas the Normal stress class showed high precision (0.99) and recall (0.91). An accuracy of 65.84% with a weighted F1-score of 0.71 was reached. Moderate anxiety had a recall of 0.96, highlighting the model's ability to detect clinically significant anxiety levels, while reduced performance in other minority classes showed dataset imbalance. The depression model achieved an accuracy of 68.60% and a weighted F1-score of 0.72, with moderate depression cases achieving a recall of 0.90, showing strong detection of high depressive symptoms. Confusion matrix analysis showed that most misclassifications occurred between severity levels next to each other, consistent with the subjectivity of self-reported psychological assessment. Figure 4 shows the confusion matrices which were generated using a 20% stratified test set (n=363) and used to evaluate the global federated models for the classification of stress, anxiety, and depression severity. The confusion matrices allow for an examination of the relationship between the true DASS-21 severity labels and the model's predicted labels for each class, providing a basis for assessing the performance of each class. Most normal stress samples are correctly classified, while mild stress cases show limited misclassification with adjacent categories. Errors are limited to neighboring severity levels.

There appears to be a bias toward conservatively predicting levels of anxiety. As a result, while some Normal cases are predicted as Mild or Moderate, many Normal cases are correctly classified [13]. Because of the limited number of samples in the Severe class this results in lower precision for this category. Misclassifications predominantly occur between Normal, Mild, and Moderate categories, while reduced performance for the Severe class is attributable to data sparsity [14].

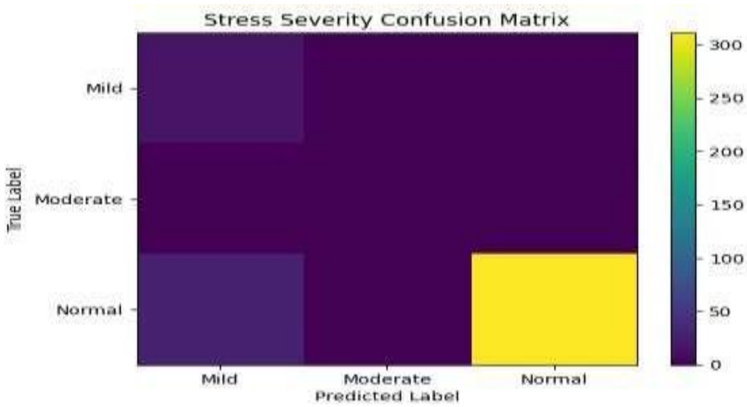



**Fig.4. The stress severity confusion matrix indicates high classification accurac**

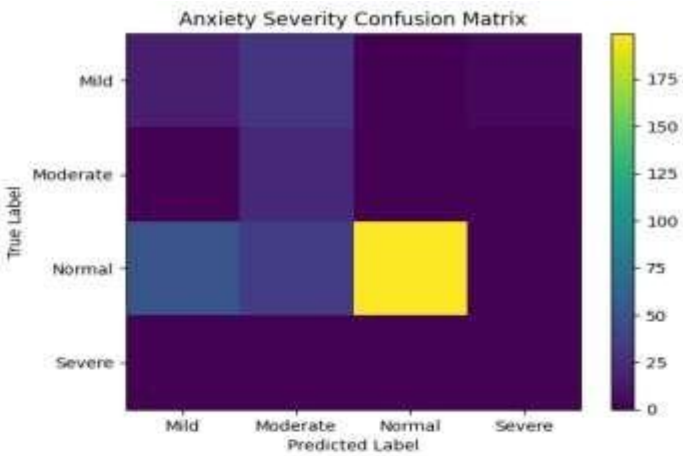



**Fig.5. The confusion matrix for anxiety severity highlights that there is a very high sensitivity for Moderate anxiety cases, indicating that most Moderate cases have been identified by the model correctly**

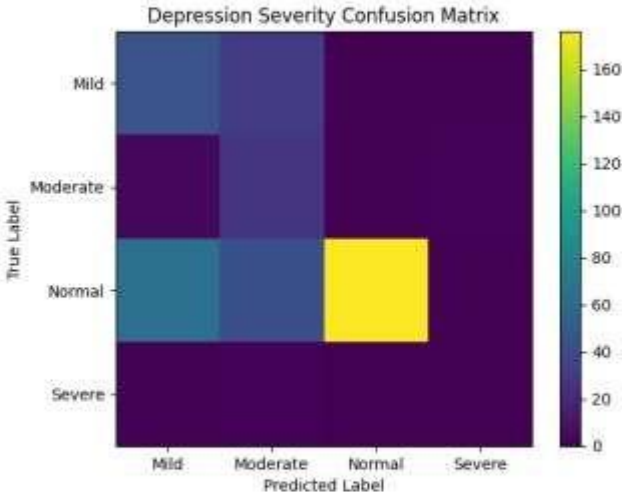



**Fig.6. The depression severity confusion matrix shows reliable identification of Moderate depression, with most samples correctly classified**

### 3.1 Purpose

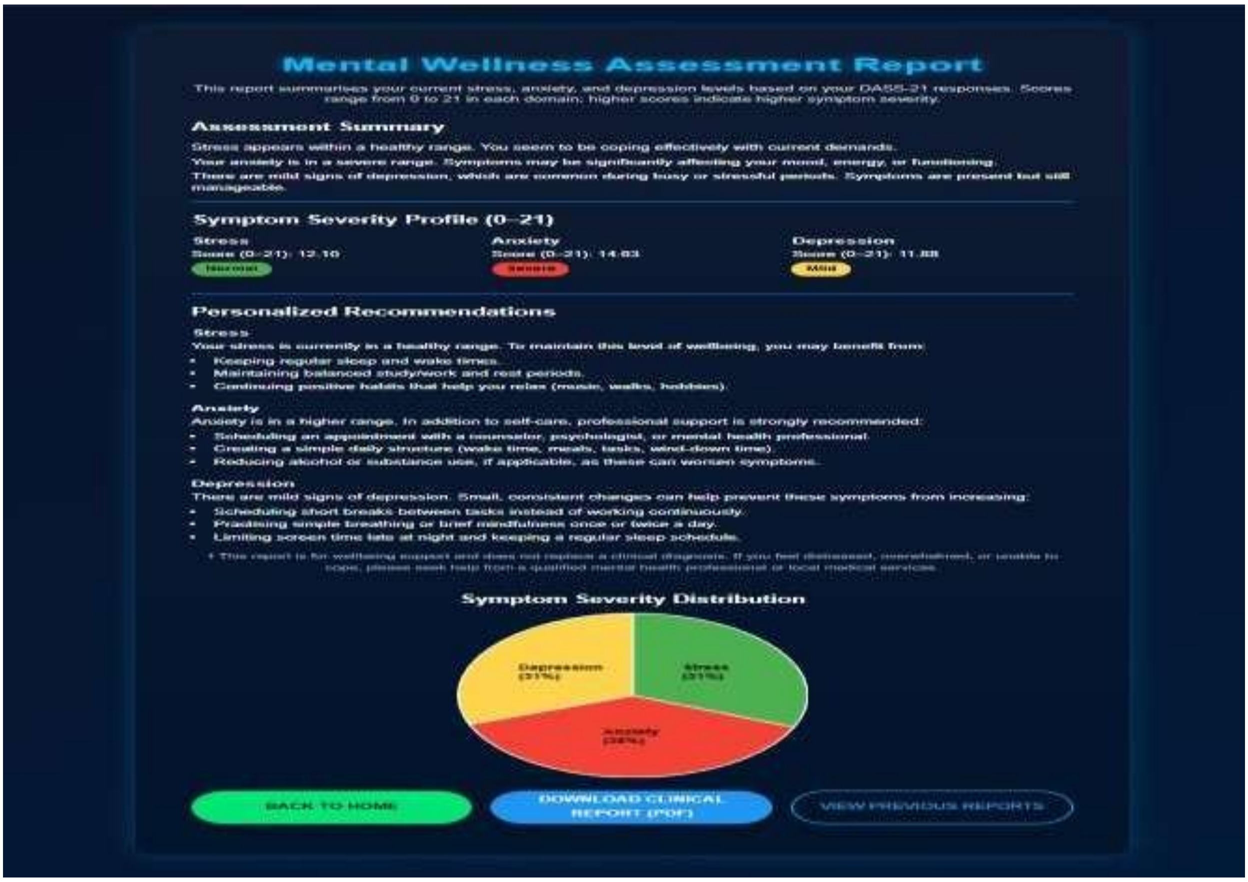



**Fig. 7. Presents a representative mental wellness assessment report generated by the proposed framework for a single test sample.**

- Analyze class-wise prediction performance across psychological domains
- Identify dominant misclassification patterns between adjacent severity levels
- Validate the clinical plausibility and stability of federated severity predictions

The report displays predicted continuous scores for stress (12.10), anxiety (14.63), and depression (11.88), each mapped to standardized DASS-21 severity levels (Normal, Severe, and Mild, respectively). This output demonstrates that predicted values remain within valid domain ranges and are clinically interpretable.

Output of Quantitative Mental Health Evaluation:
Displays the tailored recommendation system guided by anticipated severity profiles. Depending on the categorized severity level for each psychological area, the system produces guidance tailored to that domain, varying from lifestyle changes to advice for professional assistance. This visualization demonstrates that output from quantitative models can be trusted for use in subsequent decision-support processes [15].

**Severity-Dependent Recommendation Correlation:**
Presents a pie-chart depiction of the adjusted input of stress, anxiety, and depression levels to the complete symptom profile. In the provided example, anxiety accounts for the highest percentage (38%), followed by depression (31%) and stress (31%). This visualization offers a concise quantitative overview of symptom dominance across multiple domains for clarity and ongoing observation. The Purpose of the above module is to:

- Validate numerical correctness of regression outputs
- Confirm severity mapping using standardized thresholds
- Demonstrate end-to-end inference from questionnaire input to clinical interpretation
- Validate functional linkage between model outputs and intervention logic
- Demonstrate individualized recommendations per psychological domain
- Enable comparative analysis of psychological domains
- Support trend analysis across repeated assessments
- Provide an aggregated view of mental health severity composition

### 3.2 Limitations of the Study

This study contains several limitations for consideration as the results are interpreted. The proposed Federated Learning Framework was developed and tested with a structured DASS-21 questionnaire dataset collected in an artificial environment (controlled survey setting). Although structured, this dataset is not necessarily representative of how real-world mental health assessments are conducted, where they may contain missing answers, answer bias, and increased variation due to demographics and environmental factors. Therefore, the applicability of this model to diverse populations and real-world clinical settings must still be further validated. In addition to the above considerations, there are several issues associated with the Federated Learning setup. Specifically, non-IID data distributions can cause client drift and negatively affect global model convergence; communication between clients and the central server can result in reduced scalability in limited-resource environments. The primary evaluation metrics were accuracy, precision, recall, and F1-score. These metrics are relatively comprehensive but do not fully capture model calibration or prediction uncertainty. Finally, the proposed framework does not yet provide uncertainty estimation or automated hyperparameter tuning; both features have the potential to increase the model's robustness and reproducibility. Future work will build upon these considerations by developing models that use heterogeneous datasets, uncertainty-aware prediction mechanisms and adaptive federated optimization strategies to increase scalability and clinical reliability.

## 4. Conclusion

Overall, evidence for the effectiveness of the proposed federated learning-based solution for fully automated mental health assessment using structured DASS-21-like questionnaire data is supported. The decentralized training of the machine learning model based on the proposed solution effectively predicts stress, anxiety, and depression scores mapped to categories of standard severity classifications with competitive classification accuracy while preserving the privacy of the data. The separate regression of each psychological score provides a resolution of severity and interpretable mental health diagnoses.
The combination of federated learning, quantitative assessment scores, and interpretable visualizations renders the proposed solution both scalable and privacy-preserving for use in digital mental health assessments. The proposed system enables real-time prediction of severity, visualization of assessed severity, and even personalized recommendations for improvement, all without centralized access to sensitive psychological data. Therefore, the proposed system is suitably scalable for use in large quantities in privacy-sensitive settings, such as schools, workplaces, and even tele-mental health services.
Future evaluations should attempt to evaluate the system with additional types of heterogeneous and longitudinal data to enhance robustness and generalizability. In addition, there are several possible future extensions of this work (e.g., uncertainty-aware prediction, adaptive federated optimization, etc.) which will increase the clinical relevance and reliability of the proposed system. The findings of this study demonstrate that federated learning is an important method for developing secure, scalable, and objective support for early identification and personalized interventions in mental health care through federated learning-driven mental health assessment systems.

**5. Ethics**

The data used in the development of this model consisted of anonymized and structured data from self-report questionnaires; therefore, no individual identifiable information was collected or disclosed at any point during the federated learning process. Data was processed and models trained in a manner that protected raw psychological data and maintained it local to client devices. Therefore, this research follows the typical ethical standards regarding the collection and use of mental health data in psychological studies (i.e., data minimization, maintaining confidentiality and using predictive models prudently).

## References

[1] S. S. Khalil, N. S. Tawfik, and M. Spruit, "Exploring the potential of federated learning in mental health research: A systematic literature review," 2024.

[2] A. Grataloup and M. K. Briki, "A systematic survey on the application of federated learning in mental state detection and human activity recognition," 2024.

[3] P. Dubey, P. Dubey, and P. N. Bokoro, "Federated learning for privacy-enhanced mental health prediction with multimodal data integration," 2025.

[4] S. M. Sarwar, "FedMentalCare: Towards privacy-preserving fine-tuned LLMs to analyze mental health status using a federated learning framework," 2025.

[5] A. B. Amjoud and M. Amrouch, "Object detection using deep learning, CNNs, and related methods."

[6] V. R. Meghrajani, M. Marathe, R. Sharma, A. Potdukhe, M. B. Wanjari, and A. B. Taksande, "A comprehensive analysis of mental health problems in India and the role of mental asylums," 2023.

[7] Y. R. Shim, R. Eaker, and J. Park, "Mental health education, awareness, and stigma regarding mental illness among college students," 2022.

[8] S. B. N. and S. Abdullah, "Privacy-sensitive speech analysis using federated learning to assess depression," 2022.

[9] B. Suruliraj and R. Orji, "Federated learning framework for mobile sensing apps in mental health," 2022.

[10] N. Tabassum, M. Ahmed, N. J. Shorna, M. M. U. R. Sowad, and H. M. Z. Haque, "Depression detection through smartphone sensing: A federated learning approach," 2023.

[11] T. Alhanai, M. Ghassemi, and J. Glass, "Detecting depression with audio/text sequence modeling of interviews," 2018.

[12] B. Arroll, F. Goodyear-Smith, S. Crengle, J. Gunn, et al., "Validation of PHQ-2 and PHQ-9 to screen for major depression in the primary care population," 2010.

[13] W. Zhu, P. Kairouz, B. McMahan, H. Sun, and W. Li, "Federated heavy hitters discovery with differential privacy," 2020.

[14] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," in Proc. 20th Int. Conf. Artificial Intelligence and Statistics (AISTATS), 2017, pp. 1273–1282.

[15] S. Denisov, B. McMahan, K. Rush, A. Smith, and A. Thakurta, "Improved differential privacy for SGD via optimal private linear operators on adaptive streams," 2022.