Nigerian Journal of Technology (NIJOTECH)
Vol. 44, No. 4, December, 2025, pp. 634 - 646
www.nijotech.com

Print ISSN: 0331-8443
Electronic ISSN: 2467-8821
https://dx.doi.org/10.4314/njt.v44i4.10

AUGMENTED MULTI-LABEL CLASSIFICATION FOR THE EARLY DETECTION OF CO-OCCURRING MENTAL HEALTH DISORDERS

AUTHORS:
S. Kavita[1*], K. Mansi[2], K. Shruti[2], C. Rahul[2], M. Joshua[2], D. Smita[1]

AFFILIATIONS:
[1]Department of Computer Engineering, Fr. C. Rodrigues Institute of Technology, Navi Mumbai, India
[2]Department of Computer Engineering, Fr. C. Rodrigues Institute of Technology, Navi, Mumbai, India

***CORRESPONDING AUTHOR:**
Email: kavita.shelke@fcrit.ac.in

ARTICLE HISTORY:
Received: June 25, 2025.
Revised: October 12, 2025.
Accepted: October 20, 2025.
Published: January 03, 2026.

KEYWORDS:
Multi-label Classification, Generative Adversarial Network, Mental health disorders, Early detection

ARTICLE INCLUDES:
Peer review

DATA AVAILABILITY:
On request from author(s)

EDITORS:
Chidozie Charles Nnaji

FUNDING:
None

Abstract

Mental health disorders such as anxiety, depression, and schizophrenia often go undiagnosed due to limited awareness, social stigma, and reliance on subjective clinical evaluations. Traditional screening methods can be time-consuming, leading to delayed interventions and worsening conditions. This system aims to provide an early screening tool that helps individuals assess their mental health status and assists healthcare professionals in identifying disorders quickly and accurately. The system employs a multi-label classification approach to predict multiple co-existing mental health disorders simultaneously. The dataset is created using psychiatrist-approved questionnaires, and since real-world mental health data is often limited and biased, Generative Adversarial Networks (GANs) are used to generate synthetic data for improved model training. This enhances generalizability and reduces bias in predictions. By providing a user-friendly AI-powered screening tool, the system helps reduce the taboo around mental health conditions and bridges the gap between individuals and mental health professionals. It ensures faster, data-driven diagnosis, allowing for timely interventions and better treatment planning, ultimately improving mental healthcare accessibility and efficiency. The experimental results indicate that the Random Forest model achieved the best overall performance, with an F1-score weighted average of 0.40 and strong label-wise performance, particularly for obsessive-compulsive disorder (OCD) (F1 = 0.69), Post-Traumatic Stress Disorder (PTSD) (F1 = 0.62), and Normal (F1 = 0.38), demonstrating its effectiveness in multi-label mental health disorder detection.

1.0 INTRODUCTION

Mental health disorders such as anxiety, depression, and stress affect a vast number of people worldwide. Despite their high prevalence, identifying these disorders early remains a major challenge, primarily due to the dependence on self-reported symptoms, clinical interviews, and subjective evaluations. Traditional diagnostic methods often suffer from time delays, inconsistencies, and limitations caused by human interpretation, leading to postponed treatment and worsening symptoms. Recent advancements in Artificial Intelligence (AI) and Machine Learning (ML) have opened avenues for the development of data-driven tools that can enhance early detection, reduce subjective biases, and improve the reach of mental health services. AI models can analyze patterns in survey responses and physiological signals to identify early signs of mental health conditions. Various studies have examined the application of AI and ML for mental health

HOW TO CITE:
Kavita, S., Mansi, K., Shruti, K., Rahul, C., Joshua, M., and Dr. Smita, D. "Augmented Multi-Label Classification for the Early Detection of Co-Occurring Mental Health Disorders", *Nigerian Journal of Technology*, 2025. 44(4), pp. 634 - 646. https://dx.doi.org/10.4314/njt.v44i4.10

prediction. Kim et al. [1] conducted universal mental health screenings in schools to detect early signs of anxiety, depression, and behavioral issues. Using machine learning models like logistic regression and decision trees, the study attained over 80% accuracy, sensitivity, and specificity. The study recommends including additional factors such as family history and integrating screening with counseling services. M. Nadeem et al. [2] compared ML models such as SVM, LASSO, LSTM, CNN+LSTM, RF, Logistic Regression, ANN, and XGBoost for diagnosing various disorders, including ADHD, depression, anxiety, PTSD, anorexia, stress, schizophrenia, and Alzheimer's. CNN+LSTM achieved 98.3% accuracy for stress, and LSTM achieved 91.26% for anxiety.

Challenges noted include small sample sizes, dataset bias, and lack of generalizability. Ahuja Ravinder et al. [3] explored mental stress among college students based on exam pressure and internet use. Algorithms like Linear Regression, Naive Bayes, Random Forest, and SVM were used, with SVM achieving the highest accuracy of 85.71%. The study identified sample size and bias from self-reporting as limitations. Jung-Yoon Kim et al. [4] proposed a non-invasive depression detection system for elderly individuals using PIR motion sensors. Neural Networks achieved up to 96% accuracy. Despite promising results, the study had a small sample size and plans for broader validation. AB Osman et al. [5] reviewed ML and DL applications in mental health diagnosis using datasets like DAIC-WOZ, EHR, and social media. DL models, such as LSTM, demonstrated superior performance in complex data analysis, but challenges like data privacy and generalizability remain. A, Seal et al. [6] introduced DeprNet, a CNN-based model for detecting depression from EEG data. It performed well (AUC 0.999) in record-wise evaluations, with slightly lower results in subject-wise testing. The study warns against overfitting due to limited diversity. C. Wijayarathna et al. [7] reviewed stress detection methods in gameplay settings. While existing systems show promise, many use intrusive or costly equipment. The study calls for adaptation of lab techniques for real-world scenarios. M. Ravi Kumar et al. [8] utilized Logistic Regression and Naive Bayes to detect depression from Twitter posts, achieving 97.31% accuracy. The limitation lies in relying solely on text data, lacking context. Future work may include multimodal data integration. B.H. Bhavani et al. [9] examined depression detection during COVID-19 using questionnaire responses. LSTM achieved 100% accuracy, although potential overfitting and self-reporting bias were identified. Broader validation is recommended. Mashrura Tasnim et al. [10] predicted depression, anxiety, and stress using speech features and CNN models. Though effective, the dataset imbalance skewed towards normal scores was a concern. Balanced datasets and personalized models are proposed. Anu Priya et al. [11] used ML on DASS-21 questionnaire data from diverse individuals. Naive Bayes achieved the highest accuracy; Random Forest scored the highest in the F1 measure. Dataset imbalance and reliance on self-reporting were cited as weaknesses. Future directions include integrating clinical data and resampling methods like SMOTE and ADASYN. Mario Ezra Aragon et al. [12] developed BoSE and D-BoSE models using social media data to detect depression and anorexia. These models captured temporal emotional patterns but faced challenges related to limited symptom coverage and model interpretability. Guo Y. et al. [13] leveraged social media platforms for early mental illness detection using SGL-CNN and MGL-CNN models. This improved feature extraction and addressed limitations of earlier systems, enhancing scalability and accuracy. Khoo et al. [14] addressed the shortage of real-time monitoring systems by creating a wearable device capable of continuous mental health tracking. The study emphasized usability and early intervention. Sharma et al. [15] demonstrated effective early diagnosis of multiple mental health disorders by integrating behavioral and speech features, indicating the value of multi-modal data for improved screening accuracy. Kiran et. Al [16] emphasized that timely identification and intervention can significantly improve treatment outcomes, highlighting the need for accessible early screening tools. Yadav et al. [17] introduced an Emotion-Aware Ensemble Learning framework for corporate professionals, showing that combining multiple data sources enhances diagnostic performance. Rahman et al. [18] provided a systematic review of ML-based mental health detection methods and found that although various models show promise, issues such as dataset limitations and generalizability still need attention.

Similarly, Abdullah et. Al [19] showed that integrating machine learning with ensemble techniques and large language models can improve the prediction of future mental health risks from social media data, further reinforcing the potential of AI in proactive mental health monitoring. Key challenges in early mental health disorder detection include:

https://dx.doi.org/10.4314/njt.v44i4.10

1. Lack of large, high-quality training datasets. Mental health data is often limited, unbalanced, and sensitive, leading to reduced model performance and generalizability.
2. Most current models detect only one disorder at a time, ignoring co-morbid conditions like anxiety and depression occurring simultaneously.

To address these, the proposed system aims to incorporate extensive datasets and support multi-label classification for comprehensive diagnosis. Synthetic data generation via Generative Adversarial Networks (GANs) will enhance dataset diversity, promote better model training, while ensure privacy.

This system intends to bridge the gap in early mental health screening, offering accessible tools for users hesitant to seek professional help due to stigma or resource limitations. Multi-label classification enables simultaneous diagnosis of multiple conditions, while GAN-based data augmentation addresses bias and scarcity issues. Objectives of the proposed system:

1. Design a multi-label classification model for detecting multiple co-occurring mental disorders.
2. Develop datasets from psychiatrist-approved questionnaires to improve diagnostic accuracy.
3. Use GANs for synthetic data generation to enhance model robustness and reduce bias.
4. Provide a user-friendly platform for non-judgmental self-assessment.
5. Assist clinicians by reducing diagnosis time and enabling early, personalized treatment.
6. Integrate AI advancements with mental health expertise to enhance early detection, accessibility, and clinical decision-making.

This approach seeks to improve the accuracy and scalability of mental health assessments, supporting both individuals and healthcare professionals in addressing mental health challenges more effectively.

2.0 METHODOLOGY

2.1 System Model

The proposed system is designed as an AI-driven diagnosis model capable of identifying co-existing mental illnesses through multi-label classification. The primary driving force behind this design is the inherent nature and the overlap of such psychiatric disorders as Generalized Anxiety Disorder (GAD), Major Depressive Disorder (MDD), OCD, and PTSD, which tend to co-exist frequently in an individual. To address this issue, the system leverages Conditional Tabular Generative Adversarial Networks (CTGAN) in generating fake data and integrates this with ensemble machine learning techniques for accuracy in predictions and generalizability.

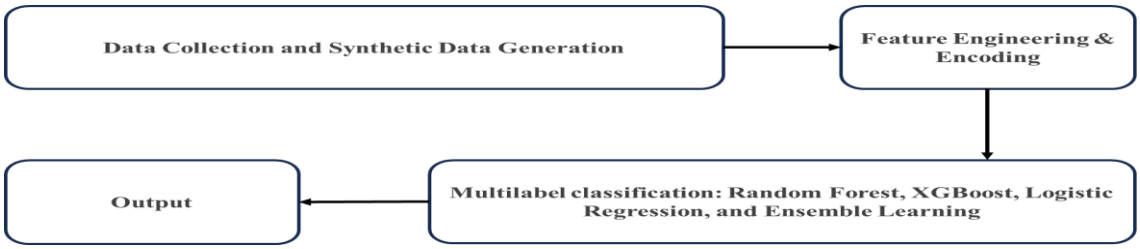

Figure 1. Block diagram of the CTGAN for early detection of mental health disorders

The architectural sequence, as depicted in Figure 1, explains the major stages of the system pipeline. It begins with acquiring the user response through a structured mental health questionnaire. Such inputs are preprocessed and feature-engineered and then input to a CTGAN module that enriches the dataset by generating synthetic records that are real. Such a final balanced dataset is used to train different machine learning models in a multi-label classification setting. The final step is real-time deployment, where new user inputs are processed and classified by the learned models to predict the presence or absence of a combination of several mental health diseases simultaneously.

The solution addresses several concerns: it avoids the issue of dealing with small and unbalanced sets, makes it easier to detect multiple disorders through a single assessment, and increases accessibility by enabling integration into telehealth applications and chatbots to make remote assessment possible.

2.2 Data Collection and Synthetic Data Generation

The system begins with the acquisition of mental health-related survey data consisting of categorical responses to symptom-based questions. Given the sensitivity and limited availability of such data,

https://dx.doi.org/10.4314/njt.v44i4.10
http://creativecommons.org/licenses/by-nc-nd/4.0/

especially for minority classes, the system incorporates synthetic data generation using CTGAN [15]. CTGAN is particularly effective in handling imbalanced categorical tabular data and can model complex relationships among discrete variables, generating high-quality synthetic data that closely resembles real records. This augmentation ensures that all classes, including rare disorder combinations, are well-represented, enhancing the generalizability of the model. In real-world deployment, the system accepts new user input in the form of completed symptom questionnaires. These responses are one-hot encoded in real-time and fed into the trained model, which outputs a binary prediction vector indicating the presence or absence of each mental health disorder. The output probabilities can also be used to indicate the confidence level of each prediction, providing valuable decision support for clinicians or mental health platforms.

Moreover, the system can be integrated into mental health applications, chatbots, or telemedicine platforms to provide preliminary assessments, helping triage patients for further psychological evaluation. This study employs a three-pronged methodological approach combining synthetic data generation using Conditional Tabular Generative Adversarial Networks (CTGAN) [20], Random Forest classification [21], and multi-label classification [22] techniques to effectively detect co-occurring mental health disorders based on questionnaire responses. The process involves preprocessing the categorical dataset, augmenting data using CTGAN to mitigate class imbalance, and training a multi-output random forest classifier to make multi-label predictions. The raw dataset consisted of categorical responses to 15 mental health-related questions. These responses include options such as “Not at all”, “Few days”, “More than half the days”, and “Nearly every day”. To transform these inputs into a machine-readable format, Ordinal Encoding was applied to all features, converting each categorical answer into a numeric rank. The target column, which contains multiple comma-separated disorder labels per sample (e.g., “PTSD, OCD, MDD”), was encoded using Multilabel Binarizer [23-24] to convert the multi-label values into a binary matrix where each column represents a disorder and each row a sample. The dataset exhibited significant class imbalance, particularly in disorders like GAD and MDD, which had fewer samples. To address this, we employed Conditional Tabular GAN (CTGAN) — a GAN-based data synthesizer tailored for tabular data. CTGAN models the distribution of categorical features using conditional vectors and generates synthetic samples by learning patterns from minority classes [5]. By augmenting the dataset with synthetic but realistic samples, CTGAN helps balance the label distribution, thereby improving model generalizability and reducing bias toward majority classes.

A custom dataset comprising 500 tuples was successfully developed to address the complex problem of mental health disorder detection using multi-label classification. Each tuple rep- resented an individual’s responses to a structured questionnaire, with labels assigned for five possible conditions: Generalized Anxiety Disorder (GAD), Major Depressive Disorder (MDD), obsessive-compulsive disorder (OCD), Post-Traumatic Stress Disorder (PTSD), and Normal (no disorder). The dataset effectively captured overlapping symptoms across these categories, making it suitable for real-world diagnostic applications where comorbidity is common. The questionnaires and the corresponding response structure were verified and approved by a certified psychologist to ensure clinical validity and ethical compliance. Since the dataset represents all unique combinations of response patterns across the questionnaire, the 500 tuples comprehensively cover the possible variations, making it statistically sufficient for training multi-label models. Additionally, the CTGAN-generated samples were cross-checked against the original dataset distribution to confirm realism and consistency before model training.

Table 1 presents 10 entries from the generated dataset, which includes 15 attributes and the target column, 'Disease'. Each sample can have more than one diagnosis. Each attribute contains only 4 options as specified earlier. Figure 2 shows the frequency of labels: GAD, MDD, PTSD, OCD, and Normal in the generated dataset. It signifies uniform distribution of all the labels and availability of all labels for training.

https://dx.doi.org/10.4314/njt.v44i4.10

Table 2: Custom dataset generated using CTGAN

1. Anxiety	2. Panic Attacks	3. Depression	4. Interest and Pleasure	5. Energy Levels	6. Concentration	7. Sleep Disturbances	8. Appetite Changes	9. Self-Perception	10. Obsessive Thoughts	11. Compulsive Behaviors	12. Flashbacks	13. Social Withdrawal	14. Irritability and Anger	15. Suicidal Thoughts	Disease
More than half the days	More than half the days	Not at all	More than half the days	Few days	Nearly every day	Nearly every day	Nearly every day	More than half the days	More than half the days	Nearly every day	Few days	Few days	More than half the days	Not at all	GAD,OCD
More than half the days	More than half the days	Not at all	More than half the days	Not at all	Nearly every day	Nearly every day	Nearly every day	More than half the days	More than half the days	Nearly every day	Nearly every day	Few days	Nearly every day	Not at all	PTSD,OCD,MDD
Not at all	Few days	Not at all	More than half the days	Not at all	Not at all	More than half the days	Nearly every day	Not at all	Not at all	Not at all	Nearly every day	Few days	More than half the days	Not at all	OCD, GAD
Not at all	Not at all	Few days	Not at all	Not at all	More than half the days	More than half the days	Few days	Not at all	Not at all	Not at all	Not at all	Few days	Not at all	Not at all	MDD
More than half the days	Not at all	Not at all	Nearly every day	Not at all	Nearly every day	Not at all	Nearly every day	Not at all	More than half the days	Not at all	Few days	Few days	Few days	More than half the days	OCD, PTSD, MDD, GAD
Nearly every day	More than half the days	Few days	Nearly every day	Not at all	Nearly every day	Nearly every day	Nearly every day	Not at all	More than half the days	Not at all	Not at all	Nearly every day	Nearly every day	Few days	PTSD, MDD
More than half the days	Few days	Few days	Few days	More than half the days	More than half the days	Few days	More than half the days	More than half the days	Few days	More than half the days	Nearly every day	Few days	More than half the days	Not at all	MDD,PTSD
Few days	More than half the days	Few days	Nearly every day	Few days	Few days	More than half the days	More than half the days	More than half the days	More than half the days	Nearly every day	Not at all	More than half the days	Nearly every day	Not at all	PTSD, GAD,MDD
More than half the days	More than half the days	Not at all	More than half the days	Not at all	Nearly every day	Nearly every day	Nearly every day	Not at all	More than half the days	Nearly every day	Nearly every day	Few days	More than half the days	Few days	OCD, PTSD
More than half the days	Few days	Not at all	More than half the days	Not at all	Not at all	Nearly every day	Nearly every day	Nearly every day	More than half the days	Nearly every day	Few days	Few days	Nearly every day	Few days	OCD,MDD
More than half the days	Not at all	Not at all	More than half the days	Nearly every day	More than half the days	More than half the days	Nearly every day	Not at all	More than half the days	Nearly every day	Nearly every day	Few days	Few days	Not at all	OCD, PTSD, GAD

https://dx.doi.org/10.4314/njt.v44i4.10
http://creativecommons.org/licenses/by-nc-nd/4.0/

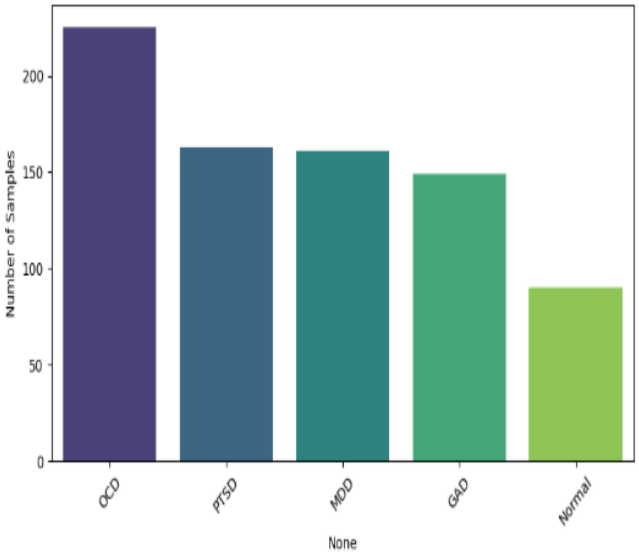

Figure 2: Label frequency in dataset

2.3 Feature Engineering and Encoding

The features in the dataset correspond to users' responses to mental health symptom questions, which are inherently categorical. To make these features suitable for machine learning models, they are transformed using Ordinal Encoding. This encoding converts ordered categories into integers. In the proposed system, the labels" Not at all"," Few days"," More than half the days", "Nearly every day" are encoded as 0,1,2,3, respectively. It is highly preferred when the categorical feature has a meaningful order (like ratings, levels, stages). As we are using tree-based models like XGBoost, Random Forest, etc., it is used for effectively increasing the dimensionality while preserving interpretability. The resultant feature matrix serves as input for model training.

2.4 Multi-Label Classification Framework

Given the nature of mental health conditions, where individuals may exhibit symptoms of more than one disorder simultaneously, the problem is formulated as a multi-label classification task. We use the Multioutput Classifier wrapper from scikit-learn, which enables a single estimator to handle multiple binary classification tasks, one for each target label [7]. To ensure the robustness, accuracy, and generalizability of the proposed multi-label mental health disorder detection system, several machine learning algorithms were evaluated. The selected models include Random Forest, XGBoost, Logistic Regression, and Ensemble Learning techniques. Each algorithm was chosen based on its unique capabilities and advantages in handling multi-label, imbalanced, and high-dimensional data, which are common characteristics in psychological datasets.

A. Random Forest is an ensemble learning technique based on decision trees. It works by training multiple decision trees and aggregating their outputs via majority voting (for classification tasks).

- Handling Categorical Features: Random Forest performs well on datasets with categorical features, such as symptom frequency labels (" Few days"," Nearly every day", etc.).
- Multi-label Capability: When used with a Multioutput Classifier wrapper, it effectively handles multi-label classification tasks.

B. XGBoost (Extreme Gradient Boosting) is a powerful gradient boosting framework that has demonstrated superior performance in many structured data problems. It was selected for the following reasons:

- High Accuracy: XGBoost typically outperforms other models in terms of accuracy and AUC, particularly on structured/tabular data.
- Handling Imbalanced Data: It offers advanced regularization and loss functions that make it resilient to class imbalance, which is a common issue in multi-label mental health prediction.

C. Despite its simplicity, Logistic Regression is a strong baseline model, especially in binary or multi-label classification problems. It was incorporated for the following reasons:

- One-vs-Rest Extension: When extended with One-Vs-Rest Classifier, logistic regression becomes capable of handling multi-label problems efficiently.
- Good with Linearly Separable Data: It performs well when there is a linear relationship between the symptoms and the presence of disorders.

D. Ensemble learning combines multiple models to improve prediction performance by reducing variance (bagging), bias (boosting), or improving predictions via voting mechanisms. In this study:

https://dx.doi.org/10.4314/njt.v44i4.10
http://creativecommons.org/licenses/by-nc-nd/4.0/

- Ensembles of Random Forest and XGBoost were tested to capture both low-variance and high-bias error components.
- Voting Classifiers and Stacking methods were considered to merge the strengths of different base classifiers.

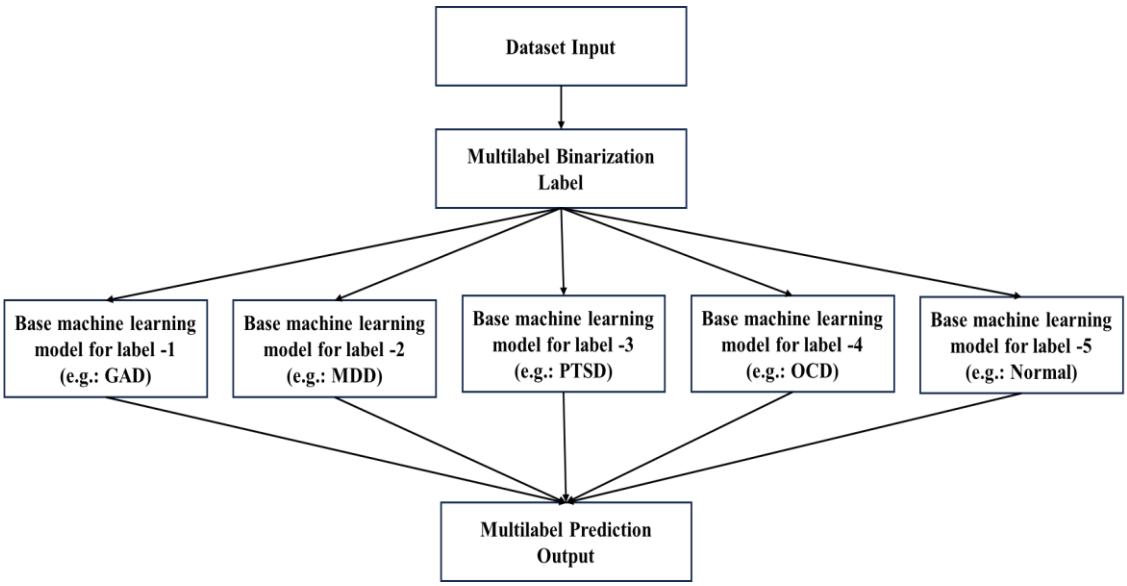

Figure 3: Multilabel classification with a base model obtained and compared to suggest the best model for the proposed system

Unlike traditional binary or multi-class classification, this proposed system employs a multi-label classification paradigm to simultaneously predict multiple mental health conditions for a single instance. This reflects real-world conditions where disorders such as Major Depressive Disorder (MDD), Generalized Anxiety Disorder (GAD), obsessive-compulsive disorder (OCD), and Post-Traumatic Stress Disorder (PTSD) frequently co-occur. To operationalize this, we utilize the Multioutput Classifier wrapper with all the above-mentioned base classifiers. Figure 3 shows the working of the training machine learning model. Upon classifying using these base models, the results were

Random Forest and XGBoost were chosen over deep learning due to the structured, categorical nature of the dataset. They handle smaller, imbalanced data efficiently, offer interpretable feature importance, and provide robust multi-label classification performance.

3.0 RESULTS AND DISCUSSION

3.1 Experimental Setup

In order to train and run machine learning models, the hardware used includes an Intel Core i5 processor, 16 GB of RAM for handling large datasets, and a 500 GB SSD for storing datasets, trained models, and necessary software libraries. A high-speed internet connection is also essential for downloading libraries, models, and accessing cloud services when needed. The used system runs on Windows. Python is the primary programming language used, with development typically conducted in environments of Google Colaboratory. Commonly used libraries and tools include Pandas and NumPy for data preprocessing, Matplotlib and Seaborn for visualization, Hyperopt for hyperparameter optimization, and Keras (when using TensorFlow as the backend) for simplified model building.

3.2 Performance Measures

The preprocessed data (real + synthetic) is split into training and testing sets using stratified sampling to preserve label distribution. To ensure the model's ability, the training phase primarily utilized CTGAN-generated synthetic samples, while the testing data was collected independently from students and teachers through structured questionnaires. This external testing ensures that the model's predictions remain consistent when applied to real-world responses, not just synthetic patterns. Since CTGAN captures diverse response distributions while the real responses reflect natural behavioral variability, the combination of both sources strengthens the model's generalizability to broader populations. The model is trained on this dataset and evaluated using metrics

https://dx.doi.org/10.4314/njt.v44i4.10
http://creativecommons.org/licenses/by-nc-nd/4.0/

tailored for multi-label tasks. These include Subset accuracy, hamming loss, Precision, Recall, F1-score, Per-label Accuracy, Micro-average, Macro-average, and weighted-average metrics, and Receiver-Operating Characteristic Curve.

3.3 Results

- **Subset Accuracy:** Measures the fraction of samples where all labels are correctly predicted.

$$\text{Subset Accuracy} = \frac{1}{N}\sum_{i=1}^{N} \mathbb{1}[y_i = \hat{y}_i] \quad (1)$$

Where N is the total number of samples, y_i are true label set for the i^{th} sample, $\hat{y}_i$ are the predicted label set for the i^{th} sample, and $\mathbb{1}$ is an indicator function, which returns 1 if the condition is true, 0 otherwise.

- **Hamming Loss**: Computes the fraction of misclassified labels.

$$\text{Hamming Loss} = \frac{1}{N.L}\sum_{i=1}^{N}\sum_{j=1}^{L} \mathbb{1}[y_{ij} \neq \hat{y}_{ij}] \quad (2)$$

Table 2: Model performance comparison: subset accuracy and hamming loss

Model	Subset Accuracy	Hamming Loss
Logistic Regression	0.06	0.328
XGBoost	0.25	0.28
Ensemble Model	0.24	0.274
Random Forest	0.23	0.26

- **Micro/Macro Averaged Precision, Recall, and F1-Score:** These metrics offer a balanced view of the model's performance across both frequent and rare labels.

$$\text{Precision}_{\text{Micro}} = \frac{\sum_{j=1}^{l} TP_j}{\sum_{j=1}^{L}(TP_j + FP_j)} \quad (3)$$

$$\text{Recall}_{\text{Micro}} = \frac{\sum_{j=1}^{l} TP_j}{\sum_{j=1}^{L}(TP_j + FN_j)} \quad (4)$$

$$\text{F1}_{\text{Micro}} = \frac{2.\text{Precision}_{\text{Micro}}.\text{Recall}_{\text{Micro}}}{2.\text{Precision}_{\text{Micro}} + \text{Recall}_{\text{Micro}}} \quad (5)$$

$$\text{Precision}_{\text{Macro}} = \frac{1}{L}\sum_{j=1}^{L}\frac{TP_j}{TP_j + FP_j} \quad (6)$$

$$\text{Recall}_{\text{Macro}} = \frac{1}{L}\sum_{j=1}^{L}\frac{TP_j}{TP_j + FN_j} \quad (7)$$

- **Per-label Accuracy:** Evaluates how well each disorder is predicted.

$$\text{Accuracy} = \frac{TP_j + TN_j}{TP_j + TN_j + FP_j + FN_j} \quad (8)$$

Random Forest outperformed other models primarily because of its ensemble nature and its capability to manage categorical and imbalanced data effectively. Its random feature selection and bagging mechanism reduced overfitting and enhanced stability in capturing distinct symptom patterns across multiple disorders, contributing to higher and more consistent AUC scores.

- **ROC-AUC Curves:** Used to visualize the trade-off between true positive rate and false positive rate across thresholds for each disorder. For a single label j,

$$AUC_j = \int_0^1 TPR_j(f)\,d(FPR_j(f)) \quad (9)$$

Where:

$$TPR_j = \frac{TP_j}{TP_j + FN_j} \quad (10)$$

$$FPR_j = \frac{FP_j}{FP_j + TN_j} \quad (11)$$

This comprehensive evaluation ensures both global and label-specific insights into the model's performance. These performance metrics were analysed for machine learning models like random forest, XGBoost, logistic regression, and ensemble techniques using XGBoost and random forest. Further insights were obtained through ROC-AUC analysis given in figures, which provides a more nuanced understanding of the model's discriminative ability using random forest as the base model. The AUC scores were as follows: PTSD (0.84), OCD (0.77), Normal (0.79), GAD (0.56), and MDD (0.44). These values confirm that the classifier can effectively distinguish between the presence and absence of PTSD, OCD, and Normal, but it struggles

https://dx.doi.org/10.4314/njt.v44i4.10

significantly with GAD and MDD. The lower AUC values for GAD and MDD suggest that these disorders exhibit overlapping and subtle symptom patterns, often co-occurring with others. This overlap reduces the distinctiveness of predictive features, making it difficult for the model to establish clear decision boundaries between these classes. These results collectively demonstrate the inherent complexity of multi-label classification in mental health detection.

The overlapping and subtle nature of symptoms across disorders like GAD and MDD reduces the model's ability to draw clear boundaries between classes. Additionally, the distribution of samples across labels and the relatively small dataset size may have contributed to the variability in model performance. From a clinical perspective, while the model performs well in detecting more distinguishable disorders such as PTSD and OCD, cautious interpretation is necessary for GAD and MDD predictions. False negatives could delay timely support for individuals in need, whereas false positives may cause undue stress or misdirection of clinical attention. The ROC - AUC curves for XGBOOST, Logistic Regression, ensemble technique, and Random Forest are given in Figures 4, 5, 6 and 7.

Table 3: Performance metrics of different models

Model	Label	Precision	Recall	F1-Score	Support
Logistic Regression	GAD	0.67	0.11	0.19	36
	MDD	0.2	0.03	0.06	29
	Normal	0.75	0.2	0.32	15
	OCD	0.38	0.34	0.36	44
	PTSD	0.12	0.04	0.06	26
	micro avg	0.39	0.16	0.23	150
	macro avg	0.43	0.14	0.2	150
	weighted avg	0.41	0.16	0.2	150
	samples avg	0.21	0.15	0.16	150
XGBoost	GAD	0.53	0.28	0.36	36
	MDD	0.24	0.17	0.2	29
	Normal	0.41	0.47	0.44	15
	OCD	0.69	0.61	0.65	44
	PTSD	0.68	0.5	0.58	26
	micro avg	0.54	0.41	0.47	150
	macro avg	0.51	0.41	0.45	150
	weighted avg	0.54	0.41	0.46	150
	samples avg	0.45	0.42	0.42	150
Ensemble Model	GAD	0.53	0.28	0.36	36
	MDD	0.26	0.17	0.21	29
	Normal	0.38	0.33	0.36	15
	OCD	0.71	0.61	0.66	44
	PTSD	0.72	0.5	0.59	26
	micro avg	0.56	0.4	0.47	150
	macro avg	0.52	0.38	0.44	150
	weighted avg	0.55	0.4	0.46	150
	samples avg	0.43	0.4	0.4	150
Random Forest	GAD	0.36	0.11	0.17	36
	MDD	0.09	0.03	0.05	29
	Normal	0.67	0.27	0.38	15
	OCD	0.76	0.64	0.69	44
	PTSD	0.81	0.5	0.62	26
	micro avg	0.62	0.33	0.43	150
	macro avg	0.54	0.31	0.38	150
	weighted avg	0.53	0.33	0.4	150
	samples avg	0.38	0.33	0.34	150

https://dx.doi.org/10.4314/njt.v44i4.10
http://creativecommons.org/licenses/by-nc-nd/4.0/

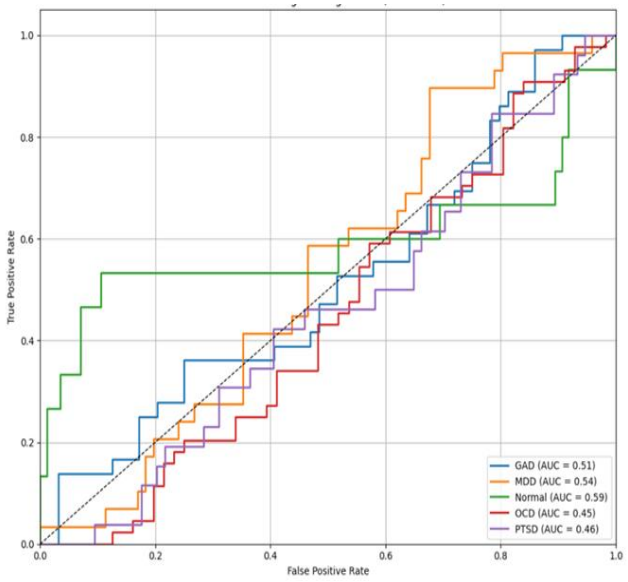

Figure 4: ROC-AUC of logistic regression model

The model using Logistic Regression predicts the labels with an area under the curve for GAD as 0.51, MDD as 0.54, PTSD as 0.46, OCD as 0.45, and Normal as 0.59. All these values suggest that nearly half of the labels are misclassified.

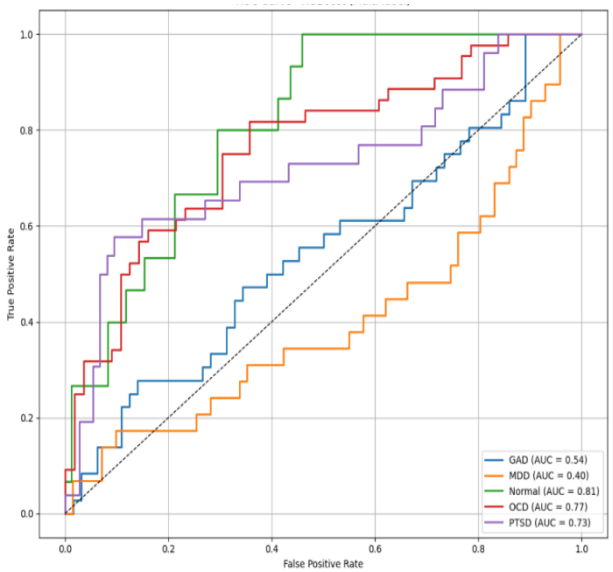

Figure 5: ROC-AUC of XGboost model

When this dataset is trained with XGBoost, it predicted the labels OCD, PTSD, and Normal with AUC as 0.77,0.73,0.81, respectively. It could classify MDD and GAD with 0.40 and 0.51 AUC, respectively.

On combining Random Forest and XGBoost using the Ensemble technique, the model predicted the labels OCD, Normal, and PTSD with AUC 0.77,0.80,0.80, respectively, and labels GAD, MDD with 0.55 and 0.41, respectively.

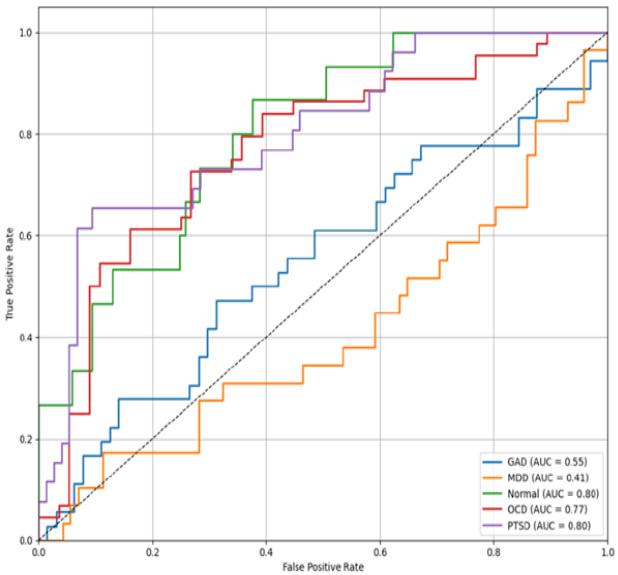

Figure 6: ROC-AUC of ensemble model

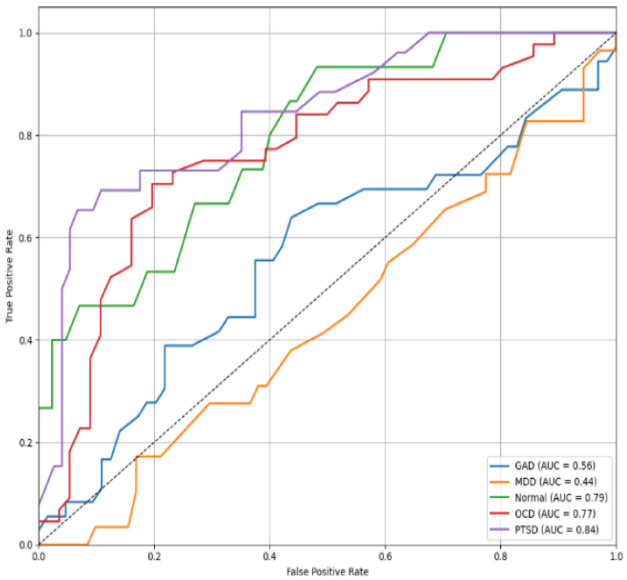

Figure 7: ROC-AUC of Random Forest

Another model Random Forest, predicted the labels OCD, PTSD, Normal, GAD, MDD with AUC as 0.77,0.84,0.79,0.56,0.44, respectively.

4.0 CONCLUSION AND FUTURE SCOPE

In conclusion, a multi-label classification model was successfully developed using a dataset specifically tailored for detecting multiple mental health conditions simultaneously. Despite the complexity of the task and modest subset accuracy, the model demonstrated promising per-label performance for

https://dx.doi.org/10.4314/njt.v44i4.10

PTSD, OCD, and Normal cases. In this study, we evaluated the performance of multiple machine learning models—XGBoost, Logistic Regression, Random Forest, and an ensemble of XGBoost and Random Forest for multi-label classification of mental health disorders. The assessment was based primarily on ROC-AUC analysis, which provided a nuanced understanding of each model's ability to distinguish between the presence and absence of various disorders. Among all models tested, the Random Forest classifier demonstrated the most consistent and superior performance, particularly for PTSD (AUC = 0.84), OCD (AUC = 0.77), and Normal (AUC = 0.79).

These results indicate that Random Forest effectively captures patterns associated with more distinguishable mental health conditions. Although the model exhibited lower AUC values for GAD (0.56) and MDD (0.44), this trend was consistent across other models, highlighting the intrinsic difficulty in detecting disorders characterized by overlapping or subtle symptoms. The ensemble approach offered more balanced predictions across labels but did not surpass Random Forest in terms of peak performance for the most identifiable disorders. Logistic Regression and XGBoost, meanwhile, struggled with overall classification, particularly for GAD and MDD, with AUC scores approaching the threshold of random chance. Based on these findings, Random Forest emerges as the most suitable model for this mental health detection system. Its robustness, ability to handle noisy and categorical data, and strong discriminative power for key disorders make it a compelling choice for practical implementation.

The results underscore the potential of machine learning in supporting mental health diagnostics, while also revealing the challenges that arise from symptom overlap and data imbalance. With further enhancements in data quality and model design, such systems can play a significant role in promoting mental health awareness and enabling early screening in clinical and community settings. Mental health conditions often exhibit overlapping symptoms and are influenced by a variety of social, psychological, and biological factors, making them inherently difficult to distinguish, especially from limited or ambiguous data. The model's difficulty in classifying GAD and MDD could be due to feature sparsity, insufficient training data for these specific disorders, or the lack of subtle linguistic and behavioral cues that are often better captured in multimodal settings (e.g., speech, text, facial expression analysis). With further refinement, this model can serve as a valuable screening tool for mental health professionals. In clinical settings, such systems could assist in the early identification of potential disorders, enabling healthcare providers to initiate timely intervention and personalized care plans. Integrating this technology into digital health platforms or mobile applications could allow individuals to complete preliminary assessments remotely, reducing the burden on mental health infrastructure and promoting early awareness.

While the study demonstrates promising results, it is based on a relatively small dataset with a high proportion of synthetic samples, which may limit generalizability. The absence of large-scale external validation also constrains direct clinical applicability at this stage. Future improvements, including integration with multimodal data sources and real-world trials, will help strengthen reliability and support smoother adoption into clinical workflows.

To enhance real-world applicability, future work should focus on expanding the dataset to include more samples and achieving better class balance. Incorporating additional contextual features, such as medical history or behavioral patterns, could also improve prediction accuracy. Moreover, exploring advanced models like deep learning architectures or transformer-based approaches may help capture subtle patterns and correlations missed by traditional classifiers.

REFERENCES

[1] Kim, J. Kim, D. Kamphaus, R. "Early Detection of Mental Health Through Universal Screening at Schools". Georgia Educational Researcher, 2022, doi: 10.20429/ger.2022.190104

[2] Nadeem, M. Rashid, J. Moon, H. Dosset, A. "Machine Learning for Mental Health: A Systematic Study of Seven Approaches for Detecting Mental Disorders", International Technical Conference on Circuits/Systems, Computers, and Communications (ITC-CSCC), Jeju, Korea, Republic of, 2023, pp. 1-6, doi: 10.1109/ITC-CSCC58803.2023.10212609.

[3] Ahuja R. and Alisha B. "Mental Stress Detection in University Students Using Machine Learning Algorithms." Procedia Computer Science, 152, pp. 349-353, 2018, doi: 10.1016/j.procs.2019.05.007

[4] Kim, J. Liu, N. Tan, H. Chu, C. "Unobtrusive monitoring to detect depression for elderly with chronic illnesses", IEEE Sensors Journal. 17, (17), pp 5694-5704, 2017, doi: 10.1109/JSEN.2017.2729594

[5] Osman, A.B., Tabassum, F., Patwary, M.J., Imteaj, A., Alam, T., Bhuiyan, M.A., & Miraz, M.H. "Examining Mental Disorder/Psychological Chaos through Various ML and DL Techniques: A Critical Review", Annals of Emerging Technologies in Computing, 2022, doi: 10.33166/AETiC.2022.02.005

[6] Seal, A. Bajpai, R. Agnihotri, J. Yazidi, A. Herrera-Viedma, E. Krejcar, O. " DeprNet: A Deep Convolution Neural Network Framework for Detecting Depression Using EEG" IEEE Transactions on Instrumentation and Measurement, p 1, 2021, doi: 10.1109/TIM.2021.3053999

[7] Wijayarathna, C. Lakshika, E. "Toward Stress Detection During Gameplay: A Survey," *in IEEE Transactions on Games*, 15 (4), pp. 549-565, 2023, doi: 10.1109/TG.2022.3216404

[8] Kumar, R. Pooja, K. Udathu, M. Lakshmi, J. Santhosh, C. "Detection of Depression Using Machine Learning Algorithms", *International Journal of Online and Biomedical Engineering*, 18, pp 155-163, 2022.

[9] Bhavani, B. Naveen, N. "An Approach to Determine and Categorize Mental Health Condition using Machine Learning and Deep Learning Models," Engineering, Technology & Applied Science Research, 14, pp 13780-13786, 2024, https://doi.org/10.48084/etasr.7162

[10] Mashrura, T. Ramon, D. Eleni, S. "A Machine learning Model For Detecting Depression, Anxiety, and Stress from Speech," IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Seoul, Korea Republic, 2024, pp. 7085-7089, 2024, doi: 10.1109/ICASSP48485.2024.10446567

[11] Anu, P., Shruti, G. Neha, T. "Predicting Anxiety, Depression and Stress in Modern Life using Machine Learning Algorithms" Procedia Computer Science, 167, 2019, doi: 10.1016/j.procs.2020.03.442

[12] Mario, A. Adrian, L. Luis, G. Manuel, M. "Detecting Mental Disorders in Social Media Through Emotional Patterns - The Case of Anorexia and Depression" IEEE Transactions on Affective Computing, 14 (1) pp. 211-222, 2023, doi: 10.1109/TAFFC.2021.3075638

[13] Guo, Y. Zhang, Z. Xu, X., "Research on the detection model of mental illness of online forum users based on convolutional network", *BMC Psychology*, 11, p 424, 2023, https://doi.org/10.1186/s40359-023-01460-4

[14] Khoo, L.S. Lim, M.K. Chong, C.Y. McNaney, R. "Machine Learning for Multimodal Mental Health Detection: A Systematic Review of Passive Sensing Approaches", *Sensors*, 24 (2), p 348, 2024, https://doi.org/10.3390/s24020348

[15] Sharma, K., Ahmed, I. A. Ahmad, K. Ghanshyam, G. Tejani, F. A. Seyed J. M. "Early Detection of Mental Health Disorders Using Machine Learning Models Using Behavioral and Voice Data Analysis." *Scientific Reports*, 15(1), pp. 1-19, 2025, https://doi.org/10.1038/s41598-025-00386-8.

[16] Kiran, V. K. Tiwari, G. "Early Identification and Management of Mental Disorders". *International Journal of Indian Psychology*, 11(4), pp 697-701, 2023, doi: 10.25215/1104.060

[17] Yadav, G. Bokhari, M. Alzahrani, S. Alam, S. Shuaib, M. "Emotion-Aware Ensemble Learning (EAEL): Revolutionizing Mental Health Diagnosis of Corporate Professionals via Intelligent Integration of Multi-Modal Data Sources and Ensemble Techniques", IEEE Access, p 1, 2025, doi: 10.1109/ACCESS.2025.3529032

[18] Abd Rahman, R. Omar, K. Mohd Noah, S. A. Mohd D. Mohd Shahrul N. Al-Garadi, M. "Application of Machine Learning Methods in Mental Health Detection: A Systematic Review", IEEE Access, 8. pp. 183952-183964, 2020, 10.1109/ACCESS.2020.3029154.

[19] Abdullah, M. & Negied, N. " Detection and Prediction of Future Mental Disorder From Social Media Data Using Machine Learning, Ensemble Learning, and Large Language Models", IEEE Access, p. 1, 2024, 10.1109/ACCESS.2024.3406469.

[20] Kim, J. Liu, N. Tan, H. X. Chu, C. "Unobtrusive monitoring to detect depression for elderly with chronic illnesses", IEEE Sensors Journal, 17, (17),

pp. 5694-5704, 2017. doi: 10.1109/JSEN.2017.2729594

[21] Lei X. Maria S. Alfredo C. Kalyan V. "Modeling tabular data using conditional GAN." Proceedings of the 33rd International Conference on Neural Information Processing Systems. Curran Associates Inc., Red Hook, NY, USA, Article 659, pp. 7335–7345, 2019.

[22] Breiman, L. "Random Forests." Machine Learning. 45, pp. 5-32, 2001, 10.1023/A:1010950718922.

[23] Jesse, R. Bernhard, P. Geoff, H. Eibe, F. "Classifier chains for multi-label classification", Machine Learning. 85 (3), 333–359, 2011, https://doi.org/10.1007/s10994-011-5256-5.

[24] Zhang, M. Zhou, Z. "A Review On Multi-Label Learning Algorithms. Knowledge and Data Engineering", IEEE Transactions on Knowledge and Data Engineering, 26, pp. 1819-1837, 2014, 10.1109/TKDE.2013.39.