Nanotechnology Perceptions
ISSN 1660-6795
www.nano-ntp.com

An Intelligent Hybrid Framework For Classification Of Anxiety And Depression Using Data Augmentation

Vaibhav Sharma[1], Sanjay Sharma[2*]

[1,2]School of Engineering & Technology, Shri Guru Ram Rai University, Dehradun, Uttarakhand-248001,India
[1]vsdeveloper10@gmail.com,Orcid-0000-0002-1404-2012
[2]sanjaypokhriyal@yahoo.com,Orcid-0000-0001-7625-5091*
*Corresponding author- [2]*sanjaypokhriyal@yahoo.com*

Mental health problems are significant challenges for individuals or society as well. The impact of mental health disorders is a kind of slow poison that affects the country's growth substantially. The accurate and timely diagnosis of these disorders is crucial for effective intervention and treatment. The current research proposes an innovative Intelligent Hybrid Framework for the classification of mental health disorders such as depression and anxiety by utilizing advanced machine learning techniques and data augmentation. The proposed framework provides the classification of mental health disorders. It enhances the accuracy and robustness of the classification process with the involvement of synthetic data values generated from the existing dataset. The hybrid framework capitalizes on the strengths of different classifiers and incorporates ensemble learning principles to create a more resilient and reliable classification system. The outcomes indicate the classification of depressive diseases and normal text. The proposed model provides 99.1% accuracy of the system for the classification of data.

Index Terms—Hybrid Machine Learning, Mental illness, Anxiety, Depression, Data augmentation.

I. INTRODUCTION

There is a broad spectrum of mental disorders that fall under the umbrella of mental disease and impact a person's thoughts, feelings, behavior, and general state of mind. A person's ability to operate daily, their relationships, and their quality of life can all be materially impacted by these conditions. There is a wide range of mental diseases, each with its own unique set of symptoms, duration, and severity. Certain common mental health illnesses, like obsessive-compulsive disorder, ADHD, depression, anxiety, schizophrenia, bipolar disorder, and post-traumatic stress disorder, are more common than others.Figure 1 shows seven common types of mental illness. Although, all mental illness conditions are very critical for human beings, however, anxiety and depression are two very dangerous conditions for individuals [1].

Anxiety is a normal and adaptive human response to stress or perceived threats. It can be a temporary emotional state or a typical scenario that can create fear and unhappiness among the people. Anxiety is a natural part of life and can serve a protective function, excessive or uncontrolled anxiety can interfere with daily functioning and overall well-being [2]. Anxiety can be classified into numerous forms, including panic disorder, social anxiety, specific phobias, and post-traumatic stress disorder. Depression is a mental health condition marked by persistent feelings of melancholy, hopelessness, and a loss of interest or pleasure in activities. It extends beyond usual sensations of sadness or mood swings and has a considerable impact on an individual's thoughts, emotions, and everyday functioning [3]. The most prevalent types of depression include persistent depressive disorder (dysthymia), bipolar disorder, major depressive disorder, and seasonal affective disorder.

Classifying mental illnesses using machine learning faces several challenges, reflecting the complexity and multifaceted nature of mental health. Subjectivity or heterogeneity is a kind of mental health condition with symptoms varying widely among individuals [4]. Each mental health problem varies greatly, making it difficult to develop a one-size-fits-all strategy. The availability of high-quality, diversified, and representative mental health datasets is a substantial difficulty. Ethical and privacy considerations are extremely sensitive, and maintaining individual privacy is critical. Ensuring ethical considerations, obtaining informed consent, and implementing strong security measures are crucial. Mental health disorders can have multiple temporal variations. Changes in symptoms over time, such as the episodic nature of some illnesses, present obstacles for model training and interpretation [5].

Mental health difficulties may result in underreporting or distortion of symptoms. Labeling biases can reduce model accuracy and contribute to inequities in mental health care. The lack of consistent protocols for data collection, evaluation tools, and diagnostic criteria across research and datasets impedes the creation of generally applicable models. Mental health issues can change over time, and people may react differently to interventions [6]. Cultural differences in expressing and experiencing mental health symptoms may not be adequately captured by models trained on specific datasets, limiting their generalizability across diverse populations.

The classification of mental illnesses can be performed by using machine learning and other techniques to analyze the behavior or patterns based on the input data. Supervised learning mechanisms are suitable for binary classification and generate the pattern for predicting the specific mental condition for high-dimensional datasets. The ensemble of distinct techniques can handle the categorical data and perform complex relationships. Deep learning techniques such as CNN, RNN, ANN, and LSTM can build an intricate pattern from complex data [7]. These techniques are suitable to analyze the spatial relationships in brain imaging or mental health.

Unsupervised learning algorithms can also be utilized to create similar instance groups without any predefined labels. Autoencoders can learn efficient representations of data and potentially reveal latent factors related to mental health. The most instructive and pertinent traits that can help with the categorization process can be found using feature engineering. Principal component analysis and t-distributed Stochastic Neighbor Embedding (t-SNE), two feature transformation methods, can aid in dimensionality reduction. Data augmentation generates synthetic data to increase the diversity of the dataset, particularly important when dealing with limited or imbalanced datasets. Data augmentation expands the dataset and reduces the risk of overfitting and improving model generalization.

To address the challenges and absence of a hybrid complex model, there is a requirement for responsible and ethical development practices for the classification of mental health. The major contribution of the current research is as follows:

- An intelligent hybrid framework is proposed for the classification of anxiety and depression mental illness.

- To extract the features that are best associated with anxiety and depression mental illness.
- To apply a data augmentation mechanism for the generation of new modified data using existing datasets.
- To validate and optimize the accuracy rate and reduction in the time complexity.

The remaining paper can be categorized into the following sections: section II focuses on the latest research which has been conducted for the classification of mental illness. Section III proposed an intelligent hybrid framework for the classification of anxiety and depression with data augmentation. Section IV discussed the experimental testbed details and dataset description. Section V demonstrates the results and comparative analysis with the existing outcomes. Finally, section VI concludes the research with significant contributions.

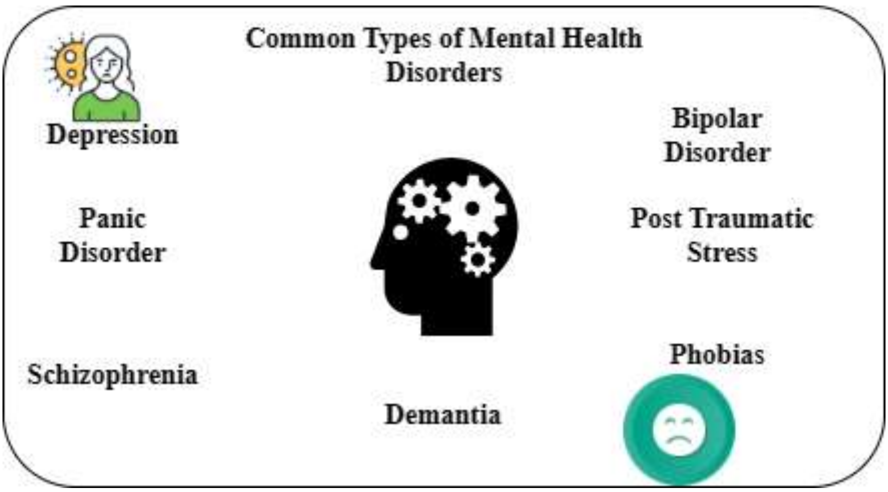


Fig. 1. Common Types Mental illness

II. BACKGROUND

Mental health illness refers to a wide range of conditions that affect a person's mood, thinking, behavior, or overall mental well-being. The severity of these conditions varies and they may make it difficult for a person to go about their everyday life as usual. A few prevalent categories of mental health conditions include, a frequent and significant mental health condition called depression is marked by enduring melancholy and hopelessness as well as a loss of interest in or enjoyment from daily activities. It influences how a person feels, thinks, and goes about their everyday business. It frequently results in physical and emotional symptoms that have a major negative influence on a person's quality of life. It is typified by enduring depressive and hopeless feelings as well as a loss of interest in or enjoyment from activities. It can have a big effect on someone's energy, attitude, and capacity to do everyday chores. A group of mental illnesses can be represented in terms of anxiety disorders and characterized by excessive and persistent feelings of anxiety, fear, worry, or apprehension that are disproportionate to the actual threat or situation. These disorders can significantly impair daily functioning and quality of life include conditions such as generalized anxiety disorder, social anxiety disorder, panic disorder, and phobias.

Previously recognized as manic-depressive illness, bipolar disorder is a long-term mental health ailment marked by intense mood fluctuations, including sadness and manic or hypomanic emotional highs. These mood episodes can vary in severity and duration, and they significantly affect a person's mood, energy levels, behavior, and functioning. It is characterized by sharp mood fluctuations, with manic episodes (high spirits, heightened vigor) and depressive episodes (low spirits, decreased interest). These emotional outbursts can range in length and severity. Persistent and severe mental illness, schizophrenia is typified by abnormalities in thinking, feeling, seeing, and acting. It is one of the most severe and debilitating mental diseases, impacting about 1% of people globally. Delusions, disordered thought patterns, hallucinations, and poor social functioning are possible symptoms. The condition known as post-traumatic stress disorder (PTSD) arises from going through or seeing a terrible event. Flashbacks, nightmares, hypervigilance, and avoiding triggers that remind the victim of the trauma are possible symptoms. Recurring thoughts, or obsessions, and repetitive actions, rituals, or compulsions, that a person feels compelled to carry out are symptoms of obsessive-compulsive disorder (OCD). OCD can be distressing and severely interfere with day-to-day activities. The symptoms of attention-deficit/hyperactivity disorder (ADHD), which can last into adulthood, include impulsivity, hyperactivity, and inattention. These symptoms typically start in childhood. ADHD can affect academic, occupational, and social functioning. Several key factors can affect the exacerbation of mental health illnesses such as:

Biological Factors:

- Genetics: Having a family history of mental health issues can make one more likely to experience related problems.
- Medical diseases: Psychiatric symptoms can arise because of medical diseases that affect mental health, such as thyroid issues, neurological disorders, or chronic pain.
- Neurochemistry: Disturbances in neurotransmitters, including norepinephrine, serotonin, and dopamine, may be a factor in mood disorders like anxiety and depression.
- Brain Structure and Function: Deficits in specific brain regions' size, connection, or activity, as well as other abnormalities, have been linked to mental health conditions like bipolar disorder and schizophrenia.

Psychological Factors:

- Adversity and Trauma: Being exposed to traumatic situations, such as violence, abuse, neglect, or loss, can make mental health conditions like depression or post-traumatic stress disorder (PTSD) more likely to occur.
- Stress: Prolonged stress at work, school, in relationships, or due to financial hardships can cause or exacerbate mental health issues.
- Personality features: Some personality features, including neuroticism, perfectionism, or poor self-esteem, can make people more vulnerable to mental health issues.
- Coping Mechanisms: Avoidance, substance abuse, and rumination are examples of maladaptive coping mechanisms that can help keep mental health symptoms alive.

Social and Environmental Factors:

- Socioeconomic Status: Socioeconomic factors are the major concerns for mental health issues which include unstable financial conditions, lack of accessibility, and a smaller number of resources.
- Social Support: Unreliability, untrustful, and stress in daily life are facilitating social support.
- Cultural and Societal Factors: Access to cultural beliefs, social attributes, stigma, and discrimination can generate the challenges of mental health.
- Environmental Exposures: The environmental exposures may generate risks such as pollutants, poisons, and substances which is related to mental health.

Lifestyle Factors:

- Diet and Nutrition: Unhealthy habits of the person can lead to nutrient deficiencies and health challenges that may affect the mood or mental health of the person.
- Physical Activity: In the current modern world, there are fewer physical activities done by people that affect the health as well as the mental health of the person.
- Substance Use: The use of alcohol, drugs, medicines, and cigarettes can elevate the signs of a person's mental health.

It attempts to investigate the data pertaining to depression and anxiety disorders by looking at the causes of mental disease. The current article is used to examine the best-fitted machine learning model that can predict mental health and classify the same in distinct issues. The feature selection for the model building is also a critical factor for finding out the better accuracy and assessment of model prediction.

III. LITERATURE REVIEW

A literature review of depressive and anxiety disorders might look at a variety of subjects, including prevalence, causes, symptoms, diagnosis, treatment operating options, and impacts on people and society. A systematic evaluation of recent searches on the appliance of machine learning to foresee mental health problems was carried out by Chung et al. [8]. Utilizing the PRISMA technique, the writers executed a comprehensive hunt of reliable bases and found 30 relevant research strides to enter. The writers examined the limitations and challenges associated with the appliance of machine learning to mental health concerns by researchers. Among the challenges include the complexity of mental health conditions, ethical dilemmas, model interpretability, and data quality. Despite these obstacles, writers stress machine learning holds the potential to tremendously transform mental health care by enhancing diagnostic precision and providing personalized, data-driven treatment.

A total of 33 papers on the diagnosis of mental health conditions such as schizophrenia, depression, anxiety, bipolar disorders, trauma stress disorder (PTSD), anorexic nervosa, and attention hyperactivity disorder (ADHD) were thoroughly dissected by Lyortsuun et al. [9]. The objective of the study was to illuminate many applications of machine learning and deep learning in mental health diagnosis. Anxiety, bipolar disorder, dementia, and psychosis are among the mental disorders that Nasrullah et al. [10] predicted using multiclass classification. The authors employed an ensemble model for multiclass classification, which combines multiple base models to make predictions. The accuracy of the model exceeds 92% in predicting the class labels and mental illness prediction.

For college students, Du et al. [11] developed a unique deep learning-based mental health monitoring system (DL-MHMS). With the help of effective CNN-based technology, the suggested model was able to categorize college students' EEG signals into three groups: positive, negative, and normal. Attaining the greatest classification accuracy and F1 scores of 97.54% and 98.35%, respectively, the suggested framework achieves impressive performance metrics. Reductions in sleeping difficulties (21.19%), low levels of depression (18.11%), and attention to suicide (28.14%) are other indications of its notable benefits over current models. Further demonstrating the model's efficacy in promoting and assessing mental health among college students, increases the self-esteem ratio by 98.42% and the personality development ratio by 97.52%. Sutranggono et al. [12] explored the potential of social media data for insights into mental illness. Experimental results demonstrated the efficacy of the multi-class and multi-level algorithms in detailed classification, with promising results at each classification level. The Robust Optimized BERT Pre-training approach (RoBERTa) classifier achieves the highest accuracy at both levels with 0.98 at level 1 and 0.95 at level 2 classifications.

Gentili et al. [13] addressed the challenge of imbalanced datasets, particularly prevalent in healthcare datasets like those derived from Electronic Health Records (EHRs). They applied several different data balancing approaches to the original data, such as random under-sampling, oversampling, and Synthetic Minority Oversampling Techniques for Nominal and Continuous (SMOTE-NC). Studies are necessary to improve the effectiveness of current methods since scientific research is still based on imbalanced datasets. Four BERT models created by Hugging Face were compared with conventional machine learning methods frequently employed in automatic depression detection by Pourkeyvan et al. [14]. The findings of the research revealed that the new BERT models consistently outperformed the previous approaches and achieved an accuracy rate of up to 97% in identifying depression.

Mazumdar et al. [15] presented a novel screening tool, the Indian Scale for Assessment of Autism (ISAA), specifically designed for the Indian population. The study utilized the ISAA along with additional significant features to develop a classification model for autism spectrum disorder (ASD). A number of machine learning models, including Random Forrest, K-Nearest Neighbours (KNN), and Support Vector Machinery (SVM), are evaluated on their performances following training. Higher weighted F1 scores of 85.7% and accuracy of 90% were attained by the SVM algorithm. Zhou et al. [16] investigate anxious depression, a common subtype of major depressive disorder (MDD) associated with significant impairments in social functions and adverse resultants. A total of 260 MDD patients and 127 healthful controls participated in the study, which included resting-state functional magnitudes stream imaging and three-dimensional T1-weighted structural scanning. A diagnostic predictive random forest model including imaging data uses an area under the curve (AUC) of 0.802, as validated by 10-fold cross-validations, to distinguish anxious from non-anxious MDD patients.

As Fang Hand et al. [17] examine the mechanisms underly depressive disorders (DD), generalized anxiety disorders (GAD), and healthful controls (HC), they create a diagnostic paradigm for triple classifications. Electroencephalogram (EEG) signals from 38 HC individuals, 45 GAD patients, and 42 DD patients were used to calculate brain functional connects using the Faze Lag Indices (PLI). The explanation examines the differences in functional connections across the three groups and examines the effects of time window feature computations on classification accuracies. Several machine learning models, including LightGBM, CatBoots, XGBoost, and ensemble models, were utilizations to identify the EEG data. According to the resultants, the ensemble models achieve the highest accuracies of

97.33% in the classifications for all three groups, with a 12-second time window producing the best resultants. Following extensive reviews, some research gaps were discovered regarding the predictions of mental diseases, such as anxiety and depression. In the following section, a proposed framework to enhance performances is described, which combines data augmentation techniques and machine learner classifiers.

IV. PROPOSED SYSTEM FRAMEWORK

The proposed framework consists of distinct modules (Figure 2) from data augmentation to model building. The classification of mental health diseases can be done by using different mechanisms. However, the utilization of data augmentation is required in the proposed model for model building.

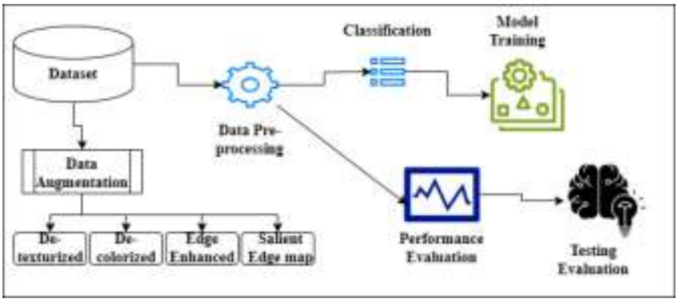


Fig. 2. Proposed System Framework

A. Data Augmentation

Data augmentation is a technique that can generate data artificially and enhance the quantity of data for diversifying the results. This technique can be utilized with original data and modification in the existing data for robustness, performance, and data quality. This technique also reduces the chances of overfitting and underfitting. The data augmentation technique can enhance the diversity of the data values and variability of the range in the dataset. There are several libraries available such as pytorch, keras, and open CV for the inclusion of data in an artificial manner. Firstly, it includes transformations such as translation, flipping, and rotation of the images and then applies some facial expressions to provide the variability [18]. The image saturation, contrast, color variations, and brightness of the image can enhance the variations in the image. Finally, some random distortion, noise, and uncertainties in the image for the inclusion of visual representations.

B. Data Pre-processing

The dataset is integrated which is collected from multiple sources and focused the anxiety and depression. The data pre-processing is the critical step utilized for the preparation of the dataset and correct classification of the attributes or features. The dataset is prepared by reducing the unrelated features from the categorization. The extraction of data from the relevant features and fill the missing values by taking the mean values of the column [19]. The text input must be converted to the numerical format with bag-of-words or word2vec mechanisms. Data balancing is also required which can deal with oversampling or undersampling of the dataset. Finally, the data with scaling and normalization techniques. The normalization of the dataset can be achieved by using equation (1):

$$Y_N = \frac{Y - Y_{min}}{Y_{max} - Y_{min}} \qquad (1)$$

Where the normalized value can be identified with the maximum and minimum value of any attribute. Further, the dataset is split into different subsets such as training, testing, and validation. The division of the dataset is random in nature [20]. The integrity of the dataset split is ensured during the distribution of samples across distinct classes.

C. Classification

Classifying test samples is a critical step in the machine learning pipeline, providing an opportunity to evaluate a model's performance and ensure it generalizes well to unseen data. This process involves applying a trained classification model to test samples to predict their class labels or categories. The classification of data depends on several steps for the classification of mental health and anxiety. The very first step is the removal of outliers and noise in the existing dataset. The selection of features and normalization must be utilized for classification purposes [21]. The identification of text data used several keywords such as restless, anxious, worried, anxious, nervous, depression, etc. The statistical test analysis is also done for the evaluation of the sample's significance.

D. Model Training

The classification is done with the help of distinct machine learning mechanisms such as linear classification, logistic regression, SVM, and neural networks.
The linear classification can be evaluated with the following equation (2):

$$y = w.x + c \qquad (2)$$

Where y is a dependent variable that can be evaluated by using the weight of the vectors, x is the feature or attribute for evaluating the y, and c is the bias that exists during the model building [22]. Logistic regression is another important binary classification method with the probability of getting output in a sample. It also used the sigmoid function or the linear classification with the following equation (3):

$$y = \sigma(w.x + c) \qquad (3)$$

Where σ is the sigmoid function can be evaluated with the probability belongs to class 1. And it can be represented in equation (4):

$$\sigma(z) = \frac{1}{1 + e^{-z}} \qquad (4)$$

The main objective of SVM is to evaluate the hyperplane which is best suited for the separation of two distinct classes. The boundary of SVM can be represented in the following equation (5):

$$y = \text{sign}\,(w.x + c) \qquad (5)$$

The sign function is used to determine the side of the hyperplane. The feedforward neural network is also utilized to evaluate for classification purpose. The mathematical function of the feedforward neural network can be represented in the following equation (6):

$$y^l = \sigma\left(w^l.x^{l-1} + c^l\right) \qquad (6)$$

Where σ represent the sigmoid function or activation function, w^l weight matrix for layer l, c^l is the bias vector for layer l.

E. Performance Evaluation

The evaluation of performance for the classification of anxiety and depressionusing data augmentation involves a combination of statistical metrics and cross-validation techniques. There are several metrics can be evaluated by using robustness and generalization of the classification model [23]. The accuracy of the model is indicated by the percentage of correctly identified samples among all samples. The precision matrix is produced by dividing the total number of positive predictions produced by the model by the ratio of true positive predictions. The ratio of actual positives in the data to true positive forecasts is known as recall. A table that counts true positives, true negatives, false positives, and false negatives is called a confusion matrix. It offers information on potential misclassifications of data by the model.

The dataset is divided into 'k' subsets, or folds, using a technique called k-fold cross-validation. The model is trained on 'k-1' folds, and the remaining fold is tested. The same process is repeated 'k' times which ensures robustness and generalization.After k-fold cross-validation, evaluation of augmentation data is also required [24]. The augmentation robustness ensures that the model does not overfit on the generated data. The model must be tested by using test data and maintain the performance in distinct scenarios. Finally, the comparative analysis is done to evaluate the performance of the models with the hybrid model and traditional neural network. The performance is evaluated and compared with the other models for the effectiveness and reliability of the model.

V. EXPERIMENTAL TESTBED AND DATASET DESCRIPTION

The dataset is available on an open-source system which has two columns, one belongs to the text, and the other belongs to the labels. This dataset is available in Excel format and comprises around 6500 data from social media, Facebook comments, posts, etc. The data belongs to distinct age groups of people and students with different categories of jobs [25]. The data is being collected with different keywords such as depression, anxiety, worried, etc. There are two columns available in the dataset whereas the text column contains text data, that is collected from distinct sources such as social media posts, forum discussions, or survey responses. The text data describes personal experiences, emotions, thoughts, or behaviors. The text column contains information about the person who wrote the messages or tweets on social media. The second column is for labels with the values 0 or 1, whereas 1 indicates that there is at least one keyword that belongs to anxiety or depression [26]. While 0 indicates that no keyword belongs to the anxiety. Figure 3 shows the dataset statistics with a number of variables, observations, duplicate rows, and other information. Figure 4 shows the dataset variable cardinality, unique value percentage, missing values, and memory size. It also counts the number of zeros and ones in the label column.

Dataset statistics

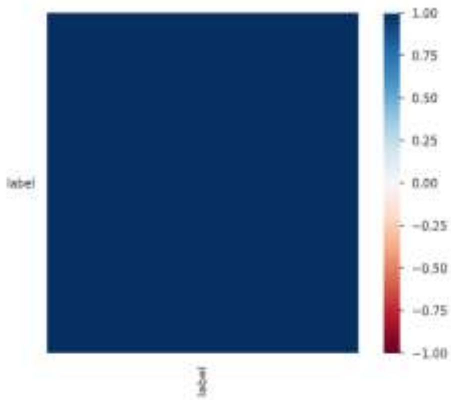
Number of variables	2
Number of observations	6982
Missing cells	12
Missing cells (%)	0.1%
Duplicate rows	44
Duplicate rows (%)	0.6%
Total size in memory	109.2 KiB
Average record size in memory	16.0 B

Variable types

Categorical	2

Fig. 3. Dataset Statistics

Fig. 4. Dataset variables with text and label columns



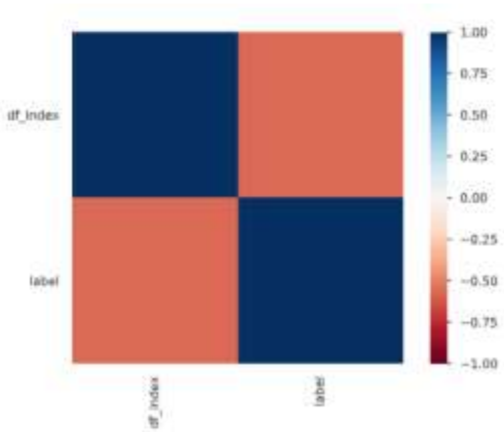


Fig. 5.Spearman’s correlation among variables

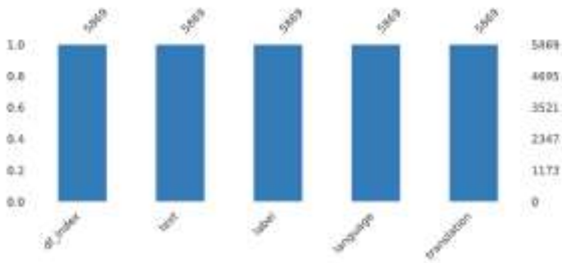


Fig. 6.Missing Values in dataset

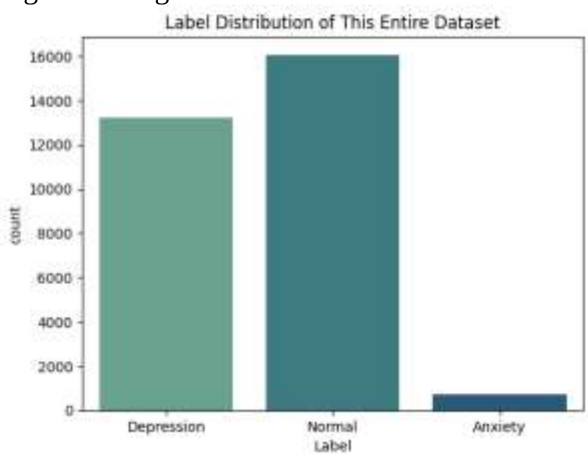


Fig. 7. Label distribution

Figure 5 shows the correlation among the variables with both the attributes and generates a heatmap. The heatmap indicates the strong correlation among the variables. Figure 6 shows the missing values that exist in the dataset for distinct attributes. Figure 7 shows the distribution of labels after classification among depression, normal, and anxiety. It also indicates the high number of normal labels and lowest number of anxiety labels exist in the dataset.

VI. RESULTS AND DISCUSSION

This section describes the results and analysis done with the help of the proposed framework. Figure 8 shows the label column distribution of the dataset. It indicates that 90% of the labels denote the depressive and 10% anxiety data which is quite imbalanced. The dataset is balanced by using k-fold cross-validation. Figure 9 shows the top 5 rows of the dataset with text, labels, total words, and characters. Figures 10 and 11 show the density of the data with reference to the total number of words and characters. These graphs clearly indicate that the density of label 0 is much greater than that of label 1.

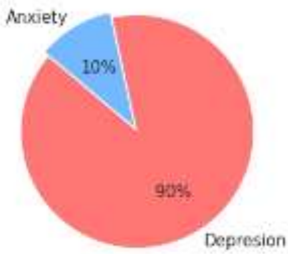


Fig. 8.Dataset label distribution

	text	label	Total Words	Total Chars
0	oh my gosh	1.0	3	8
1	trouble sleeping, confused mind, restless hear...	1.0	10	55
2	All wrong, back off dear, forward doubt. Stay ...	1.0	14	65
3	I've shifted my focus to something else but I'...	1.0	11	51
4	I'm restless and restless, it's been a month n...	1.0	14	59

Fig. 9. Top 5 rows of dataset

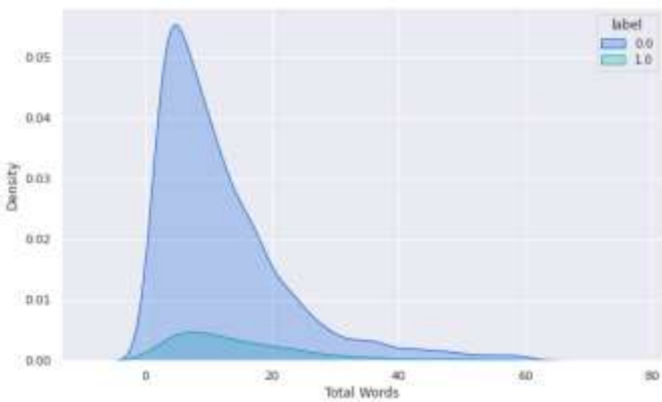


Fig. 10.Density Vs total words

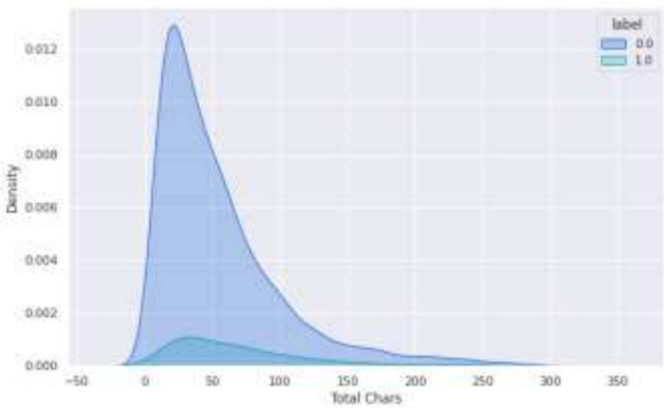


Fig. 11.Density Vs total characters

Fig. 12.Normal keywords in dataset

Fig. 13.Keywords with anxiety and depression

Figures 12 and 13 show the type of keywords with or without anxiety and depression. There are distinct keywords such as want, make, can’t, real, good, etc. clearly indicate the normal expressions of the person. While, some keywords like restless, anxious, nervous, anxiety, etc. represent the depressive words available in the dataset. Figures 14 and 15 show the most frequent word used when the person is not depressed and depressed respectively. Restless is the word that occurred the highest number of times in the dataset. Similarly, want is also a keyword that occurs frequently in the dataset.

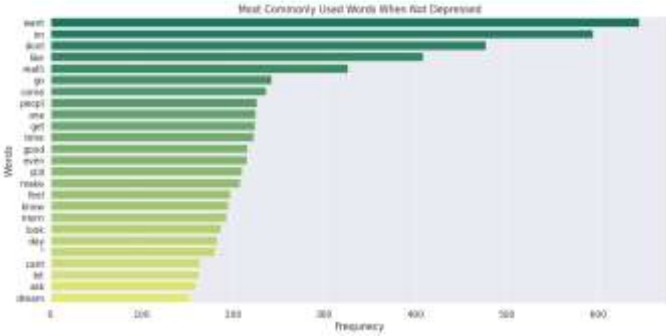


Fig. 14.Most frequently words when not depressed

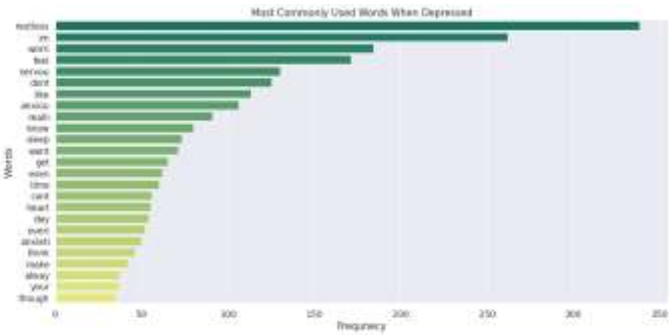
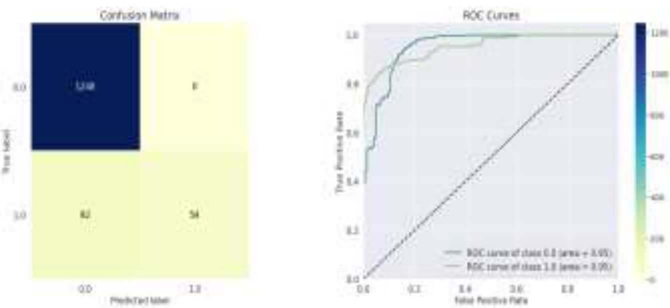
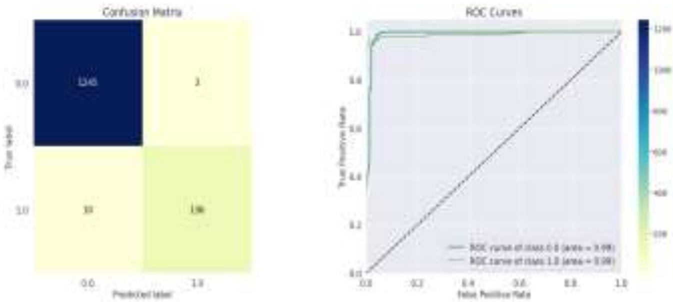


Fig. 15.Most frequently words when depressed



Prediction	Anxiety	Depression
Actual		
Anxiety	4881	1
Depression	433	161

Fig. 16.Confusion matrix and RoC curve with Multinomial NB



Prediction Actual	Anxiety	Depression
Anxiety	4991	1
Depression	1	583

Fig. 17.Confusion matrix and RoC curve with Random Forest

Figure 16 shows the confusion matrix and RoC curve with multinomial naïve Bayes and random forest algorithm. The confusion matrix indicates a higher number of true negative classifications with fewer false positives and negatives. The RoC curve with random forest covered more area under the curve for both the classes, class 0 and 1. The RoC curve with multinomial NB generates 95% area under the curve while random forest generates 99% area under the curve for both the classes. The accuracy for multinomial NB is 93.4% and the precision is 100% which indicates the pretty good performance of the model. The confusion matrix for test data shows the anxiety and depression classification. The random forest algorithm generates 99.1% accuracy of the model with a 97.8% precision score in Figure 17. Table 1 shows the LSTM model sequential with distinct parameters and dense layers. Figure 18 shows the classification report for the proposed framework that shows the accuracy parameters with macro and weighted averages. It also shows the number of suicide and non-suicide cases who express their feelings on social media with depressive keywords.

Table 1. LSTM Model: "sequential_1"

Layer (type)	Output Shape	Param #
embedding_1 (Embedding)	(None, 40, 300)	81560700
lstm_1 (LSTM)	(None, 40, 20)	25680
global_max_pooling1d_1	(Glob (None, 20)	0
dense_2 (Dense)	(None, 256)	5376
dense_3 (Dense)	(None, 1)	257

Total params: 81,592,013
Trainable params: 31,313
Non-trainable params: 81,560,700

	Precision	recall	f1-score	support
non-suicide	0.92	0.95	0.93	23209
suicide	0.95	0.91	0.93	23206
accuracy			0.93	46415
macroavg	0.93	0.93	0.93	46415
weightedavg	0.93	0.93	0.93	46415

Fig. 18.Classification report for the model

VII. CONCLUSION

There are several problems and issues with the human being in today's world. Although, the physical problem can be monitored and solved by proper treatment or guidance. But, no one is talking about mental health. Mental health such as anxiety and depression are significant challenges for individuals and societies. The timely identification and detection of these diseases are equally important than other challenges. The current research proposes an intelligent hybrid framework for the timely diagnosis of these diseases. By integrating advanced machine learning techniques with data augmentation methods, the proposed framework offers a promising approach to improving the accuracy and efficiency of anxiety and depression classification. Data augmentation is utilized for synthetic data generation and to improve the quality of the dataset. The results indicate the number of suicide and non-suicide cases with 93% accuracy. This ensures that the classification can effectively differentiate between different levels of anxiety and depressive comments and texts. The iteratively refining the classification framework based on new data generated from the data augmentation or feedback ensures the optimization and relevance to the clinical purposes. The future scope of the proposed work is to design an ensemble learning mechanism for the classification model. Also, the dataset can be collected in real-time for better prediction and treatments.

ACKNOWLEDGMENT

The authors declare that, there is no financial support from any of the institutions or personal relationship to affect the quality of the paper.

REFERENCES

[1] J. A. Harrington and V. Blankenship, "Ruminative Thoughts and Their Relation to Depression and Anxiety1," Journal of Applied Social Psychology, vol. 32, no. 3. Wiley, pp. 465–485, Mar. 2002. doi: 10.1111/j.1559-1816.2002.tb00225.x.

[2] J. W. G. Tiller, "Depression and anxiety," Medical Journal of Australia, vol. 199, no. S6. AMPCo, Oct. 2013. doi: 10.5694/mja12.10628.

[3] L. Ishihara and C. Brayne, "A systematic review of depression and mental illness preceding Parkinson's disease," ActaNeurologicaScandinavica, vol. 113, no. 4. Hindawi Limited, pp. 211–220, Apr. 2006. doi: 10.1111/j.1600-0404.2006.00579.x.

[4] A. Muhammad et al., "Classification of Anxiety Disorders using Machine Learning Methods: A Literature Review," Insights of Biomedical Research, vol. 4, no. 1. sPage.direcT, Nov. 07, 2020. doi: 10.36959/584/455.

[5] G. Cho, J. Yim, Y. Choi, J. Ko, and S.-H. Lee, "Review of Machine Learning Algorithms for Diagnosing Mental Illness," Psychiatry Investigation, vol. 16, no. 4. Korean Neuropsychiatric Association, pp. 262–269, Apr. 25, 2019. doi: 10.30773/pi.2018.12.21.2.
[6] R. A. Rahman, K. Omar, S. A. Mohd Noah, M. S. N. M. Danuri, and M. A. Al-Garadi, "Application of Machine Learning Methods in Mental Health Detection: A Systematic Review," IEEE Access, vol. 8. Institute of Electrical and Electronics Engineers (IEEE), pp. 183952–183964, 2020. doi: 10.1109/access.2020.3029154.
[7] I. Elujide, S. G. Fashoto, B. Fashoto, E. Mbunge, S. O. Folorunso, and J. O. Olamijuwon, "Application of deep and machine learning techniques for multi-label classification performance on psychotic disorder diseases," Informatics in Medicine Unlocked, vol. 23. Elsevier BV, p. 100545, 2021. doi: 10.1016/j.imu.2021.100545.
[8] J. Chung and J. Teo, "Mental Health Prediction Using Machine Learning: Taxonomy, Applications, and Challenges," Applied Computational Intelligence and Soft Computing, vol. 2022. Hindawi Limited, pp. 1–19, Jan. 05, 2022. doi: 10.1155/2022/9970363.
[9] N. K. Iyortsuun, S.-H. Kim, M. Jhon, H.-J. Yang, and S. Pant, "A Review of Machine Learning and Deep Learning Approaches on Mental Health Diagnosis," Healthcare, vol. 11, no. 3. MDPI AG, p. 285, Jan. 17, 2023. doi: 10.3390/healthcare11030285.
[10] S. Nasrullah and A. Jalali, "Detection of Types of Mental Illness through the Social Network Using Ensembled Deep Learning Model," Computational Intelligence and Neuroscience, vol. 2022. Hindawi Limited, pp. 1–6, Mar. 26, 2022. doi: 10.1155/2022/9404242.
[11] C. Du, C. Liu, P. Balamurugan, and P. Selvaraj, "Deep Learning-based Mental Health Monitoring Scheme for College Students Using Convolutional Neural Network," International Journal on Artificial Intelligence Tools, vol. 30, no. 06n08. World Scientific Pub Co Pte Ltd, Dec. 2021. doi: 10.1142/s0218213021400145.
[12] A. N. Sutranggono, RiyanartoSarno, and Imam Ghozali, "Multi-Class Multi-Level Classification of Mental Health Disorders Based on Textual Data from Social Media," Journal of Information and Communication Technology, vol. 23, no. 1. UUM Press, Universiti Utara Malaysia, pp. 77–104, Jan. 30, 2024. doi: 10.32890/jict2024.23.1.4.
[13] E. Gentili et al., "Machine learning from real data: A mental health registry case study," Computer Methods and Programs in Biomedicine Update, vol. 5. Elsevier BV, p. 100132, 2024. doi: 10.1016/j.cmpbup.2023.100132.
[14] A. Pourkeyvan, R. Safa, and A. Sorourkhah, "Harnessing the Power of Hugging Face Transformers for Predicting Mental Health Disorders in Social Networks," IEEE Access, vol. 12. Institute of Electrical and Electronics Engineers (IEEE), pp. 28025–28035, 2024. doi: 10.1109/access.2024.3366653.
[15] A. Mazumdar, B. Chatterjee, M. Banerjee, and S. Shanker, "Machine learning based autism screening tool—a modified approach," Multimedia Tools and Applications. Springer Science and Business Media LLC, Feb. 23, 2024. doi: 10.1007/s11042-024-18519-y.
[16] E. Zhou et al., "Prediction of anxious depression using multimodal neuroimaging and machine learning," NeuroImage, vol. 285. Elsevier BV, p. 120499, Jan. 2024. doi: 10.1016/j.neuroimage.2023.120499.
[17] J. Fang et al., "Exploring Abnormal Brain Functional Connectivity in Healthy Adults, Depressive Disorder, and Generalized Anxiety Disorder through EEG Signals: A Machine Learning Approach for Triple Classification," Brain Sciences, vol. 14, no. 3. MDPI AG, p. 245, Mar. 01, 2024. doi: 10.3390/brainsci14030245.
[18] A. Nadeem, M. Naveed, M. Islam Satti, H. Afzal, T. Ahmad, and K.-I. Kim, "Depression Detection Based on Hybrid Deep Learning SSCL Framework Using Self-Attention Mechanism: An Application to Social Networking Data," Sensors, vol. 22, no. 24. MDPI AG, p. 9775, Dec. 13, 2022. doi: 10.3390/s22249775.

[19] L. Yang, D. Jiang, and H. Sahli, “Feature Augmenting Networks for Improving Depression Severity Estimation From Speech Signals,” IEEE Access, vol. 8. Institute of Electrical and Electronics Engineers (IEEE), pp. 24033–24045, 2020. doi: 10.1109/access.2020.2970496.

[20] Asmita De and S. Mishra, “Augmented Intelligence in Mental Health Care: Sentiment Analysis and Emotion Detection with Health Care Perspective,” Augmented Intelligence in Healthcare: A Pragmatic and Integrated Analysis. Springer Nature Singapore, pp. 205–235, 2022. doi: 10.1007/978-981-19-1076-0_12.

[21] L. Zhang et al., “Hybrid representation learning for cognitive diagnosis in late-life depression over 5 years with structural MRI,” Medical Image Analysis, vol. 94. Elsevier BV, p. 103135, May 2024. doi: 10.1016/j.media.2024.103135.

[22] N. Dhariwal et al., “A pilot study on AI-driven approaches for classification of mental health disorders,” Frontiers in Human Neuroscience, vol. 18. Frontiers Media SA, Apr. 10, 2024. doi: 10.3389/fnhum.2024.1376338.

[23] S. Verma, C. Sharma, G. Aggarwal, and P. Upadhya, “Artificial Intelligence-Based Approach for Classification and Prediction of Mental Health,” 2024 14th International Conference on Cloud Computing, Data Science & Engineering (Confluence). IEEE, Jan. 18, 2024. doi: 10.1109/confluence60223.2024.10463203.

[24] P. Moon and P. Bhattacharyya, “We Care: Multimodal Depression Detection and Knowledge Infused Mental Health Therapeutic Response Generation.” arXiv, 2024. doi: 10.48550/ARXIV.2406.10561.

[25] S. Bhatnagar, J. Agarwal, and O. R. Sharma, “Detection and classification of anxiety in university students through the application of machine learning,” Procedia Computer Science, vol. 218. Elsevier BV, pp. 1542–1550, 2023. doi: 10.1016/j.procs.2023.01.132.

[26] G. Singh, O. ChetiaPhukan, R. Gupta, and A. Nayyar, “Hybrid deep learning model for wearable sensor-based stress recognition for Internet of Medical Things (IoMT) system,” International Journal of Communication Systems, vol. 37, no. 3. Wiley, Nov. 06, 2023. doi: 10.1002/dac.5657.