Journal of Computing & Biomedical Informatics
ISSN: 2710 - 1606

Volume 11 Issue 02
2026

Research Article
https://doi.org/10.56979/1102/2026/1481

A Federated Learning Model for Privacy-Preserving and Personalized Mental Health Monitoring using Big Data Analytics

Aisha Abdalla Alansari[1], Muhammad Adnan Khan[2,3,4], Saif Jasim Almheiri[1], Arselan Ashraf[5*], Sheikh Tahir Bakhsh[6], and Muhammad Amir Khan[7]

[1]School of Computing, University of Sharjah, University City, Sharjah, United Arab Emirates.
[2]International Center for Materials Sciences and Technology, Western Caspian University, Baku, AZ1001, Azerbaijan
[3]Riphah School of Computing, Faculty of Computing, Riphah International University, Lahore Campus, Lahore, 54000, Pakistan.
[4]Faculty of Management and Hospitality at Spectrum International University College (SIUC), Selangor, Malaysia.
[5]Faculty of Computing and Informatics, Multimedia University, Cyberjaya, 63100, Malaysia.
[6]Cardiff School of Technologies, Cardiff Metropolitan University, Cardiff, UK.
[7]Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia.
[*]Corresponding Author: Arselan Ashraf. Email: arselan@mmu.edu.my

Received: March 08, 2026 **Accepted:** July 05, 2026

Abstract: Mental health disorders are rising globally, posing challenges in early detection, personalized care, and data privacy. The conventional centralized Machine Learning (ML) models involve the transfer of sensitive patient information to central servers, a factor that jeopardizes privacy and reduces user confidence. In addition, these models normally do not consider individual differences in behavior and thus generalized predictions are made, which are less accurate. To address these shortcomings, this research proposed a Federated Learning (FL)-based model of mental health monitoring that allows training models without raw data sharing in distributed medical customers. The model incorporates the concept of Big Data Analytics (BDA) to handle heterogeneous mental health records and assist with user-level customization and meaningful evaluation based on overall performance measures. The simulation outcomes indicate that the proposed model has a global accuracy of 99.53% and a miss rate of just 0.47% which beats conventional centralized methods. This research highlights a scalable, secure, and personalized solution for mental healthcare, advancing the integration of federated learning and big data in digital health systems.

Keywords: Federated Learning; Mental Health; Mental Health Monitoring; Big Data Analytics

1. Introduction

Over the last few years, the frequency of mental health problems has escalated drastically as society, economy, and the environment present more stressors [1]. Occupational stress, financial difficulties, family issues, and exposure to violence have greatly contributed to the development of psychological complexes like depression, anxiety, and bipolarity [2,3]. The World Health Organization (WHO) reported a total of 450 million people across the world with mental health challenges, which constitutes about 13 percent of the disease burden in the world [4,5]. Depression in itself has more than 350 million individuals, ranked as one of the major causes of suicide, shortening life span, and quality of life [6]. The negative effects of psychiatric illnesses are widespread, and they are on the increase. WHO approximates that approximately 1 out of 8 individuals worldwide has some form of mental illness, and in 2019, approximately 970 million were estimated to be living with mental illnesses [7]. Such troubling data points to the increasing sense of urgency to create more affordable, accessible, and scaled mental care across the board rather than traditional models of care.

Research consistently demonstrates a strong association between psychological distress and technology-related behaviors among young people and students. During the COVID-19 pandemic, healthcare students experienced high levels of anxiety and depression, which significantly affected their academic performance, concentration, motivation, and overall learning process, highlighting the adverse impact of mental health challenges on educational outcomes [8]. Similarly, Fekih-Romdhane et al. reported that technology addictions are positively associated with schizotypal traits, with depression, anxiety, and stress acting as significant mediating factors, emphasizing the complex psychological consequences of excessive technology use [9]. Furthermore, Bitar et al. found that cyberbullying perpetration is significantly associated with increased levels of anxiety, depression, and suicidal ideation among adolescents, suggesting that problematic online behaviors not only affect victims but also have detrimental mental health consequences for perpetrators [10]. Collectively, these studies underscore the critical need for interventions that address anxiety, depression, and unhealthy technology use to promote psychological well-being and improve educational and social outcomes. Historical mental health monitoring strategies, including face-to-face-based psychological examinations, self-filled surveys, and custodial Electronic Health Records (EHRs), have been central in determining and treating mental illness. These approaches, however, are prone to late intervention, small-scale, and low reporting as they are associated with the stigma and the need to protect privacy. Additionally, data collection frameworks are centralized, posing critical risks in terms of data security, consent, and institutional biases when handling mental health data, in particular, sensitive data types [11]. Such deficiencies are a threat to proactive care and early identification, which are essential to addressing long-term mental health outcomes. To explain the workflow that usually takes place between mental health patients, healthcare facilities, and service providers, it is graphically represented in Figure 1.

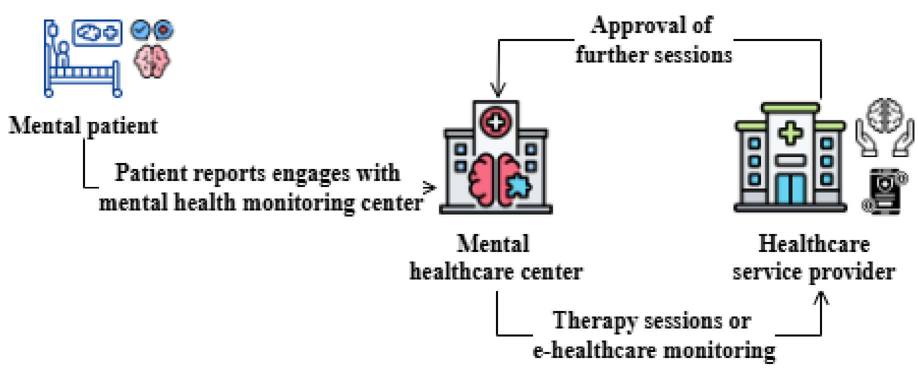

Figure 1. Interaction between mental patients, hospitals, and healthcare service providers

Figure 1 illustrates the traditional communication loop among a mental health patient, a mental healthcare center, and the healthcare service provider [12]. The patient is initially presented with symptoms or contacts the mental health monitoring center. According to the preliminary assessment, the mental healthcare facility organizes therapy sessions or refers the patient to a healthcare service provider to receive further advice and other sessions. The service provider can then approve ongoing care or prescribe remote e-health monitoring. Though a workflow of this nature enables organizing care delivery, it also suggests the centralized, sequential nature of traditional mental health services, which can lead to delays in diagnosis, a lack of personalization, and problems with data privacy and accessibility.

To overcome such gaps, the medical domain increasingly adopts Artificial Intelligence (AI) and ML [13,14] to support real-time monitoring, prognostic diagnosis, and personal therapeutic suggestions [15,16]. Mental health surveillance is one of the areas where AI-powered technologies can be utilized to identify signs of psychological stress in the digital footprint of wearable sensor records, smartphone usage history, and clinical records, and thus enable the identification of emotional stressors early on. However, the vast majority of extant artificial-intelligence systems are still based on centralized data repositories, making them susceptible to unintentional information leaks and privacy violations [17,18]. In addition, these architectures often do not take into account the individualized paradigm of mental-health assessment, and thus the prognostications are broad brushes that do not match the heterogeneity of individual users.

These restrictions outline a curious gap in modern mental health surveillance infrastructures, that is, a lack of scalable, privacy-protective, and individualized system that could operate within decentralized health care environments. This gap, in its turn, gives a strong incentive to the implementation of FL that offers a safer and more resilient alternative to the traditional centralized AI-based paradigms.

These boundaries characterize a permeable gap in modern mental health surveillance frameworks, i.e., lack of a scalable, privacy-sensitive, and individualized framework that can work at decentralized healthcare ecosystems. This gap, in its turn, gives a strong incentive to the implementation of Federated Learning that offers a safer and more resilient alternative to the traditional centralized AI-based paradigms.

FL has also emerged as an appealing approach to privacy-preserving and collaborative model training in healthcare to resolve these issues [19-24]. As opposed to traditional distributed methods, FL allows many institutions, or edge devices, to cooperate in training ML models without high volumes of raw data being shared. This decentralized system allows not only for preserving user privacy but also for managing the data heterogeneity and improving personalization. Together with BDA, FL can be effectively used to process diverse and large mental health data across institutions and provide scalable and context-specific monitoring; this can happen without disregarding data protection laws and regulations [25-29]. A privacy-preserving mental healthcare monitoring based on an FL architecture is illustrated in Figure 2.

Figure 2 depicts the FL architecture in mental health monitoring, whereby local models are trained on the mental health clients' devices using their privacy data. The local models are then transmitted to a central server to undergo secure aggregation, maintaining privacy by design whilst allowing it to learn collaboratively across users. This configuration enables privacy-preserving, scalable, and real-time mental healthcare support to take place without centralized data exchanges.

The study introduced a new FL structure [30-34] that suits privacy-sensitive and personalized monitoring of mental health via BDA. The proposed model mitigates these limitations of currently available approaches by integrating federated intelligence, BDA-driven preprocessing and scalable data handling, and individual-level modelling to provide a fully integrated and secure platform of mental health monitoring. The model has undergone rigorous validation thus providing a solid basis to its use in practical environments both in clinical practice and digital health ecosystems.

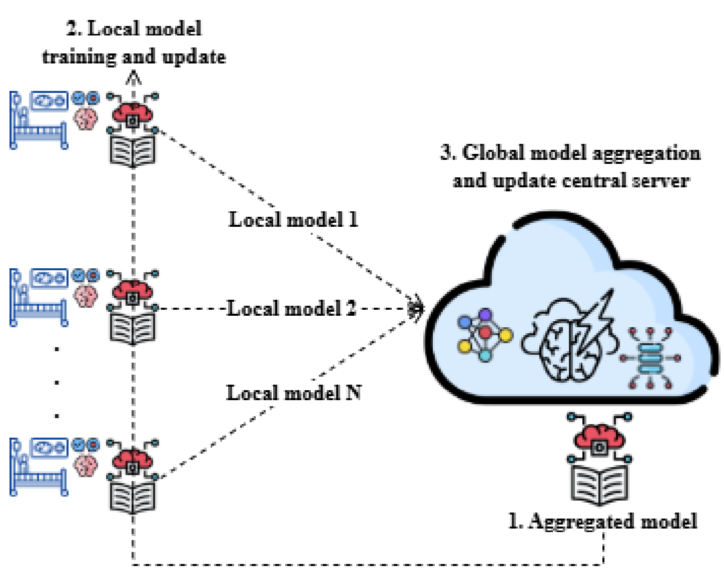

Figure 2. FL architecture for privacy-preserving mental healthcare monitoring

2. Literature Review

The increasing number of mental health disorders has urged many researchers to study new ways to resolve them, focusing on the efficient diagnosis, constant observation, and individual treatment of the victims. Some studies specifically addressed the issues of early symptom identification, underreporting based on social stigma, lack of real-time insights, and serious data privacy concerns. To this, machine-learning approaches have been utilized more frequently to learn behavioral patterns, anticipate risks of mental-health, and aid clinical decision-making. However, so far most of these solutions are based on centralized data systems, creating massive problems in terms of privacy, data silos, and the overall absence of personalization. The section therefore provides a critical analysis of traditional approaches, artificial intelligence, and emerging trends in federated learning systems in mental-health monitoring.

2.1. Traditional and Conventional Methods

Analyses of electroencephalographic (EEG) signal, self-reported questionnaires and manual clinical evaluation have historically been relied upon as early methods of mental health monitoring. The authors in the contribution cited as [35] gave an extensive overview of approaches that use EEG-based signal analysis to measure mental stress, thus identifying EEG as a promising modality of early mental distress detection. The paper highlighted various salient issues such as lack of consistency in the previous studies because of difference in experimental design, choice of cortical regions, stressor types, preprocessing protocol, feature extraction techniques and classification software. The researchers also stressed the paramount significance of a prudent choice of features and stated that EEG signals have intrinsic complexity that requires careful treatment. In an attempt to maximize the accuracy and reliability of detecting mental stress, the paper suggests a combination of cortical activation measures with brain-connectivity networks, as well as the use of deep learning methods. Regardless of these contributions, the review admits the long-standing limitation due to the absence of standardised methodologies in this field.

2.2. AI-Based Solutions for Mental Health Monitoring

The recent events have highlighted the effectiveness of artificial intelligence and machine learning in identifying trends related to mental health using online media and behavioral information. In [36], the authors outline a machine-learning-supported model that can be used to conduct proactive surveillance of mental health, specifically focusing on Twitter as the knowledge source. The research develops one of the ideas of a so-called social sensor cloud, where users of social media act as sensing agents by addressing the flaws of traditional systems of health administration, particularly when it comes to mass mental-health surveillance and at-risk screening. The framework has a pattern-based data-cleaning mechanism, which uses regular expressions, to improve the accuracy of subsequent sentiment analysis. In predictive modelling Long Short-Term Memory networks have been exploited to pick out people at increased risk, due to the events that are delineated in nature. Empirical studies prove that the model is better than traditional methods, thus validating its ability to mitigate mental-health disorders at early and scalable stages.

In [37], an AI-based model is outlined to track mental health through identifying suicidal ideation on social media, namely, Twitter, which will fulfill the urgency of timely interventions in the global context of increasing suicide rates and the inefficacy of traditional monitoring systems. As an approach to overcoming these obstacles, the framework uses sophisticated AI methodologies- e.g., NLP and ML- and strict data preprocessing, e.g., tokenization, stemming and feature extraction based on TF-IDF and count vectorization. A Random Forest classifier was chosen due to its capability to operate with high-dimensional text and the emotional aspects of such data, and its strong performance with 85% accuracy, 88% precision, 83% recall, and an AUC-PR of 0.93, shows that it is reliable in real-world use. The system can be used in the analysis of scalable mental-health intervention since it was established to process real-time data to detect the suicide risk patterns with the minimal proportion of false positives, but it can be platform-specific which could limit its applicability to other populations and social media settings.

The authors of [38] discussed the aspect of detecting depression using user-generated material on social media, specifically the analysis of tweets through machine-learning approaches. To address the rising need to identify mental health disorders at an early stage, the researchers presented a large set of 632,000 tweets and put it through a thorough preprocessing and careful feature filtering. The models that were experimented with included not only the classic classifiers such as logistic regression and Naive Bayes but also state-of-the-art transformer-based models, such as DistilBERT, DeBERTa, SqueezeBERT, and RoBERTa, all of them trained and properly evaluated. Among the models that were tested, RoBERTa showed the best performance with an accuracy of 0.981 and an average cross-validation accuracy of 0.97 during ten folds. These results highlight the inclination of models based on transformers to outperform the conventional in the field of depression-sensing using social media data. However, the research admits that the use of publicly available twitters may introduce bias and limit the ability of generalizing findings to different demographic groups or other online platforms.

In [39] an all-inclusive machine-learning and natural-language-processing model was worked out to estimate the affective reactions of the masses expressed in the form of social-media publishes with respect to significant events. The system is driven by the urgency to track the emotional responses during a crisis, like the COVID-19 pandemic, and inform prompt governance responses, such as specifically mental-health responses. This dataset consisted of 15,455 tweets that were categorized as eleven affective groups based on the Plutchik psycho-color wheel and coded using Bag-of-Words and TF-IDF. It used predictive modelling based on the classifiers of Random-Forest and logistic-regression estimators and provided the overall accuracy of 95.5%. Analysis of the Twitter feed on COVID-19 vaccines demonstrated the effectiveness of the model in monitoring population-wide trends in affect. Although the framework shows a strong predictive accuracy, it may have a limited scope to generalize to other people or platforms based on the reliance on a single dataset and technological reliance.

According to [40], a novel deep learning framework designed to overcome the limitations of traditional EEG-based mental health monitoring methods. Prior approaches often relied on manually engineered features and basic classifiers, which struggled to capture the intricate spatiotemporal relationships within EEG signals, leading to poor adaptability and reduced classification accuracy. To address these gaps, the EEG Mind-Transformer integrates three key components: a Dynamic Temporal Graph Attention Mechanism (DT-GAM) for temporal dependency modelling, a Hierarchical Graph Representation and Analysis (HGRA) module to capture local and global brain interactions, and a Spatial-Temporal Fusion Module (STFM) for comprehensive EEG signal representation. Experimental evaluations across multiple datasets demonstrate robust performance, achieving 92.5% accuracy, 91.3% recall, 90.8% F1-score, and a 94.2% AUC. This architecture presents a significant advancement in EEG-based mental health monitoring by improving generalizability, accuracy, and interpretability in clinical and research applications.

In [41] and [42], the literature highlights the incorporation of privacy-conscious intelligence in the modern healthcare systems through specific methodological approaches. Specifically, in [41], the concept of FL is cleverly combined with the fuzzy ensemble deep learning architectures; namely, Temporal Convolutional Networks (TCN), Long Short-Term Memory (LSTM) networks, and Gated Recurrent Units (GRU) to enable highly precise, but privacy-aware, recognition of emotion based on electroencephalographic (EEG) data. In [42] takes the discussion a step further by applying an optimised differential privacy algorithm, thus, ensuring the securitisation of release of healthcare data and preserving a significantly higher utility of the underlying data. Taken together, these studies provide an insight into the urgency of striking the right balance between strong privacy controls and reliable data analysis capabilities in the Internet-of-Medical-Things (IoMT) and Connected IoT (CIoT) systems.

Table 1. Comparative analysis of previously published approaches

Ref.	**Model Used**	**Objective**	**Preprocessing Techniques**	**Predictive Model**	**Privacy-Preserving (FL)**	**Scalability**
[35]	EEG-based methods (Review)	Review EEG methods for	Emphasized EEG cleaning, feature extraction	Deep learning suggested	✗	Limited

		mental stress detection				
[36]	Social Sensor Cloud + LSTM	Proactive mental health monitoring via Twitter	Regular expression-based tweet cleaning	LSTM	✗	High
[37]	AI + NLP + RF	Detect suicidal ideation via Twitter posts	Tokenization, stemming, TF-IDF, count vectorization	Random Forest	✗	High
[38]	ML & Transformer Models (e.g., RoBERTa)	Detect depression from tweets	Preprocessing and feature selection	RoBERTa (best performer)	✗	High
[39]	ML-NLP with RF & LR	Detect emotional trends from social media during events	Bag of Words, TF-IDF	Random Forest, Logistic Regression	✗	High
[40]	EEG Mind-Transformer	Enhance EEG-based mental health monitoring with deep learning	EEG signal representation via spatiotemporal modeling	Deep Learning (Graph Attention + Fusion Modules)	✗	High
[43]	ML-based emotion prediction	Social media-based emotion detection	Not explicitly detailed	Not specified	✗	Moderate
[44]	Transfer Learning (Optimized)	Efficient real-time mental health monitoring	Algorithm-specific optimization	Optimized Transfer Learning with residual and gradient descent	✗	Moderate
[45]	ML models (LR, RF, DT, KNN)	Predict mental health status via questionnaire	Data normalization, QA curation	Logistic Regression, KNN, Decision Tree, RF	✗	Moderate
[46]	DNN, SVM, RF	Forecast mental well-being using diverse factors	Feature extraction from behavioral and contextual data	DNN, SVM, Random Forest	✗	High
Proposed Model (FL with BDA)		Privacy-preserving and mental healthcare monitoring	Data harmonization, client-side preprocessing, feature selection	Federated ML models	✓	High

Table 1 also provides a comparative overview of AI-based mental-health monitoring paradigms that exist, highlighting their goals, methodological basis, and other limitations thereof. Even though most of these frameworks have admirable levels of acuity and real-time sensitivity, they often lack privacy preservation, heterogeneous population scalability, or regulatory framework compliance. The proposed federated mental-healthcare monitoring architecture addresses these shortcomings by balancing the federated learning and blockchain-based data sharing to enable the provision of a secure, scalable and personalized continuum of care. This paradigm is therefore a more credible and operationally feasible solution than the antecedent methodologies.

3. Limitations of Prior Published Works

Although there is extensive academic research on AI-based monitoring of mental health, many of these central limitations remain unanswered in the literature [35–46]:

3.1. Lack of Privacy-Preserving Mechanisms

Majority of modern models rely on centralized data structures (e.g., [35-39, 45]), thus spawning significant risks of data leakage, user profiling and regulatory non-compliance. Even though [44,46] attempt to introduce real-time monitoring and forecasting, all these attempts do not involve FL, which is an essential element to maintain data privacy in distributed mental health applications.

3.2. Limited Scalability, Real-Time Capability, and Regulatory Integration

Numerous studies (e.g., [35, 40, 43]) do not have the potential to be scaled to various client settings or offer real-time analytics. In addition, most of the models do not consider regulatory compliance (e.g. HIPAA, GDPR) and do not promise reliable, explainable results, which reduces their use in clinical or national mental health systems.

4. Contribution of the Proposed Model

To overcome the shortcomings observed in existing literature, the proposed model offers the following key innovations:

4.1. Federated Learning-Based Privacy-Preserving Framework

Unlike prior models that rely on centralized data processing, the proposed model employs FL to ensure that user data remains on local clients. This enhances privacy, supports regulatory compliance (e.g., GDPR), and mitigates data leakage risks, addressing limitations found in [30–39].

4.2. Real-Time, Scalable Mental Health Monitoring with Big Data Support

The model integrates BDA for client-side preprocessing, feature harmonization, and efficient model aggregation. This enables scalable, real-time monitoring across decentralized environments, making it more adaptive and applicable to practical healthcare infrastructures compared to models in [36] [38] [45].

5. Proposed Methodology

Conventional centralized machine learning methods to observe mental health are often faced with barriers related to privacy, lack of scalability, and a lack of flexibility to nonhomogeneous populations. In comparison, Federated learning offers a decentralized model that allows joint model training to many clients without the need of raw data exchange. However, the challenges of federated learning include the communication overhead, the heterogeneity of the data, and the problem of maintaining the steady performance of models among the participants.

To address the defined difficulties, the current paper suggests a privacy-sensitive federated learning system enhanced with the capabilities of BDA, specifically designed to the field of mental health surveillance. The designed solution will allow a consortium of healthcare facilities to jointly train a common worldwide model and strictly protect patient confidentiality and complete compliance with existing data protection laws. The methodology provides secure, scalable, and intelligence-driven big-data solutions to mental health applications by localizing raw data on an institutional level and limiting inter-facility interactions to model parameters. The system, as shown in Figures 3 and 13, coordinates local model training, secure aggregation and global parameter updates in a client-server system, thus enabling the implementation of regulation-consistent mental health monitoring devices.

Figure 3 depicts the client-side of the proposed model, which is designed to process and analyze the data that relates to mental health which is collected through a distributed array of healthcare sensors and digital interfaces. It starts by acquiring data about a multitude of sources, such as Internet-of-Medical-Things (IoMT) sensors and mobile health applications, and captures such vital attributes as patient identification, age, diagnostic codes, and behavioral indicators. As detailed in Table 2, this set of data enables real-time, trustworthy, client-level monitoring of mental healthcare. The local node uses the techniques of BDA to preprocess the data received, which entails cleansing, normalization, and transformation of the data to ensure the highest quality and consistency of the data. The resulting filtered dataset is subsequently used to train models on-device thus maintaining privacy before any communication with the global aggregation process of the FL architecture.

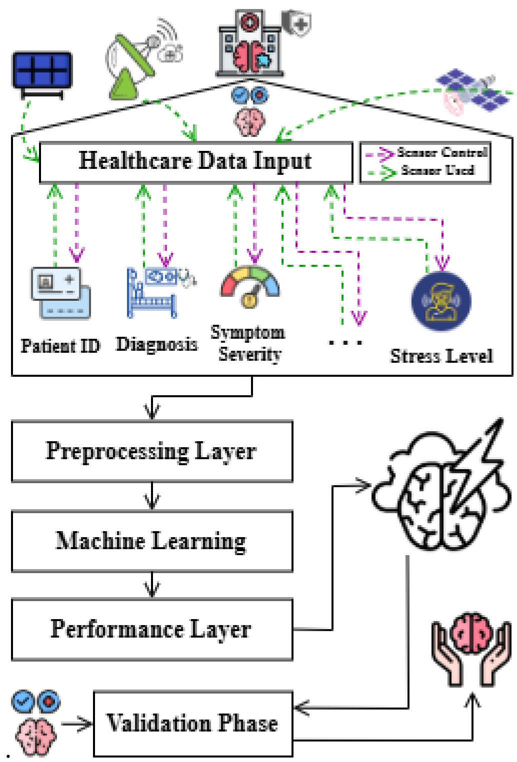

Figure 3. Proposed federated mental healthcare monitoring model (local client)

Table 2. Dataset Features Description

Features	Type	Description
Patient ID	Integer	Unique identifier assigned to each patient.
Age	Integer	Age of the patient in years.
Gender	Categorical	Biological gender of the patient (e.g., Male, Female).
Diagnosis	Categorical	Clinically diagnosed mental health condition.
Symptom Severity (1-10)	Integer	Severity level of symptoms rated on a scale from 1 (low) to 10 (high).
Mood Score (1-10)	Integer	Self-reported mood score from 1 (low) to 10 (positive).
Sleep Quality (1-10)	Integer	Quality of sleep rated from 1 (poor) to 10 (excellent).

Physical Activity (hrs/week)	Integer	Number of hours the patient exercises per week.
Medication	Categorical	Type of prescribed medication (e.g., Antidepressants, Mood Stabilizers).
Therapy Type	Categorical	Type of therapy being administered (e.g., CBT, Psychodynamic).
Treatment Start Date	Date	Date on which the mental health treatment began.
Treatment Duration (weeks)	Integer	Total duration of treatment in weeks.
Stress Level (1-10)	Integer	Perceived stress level from 1 (low) to 10 (high).
Outcome	Categorical	Overall result of treatment (e.g., Improved, Deteriorated).
Treatment Progress (1-10)	Integer	Progress rating from 1 (no progress) to 10 (fully recovered).
AI-Detected Emotional State	Categorical	Emotion detected by AI from behavioral/biometric cues (e.g., Anxious, Calm).

Once acquired, the data is sent to the preprocessing module where the data is subjected to strictly defined series of cleansing, normalization, and transformation processes. This step includes imputation methods, feature-level normalization, and noise reduction techniques, such as but not limited to label encoding, creative treatment of missing data, MinMax scaling, and Exploratory Data Analysis (EDA). These processes form an important part of the BDA pipeline, consequently enhancing the integrity, consistency, and applicability of expanse mental-health corpora on the whole. As a result, the resulting prepared data reaches a reliability condition, structural consistency, and optimization, making it fit to succeeding model training and assessment in the larger mental-health monitoring ecosystem.

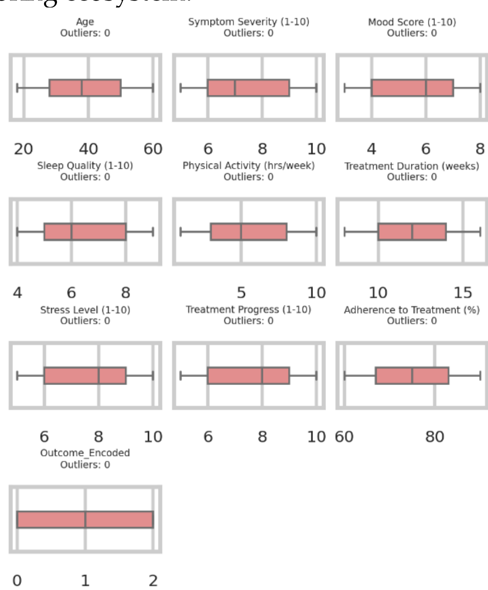

Figure 4. Outlier Detection in Mental Healthcare Data Using the Z-Score Method

Figure 4 presents an outlier analysis of the different numerical variables in the mental healthcare dataset using Z score method. A brief boxplot of a specific attribute (e.g., age, mood score, sleep quality, etc.) is given in each subplot, thus showing the underlying distribution, as well as all possible anomalies. The Z-score threshold of three was used, and the outcomes indicate that there were no significant outliers in all the parameters monitored. Besides this univariate screening, correlation heatmaps and pair-wise scatter plots were used to understand feature-interaction patterns to acquire multivariate behaviour of anomaly and inter-feature inconsistencies. Such a combined modelling ensures the reliability and statistical consistency of the dataset to be further trained in the context of the suggested federated mental health monitoring framework.

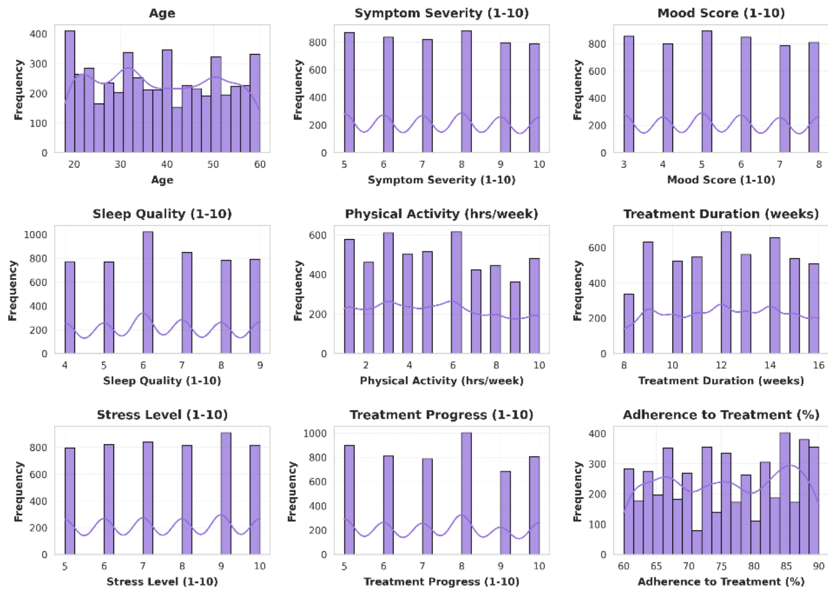

Figure 5. Distribution Analysis of Numerical Features using Histograms with KDE

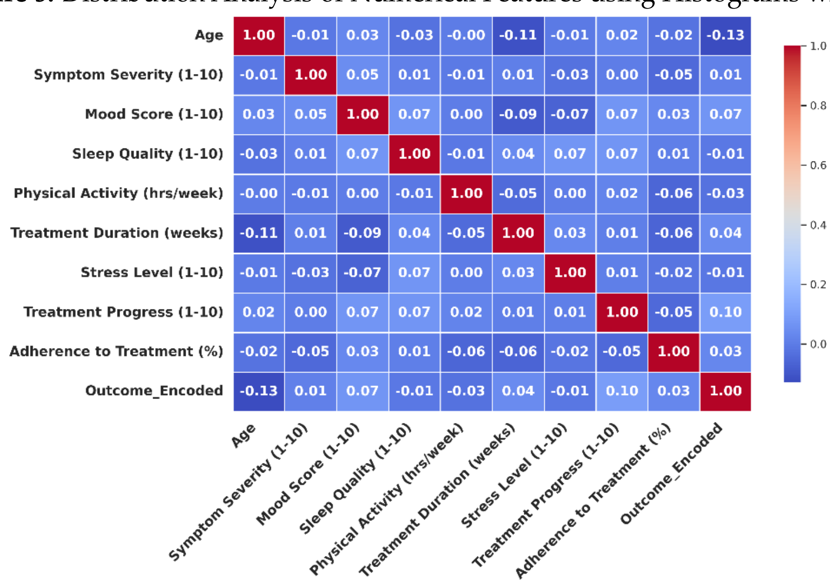

Figure 6. Feature Correlation Heatmap for Mental Health Attributes

Figure 5 presents the histograms along with Kernel Density Estimation (KDE) curves for the key numerical features in the mental healthcare dataset. This visualization is useful in understanding distributional features of the data, such as skewness and modality of the major features like age, mood score, treatment duration and level of stress. Most of these variables are multimodally distributed hence portraying varying mental health trends among the cohort. The curves of the kernel density estimations provide a smooth approximation of the trends of the frequencies which is useful in identification of central tendencies and dispersion. Based on such distributional knowledge, principled preprocessing decisions can be made, and the effective training of predictive models in the federated model is informed.

Figure 6 shows a heatmap of Pearson correlation coefficients between the numerical variables in the mental healthcare dataset. The intensity of colors depicts the strength and direction of the linear relationships between pairs of variables. Remarkably, most features show low correlations (near 0), which means that there is a weak multicollinearity, and the input variables remain independent of each other in ML models. This is to make sure that no one feature is over-representing the predictive process and thus models are trained more cheaply and generalized. These insights into correlation can help to make sound choices when it comes to choosing features and dimensionality reduction particularly under federated conditions.

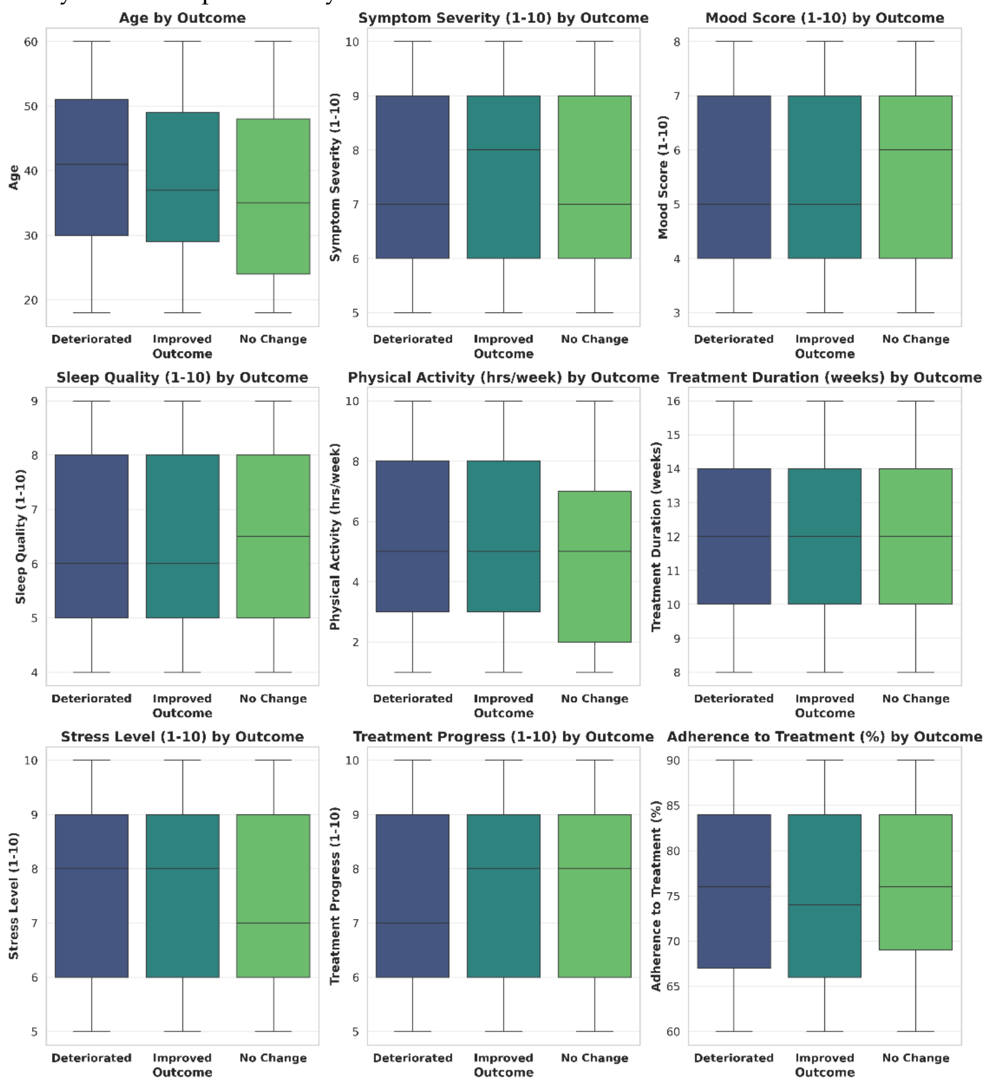

Figure 7. Class-wise Boxplot Comparison for Mental Health Features

Figure 7 visualizes how different numerical features vary across the three outcome classes: No Change, Deteriorated Outcome, and Improved. The subplots are the stratification of one feature by outcome category to visualize the relationship between treatment effectiveness and the changes in the variables of symptom severity, sleep quality, physical activity, and treatment adherence. The steady changes in median and interquartile ranges indicate some minor yet significant variations in patient behaviors and conditions among outcomes. The comparative insights will be critical to interpret the model and test whether features are relevant in mental health prediction trajectories.

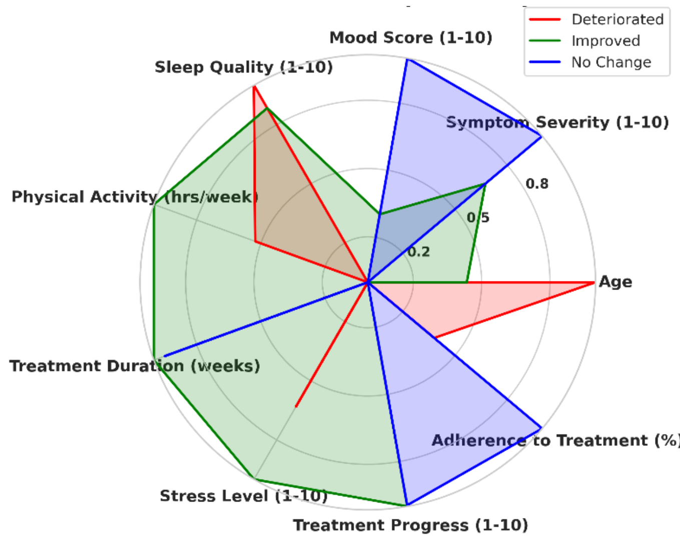

Figure 8. Radar Chart – Multivariate Comparison of Outcomes

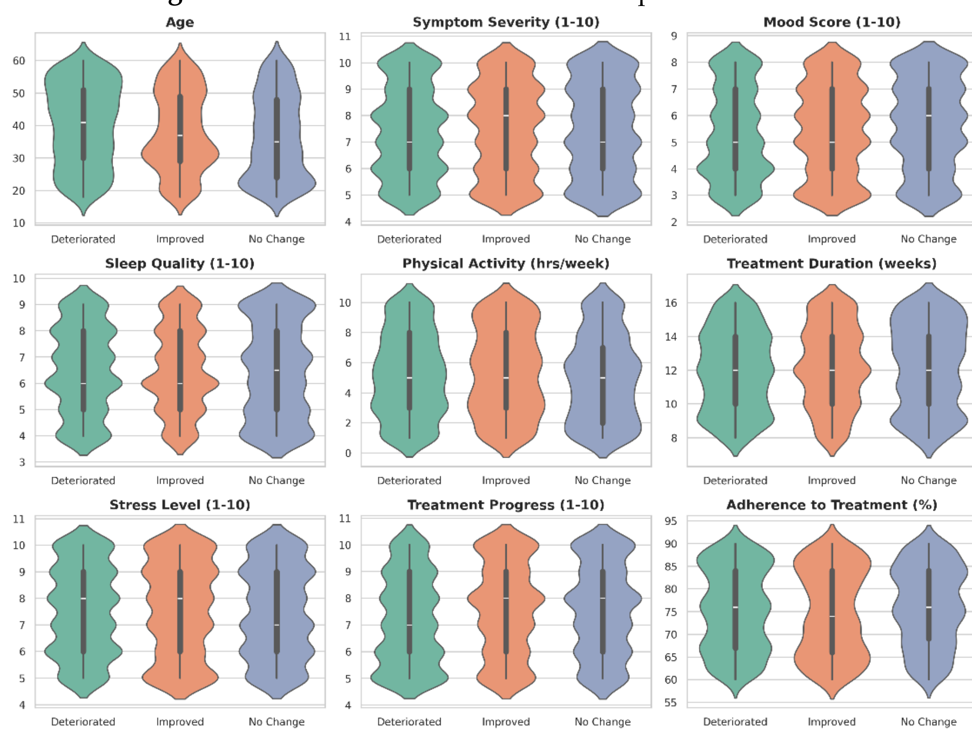

Figure 9. Violin Plots – Distributional Insights by Mental Health Outcome

Figure 8 presents a radar chart that enables visual comparison of key normalized mental health features across three patient outcome categories: Deteriorated, Improved, and No Change. Every spoke is a feature (e.g., Stress Level, Physical Activity, Sleep Quality), and the form of every line is the profile of the patients in that category result. As an example, the patients with Improved outcomes demonstrate more Symptom Severity and Physical Activity, whereas the patients with No Change have higher Mood Scores and Treatment Adherence. The Deteriorated group exhibits increased Stress Levels and Age. This visualization is a clear indication of the fact that various attributes change together in the context of the classes of outcomes, providing valuable information on patient stratification and individual intervention planning.

Figure 9 visualizes the distribution of numerical mental health features across three outcome classes (No Change, Deteriorated, Improved) using violin plots. These plots combine boxplot structure with KDE, offering a dual perspective on both central tendency and distribution shape. For instance, the distribution of Physical Activity shows a wider spread for the Improved group, while Stress Level appears more elevated and narrowly distributed for Deteriorated cases. This enables practitioners to discern subtle variations and asymmetries in feature behavior across outcomes, supporting more informed feature selection and patient profiling strategies.

Figure 10 presents two advanced dimensionality reduction techniques—UMAP (left) and t-SNE (right)—to visualize the high-dimensional mental health dataset in a 2D space. Every point is a patient record, coded in terms of mental health outcome (No Change, Deteriorated, Improved). In spite of the complexity and noisiness of mental health data which result in some overlapping, there are minor clustering trends. UMAP exposes localized groupings, but retains global and local structures, whereas t-SNE emphasizes more fine separations, but with a more widely spaced layout. The projections are useful to evaluate the intrinsic separability of the results and determine the selection of appropriate modeling strategies.

Figure 11 presents a dual visualization of the distribution of mental health outcomes using both a bar chart and a pie (donut) chart. The bar chart on the left illustrates the count of instances across the three outcome categories: No Change, Deteriorated, and Improved. It reveals a relatively balanced dataset, with "Improved" having a slightly higher representation. The adjacent donut chart provides the percentage distribution, indicating that 34.8% of patients showed improvement, 33.1% experienced deterioration, and 32.1% remained unchanged. This balanced class distribution is crucial for developing unbiased and robust classification models, ensuring that no single outcome class dominates the training process.

Figure 12 illustrates the pairwise relationships among key numerical features in the mental health dataset. Each data point is color-coded based on the patient's outcome category (No Change, Deteriorated, or Improved). The lower triangle displays scatter plots highlighting interactions between feature pairs, while the diagonal elements show the KDE for each feature. This visualization helps identify potential correlations and class separability, particularly among features like Symptom Severity, Treatment Progress, and Adherence to Treatment, which may be useful for outcome prediction and feature relevance analysis.

Figure 10. Dimensionality Reduction using UMAP and t-SNE

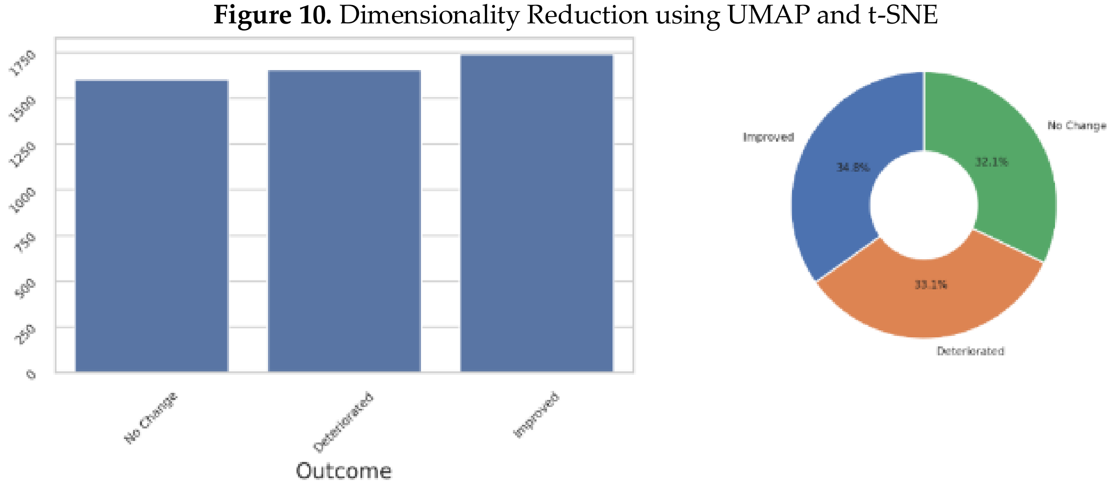

Figure 11. Outcome Distribution Visualization using Countplot and Pie Chart

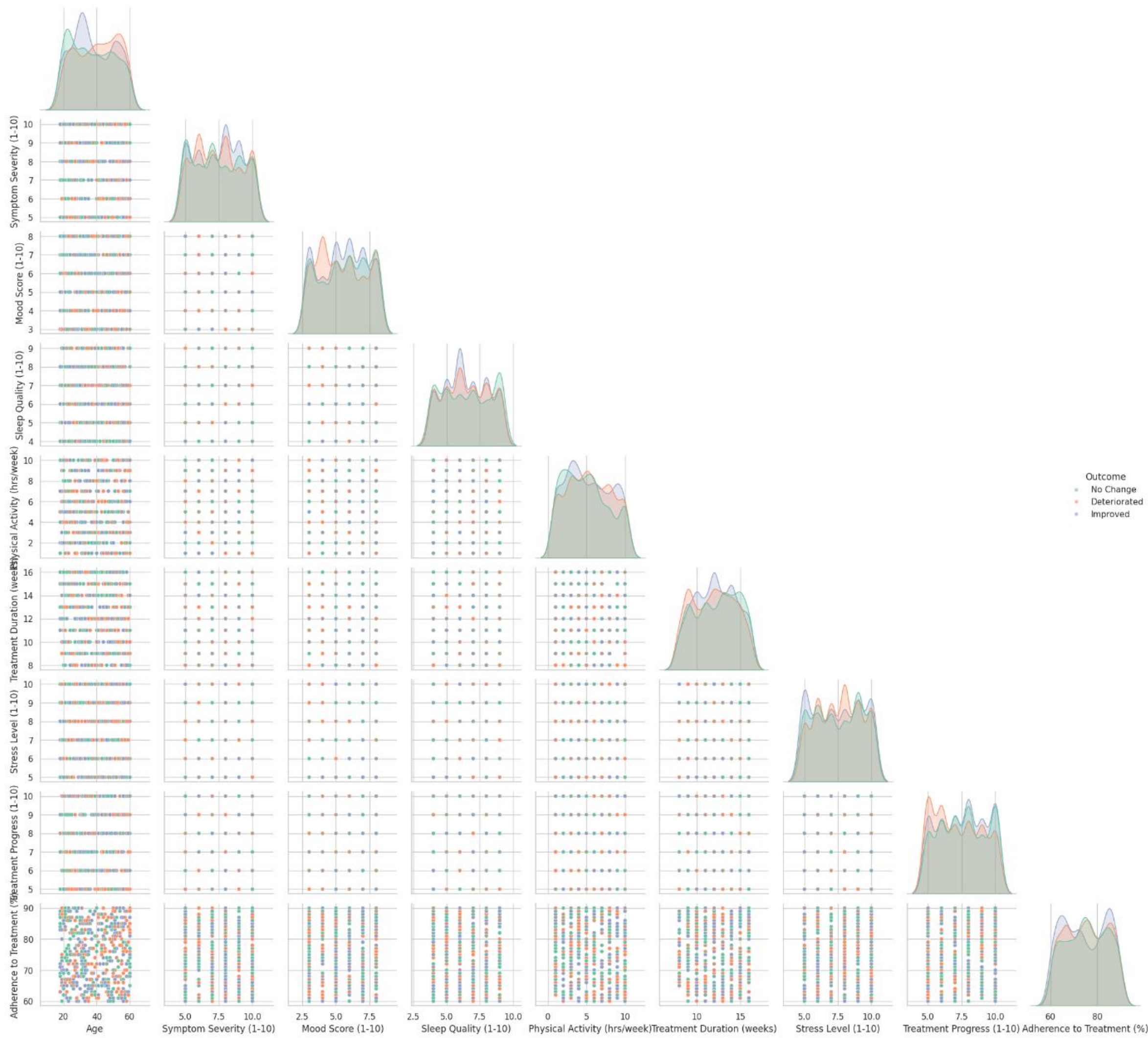

Figure 12. Pairplot of Numerical Features Colored by Outcome

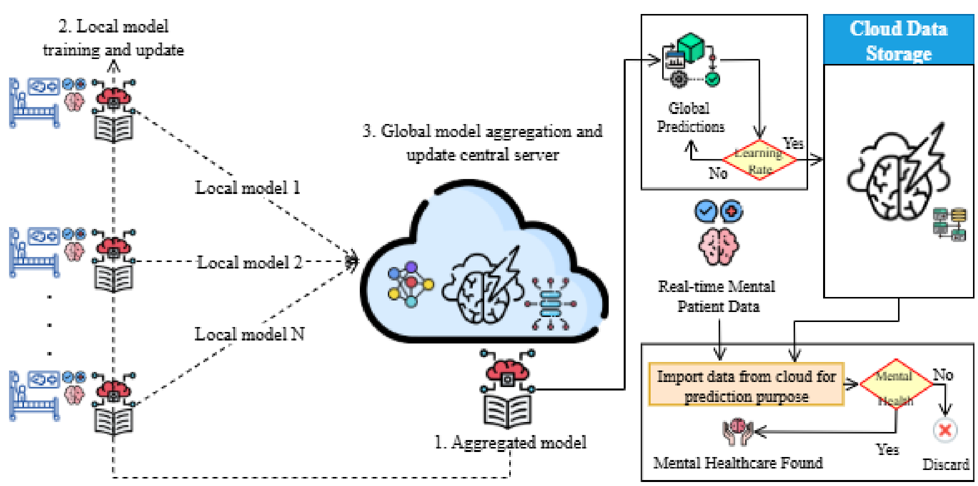

Figure 13. Proposed federated mental healthcare monitoring model (global server model)

After preprocessing, the dataset is partitioned into training (70%) and testing (30%) subsets. The dataset, comprising 5,000 patient records sourced from Kaggle, is further distributed across simulated decentralized client nodes to emulate real-world Federated Learning conditions, where data remains locally stored at individual healthcare institutions. In line with real-world healthcare environments, data across client nodes is intentionally maintained in a Non-IID manner, where each client represents a different institution with heterogeneous patient distributions and varying clinical patterns. The training data is utilized at the client level to build ML models that support privacy-preserving and accurate local predictions. These models undergo iterative training based on a convergence criterion defined by performance stabilization across successive communication rounds, until an optimal learning rate and convergence threshold are achieved. If performance criteria are unmet, the model continues to train on localized patient data to improve reliability and ensure consistency in federated aggregation. This simulated train-test strategy reflects practical FL deployments rather than a conventional centralized evaluation process.

Upon completion of training, the local model is stored securely on the client's cloud infrastructure. In subsequent communication rounds, this model is evaluated using the reserved testing dataset, and real-time mental patient data to ensure generalization and robustness in predicting mental healthcare outcomes. This validation step confirms the suitability of the trained model for real-time mental healthcare monitoring.

While local models support decentralized learning and uphold data privacy, global aggregation enhances clinical decision-making by creating insights from distributed clients. As illustrated in Figure 13, this federated model, empowered by BDA, enables large-scale integration of healthcare knowledge while maintaining data confidentiality. Each client independently processes healthcare data and communicates encrypted model updates to the global server in a traceable and verifiable manner, ensuring both scalability and regulatory compliance.

Figure 13 illustrates the end-to-end flow of the proposed FL-based mental healthcare monitoring framework across multiple decentralized clients (hospitals). Each client possesses local patient data D_i, and trains its model M_i independently, without sharing raw data, ensuring data privacy. Empowered by BDA, this model facilitates scalable integration of heterogeneous health records across institutions while maintaining secure, collaborative, and privacy-aware learning.

5.1. Local Model Training and Update

Each institution C_i optimizes its local model M_i by minimizing a regularized empirical loss function (i.e., reducing local prediction error while remaining aligned with the global model during decentralized training):

$$\min_{\mathbf{W}_i} F_i(\mathbf{W}_i) = \frac{1}{|D_i|} \sum_{j=1}^{|D_i|} \ell(M(\mathbf{x}_{ij}; \mathbf{W}_i), y_{ij} + \lambda ||\mathbf{W}_i - \mathbf{W}_g||^2 \quad (1)$$

Where, $\ell(\cdot)$ is loss function (e.g, cross-entropy), $\mathbf{W}_i$ is local weights, $\mathbf{W}_g$ current global model weights, and λ is regularization coefficient ensuring alignment with global weights (to avoid excessive deviation from the global model and improve stability).

Clients update their local weights using gradient descent (iteratively refining the model using only local client data):

$$\mathbf{w}_i^{(t+1)} = \mathbf{w}_i^{(t)} - \alpha \nabla F_i(\mathbf{w}_i^{(t)}) \quad (2)$$

5.2. Client Evaluation and Reliability Scoring

To ensure only high-quality local updates influence the global model, each client is assigned a reliability score $r_i \in [0,1]$ that considers both model accuracy and update quality:

$$r_i = \gamma . \frac{A_i^{val}}{\bar{A}} + (1-\gamma).(1 - \frac{||\mathbf{W}_i - \mathbf{W}_g||_2}{||\mathbf{W}_g||_2 + \epsilon}) \quad (3)$$

Where, A_i^{val} is validation accuracy of M_i, $\bar{A}$ is mean accuracy across clients, γ is weight controlling accuracy vs divergence influence, and ϵ is stability constant.

5.3. Global Aggregation and Best Model Selection

The central server performs weighted aggregation based on the reliability scores:

$$\mathbf{w}_g^{(t+1)} = \sum_{i=1}^{N} \frac{r_i}{\sum_{j-1}^{N} r_j} . \mathbf{w}_i^{(t+1)} \quad (4)$$

The best-performing model is selected as the global model $\mathbf{w}_g^{(t+1)}$. This update represents the final aggregated model used for real-time predictions, assuming convergence criteria are met.

5.4. Convergence Monitoring and Re-training

A convergence threshold Δ_t is defined to assess global model stability:

$$\Delta_t = |L_g^{(t)} - L_g^{(t-1)}| < \varepsilon \quad (5)$$

If convergence is not achieved, clients with $A_i^{val} < \tau$ are required to re-train with updated hyperparameters (e.g., reduced learning rate), repeating the cycle until an optimal global model is formed.

5.5. Final Prediction

Once optimized, the global model performs real-time inference on new mental health records $\mathbf{x}_{new}$:

$$\hat{y} = \sigma(M_g(\mathbf{x}_{new}; \mathbf{w}_g^{(t+1)})) \text{ where } \sigma(z) = \frac{1}{1+e^{-z}} \quad (6)$$

The decision is made based on a classification threshold θ, tuned to balance sensitivity and specificity in clinical practice.

5.6. Global Model Storage and Real-Time Monitoring

Once the aggregated model satisfies performance criteria, it is stored on a secure cloud infrastructure. Real-time mental patient data is then fetched from the cloud and analyzed using the trained global model:

$$\hat{y} = \sigma(M_g(\mathbf{x}_{new}; \mathbf{w}_g^{(t+1)})) \text{ where } \sigma(z) = \frac{1}{1+e^{-z}} \quad (7)$$

If the prediction probability exceeds a predefined threshold θ, the patient is flagged for mental healthcare follow-up; otherwise, the case is discarded as non-critical.

This hierarchical FL model enables scalable, secure, and collaborative mental health diagnostics while keeping sensitive data decentralized. The figure emphasizes traceable transitions from local client training to global model optimization and deployment for real-time monitoring, supporting regulatory compliance and practical deployment across healthcare institutions. Table 3 summarizes the core pseudocode outlining this federated workflow.

Table 3. Pseudocode of the Proposed Federated Mental Healthcare Monitoring Model

Step	Process
	Client Side (Smart Node/EAV)
1	Start
2	Data Collection: Acquire mental health-related data from IoMT sensors and digital health apps, including features like age, gender, symptom severity, mood score, etc..
3	Preprocess Data: ☑ Label encoding ☑ Missing value handling ☑ MinMax scaling ☑ Exploratory Data Analysis
4	Form Feature-Label Dataset: Create local dataset $D_i = \{(x_j^i, y_j^i)\}_{j=1}^n$ Where $x_j^i \in \mathbb{R}^d$: input features, $y_j^i \in Y$: Class label
5	Local Model Training: $\mathbf{w}_i^{(t+1)} = \mathbf{w}_i^{(t)} - \alpha \nabla F_i(\mathbf{w}_i^{(t)})$ Optimize local weights based on local data
6	Evaluate Local Model: Validate trained model using client validation set and compute metrics (accuracy, loss)
7	

	Transmit to Server: Securely send local weights $\mathbf{w}_i^{(t+1)}$ and validation accuracy A_i^{val} to the global server
	Global Server Side
8	Receive Client Updates: Collect $\mathbf{w}_i^{(t+1)}$ and A_i^{val} from all N clients
9	Client Reliability Scoring: Compute reliability score $r_i \in [0,1]$: $r_i = \gamma . \frac{A_i^{val}}{\bar{A}} + (1-\gamma).(1 - \frac{\|\mathbf{W}_i - \mathbf{W}_g\|_2}{\|\mathbf{W}_g\|_2 + \epsilon})$ where γ: weight factor, ϵ: stability constant
10	Weighted Model Aggregation: Aggregate using: $\mathbf{w}_g^{(t+1)} = \sum_{i=1}^{N} \frac{r_i}{\sum_{j-1}^{N} r_j} . \mathbf{w}_i^{(t+1)}$
10.1	Convergence Check: $\Delta_t = \lvert L_g^{(t)} - L_g^{(t-1)} \rvert < \varepsilon$ If convergence threshold Δ_t is not achieved, clients with $A_i^{val} < \tau$ are required to re-train.
11	Global Model Deployment: If the performance threshold is satisfied, store $\mathbf{w}_g^{(t+1)}$ on a secure cloud infrastructure
12	Real-Time Mental Health Monitoring: For a new instance, $\mathbf{x}_{new}$, predict: $\hat{y} = \sigma(M_g(\mathbf{x}_{new}; \mathbf{w}_g^{(t+1)}))$ If $\hat{y} > \theta$, flag for follow up
14	Stop

6. Simulation Results

The proposed FL model is applied to a real-world mental health dataset, containing both numerical and categorical features such as age, symptom severity, and treatment adherence. After preprocessing, the data was partitioned into training and testing sets in a 70:30 ratio. The training set was used to simulate client-side learning under the FL setting, while the test set was used for evaluating the global model's generalization capability. The simulation was executed in Google Colab using a fully custom-built client-server architecture written in Python, designed to emulate a federated healthcare setting without using external FL frameworks. This setup enables decentralized training while preserving patient data privacy. Each client trained its model independently using various ML algorithms, ensuring privacy-preserving collaboration, with performance evaluated through comprehensive diagnostic metrics aligned with Equations 8–15.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{8}$$

$$Sensitivity\ (TPR - Recall) = \frac{TP}{TP+FN} \tag{9}$$

$$Specificty\ (TNR) = \frac{TN}{TN+FP} \tag{10}$$

$$Miss - rate\ (FNR) = \frac{FN}{TP+FN} \tag{11}$$

$$Precision = \frac{TP}{TP+FP} \tag{12}$$

$$F1 - Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{13}$$

$$Cohen's\ Kappa = \frac{p_o - p_e}{1 - p_e} \tag{14}$$

$$MCC = \frac{(TP \times TN) - (FP \times FN)}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \quad (15)$$

Table 4. Performance analysis of the proposed federated mental healthcare monitoring model (local client)

	Performance Matrices							
	Stacking Classifier		**CatBoost Classifier**		**MLP Classifier**		**KNN**	
	Train (3500)	Test (1500)	Train (3500)	Test (1500)	Train (3500)	Test (1500)	Train (3500)	Test (1500)
Accuracy	0.9986	0.9953	0.9969	0.9907	0.9563	0.9313	0.9980	0.9913
Sensitivity (TPR)	0.9986	0.9953	0.9969	0.9907	0.9563	0.9313	0.9980	0.9913
Specificity (TNR)	0.9993	0.9976	0.9984	0.9953	0.9780	0.9654	0.9990	0.9957
Miss-rate (FNR)	0.0014	0.0047	0.0031	0.0093	0.0437	0.0687	0.002	0.0087
Precision	0.9986	0.9954	0.9969	0.9907	0.9565	0.9324	0.9980	0.9914
F1-Score	0.9986	0.9953	0.9969	0.9907	0.9563	0.9313	0.9980	0.9913
Cohen's Kappa	0.9979	0.9930	0.9953	0.9860	0.9344	0.8969	0.9970	0.9870
Matthews Corr. Coef.	0.9979	0.9930	0.9953	0.9860	0.9345	0.8975	0.9970	0.9871

Table 4 shows that during the training phase on 3500 samples per client, all four models—Stacking Classifier, CatBoost, MLP, and KNN—demonstrated exceptionally high performance. The Stacking Classifier achieved nearly perfect results across all metrics, with 99.86% accuracy, 99.86% sensitivity, and a Cohen's Kappa of 0.9979, indicating extremely strong agreement with true labels. CatBoost and KNN followed closely with similar high values, both exceeding 99.6% accuracy, while MLP recorded a comparatively lower accuracy of 95.63% and slightly weaker agreement (Kappa = 0.9344). Overall, training results suggest that the models were well-fitted to their respective local client datasets, with ensemble methods (Stacking and CatBoost) showing the highest effectiveness.

When evaluated on the unseen test set (1500 samples), generalization capabilities of the models became more distinguishable. The Stacking Classifier retained its superior performance, achieving 99.53% accuracy, 99.53% sensitivity, and a Cohen's Kappa of 0.9930, highlighting strong consistency and reliability. CatBoost also performed well with 99.07% accuracy and 0.9860 Kappa, showing robust but slightly less consistent generalization. KNN delivered impressive results too, reaching 99.13% accuracy and 0.9870 Kappa, outperforming MLP, which dropped to 93.13% accuracy and 0.8969 Kappa, suggesting it was less effective in handling test variability.

Among all models, the Stacking Classifier consistently outperformed others during both training and testing, making it the most accurate, balanced, and dependable choice for global model selection in the proposed FL-based mental healthcare framework. Following client-level evaluations, the proposed framework applies FL to securely aggregate distributed knowledge, yielding a global model with improved generalization and privacy preservation, as validated by the results in Table 5.

Table 5. Performance analysis of the proposed model for mental healthcare monitoring (global client)

	Performance Matrices							
	Accuracy	Sensitivity (TPR)	Specificity (TNR)	Miss-rate (FNR)	Precision	F1-Score	Cohen's Kappa	Matthews Corr. Coef.
Proposed global model for mental healthcare monitoring	0.9953	0.9953	0.9976	0.0047	0.9954	0.9953	0.9930	0.9930

After applying FL, Table 5 shows that the global model achieved outstanding performance across all key metrics. It reported a high level of accuracy of 99.53 with both sensitivity and precision of 99.53 and 99.54 respectively- indicating that the model accurately identifies true cases with minimal false positives. Its diagnostic reliability is also validated by the specificity of 99.76% and the low miss-rate of only 0.47%. Besides, a significant agreement with the ground truth can be seen in the Cohens Kappa (0.9930) and Matthews Correlation Coefficient (0.9930), which indicate that the federated approach can provide both privacy and near-optimal predictive accuracy of real-world mental health monitoring.

Table 6. Comparison of the proposed mental healthcare monitoring model with prior approaches

References	Model	Accuracy (%)	Miss-rate (%)
Das et al., 2024 [47]	SVM, CNN, LSTM, DT	DAIC-WOZ dataset: 90.26 MODMA dataset: 90.47	DAIC-WOZ dataset: 9.74, MODMA dataset: 9.53
Chung et al., 2023 [48]	Different ML Models	88.80	11.2
Sara et al., 2024 [49]	KNN	91.45	8.55
Geng et al., 2023 [50]	SVM, ERTC	SVM: 79.29, ERTC: 86.32	SVM: 20.71, ERTC: 13.68
Vasha et al., 2023 [51]	SVM	73.15	26.85
Kim et al., 2023 [52]	DL	78.14	21.86
Hilbert et al., 2024 [53]	RF, LR	46.5–60	53.5-40
Wang et al., 2024 [54]	ANN	62	38
Dougherty et al., 2025 [55]	LR, NLP	85	15
Lønfeldt et al., 2023 [56]	ML	70	30
Ricka et al., 2023 [57]	ML	86	14
Fernandez et al., 2024 [58]	LGBM	86.77	13.23
Sadoun et al., 2024 [59]	LGBM, RF	91	9
Proposed global model for mental healthcare monitoring	Stacking Classifier	99.53	0.47

Table 6 illustrates the comparative performance of several traditional ML models [47-59] for mental healthcare prediction. While models like SVM, CNN, LSTM, DT, and ANN demonstrate reasonable effectiveness, they have comparatively lower accuracies than other methods. The proposed FL-based global model, conversely, outperforms all other baseline classifiers by a significant margin, achieving the highest accuracy of 99.53 and strong sensitivity and specificity. This confirms the strength, generalization, and clinical reliability of the suggested federated architecture within delicate healthcare applications.

7. Conclusion

Mental health monitoring faces critical challenges, including fragmented data sources, privacy constraints, and the need for scalable, high-performance diagnostic systems. Conventional centralized systems are not effective in dealing with these issues particularly when sensitive healthcare information is shared among many institutions. In order to address such obstacles, this study introduces a FL-based model, combined with BDA that allows training models in a decentralized manner, preserving privacy, and utilizing complex mental health data. The proposed system guarantees the institutional autonomy, data confidentiality, and real-time diagnostics, by simulating heterogeneous client environments. The simulation results demonstrate that the proposed global FL model outperformed all local methods [48-60], achieving 99.53% accuracy and a miss-rate of only 0.47%, making it highly suitable for real-world deployment in mental healthcare monitoring, where trust, privacy, and performance are essential.

7.1. Limitations and Future Work

While the current model emphasizes privacy and collaborative learning, it does not deeply integrate interpretability methods. Additionally, the reported performance is based on a controlled experimental setup and may vary in real-world clinical environments. Future work will focus on enhancing model transparency by embedding Explainable AI techniques (e.g., SHAP, LIME) into the federated setup to support explainable decision-making and clinician trust.

Authors' Contribution Statement:

A.A.A., S.J.A., and M.A.K have collected data from different resources and contributed to writing—original draft preparation. A.A., and M.A.K drafted pictures & tables. M.A.K., S.J.A., and M.A.K. performed formal analysis and Simulation, performed revision, and improved the quality of the draft. S.T.B., and M.A.K.; writing—review and editing, performed supervision. A.A. received the funding. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: In our study, the data are available in a famous repository, Kaggle, and it is publicly available on the following link: https://www.kaggle.com/code/hainescity/mental-health-diagnosis-and-treatment-eda/input

Conflicts of Interest: The authors declare no conflict of interest.

Funding Declaration: The authors would like to acknowledge the Research Management Centre (RMC), Multimedia University (MMU), for its financial support in facilitating the publication of this paper.

Acknowledgement: The authors would like to acknowledge the Research Management Centre (RMC), Multimedia University (MMU), for its financial support in facilitating the publication of this paper

Consent to Publish declaration: Not applicable.

Ethics and Consent to Participate declarations: Not applicable.

Clinical Trial Number: Not applicable.

References

1. Nie, P., Zhao, K., Ma, D., Liu, H., Amin, S., & Yasin, I. (2024). Global Climate Change, Mental Health, and Socio-Economic Stressors: Toward Sustainable Interventions across Regions. *Sustainability*, *16*(19), 8693. https://doi.org/10.3390/su16198693
2. Kirkbride, J. B., Anglin, D. M., Colman, I., Dykxhoorn, J., Jones, P. B., Patalay, P., ... & Griffiths, S. L. (2024). The social determinants of mental health and disorder: evidence, prevention and recommendations. *World psychiatry*, 23(1), 58-90. https://doi.org/10.1002/wps.21160
3. Yang, Y., Niu, L., Amin, S., & Yasin, I. (2024). Unemployment and mental health: a global study of unemployment's influence on diverse mental disorders. *Frontiers in Public Health*, *12*, 1440403. https://doi.org/10.3389/fpubh.2024.1440403
4. Rivas, M. E. R. (2021). Attitudes towards people with mental illness among medical students: Qualitative research. *Psychology, Society & Education*, *13*(2), 21-33.
5. Meshram, D., & Kasturkar, P. (2021). Assessment of Common Mental Health Problems among General Population. *Journal of Pharmaceutical Research International*, *33*(36B), 210-217.
6. Jin, W., Felix, M. S., Paek, S. C., & Lamy, F. R. (2024). Social activities and depressive symptoms among elderly based on rural and urban differences in China: a National Cross-Sectional study. *Psychology Research and Behavior Management*, 2243-2254. https://doi.org/10.2147/PRBM.S460453
7. Fan, Y., Fan, A., Yang, Z., & Fan, D. (2025). Global burden of mental disorders in 204 countries and territories, 1990–2021: results from the global burden of disease study 2021. *BMC psychiatry*, *25*(1), 486. https://doi.org/10.1186/s12888-025-06932-y
8. Basheti, I. A., Mhaidat, Q. N., & Mhaidat, H. N. (2021). Prevalence of anxiety and depression during COVID-19 pandemic among healthcare students in Jordan and its effect on their learning process: A national survey. PloS one, 16(4), e0249716. https://doi.org/10.1371/journal.pone.0249716
9. Fekih-Romdhane, F., Jahrami, H., Away, R., Trabelsi, K., Pandi-Perumal, S. R., Seeman, M. V., & Cheour, M. (2023). The relationship between technology addictions and schizotypal traits: mediating roles of depression, anxiety, and stress. BMC psychiatry, 23(1), 67. https://doi.org/10.1186/s12888-023-04563-9
10. Bitar, Z., Elias, M. B., Malaeb, D., Hallit, S., & Obeid, S. (2023). Is cyberbullying perpetration associated with anxiety, depression and suicidal ideation among Lebanese adolescents? Results from a cross-sectional study. BMC psychology, 11(1), 53. https://doi.org/10.1186/s40359-023-01091-9
11. Herath, H. M. S. S., Herath, H. M. K. K. M. B., Madhusanka, B. G. D. A., & Guruge, L. G. P. K. (2024). Data protection challenges in the processing of sensitive data. In *Data Protection: The Wake of AI and Machine Learning* (pp. 155-179). Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-76473-8_8
12. Al Omar, A., Jamil, A. K., Khandakar, A., Uzzal, A. R., Bosri, R., Mansoor, N., & Rahman, M. S. (2021). A transparent and privacy-preserving healthcare platform with novel smart contract for smart cities. *Ieee Access*, *9*, 90738-90749. https://doi.org/10.1109/ACCESS.2021.3089601
13. Shahzad, T., Saleem, M., Farooq, M. S., Abbas, S., Khan, M. A., & Ouahada, K. (2024). Developing a transparent diagnosis model for diabetic retinopathy using explainable AI. *IEEE Access*. https://doi.org/10.1109/ACCESS.2024.3475550
14. Saleem, M., Farooq, M. S., Shahzad, T., Hassan, A., Abbas, S., Ali, T., ... & Khan, M. A. (2024). Secure and transparent mobility in smart cities: Revolutionizing AVNs to predict traffic congestion using MapReduce, Private Blockchain and XAI. *IEEE Access*. https://doi.org/10.1109/ACCESS.2024.3458983
15. Hanna, M. G., Pantanowitz, L., Dash, R., Harrison, J. H., Deebajah, M., Pantanowitz, J., & Rashidi, H. H. (2025). Future of artificial intelligence (AI)-machine learning (ML) trends in pathology and medicine. *Modern Pathology*, 100705. https://doi.org/10.1016/j.modpat.2025.100705
16. Alshamrani, M. (2022). IoT and artificial intelligence implementations for remote healthcare monitoring systems: A survey. *Journal of King Saud University-Computer and Information Sciences*, *34*(8), 4687-4701. https://doi.org/10.1016/j.jksuci.2021.06.005
17. Kalluri, K. (2022). Federate Machine Learning: A Secure Paradigm for Collaborative AI in Privacy-Sensitive Domains. *International Journal on Science and Technolo-gy*, *13*(4), 1-13.

18. Dai, D., & Boroomand, S. (2022). A review of artificial intelligence to enhance the security of big data systems: state-of-art, methodologies, applications, and challenges. *Archives of Computational Methods in Engineering*, *29*(2), 1291-1309. https://doi.org/10.1007/s11831-021-09628-0
19. Ali, M., Naeem, F., Tariq, M., & Kaddoum, G. (2022). Federated learning for privacy preservation in smart healthcare systems: A comprehensive survey. *IEEE journal of biomedical and health informatics*, *27*(2), 778-789. https://doi.org/10.1109/JBHI.2022.3181823
20. Abaoud, M., Almuqrin, M. A., & Khan, M. F. (2023). Advancing federated learning through novel mechanism for privacy preservation in healthcare applications. *IEEE Access*, *11*, 83562-83579. https://doi.org/10.1109/ACCESS.2023.3301162
21. Stylianou, T. and Pantelidou, A., 2025. Big data and consumer behavior: A macroeconomic perspective through supermarket analytics. *Quantitative Finance and Economics*, *9*(3), pp.682-712.
22. Qiu, H., Zhu, K., Luong, N.C., Yi, C., Niyato, D. and Kim, D.I., 2022. Applications of auction and mechanism design in edge computing: A survey. *IEEE Transactions on Cognitive Communications and Networking*, *8*(2), pp.1034-1058. https://doi.org/10.1109/TCCN.2022.3147196
23. Ali, Y., Han, K. H., Majeed, A., Lim, J. S., & Hwang, S. O. (2025). An Optimal Two-Step Approach for Defense Against Poisoning Attacks in Federated Learning. *IEEE Access*. https://doi.org/10.1109/ACCESS.2025.3556906
24. Khan, M. A., Farooq, M. S., Saleem, M., Shahzad, T., Ahmad, M., Abbas, S., & Abu-Mahfouz, A. M. (2025). Smart buildings: Federated learning-driven secure, transparent and smart energy management system using XAI. *Energy Reports*, *13*, 2066-2081. https://doi.org/10.1016/j.egyr.2025.01.063
25. Ahmed, A., Xi, R., Hou, M., Shah, S. A., & Hameed, S. (2023). Harnessing big data analytics for healthcare: A comprehensive review of frameworks, implications, applications, and impacts. *IEEE Access*, *11*, 112891-112928. https://doi.org/10.1109/ACCESS.2023.3323574
26. Fathima, A. S., Basha, S. M., Ahmed, S. T., Mathivanan, S. K., Rajendran, S., Mallik, S., & Zhao, Z. (2023). Federated learning based futuristic biomedical big-data analysis and standardization. *Plos one*, *18*(10), e0291631. https://doi.org/10.1371/journal.pone.0291631
27. Hwang, S. O., & Majeed, A. (2024). Analysis of federated learning paradigm in medical domain: taking COVID-19 as an application use case. *Applied Sciences*, *14*(10), 4100. https://doi.org/10.3390/app14104100
28. Deebak, B. D., & Hwang, S. O. (2023). Federated learning-based lightweight two-factor authentication framework with privacy preservation for mobile sink in the social IoMT. *Electronics*, *12*(5), 1250. https://doi.org/10.3390/electronics12051250
29. Majeed, A., Zhang, X., & Hwang, S. O. (2022). Applications and challenges of federated learning paradigm in the big data era with special emphasis on COVID-19. *Big Data and Cognitive Computing*, *6*(4), 127. https://doi.org/10.3390/bdcc6040127
30. Salahat, M., Al-Zoubi, H. Q., Al-Dmour, N. A., & Ghazal, T. M. (2025). A fused weighted federated learning-based adaptive approach for early-stage drug prediction. *Scientific Reports*, *15*(1), 31763. https://doi.org/10.1038/s41598-025-13991-4
31. Anter, A. M., & Abualigah, L. (2023). Deep federated machine learning-based optimization methods for liver tumor diagnosis: A review. *Archives of Computational Methods in Engineering*, *30*(5), 3359-3378. https://doi.org/10.1007/s11831-023-09901-4
32. Abbas, S.R., Abbas, Z., Zahir, A. and Lee, S.W., 2024, December. Federated learning in smart healthcare: a comprehensive review on privacy, security, and predictive analytics with IoT integration. In *Healthcare* (Vol. 12, No. 24, p. 2587). MDPI. https://doi.org/10.3390/healthcare12242587
33. Ghazal, T. M., Hasan, M. K., Hussein, A. H., Pandey, B. K., Ahmad, M., Safie, N., & Cascone, L. (2025). Federated Learning With Small and Large Models With Privacy-Preserving Data Space for Holographic Internet of Things in Consumer Electronics. *IEEE Transactions on Consumer Electronics*. https://doi.org/10.1109/TCE.2025.3573033
34. Ahsan, S.I., Djenouri, D. and Haider, R., 2024. Privacy-enhanced sentiment analysis in mental health: Federated learning with data obfuscation and bidirectional encoder representations from transformers. *Electronics*, *13*(23), p.4650. https://doi.org/10.3390/electronics13234650

35. Katmah, R., Al-Shargie, F., Tariq, U., Babiloni, F., Al-Mughairbi, F., & Al-Nashash, H. (2021). A review on mental stress assessment methods using EEG signals. *Sensors*, *21*(15), 5043. https://doi.org/10.3390/s21155043
36. Hinduja, S., Afrin, M., Mistry, S., & Krishna, A. (2022). Machine learning-based proactive social-sensor service for mental health monitoring using twitter data. *International Journal of Information Management Data Insights*, *2*(2), 100113. https://doi.org/10.1016/j.jjimei.2022.100113
37. Allam, H., Davison, C., Kalota, F., Lazaros, E., & Hua, D. (2025). AI-Driven Mental Health Surveillance: Identifying Suicidal Ideation Through Machine Learning Techniques. *Big Data and Cognitive Computing*, *9*(1), 16. https://doi.org/10.3390/bdcc9010016
38. Bokolo, B. G., & Liu, Q. (2023). Deep learning-based depression detection from social media: Comparative evaluation of ml and transformer techniques. *Electronics*, *12*(21), 4396. https://doi.org/10.3390/electronics12214396
39. Ismail, L., Shahin, N., Materwala, H., Hennebelle, A., & Frermann, L. (2023). ML-NLPEmot: machine learning-natural Language processing event-based emotion detection proactive framework addressing mental health. *IEEE Access*, *11*, 144126-144149. https://doi.org/10.1109/ACCESS.2023.3343121
40. Liu, Z., & Zhao, J. (2025). Leveraging deep learning for robust EEG analysis in mental health monitoring. *Frontiers in Neuroinformatics*, *18*, 1494970. https://doi.org/10.3389/fninf.2024.1494970
41. Jiang, W., Zhang, Y., Han, H., Liu, X., Gwak, J., Gu, W., ... & Maple, C. (2025). Fuzzy ensemble-based federated learning for EEG-based emotion recognition in Internet of Medical Things. *Journal of Industrial Information Integration*, *44*, 100789. https://doi.org/10.1016/j.jii.2025.100789
42. Zhang, D., Xu, H., Li, P., Sun, Y., Jiang, W., Wang, X. A., & Zhou, M. (2025). Privacy Parameter Setting and Usability Optimization Algorithm for Medical Data. *IEEE Transactions on Consumer Electronics*. https://doi.org/10.1109/TCE.2025.3569752
43. Hu, T., & Lee, C. H. (2024). Monitoring the mental state of the elderly in the community based on machine learning. *Applied and Computational Engineering*, *99*, 155-160.
44. Li, P., & Liang, F. (2023). Construction of mental health monitoring system based on model transfer learning algorithm. *International Journal of Wireless and Mobile Computing*, *24*(1), 58-65. https://doi.org/10.1504/IJWMC.2023.129088
45. Thotad, P. N., Kallur, S., Mundaragi, L., & Kadam, S. H. (2023, October). Mental Health Tracker Using Machine Learning Approaches. In *2023 4th IEEE Global Conference for Advancement in Technology (GCAT)* (pp. 1-5). IEEE. https://doi.org/10.1109/GCAT59970.2023.10353336
46. Rayan, A., & Alanazi, S. (2023). A novel approach to forecasting the mental well-being using machine learning. *Alexandria Engineering Journal*, *84*, 175-183. https://doi.org/10.1016/j.aej.2023.10.060
47. Das, A.K. and Naskar, R., 2024. A deep learning model for depression detection based on MFCC and CNN generated spectrogram features. *Biomedical Signal Processing and Control*, *90*, p.105898. https://doi.org/10.1016/j.bspc.2023.105898
48. Chung, J., & Teo, J. (2023). Single classifier vs. ensemble machine learning approaches for mental health prediction. *Brain informatics*, *10*(1), 1. https://doi.org/10.1186/s40708-022-00180-6
49. Sara, S. S., Rahman, M. A., Rahman, R., & Talukder, A. (2024). Prediction of suicidal ideation with associated risk factors among university students in the southern part of Bangladesh: Machine learning approach. *Journal of affective disorders*, *349*, 502-508. https://doi.org/10.1016/j.jad.2024.01.092
50. Geng, D., An, Q., Fu, Z., Wang, C. and An, H., 2023. Identification of major depression patients using machine learning models based on heart rate variability during sleep stages for pre-hospital screening. *Computers in Biology and Medicine*, *162*, p.107060. https://doi.org/10.1016/j.compbiomed.2023.107060
51. Vasha, Z. N., Sharma, B., Esha, I. J., Al Nahian, J., & Polin, J. A. (2023). Depression detection in social media comments data using machine learning algorithms. *Bulletin of Electrical Engineering and Informatics*, *12*(2), 987-996. https://doi.org/10.11591/eei.v12i2.4182
52. Kim, A. Y., Jang, E. H., Lee, S. H., Choi, K. Y., Park, J. G., & Shin, H. C. (2023). Automatic depression detection using smartphone-based text-dependent speech signals: deep convolutional neural network approach. *Journal of medical Internet research*, *25*, e34474. https://doi.org/10.2196/34474
53. Hilbert, K., Böhnlein, J., Meinke, C., Chavanne, A.V., Langhammer, T., Stumpe, L., Winter, N., Leenings, R., Adolph, D., Arolt, V. and Bischoff, S., 2024. Lack of evidence for predictive utility from resting state fMRI data for individual

exposure-based cognitive behavioral therapy outcomes: A machine learning study in two large multi-site samples in anxiety disorders. *NeuroImage*, *295*, p.120639. https://doi.org/10.1016/j.neuroimage.2024.120639

54. Wang, J., Wu, D.D., DeLorenzo, C. and Yang, J., 2024. Examining factors related to low performance of predicting remission in participants with major depressive disorder using neuroimaging data and other clinical features. *Plos one*, *19*(3), p.e0299625. https://doi.org/10.1371/journal.pone.0299625
55. Dougherty, R.F., Clarke, P., Atli, M., Kuc, J., Schlosser, D., Dunlop, B.W., Hellerstein, D.J., Aaronson, S.T., Zisook, S., Young, A.H. and Carhart-Harris, R., 2025. Psilocybin therapy for treatment resistant depression: prediction of clinical outcome by natural language processing. *Psychopharmacology*, *242*(7), pp.1553-1561.https://doi.org/10.1007/s00213-023-06432-5
56. Lønfeldt, N.N., Olesen, K.V., Das, S., Mora-Jensen, A.R.C., Pagsberg, A.K. and Clemmensen, L.K.H., 2023. Predicting obsessive-compulsive disorder episodes in adolescents using a wearable biosensor—A wrist angel feasibility study. *Frontiers in Psychiatry*, *14*, p.1231024. https://doi.org/10.3389/fpsyt.2023.1231024
57. Ricka, N., Pellegrin, G., Fompeyrine, D.A., Lahutte, B. and Geoffroy, P.A., 2023. Predictive biosignature of major depressive disorder derived from physiological measurements of outpatients using machine learning. *Scientific Reports*, *13*(1), p.6332. https://doi.org/10.1038/s41598-023-33359-w
58. Fernandez, J., Martínez, R., Innocenti, B., & López, B. (2024). Contribution of EEG biosignals for Stress Detection. *Authorea Preprints*.
59. Sadoun, M. S. N., Vargas, J. M., Boularas, M. M., Boutin, A., Cottin, F., & Laleg-Kirati, T. M. (2024). Cognitive Stress Detection during Physical Activity using Simultaneous, Mobile EEG and ECG signals. *IFAC-PapersOnLine*, *58*(24), 291-296. https://doi.org/10.1016/j.ifacol.2024.11.052