RESEARCH PAPER

Short Title: Survey-Based Mental Health Prediction Mental-Health State Prediction Using Survey-Based Machine-Learning Techniques

[1]Komal Patel, [2] Dr Kaushal Jani
[1](Indus University, Ahmedabad, Gujarat, patelkomal.21.rs@indusuni.ac.in)
[2](Indus University, Ahmedabad, Gujarat, kaushaljani.ce@indusuni.ac.in)

Abstract: A global challenge continues to increase with the rise of mental health conditions, therefore a reliable and accessible screening tool to detect mental illness earlier in an individual's process of seeking professional services must be developed. Typical ways of diagnosing mental illness are subjective and may depend on self-report surveys or clinical assessments which may depend upon availability, stigma or inconsistencies when reporting symptoms. This paper describes a machine-learning classification system for mental health using structured survey data with; demographic, behavioural, psychological, and workplace indicators. The data set underwent systematic pre-processing through; data cleaning, numerical encoding, normalisation, and exploratory correlation before being used as a high-quality and interpretable input for creating the ML models. Supervised learning algorithms were used to train and evaluate each of the algorithms to determine predictive performance. The results indicate that ensemble models widely outperformed all of the other algorithms with respect to their resistance to noise and ability to identify complex feature inter-actions. All performance metrics such as accuracy, precision, recall, and F1-score confirm that the models were stable across data splits. Exploratory analysis also supported a coherent relationship between identifying psychological characteristics and the models developed from surveys to predict mental health. Overall, the proposed framework offers a practical, adaptable, and scalable approach to early mental-health screening, supporting data-driven decision-making in organizational and community well-being initiatives.

Keywords: Mental-health prediction, Random Forest classifier, Psychological assessment, Exploratory data analysis, Mental-health screening
How to cite this article: Komal Patel, Dr Kaushal Jani., Short Title: Survey-Based Mental Health Prediction Mental-Health State Prediction Using Survey-Based Machine-Learning Techniques. Int J Drug Deliv Technol. 2026;16(4s): 1-8; DOI: 10.25258/ijddt.16.1-8
Source of support: Nil.
Conflict of interest: None

1. Introduction

Mental health is a critical global public health concern that impacts people from a variety of demographic and occupational groups. Early identification of mental illness risk is essential to minimize long-term psychological, social, and economic consequences. However, traditional diagnostic approaches based on self-reporting and clinician judgment are often influenced by stigma, subjectivity, and inconsistencies in symptom interpretation. These limitations have encouraged the adoption of machine-learning (ML) techniques for scalable and objective mental health screening.

Early research established the feasibility of computational methods for identifying mental states using neuroimaging and behavioral data[1]. Machine-learning models subsequently demonstrated strong performance in structured healthcare domains, including disease detection and clinical decision support[2]. Advances in deep learning further improved diagnostic accuracy in medical imaging applications, reinforcing confidence in ML-based healthcare analytics[3]. To support clinical adoption, explainable ML approaches were introduced to improve model transparency and interpretability[4].

With the expansion of digital data sources, ML techniques were applied to large-scale behavioral datasets to predict mental health risks. Studies utilizing social-media and user-generated content showed that linguistic and behavioral patterns could indicate depression and future mental illness[5, 6]. Deep-learning models were also applied to neuropsychiatric and personality assessment tasks,

*Author for Correspondence: patelkomal.21.rs@indusuni.ac.in)

demonstrating robustness across mental health conditions[7, 8]. The COVID-19 pandemic further accelerated the use of ML in healthcare prediction and monitoring, highlighting the adaptability of structured ML frameworks[9, 10].

2. Literature Review

Structured surveys and surveys with quantitative data for mental health assessment have received a growing level of attention in recent years. Many scaling and individual studies on structured survey data have established that machine-learning-based artificial intelligence effectively classifies and distinguishes between various mental health states such as Major Depressive Disorder and Generalized Anxiety Disorder based upon standardized psychometric indicators and/or behavioral indicators[11]. The validity of multi-class deep-learning methods for differentiating between multiple classes of mental health status has also been established, although, like other machine-learning models, challenges related to generalization from datasets that are imbalanced continue[12, 13].

As the field matured, emphasis shifted toward interpretability and clinical applicability. Explainable artificial intelligence (XAI) techniques were introduced to clarify decision-making processes in mental health screening models, improving transparency and trust[14]. These approaches have been successfully applied to autism spectrum disorder detection, early diagnosis of mental illnesses, and structured healthcare analytics, demonstrating the clinical relevance of interpretable ML systems[15, 16, 17].

More recent studies explored advanced deep-learning, ensemble, and transformer-based frameworks to enhance predictive performance in mental health assessment[18, 19]. Research on depression severity detection and multi-disorder classification showed that multi-class models can effectively capture varying mental health states when trained on large datasets[20, 21]. Public datasets and explainable deep-learning approaches further supported transparency and reproducibility in mental health prediction tasks[22, 23]. Several studies emphasized the importance of interpretable risk-prediction frameworks and multimodal learning to integrate behavioral and activity-based indicators[24, 25, 26, 27]. However, many approaches remain dataset-dependent, limiting generalizability, which highlights the need for reproducible and scalable benchmarking frameworks based on structured survey data for real-world mental health screening applications[28, 29, 30].

Table 1: Summary of Selected Literature on AI-Based Mental Health Classification

Authors / Year	Data Type	Method Used	Target Disorder(s)	Key Contribution
Sharma & Chariar (2024)[11]	Survey, bibliometric data	ML review & mapping	Multiple mental disorders	Comprehensive review and bibliometric analysis of ML-based mental disorder diagnosis
Bendebane et al. (2023)[12]	Twitter text data	Multi-class deep learning	Depression, Anxiety	Early detection of depression and anxiety using social media data
Thomas (2025)[13]	Administrative health records	Machine learning models	Cognitive disorders	ML-based cognitive disorder detection using structured health data
Tutun et al. (2024)[14]	Structured screening data	Explainable AI (XAI)	General mental disorders	XAI framework for transparent mental disorder screening
Vimbi et al. (2025)[15]	Clinical & behavioral data	Explainable deep learning	Autism Spectrum Disorder	Application of XAI for autism detection
Baran & Cetin (2025)[16]	Clinical datasets	AI-driven ML models	Specific mental disorders	Early diagnosis using AI-based predictive models

Aggarwal et al. (2025)[17]	Medical image data	Explainable DL	Multiple disorders	Real-time healthcare analysis using XAI-driven image processing
Hussain et al. (2025)[18]	Text & behavioral data	Deep transformers	Depression (severity levels)	Multi-level depression severity detection framework
Kumari et al. (2025)[19]	Mixed datasets (survey & text)	ML & DL survey	Depression	Survey of ML/DL techniques for depression detection
Hermawan et al. (2025)[20]	Survey & clinical data	Machine learning models	General mental disorders	Practical ML framework for early mental health screening
Kaggle CID007 (2021)[21]	Structured survey data	Benchmark ML models	Multiple mental disorders	Publicly available dataset for mental disorder classification
Hu & Sokolova (2021)[22]	Structured COVID-19 mental health data	Explainable ML	Multiple mental states	Explainable multi-class classification framework
Wang et al. (2021)[23]	Environmental & health data	Explainable deep learning	Mental disorder risk	Risk prediction with model interpretability
Misgar & Bhatia (2024)[24]	Time-series motor activity data	Explainable DL	Psychotic disorders	Interpretable deep learning for psychotic pattern analysis
Ahmed et al. (2025)[25]	Activity & behavioral data	XAI with SHAP/LIME	Depression	Explainable framework for depression detection and severity classification
Islam et al. (2025)[26]	Text data	Transformer (RoBERTa)	Multiple mental disorders	Fine-tuned multi-class transformer model
De Souza et al. (2022)[27]	Reddit text data	Deep ensemble learning	Anxiety, Depression	Ensemble-based classification of comorbid mental disorders
Bin Saeed & Cha (2025)[28]	Multimodal social media data	Attention-based BiLSTM	General mental health issues	Multimodal early detection framework
Islam et al. (2025)[29]	Healthcare datasets	Explainable ML (SHAP, LIME)	Chronic & mental disorders	Interpretable ML using advanced explainability techniques
Kasanneni et al. (2024)[30]	Social media conversations	ML & DL comparison	Mental health disorders	Comparative evaluation of ML/DL methods

Table 1 shows AI-based mental-health studies, showing the evolution from traditional ML to deep, transformer, and explainable models and positioning this work within that progression.

3. Methodology

This study develops a machine-learning framework for mental-health classification using a structured survey dataset containing a demographic, workplace, behavioral, and psychological indicators.

3.1 Dataset

The study will use 1,250 anonymized survey responses from participants on 26 structured variables concerning demographics, working conditions, behaviour, emotional health, mental health status, and a target variable showing whether participants have received mental health treatment containing both categorical and numerical elements. The dataset has a class imbalance but does not contain missing data, and any incomplete submissions have been removed. An 80/20 training/testing split will be implemented with 1,000 responses in training set and 250 responses in testing set. The dataset is publicly available, fully anonymized, and published under the terms of an open research license, therefore additional ethical clearance is not required.

3.2 Data Preprocessing

A systematic approach to pre-processing was developed for analysing, identifying and correcting inconsistencies, missing values and format problems throughout a dataset. Missing entries in a dataset were met with either imputed appropriate methods or was deleted. Categorical variables were transformed into numerical representations to allow the use of machine learning interpretability, and continuous variables were normalized to increase convergence and eliminate dominance over other features. The final dataset had class labels checked for validity; duplicate records removed; class labels had been fully verified; and was subsequently constructed in a fully numerical array for model training and evaluation.

3.3 Exploratory Data Analysis

Following preprocessing, an exploratory data analysis (EDA) was performed on all features to explore the features' distributions and relationships across variables. Descriptive statistics and information visualizations such as histograms, bar charts, and correlation heat maps were utilized for the purpose of discovering trends, feature interaction, and class imbalances. The EDA platform was also used to identify the most appropriate methods for selecting features, managing outliers and evaluating models, while confirming that there is enough variation among features to allow for successful classification using machine-learning algorithms.

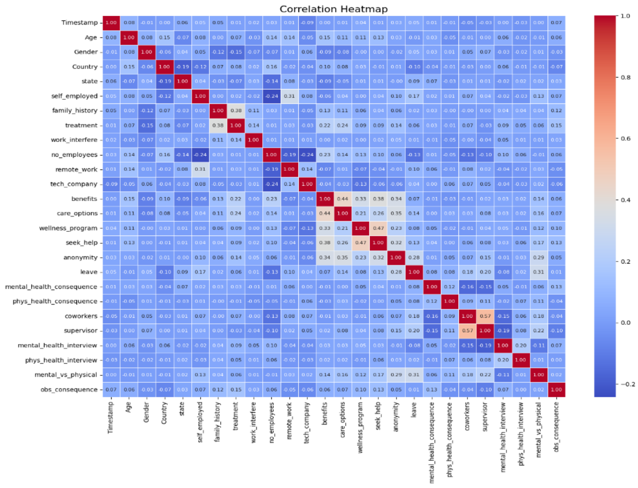

Figure 1: Correlation Heatmap showing correlation between all the attributes of dataset

Figure 1 shows the correlation heatmap of survey features, indicating mostly weak to moderate relationships and minimal multicollinearity. A few stronger correlations reflect meaningful workplace and behavioral patterns, while overall feature independence supports stable model training and informed feature engineering.

3.4 Feature Engineering

Prior to training the models an extensive amount of engineering was carried out in order to improve the representation of the data for each model. Using the exploratory data analysis results, appropriate features were selected and irrelevant features were eliminated. The encoded categorical variables were double checked to ensure that numerical values were accurately mapped, and new derived features were produced through the merging of associated indicators in order to maximize the ability for disambiguating classes. All of the selected features were merged into one input matrix, arranged in the same way for all machine learning algorithms.

3.5 Model Development

The authors used various classification techniques Logistic Regression, Decision Trees, Random Forests, Support Vector Machines, and KNN to create a machine-learning system that predicts mental health disorders. Supervised learning was utilized to train the models built on the pre-processed feature matrix and utilized hyperparameter tuning to provide an unbiased measure of how well the systems perform and to reduce the likelihood that they would overfit.

3.6 Training Strategy and Evaluation

In order to determine generalization, the training and testing datasets were separated based on train/test patterns. The features after preprocessing were trained on separate optimized parameters from each of the model algorithms using minimization of the Loss Function for parameter selection. The results were evaluated by Accuracy, Precision, Recall, F1-Score, and Confusion Matrix. In addition to comparing the classifiers, the comparative analysis also included a misclassification analysis of the classifiers to confirm the reliability and consistency of the performance of the classifier on the schizophrenia/mood disorder prediction.

3.7 Prediction Pipeline

When a machine-learning model is ready to classify new instances of mental health, the model has been trained on structured features developed from user survey responses. Prior to this stage, the collected data will go through pre-processing and feature engineering. Once the clean data has been pre-processed, the cleaned data can be entered into the predictive model to produce a prediction for a given user in the form of class label or probability based on their mental health classifications as outputted by the predictive model. Figure 2 illustrates the end-to-end machine-learning pipeline for survey-based mental-health classification. The system architecture for this prediction stage is defined in the workflow charts.

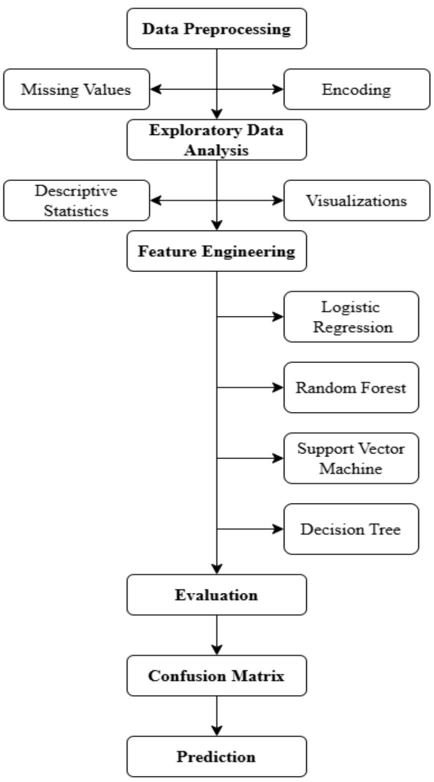

Figure 2: System Flow

4. Experiment

This section offers a thorough summary of our suggested experiments, along with thorough evaluation findings and an analysis of them.

4.1 Implementation Details

We conducted all our experiments in a python environment using scientific libraries for the purpose of conducting our analysis. Once we evaluated the structure, type and distribution of our dataset, we began the preprocessing of our dataset and created a training and testing dataset with untested records. We trained and tested several supervised classifiers including Logistic Regression, Random Forest, Decision Tree, K-Nearest Neighbors, Support Vector Machines and Naive Bayes using the scikit-learn library. Feature normalization was accomplished using Standard Scaler and categorical variables were encoded using Label Encoder. Our results are reproducible because our preprocessing was controlled and we utilized fixed random seeds in our experiments.

4.2 Result Analysis

The assessment of Machine Learning Models was completed using the test dataset. Assessment of performance metrics included Accuracy, Precision, Recall and F1-Score to provide a comprehensive view of Classifier Ability. Prior to the training of ML Models, a Correlation Heatmap and various EDA Visualizations were assessed to understand feature relationships and variable importance; this analysis guided the Feature Selection Process and Pre-Processing Steps used to determine Final Performance. As a result, these results showed Moderate Correlation between certain Psychosocial Attributes and Workplace Characteristics with the Target Labels (i.e., the potential predictive value).

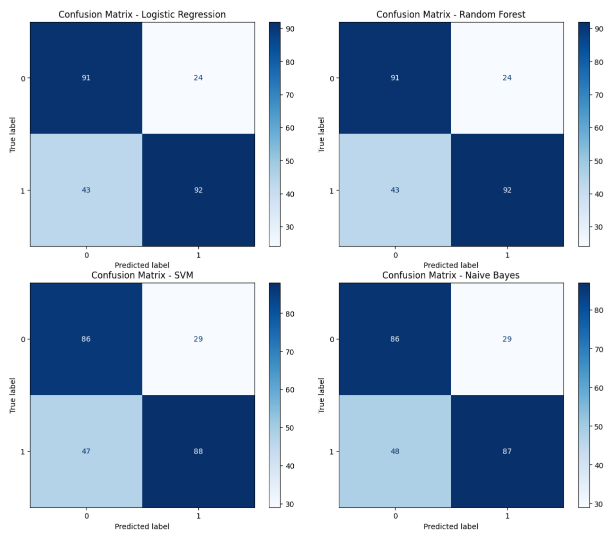

Figure 3: Confusion matrix of all applied models

Confusion matrices of Logistic Regression, Naïve Bayes, SVM, and Random Forest models are depicted on Figure 3. The confusion matrices show how well the algorithms classified mental health treatments. The algorithms with the highest and most consistent classification performance were Logistic Regression and Random Forests, as opposed to SVM and the Naïve Bayes, which resulted in higher levels of misclassification.

Table 2: Model Performance Comparison

Model	Accuracy	Precision	Recall	F1-Score	AUROC
Logistic Regression	0.73	0.793	0.681	0.733	0.784
Random Forest	0.732	0.793	0.681	0.733	0.798
SVM	0.696	0.752	0.652	0.698	0.757
Naive Bayes	0.692	0.75	0.644	0.693	0.778
Decision Tree	0.632	0.672	0.622	0.646	0.633
KNN	0.64	0.742	0.511	0.605	0.663

As indicated in Table 2, the performance results from six different types of machine-learning algorithms were evaluated on the Mental Health Classification Data Set.

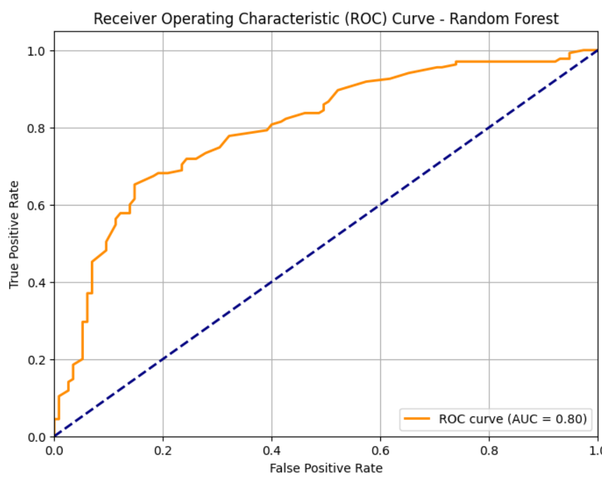

Figure 4: ROC Curve of Random Forest Model

Figure 4 shows the ROC curve of the Random Forest model, demonstrating strong class discrimination across different thresholds.

4.3 Performance of Machine Learning Models

Evaluation of the models indicates considerable variation in model performance due to the differing structures of the models. Logistic Regression created an effective and stable base level of results while Decision Tree and KNN captured non-linear structures but were greatly impacted by noise. Random Forest produced stable and reliable results through the use of ensemble methods. Support Vector Machine produced good results but required feature scaling. Based on these results, Random Forest and SVM were the two highest performing models, indicating that the use of ensemble and margin-based classifiers is most effective for mental-health datasets.

5. Discussion

Structured survey data is an appropriate input for machine-learning algorithms, which provide significant potential for assessing participants' individual mental health. Random Forest and other tree-based machine-learning models performed significantly better than SVM and KNN models due to the superior ability of tree-based algorithms to identify and model complex non-linear relationships while also reducing noise. In addition to workplace comfort as a major predictor of mental health, exploratory analyses showed that anxiety and emotional wellness were also strong predictors of each individual's mental health. The reliability of using machine learning to predict individual mental health depends on the accuracy of survey responses; thus, such applications should be viewed as complementary tools that will require frequent and extensive validation, retraining, and professional supervision prior to being utilized in practice.

6. Conclusion and Future Work

The results of this research support that using a machine learning approach with a systematic workflow for creating and utilizing the results of mental health surveys allows us to accurately identify individuals at risk for mental health disorders. Models that utilize ensemble methods produced better results due to their ability to model the intricate connections between workplace and behavioral factors, and demonstrated the potential for using organized questionnaires to perform automated screening for mental health disorders. Future efforts will concentrate on the development of a larger and more diverse dataset to allow for the integration of multiple data sources into future work; and on the development and application of large-scale learning and hybrid-domain, deep-learning based models to enhance the ability to generalize across a spectral range of patient populations and enable broader application in clinical practice.

7. References

1. Poldrack, Russell A., Yaroslav O. Halchenko, and Stephen José Hanson. "Decoding the large-scale structure of brain function by classifying mental states across individuals." Psychological science 20, no. 11 (2009): 1364-1372.

2. Li, L., Qin, L., Xu, Z., Yin, Y., Wang, X., Kong, B., ... & Xia, J. (2020). Using artificial intelligence to detect COVID-19 and community-acquired pneumonia based on pulmonary CT: evaluation of the diagnostic accuracy. Radiology, 296(2), E65-E71.

3. Lakhani, Paras, and Baskaran Sundaram. "Deep learning at chest radiography: automated classification of pulmonary tuberculosis by using convolutional neural networks." Radiology 284.2 (2017): 574-582.

4. Hatwell J, Gaber MM, Atif Azad RM. Ada-WHIPS: explaining AdaBoost classification with applications in the health sciences. BMC Medical Informatics and Decision Making. 2020 Oct 2;20(1):250.

5. Thorstad R, Wolff P. Predicting future mental illness from social media: A big-data approach. Behavior research methods. 2019 Aug 15;51(4):1586-600.

6. Fatima, Iram, Hamid Mukhtar, Hafiz Farooq Ahmad, and Kashif Rajpoot. "Analysis of user-generated content from online social communities to characterise and predict depression degree." Journal of Information Science 44, no. 5 (2018): 683-695.

7. Zeng LL, Wang H, Hu P, Yang B, Pu W, Shen H, Chen X, Liu Z, Yin H, Tan Q, Wang K. Multi-site diagnostic classification of schizophrenia using discriminant deep learning with functional connectivity MRI. EBioMedicine. 2018 Apr 1; 30:74-85.

8. Mehta, Yash, et al. "Recent trends in deep learning-based personality detection." Artificial Intelligence Review 53.4 (2020): 2313-2339.

9. Alimadadi, Ahmad, Sachin Aryal, Ishan Manandhar, Patricia B. Munroe, Bina Joe, and Xi Cheng. "Artificial intelligence and machine learning to fight COVID-19." Physiological genomics 52, no. 4 (2020): 200-202.

10. Choudhury, Avishek, Emily Renjilian, and Onur Asan. "Use of machine learning in geriatric clinical care for chronic diseases: a systematic literature review." JAMIA open 3, no. 3 (2020): 459-471.

11. Sharma, Chandra Mani, and Vijayaraghavan M. Chariar. "Diagnosis of mental disorders using machine learning: Literature review and bibliometric mapping from 2012 to 2023." Heliyon 10, no. 12 (2024).

12. Bendebane L, Laboudi Z, Saighi A, Al-Tarawneh H, Ouannas A, Grassi G. A multi-class deep learning approach for early detection of depressive and anxiety disorders using twitter data. Algorithms. 2023 Nov 27;16(12):543.

13. Thomas, G. (2025). Machine Learning Approaches for Cognitive Disorder Detection using Administrative Health Data.

14. Tutun S, Topuz K, Tosyali A, Bhattacherjee A, Li G. Explainable artificial intelligence for mental disorder screening: A computational design science approach. Journal of Management Information Systems. 2024 Oct 1;41(4):958-81.

15. Vimbi, Viswan, Noushath Shaffi, Mohamed AK Sadiq, Srinivasa Rao Sirasanagandla, VN Manjunath Aradhya, M. Shamim Kaiser, Shuqiang Wang, and Mufti Mahmud. "Application of Explainable Artificial Intelligence in Autism Spectrum Disorder Detection." Cognitive Computation 17, no. 3 (2025): 104.

16. Baran FD, Cetin M. AI-driven early diagnosis of specific mental disorders: a comprehensive study. Cognitive Neurodynamics. 2025 Dec;19(1):70.

17. Aggarwal, Lakshita, Vikram Ranjan, and Ananya Sharma. "Real time image processing and smart healthcare using Explainable Artificial Intelligence (XAI)." Image and Vision Computing (2025): 105653.

18. Hussain, Nisar, Amna Qasim, Gull Mehak, Muhammad Zain, Grigori Sidorov, Alexander Gelbukh, and Olga Kolesnikova. "Multi-Level Depression Severity Detection with Deep Transformers and Enhanced Machine Learning Techniques." AI 6, no. 7 (2025): 157.

19. Kumari, Meena, Gurpreet Singh, and Sagar Dhanraj Pande. "A Survey of Current Progress in Depression Detection Using Deep Learning and Machine Learning." Biomedical Materials & Devices (2025): 1-25.

20. Hermawan, Latius, Rizma Adlia Syakurah, Meilinda Meilinda, Deris Stiawan, Edi Surya Negara, Indri Ramayanti, Muhammad Fahmi, Muhammad Qurhanul Rizqie, and Dedy Hermanto. "Early Mental Health Detection with Machine Learning: A Practical Approach to Model Development and Implementation." Indonesian Journal of Electrical Engineering and Informatics (IJEEI) 13, no. 2 (2025).

21.https://www.kaggle.com/datasets/cid007/mental-disorder classification?utm_source=chatgpt.com

22. Hu, YuanZheng, and Marina Sokolova. "Explainable multi-class classification of the camh covid-19 mental health data." arXiv preprint arXiv:2105.13430 (2021).

23. Wang, Ce, Lan Feng, and Yi Qi. "Explainable deep learning predictions for illness risk of mental disorders in Nanjing, China." Environmental Research 202 (2021): 111740.

24. Misgar, Muzafar Mehraj, and M. P. S. Bhatia. "Unveiling psychotic disorder patterns: A deep learning model analysing motor activity time-series data with explainable AI." Biomedical Signal Processing and Control 91 (2024): 106000.

25. Ahmed, Iftikhar, et al. "Explainable AI for depression detection and severity classification from activity data: development and evaluation study of an interpretable framework." JMIR Mental Health 12.1 (2025): e72038.

26. Islam, K. M., John Fields, and Praveen Madiraju. "multiMentalRoBERTa: A Fine-tuned Multiclass Classifier for Mental Health Disorder." arXiv preprint arXiv:2511.04698 (2025).

27. De Souza, Vanessa Borba, Jeferson Campos Nobre, and Karin Becker. "DAC stacking: A deep learning ensemble to classify anxiety, depression, and their comorbidity from Reddit texts." IEEE Journal of Biomedical and Health Informatics 26.7 (2022): 3303-3311.

28. Bin Saeed, Qasim, and YoungJin Cha. "Multi-modal deep-attention-BiLSTM based early detection of mental health issues using social media posts." Scientific Reports 15.1 (2025): 35152.

29. Islam, Md Manowarul, et al. "Explainable Machine Learning for Efficient Diabetes Prediction Using Hyperparameter Tuning, SHAP Analysis, Partial Dependency, and LIME." Engineering Reports 7.1 (2025): e13080.

30. Kasanneni, Yashwanth, et al. "Effective analysis of machine and deep learning methods for diagnosing mental health using social media conversations." IEEE Transactions on Computational Social Systems (2024).