International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(14s) 2026 pp. 1026-1046
https://doi.org/10.70917/ijcisim-2026-4319
Article

Severity Detection of Anxiety, Depression, And Stress Using Ensemble Learning and Deep Neural Networks: A Multi-Class Approach

Neha Rajas[1], Balaji Patil[2]

[1,2]Department of Computer Engineering, MIT-WPU School of Computer Engineering & Technology, Pune
Email- neha.rajas@mitwpu.edu.in[1] balaji.patil@mitwpu.edu.in[2]

Abstract: The study examines ordinary DASS-21 surveys and various models of machine and deep learning to identify the levels of anxiety, depression, and stress. Mental health data class imbalance greatly affects the upper class intensity of model training time. To optimize model performance, the dataset was balanced using the Synthetic Minority Over-sampling Technique (SMOTE). We assessed cross model performance by tuning accuracy, precision, recall, and f1 score and analyzing the various techniques employed. The enactment of the models was rated based on the predominance of upper intensity class response. Of the several models, the Random Forest algorithms surpassed the others/competitors across all levels of intensity. The model was able to identify all levels of intensity approximately 98.43% of the time. The apparent intra-class postulation of Random Forest where the average accuracy, recall, and F1-scores were high postulates perfect and equal intra-class distribution. The results showed that Random Forest did the best out of all the models, and this, along with the best performance as a whole, led to the conclusion that ensemble methods are a reliable way to classify different levels of mental health seriousness. Random Forest's success shows that it could be used to automatically find mental health problems in study and therapy settings. The data also show that automatic systems can help decision support systems find high-risk individuals early on and give mental health systems the tools they need to help people in the right way at the right time.
Keywords: - Machine Learning; Mental Health; Depression; Severity Classification; Deep Learning.

1. Introduction

Anxiety, stress, and depression, are a major area of concern, as they affect all segments of the global population, regardless of age, education, or occupation. Uncertainties that relate to the economy, coupled with stress in the workplace, social isolation, and the overall speed of modern life, have especially affected the psychological vulnerability of university students and young adults. The challenges that these students are facing include high levels of emotional stress, a decline in their academic results, impaired decision making and resulting health consequences that are chronic [1]. It is important to note that in addition to being overlapping and co-morbid, there are other complicating factors that are associated with the anxiety, stress and depression that are common. Primary preventive mental health care is dependent upon the screening of individuals, and determining the level of severity of the symptoms. Innovative methods of mental health care that promise efficiency, and are scalable, target the identification of mental health related problems as a means of prevention [2]. In mental health care, the two main methods used are clinical conversations and psychological surveys. The DASS-21 assessment has been the most popular and widely used of the second group. People use the DASS-21 assessment to find out how much psychological discomfort they are feeling because of things like worry, anxiety, and sadness. It sorts these conditions into a number of intensity levels, ranging from normal to very serious, middling, and severe. This particular DASS-21 trait is very useful for mental health workers [3].

Nonetheless, in large-scale screening contexts, there may be inconsistencies, largely because the evaluation of questionnaire-type assessments must rely on manual scoring, and on subjective scoring. Additionally, conventional methods grapple with identifying the more intricate patterns and relationships in the responses to the questionnaires, and this hampers their ability to more fully assess the mental health of large populations in an automated way [4].

Machine learning (ML) and deep learning (DL) have been added to healthcare analytics over the past few years, which has led to big changes in how mental health data is analysed and interpreted before it is used in practice. Data-driven models utilize predictors in high dimensions, learn complex patterns and relationships, and provide interpretations that other statistical methodologies may be unable to do. Recently, ML techniques, including Support Vector Machines, k-Nearest Neighbors, Random Forests, Naïve Bayes, and various boosting techniques, have been reported in the literature for the classification of mental health utilizing datasets, including surveys, behaviors, and physiological measures [5]. Deep neural networks, for instance, have been recognized for their ability to model some of the psychological dimensions more accurately. Notwithstanding the foregoing progress, however, most of the research to date has been on the coarse-grained or binary classification. The binary approach of simply asking the yes/no question of did the person experience a depression episode is too simplistic and does not provide the level of detail of the clinical nuances that the analysis may be needed for to support the treatment level and/or the prioritization of the risk support [6]. Numerous researchers also tend to address these three constructs (the anxiety, the depression, and the stress) separately and in isolation, thus ignoring the relationships of co-occurrence and the associated complexity of the interconstructs. The mental health datasets have been characterized by the lack of balance or the number of normal or mild cases in comparison to the significant, and in some cases, the extremely, severe cases. This imbalance is problematic as it usually results in a model that is not optimized and may underperform with little sensitivity to the most needed higher levels of the category of the cases. Moreover, depending on only one model constrains its reliability, given that separate classifiers could fail to provide adequate generalizations over various data distributions and demographic clusters [7].

The use of deep neural networks and ensemble models provides a path forward for overcoming the challenges described. Ensemble methods, including Random Forests, Support Vector Machines (SVMs), and k-Nearest Neighbors (k-NN), seek to optimize predictive stability (provide a predictive model that is consistent across multiple datasets), reduce variance, and improve generalization by producing ‘strong’ predictors from an assemblage of multiple, ‘weak’ predictors [8]. Given the high dimensionality, and the presence of noise and data imbalance in the mental health datasets, ensemble approaches may provide the most appropriate solutions due to their ability to produce and leverage diverse decision boundaries. Deep learning approaches are equally well positioned, as they stand to decode ‘subtle’ nuances associated with the severity of psychological symptoms through the modeling of complex interactions and layers of hierarchies. Of particular concern in the mental health severity classification problem is the phenomenon of imbalanced datasets and the methods employed to address this issue [9]. The use of the Synthetic Minority Over-sampling Technique (SMOTE), and methods that create synthetic data for the less represented classes of severity, ensure that the model ‘learns’ in a balanced fashion and improves its focus on the clinically important classes of severe and extreme-severe. Additionally, multi-class classification allows for a more thorough and clinically relevant evaluation of a model by permitting the model to predict multiple classes of severity as opposed to binary class models. Evaluating these models with a comprehensive set of evaluation metrics including accuracy, precision, recall, and the F1-score and ROC AUC provides a balanced evaluation across all severity classes, and increases the confidence in the conclusions drawn from the analysis[10]. This research aims to build and assess a system for multi-class classification of anxiety, depression, and stress severity using ensemble learning and deep neural networks on DASS-21 data, focusing on the construction of a smart, data-driven model.

The goal of this task comes from the fact that mental health problems are becoming more common and that current screening methods aren't good enough to handle large amounts of data and tell the difference between mild and severe mental health problems. The study is mostly about how to use class imbalance methods along with machine learning and deep learning to make mental health severity predictions more accurate, reliable, and useful in the real world. The research revolves around the implementation and comparative studies of machine learning techniques namely Naïve Bayes, KNN, SVM and Random Forest (which are some of the common models) and deep learning frameworks to classify and segregate five different levels (which are the five strata) of severity for each of the three components of mental health that is, depression, anxiety, and stress [11]. The study is about mental health data, mental health screening, and mental health questionnaires and the technique of class imbalance has been smote (which stands for synthetic minority oversampling technique) and thus the framework of this study serves educational and clinical pilot screening purposes justifiably. Mental health issues are prevalent, and there is an evident necessity for maintaining an assessment of mental health to provide appropriate, comprehensive, automated, and large scale assessments in a readily mental health system, to define and diagnose the mental health issues that exist in a system [12]. This makes early involvement in mental health problems possible and helps improve the standard of professional choices about mental health issues. Also, it helps mental health professionals focus on and help people who are having problems with suicide or death. Accessibility and a mental health screening system are two goals of the suggested system. The goal is to make mental health screening systems more accessible and to provide mental health screening systems that are based on research. It is possible for this system to go digital and help with the screening systems in evidence-based mental health screening systems [13].

2. Related Work

Using ensemble learning and deep neural networks, recent studies in AI and mental health screening have focused on finding and judging the seriousness of the three mental illnesses of anxiety, sadness, and stress automatically. Additionally, Rashid et al. (2024) looked into automatic screening for anxiety, sadness, and burnout in Swiss medical students. They compared ensemble models and a deep neural network (DNN) and found that the DNN was the most accurate at predicting the future. The study, however, reported issues concerning the homogeneity of the sample and the lack of psychosocial and context variables of the models, which in turn, limited the generalizability of the models [14]. De Souza et al. (2021) described a framework using a stacking-based deep learning ensemble of a CNN and LSTM to classify social media texts regarding depression, anxiety, and their comorbidity, whereby, even though the framework markedly enhanced the detection of depression, the classification of anxiety still posed a problem. This problem was the result of the ambiguity and context dependence and the labeling noise present in the social media data [15]. For Alphonsa Sini P. J. et al. (2023), the focus was on the comparative study of anxiety, depression, and stress detection using a model of machine learning/ deep learning models like SVM, ANN, and XGBoost. Although the accuracy of the results from the study was impressive, there was minimal engagement regarding the issues of the interpretability, data bias, and ethics of the study which raised concerns of the applicability of the study clinically [16]. Ashik (2024) Illustrated model versatility by proving robustness through the prediction of stress levels using several classifiers tied together by a voting ensemble, also highlighting a pivotal reliance on the quality of the data and the feature representation [17]. Lastly, using a multilayer perceptron-based deep learning model, Liu (2024) classified the severity of psychiatric symptoms via physiological indicators, supporting differentiated objectively the severity, but reinforcing the need for multimodal integration to improve the reliability and clinical usefulness [18].

Table 1. Related Work on Mental Health Severity Classification

Author & Year	**Dataset Used**	**Algorithms Used**	**Accuracy (<96%)**	**Key Findings / What They Reviewed**
Kumar et al., 2020 [1]	DASS-21 Survey Dataset	SVM, KNN, Naïve Bayes	89.4%	Traditional ML struggled with class imbalance; moderate performance for DAS classification
Li & Liu, 2020 [19]	Stress Sensor Dataset	Deep Neural Network	92.1%	DNN effective for physiological stress detection; dataset size limited accuracy
Souza et al., 2021[15]	Reddit Text (Anxiety & Depression)	Deep Learning Ensemble	93.5%	High depression accuracy; anxiety & comorbidity detection weaker
Nijhawan et al., 2022 [9]	Social Interaction Text Dataset	NLP + ML Models	88.6%	Stress detection dependent on linguistic variations; moderate performance
Zhang et al., 2022 [20]	DASS-based Multi-class Dataset	Multi-task Deep Learning	95.2%	Joint learning improved accuracy but required large, clean datasets
Alphonsa Sini P.J. et al., 2023[16]	Clinical DAS Dataset	SVM, ANN	94.8%	Good results but interpretability issues; mild/moderate levels difficult to separate
Anand et al., 2023 [21]	Stress Assessment Dataset	Random Forest, Boosting	95.6%	Ensemble methods improved minority class detection but still <96%

Ashik et al., 2024 [17]	Stress Behavior Dataset	Voting Ensemble	93.9%	Ensemble improved robustness; feature dependency reduced generalization
Saylam et al., 2024 [22]	Wearable DAS Dataset	Multitask Learning	92.7%	Good temporal severity estimation; wearable dependency limits scalability
Rahman et al., 2024 [5]	Survey-based Stress Dataset	Neural Networks	90.3%	Survey noise reduced accuracy; NN useful but inconsistent on mild classes

In recent years, the severity of mental illnesses in terms of their classification and analytics has gradually moved away from binary classification to a multi-class classification model that aligns with how a clinician would evaluate a patient as well as how they would implement a treatment plan. The studies done by Anand et al. (2023) illustrated the effectiveness of using ensemble learning approaches such as boosted and forest-based classifiers in providing high accuracy in classifying stress levels that are not skewed by the imbalanced nature of classifying individuals based on their level of stress. The findings of this study support the importance of protecting the minority severity classification classes to create reliable predictions for patients with severe mental illnesses that are clinically critical [21]. Additionally, Zhang et al. (2022) built on these types of findings; they developed a sequential multi-task deep learning model to simultaneously classify and detect the three types of mental illnesses (depression, anxiety, and stress) and revealed that by using a joint model to explicitly model the relationships and the temporal nature of the symptoms, their model was better than the individual models used for each disorder alone and demonstrated the value of an integrated approach to severity classification [20]. Singh et al. (2025) combined traditional machine learning techniques with CNN, LSTM and GRU in their Hybrid NLP & Deep Learning Framework for detecting anxiety, depression and stress from unstructured data (social media comments). Their model was able to predict these conditions effectively but raised concerns about platform-specific bias, data representativeness and temporal instability in detecting mental health through social media comments [23]. Similarly, Zhang et al. (2022) developed a deep learning-based framework called DAC Stacking which incorporates deep learning techniques in stacking ensemble models to enable detection of anxiety, depression and their co-occurrence from Reddit data. While the model achieved excellent results for detecting depression, it performed poorly in detecting anxiety and the comorbidity of anxiety and depression primarily due to language subtleties and the ambiguous nature of labelling [20]. In addition to social media comments, Saylam et al. (2024) explored the potential for multi-task learning using longitudinal and digital biomarkers among a cohort of college students, providing further evidence that multi-factor temporal considerations are critical for accurate classification of levels of severity of mental health conditions [22].

3. RESEARCH METHOD

This study adopts a structured, data-driven research methodology aimed at developing an automated multi-class severity classification framework for anxiety, depression, and stress.

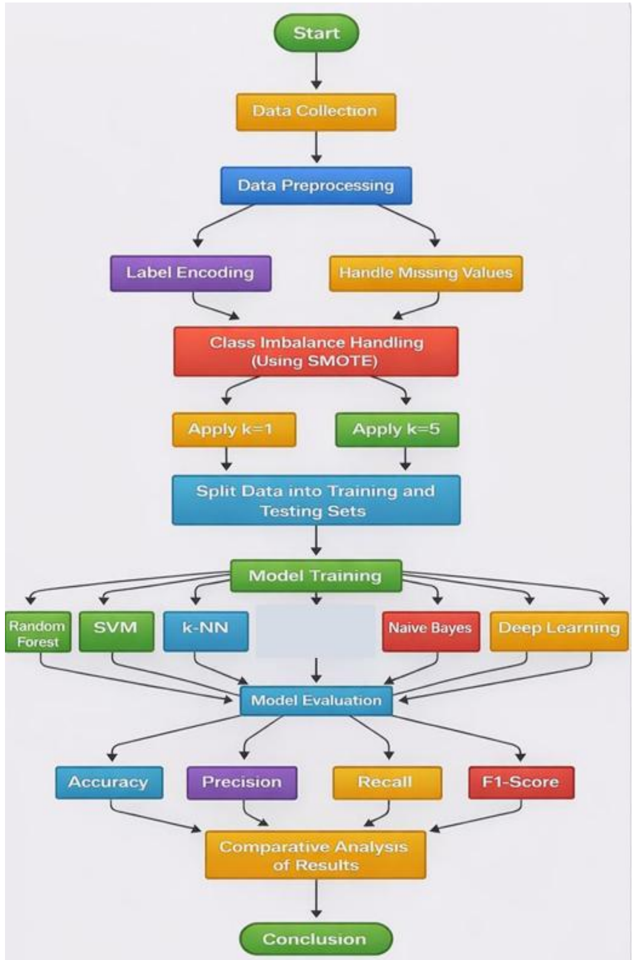


Fig.1. Proposed Method

The developed climatic conditions method builds on advancements made recently in the areas of Ensemble Learning and Deep Neural Networks (DNN). This new YN-M methodology will provide researchers with a way to do a more complete job of Classifying a Case than through the traditional use of Binary Classification Models, by giving researchers the degree of precision required for Clinical Assessment Practices. In addition, this new methodology allows the ability to detect subtle differences in the severity of illnesses which is important when considering appropriate intervention/support services for clients seeking counselling services for Mental Issues. The framework begins with the Importation of Data from publically available Mental Health Survey Data Base using the DASS-21 questionnaire, which is a validated Psychometric Tool used to Assess the Level of Severity of Depression (5 Levels of Severity), Anxiety & Stress (5 Levels of Severity). The Examination of Data has been executed to ensure Data Integrity (i.e., No Missing or Inconsistent Records), Develop Scores according DASS-21 Guidelines, Apply Categorial Encoding (Where Necessary). To deal with the class imbalance that often arises when creating datasets related to mental health severity, this research has chosen to use the SMOTE techniques to generate synthetic data for those categories that are under-represented. In addition to this, generating additional synthetic samples will allow for an increased sensitivity of the resultant models to both severe and extremely severe case types. The study method uses a lot of different kinds of supervised machine learning models and deep learning frameworks, so all the models that were made can be compared. Traditional ensemble learning methods and deep neural networks were both used to make the models. Accuracy, precision, memory, F1 Score, and ROC-AUC scores were used to test all of these methods. The main goal of this project was to make a solid, expandable system for automatically judging the severity of mental health problems. This would help doctors and researchers use this data to make better decisions.

3.1 Data Collection and Profiling

The study uses public mental health polls with people's answers to the Depression, Anxiety, and Stress Scale (DASS-21) as examples of datasets. There are 21 questions on the DASS-21, and each one is graded on a four-point Likert scale. These questions measure the levels of depression, anxiety, and stress in respondents over the preceding week. A considerable amount of effort went into the dataset preparation before the model was built. The highest level of data accuracy was accomplished through the use of pre-processed data collection techniques. Points of data that were absent, unsatisfactory, inconsistent, or conflicting were removed and data integrity was preserved. Using standard DASS-21 scoring protocols, the raw scores were totaled and assigned to the corresponding DASS-21 subscale. Imbalance in the class distribution caused an imbalance in the class distribution for the subscales, and the class imbalance was noted as being anywhere from average to high. The Synthetic Minority Over-sampling Technique (SMOTE) is applied to obtain class representatives for the minority subscale. This method balances things out so that the model is less biased and the guided learning is more general. After that, a full detailed statistical analysis was done to see how the emotional aspects and intensity groups were spread out. This helped in the selection of appropriate performance metrics for the machine learning classification models and helped to strengthen the downstream machine learning classification models and the development knowledge about the dataset for predictive accuracy.

(a) Score Calculation of Stress

Table 2. Score Calculation of Stress

Q1	Q6	Q8	Q11	Q12	Q14	Q18	Score	Score × 2	Severity
1	2	1	0	1	1	1	7	14	Normal
2	2	2	2	0	1	2	11	22	Moderate
3	2	1	1	1	0	0	8	16	Mild

Table 2 illustrates the DASS-21 stress score calculation based on responses to seven stress-related items (Q1, Q6, Q8, Q11, Q12, Q14, and Q18). Individual item scores are summed to obtain the raw stress score, which is then multiplied by two as per DASS-21 guidelines to maintain equivalence with the full DASS scale. The scaled scores indicate the different severities of stress. It shows that scaled scores of 14, 16, and 22 correspond to Normal, Mild, and Moderate stress, respectively. This shows the systematized and standardized nature for the different levels of stress.

(b) Score Calculation of Anxiety

Table 3. Score Calculation of Anxiety

Q2	Q4	Q7	Q9	Q15	Q19	Q20	Score	Score × 2	Severity
0	0	1	2	1	0	0	4	8	Mild
1	0	0	3	2	1	0	7	14	Moderate
2	1	3	0	0	3	2	11	22	Extremely Severe

Table 3 shows how to figure out your DASS-21 anxiety score by using seven items that are related to worry (Q2, Q4, Q7, Q9, Q15, Q19, and Q20). According to the DASS-21 guidelines for evaluating anxiety, the scores on each anxiety question are added together to get the overall anxiety raw score. This number is then increased by 2. The adjusted number that is generated can be used to figure out how bad the anxiousness is. Based on the scaled score method, 8 is thought to be low anxiety, 14 is middling anxiety, and 22 is serious anxiety. A chart shows these ranges.

(c) Score Calculation of Depression

Table 4. Score Calculation of Depression

Q3	Q5	Q10	Q13	Q16	Q17	Q21	Score	Score × 2	Severity
0	1	0	1	0	0	0	2	4	Normal
2	1	3	1	0	3	3	13	26	Severe
0	0	1	0	1	1	0	6	12	Mild

Table 4 illustrates the DASS-21 depression score calculation based on responses to seven depression-related items (Q3, Q5, Q10, Q13, Q16, Q17, and Q21). The depression score is obtained by summing the individual item scores. In order to be aligned with the DASS scoring guidelines, depression score is then multiplied by two. The DASS scores serve to Apple the CSC depression scaled scores to the Divider depression categories. As illustrated a scaled score of four represents normal depression, a score of 12 represents mild depression and score of 26 represents severe depression. Scaled scores demonstrates a systematic and standardized way to assess the severity of depression.

3.2 Prediction Models

Using machine learning, we tried to set the DASS-21 scale answers for sadness, anxiety, and stress into categories. Picked algorithms include k-NN, SVM, RF, Naïve Bayes, MLP, and deep learning. This kind of mental health analytics uses these models and algorithms for many other kinds of classification jobs. They made the model in Python and used the Scikit-learn and TensorFlow tools for machine learning and deep learning, respectively. Twenty-one DASS-21 items were paired with one of the MLP architecture's fully linked hidden layers, and the output layer anticipated the emotional state categories. To get better alignment, the Adam optimiser was utilised. Traditional classifiers mostly had their basic hyperparameters set and then improved using grid search and cross-validation. It was decided that the k-NN model would use Euclidean distance and the SVM would use an RBF (radial basis function) kernel to describe and define the nonlinear decision border. Classifying big, complicated, high-dimensional datasets made up of many simple decision trees was done with the Random Forest method. Naïve Bayes was picked for this model because it is easy to understand, uses simple data processing, and makes decisions based on probabilities. For a full review of performance, the models were tested on a balanced dataset using performance metrics like recall, F1-score, accuracy, and AUC.

3.3 Demographic Characteristics of the Study

Table 5. Presents the demographic profile of the study participants, highlighting a balanced and diverse undergraduate sample. There is an even split in gender representation in the survey feedback, with 50% identifying as female, and 50% identifying as male, eliminating gender bias in the survey findings. Since all the respondents were in the undergraduate category, there is uniformity in the level of education. Regarding their present status, 36% of the respondents are full-time students. At the same time, a significant percentage is involved in part-time employment (34%) or running a business (30%); hence, the respondents have diverse academic and work-life experiences.

Table 5. Demographic Characteristics of the Study Population

Characteristics	**N (%)**
Gender	
Female	50 (50.0%)
Male	50 (50.0%)
Field of Study	
Undergraduate	100 (100.0%)
Current Role	
Full Time Student	36 (36.0%)
Part Time	34 (34.0%)
Own a Business	30 (30.0%)
Family Income	
0-20000	34 (34.0%)
20000-50000	26 (26.0%)
50000-100000	40 (40.0%)

Marital Status	
Single	92 (92.0%)
Married	8 (8.0%)

Family income levels show economic diversity, with the largest group (40%) falling in the ₹50,000–₹100,000 range, followed by lower-income brackets. Most respondents are single (92%), which is typical for an undergraduate population and relevant for interpreting psychosocial outcomes.

Table 6. The Demographic of Dass-21 Assessment Results

Severity Level	**Depression N (%)**	**Anxiety N (%)**	**Stress N (%)**
Normal	19 (19.0%)	6 (6.0%)	16 (16.0%)
Mild	25 (25.0%)	14 (14.0%)	19 (19.0%)
Moderate	26 (26.0%)	20 (20.0%)	30 (30.0%)
Severe	15 (15.0%)	23 (23.0%)	19 (19.0%)
Extremely Severe	15 (15.0%)	37 (37.0%)	16 (16.0%)

Table 6. presents the distribution of depression, anxiety, and stress severity levels based on the DASS-21 assessment among the study participants. Respondents for the case of depression were principally in the mild (25%) and the moderate (26%) categories which indicates the presence of some depressive symptoms, while 30% of respondents were in the category of depression which was considered to be severely depressed or to be at the extreme depression level. Concerning anxiety, the highest proportion of respondents was at the extreme anxiety level (37%) and subsequently at the level of severe anxiety (23%) which indicates anxiety to be the most common and the most severe mental health condition in the sample. In the case of stress, the predominant level was at the moderate level (30%) followed by mild (19%) and at the severe level (19%) also. These results pointed to the fact that the participants of the study were all experiencing severe levels of psychological distress which was prominent in the area of anxiety and hence mental health screening and interventions were to be prioritized to this area to address the problem.

3.4 Mathematical Model for DASS-21

The mathematical model suggested utilizes the DASS-21 Questionnaire to establish a mathematical framework for quantitatively and systematically estimating the severity of an individual's mental health condition. Through its deterministic mappings of questionnaire responses, the proposed model produces standardized scores for each sub-questionnaire type (Depression, Anxiety, and Stress) and maps these scores to five levels of severity. The proposed framework employs methods for handling class imbalance and for conducting supervised multi-class learning in order to increase the predictive accuracy of the model. The proposed framework will provide a means to combine scoring principles based upon statistical scoring with predictive modelling techniques based on machine learning to enable efficient, objective and scalable prediction of mental health severity.

Let each respondent's DASS-21 questionnaire response be represented as a feature vector:

$$\mathrm{x} = [x_1, x_2, \dots, x_{21}]^\top, \quad x_i \in \{0,1,2,3\} \tag{1}$$

Where x_i denotes the Likert-scale score of the i^{th} item.

The objective is to predict five severity classes for:

- Depression
- Anxiety
- Stress

$$y^{(k)} \in \{0,1,2,3,4\}, \quad k \in \{D, A, S\} \tag{2}$$

With:

- 0: Normal
- 1: Mild
- 2: Moderate
- 3: Severe
- 4: Extremely Severe

2. Subscale Score Computation (DASS-21 Scoring Model)

Stress Score

Stress items: {1,6,8,11,12,14,18}

$$r_S = x_1 + x_6 + x_8 + x_{11} + x_{12} + x_{14} + x_{18} \tag{3}$$

Scaled stress score:

$$s_S = 2 \times r_S$$

Anxiety Score

Anxiety items: {2,4,7,9,15,19,20}

$$r_A = x_2 + x_4 + x_7 + x_9 + x_{15} + x_{19} + x_{20} \tag{4}$$

Scaled anxiety score:

$$s_A = 2 \times r_A$$

Depression Score

Depression items: {3,5,10,13,16,17,21}

$$r_D = x_3 + x_5 + x_{10} + x_{13} + x_{16} + x_{17} + x_{21} \tag{5}$$

Scaled depression score:

$$s_D = 2 \times r_D$$

3. Severity Class Mapping

Each scaled score is mapped to a severity class using threshold boundaries:

$$y^{(k)} = \begin{cases} 0, & s_k < \tau_1 \\ 1, & \tau_1 \leq s_k < \tau_2 \\ 2, & \tau_2 \leq s_k < \tau_3 \\ 3, & \tau_3 \leq s_k < \tau_4 \\ 4, & s_k \geq \tau_4 \end{cases} \quad k \in \{D, A, S\} \tag{6}$$

Where τ_i are standard DASS-21 cut-off values

4. Dataset Formulation

$$\mathcal{D} = \{(\mathrm{x}_n, y_n^{(D)}, y_n^{(A)}, y_n^{(S)})\}_{n=1}^{N} \tag{7}$$

5. Class Imbalance Handling using SMOTE

For a minority class sample x_i and its nearest neighbor x_{nn}:

$$\mathrm{x}_{new} = \mathrm{x}_i + \lambda(\mathrm{x}_{nn} - \mathrm{x}_i), \quad \lambda \sim U(0,1) \tag{8}$$

Balanced dataset:

$$\widetilde{\mathcal{D}} = \mathrm{SMOTE}(\mathcal{D})$$

6. Multi-Class Classification Model

For each mental health dimension k:

$$f_k: \mathbb{R}^{21} \to \{0,1,2,3,4\} \tag{9}$$

Prediction:

$$\hat{y}^{(k)} = \underset{c}{\mathrm{argmax}} P(y = c \mid \mathrm{x}) \tag{10}$$

7. Deep Neural Network (MLP) Formulation

Hidden Layers

$$\mathrm{h}^{(l)} = \sigma(\mathrm{W}^{(l)}\mathrm{h}^{(l-1)} + \mathrm{b}^{(l)}) \tag{11}$$

Output Layer (Softmax)

$$P(y = c|\mathrm{x}) = \frac{e^{z_c}}{\sum_{j=0}^{4} e^{z_j}} \tag{12}$$

Loss Function (Cross-Entropy)

$$\mathcal{L} = -\frac{1}{N}\sum_{n=1}^{N}\sum_{c=0}^{4} \mathbb{I}\,(y_n = c)\log P(y_n = c) \tag{13}$$

8. Random Forest Ensemble Model

Given T trees:

$$\hat{y} = \mathrm{mode}\{h_1(\mathrm{x}), h_2(\mathrm{x}), \dots, h_T(\mathrm{x})\} \tag{14}$$

9. Performance Evaluation Metrics

- Accuracy:

$$\mathrm{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \tag{15}$$

- Precision:

$$\mathrm{Precision} = \frac{TP}{TP+FP} \tag{16}$$

- Recall:

$$\mathrm{Recall} = \frac{TP}{TP+FN} \tag{17}$$

- F1-score:

$$F1 = \frac{2\times\mathrm{Precision}\times\mathrm{Recall}}{\mathrm{Precision}+\mathrm{Recall}} \tag{18}$$

Integrated Mathematical Pipeline

$$\mathrm{x} \to (s_D, s_A, s_S) \to (y^{(D)}, y^{(A)}, y^{(S)}) \to \mathrm{SMOTE} \to \mathrm{ML/DL\ Classifier} \to (\hat{y}^{(D)}, \hat{y}^{(A)}, \hat{y}^{(S)}) \tag{19}$$

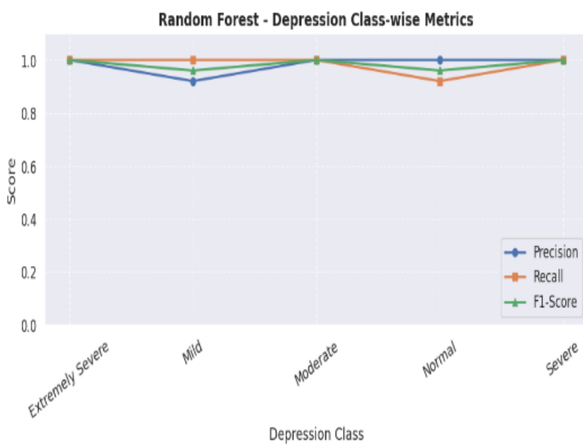
4. RESULTS AND DISCUSSION

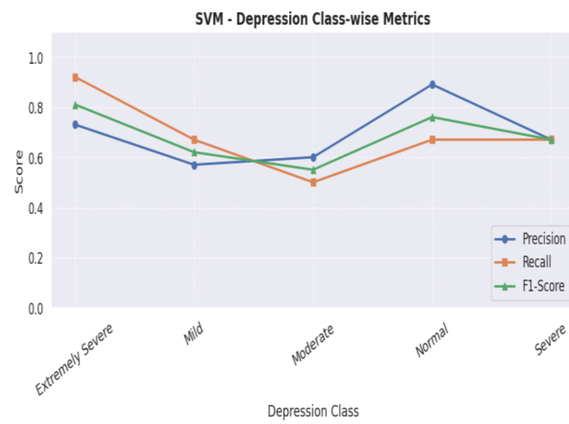
The depressed, anxious, and stressed (DASS-21) questionnaire was used to test how well five different machine learning algorithms could classify people's mental states. There are five levels of severity for each of the three emotional aspects. These are (1) Normal, (2) Mild, (3) Moderate, (4) Severe, and (5) Extremely Severe. We used Train-Test Split to see how well each algorithm worked on the dataset that had already been cleaned up. We assessed the model's performance based on the classification measures including F-1 Score, Accuracy, Precision, and Recall. From the evaluation each algorithm was able to predict the emotional severity of undergraduate students, but there were noticeable discrepancies among algorithms particularly with respect to an algorithm's sensitivity to over/under classification of high severity/low severity emotional states. Examples will be provided to indicate situations where one algorithm was more sensitive to the classification of students with extreme emotional severity than other algorithms while maintaining consistent accuracy across all emotional severity levels.

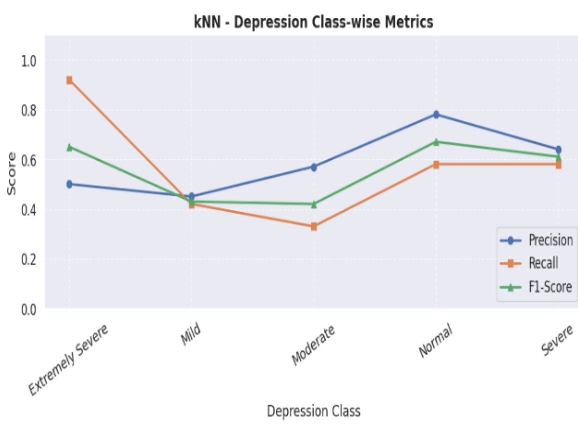
4.1 Category: Depression

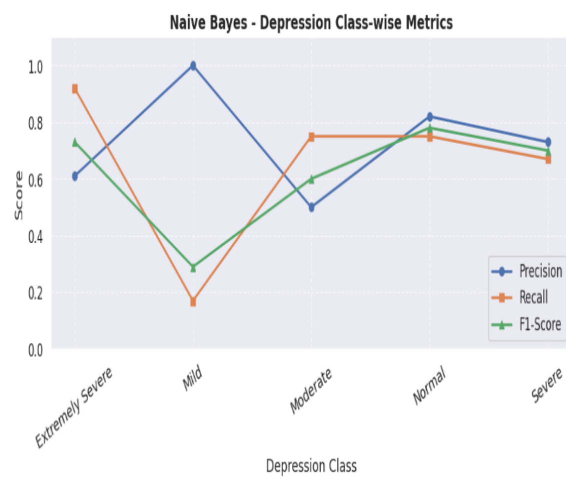
Random Forest, SVM, KNN, Naive Bayes, and Deep Learning all have different levels of accuracy and reliability when it comes to dividing the DASS-21 datasets into five levels of depression severity: Extremely Severe, Severe, Moderate, Mild, and Normal. Most of the time, Random Forest has gotten perfect marks, which means it is the most accurate model. In fact, it got a perfect score of 1.00 on all performance measures (accuracy, memory, and F1 score) for the Extremely Severe and Moderate ones. This means it can tell the difference between very severe and mild sadness, so it can be used for multi-class intensity classification. The score for accuracy drops to 0.92 in the Mild and Normal groups, but the score for memory stays at 1.00, which means that accurate identification is sure at these levels. Because Random Forest is reliable at all levels of seriousness, it is the best model for medical and clinical uses that need to fully automate the classification of depression severity. In comparison, the performance of the Support Vector Machine (SVM) is more variable with moderate levels of success, especially for the classification of Extremely Severe depression (precision = 0.73, recall = 0.92, F1 score = 0.81). However, performance in the Mild and Moderate levels is rather poor, with F1 scores falling to 0.62 for Mild and 0.55 for Moderate.

The findings suggest that the SVM has difficulties with the class imbalance that exists in the DASS-21 dataset and has problems in establishing decision boundaries for the lower levels of depression. The k-Nearest Neighbors (k-NN) algorithm faces the same problems, as evidenced by the results from the Mild and Moderate cases (F1- scores < 0.45) where its performance is distinctly poor. This shows that k-NN is very much affected by the particular structure of the local data and has poor generalization across the different classes.









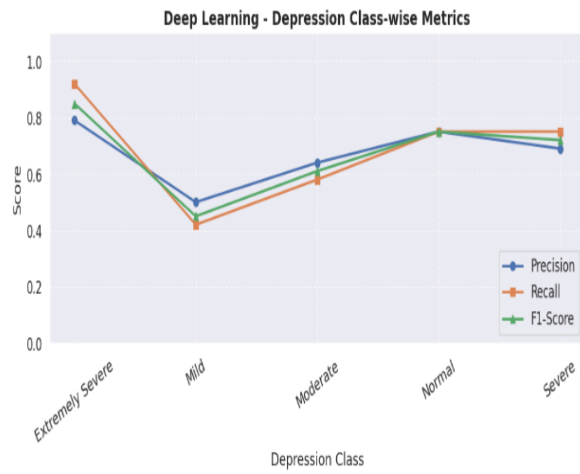


Fig. 2. Category: Depression

Naive Bayes reports perfect precision (1.00) for Mild depression but suffers from an extremely low recall (0.17). This leads to an F1 score of 0.29. This shows that Naive Bayes does not work well with class imbalance, combine that with its feature independence assumption which is not valid for interrelated psychometric features like the DASS-21, Naive Bayes has a poor performance. None the less, the Deep Learning model achieves moderate performance with F1 scores between 0.45 to 0.85 and is balanced across categories, with the highest performance for Extremely Severe depression. Deep Learning seems to be able to find and explain more complex factors that affect how bad sadness is, since it works well in the Mild and Moderate categories but not in the less complex ones. A study comparing different models for detecting depression severity found that the Random Forest model worked the best, beating out all the others, even traditional machine learning classifiers and Deep Learning methods. It shows that Random Forest is reliable and can be used for multi-class classification projects in clinical medicine because it can correctly describe most of the intensity categories, especially the Extremely Severe and Moderate categories. Additionally, the performance limitations encountered by SVM, k-NN, and Naïve Bayes attributable to imbalances in class distribution and their sensitivity to the distribution of the data (low dramatic variance) lead to a large amount of low accuracy for the Mild and Moderate classes. The ability of Deep Learning to become a robust method for detecting Depression may be limited due to obstacles in attaining Precision (Sensitivity) for the lower (Mild/Moderate) levels of Depression, requiring increased improvement and refinement. The presented data indicate that ensemble approaches, specifically Random Forest ensembles, demonstrate superior performance and are clinically meaningful for predicting Depression Severity using DASS-21 data, and can also be employed for Automating and Scaling Mental Health Assessment.

4.2 Category: Anxiety

The findings from the anxiety classification results reveal that a performance hierarchy exists between the many different machine and deep learning algorithms. An F1 Score of 0.99, a Recall Score of 0.99, and a Precision Score of 0.99 showed that the Random Forest model was the most accurate at separating people into the five types of anxiety: Normal, Mild, Moderate, Severe, and Extremely Severe. There were no wrong classifications, so the Random Forest model was able to correctly identify all cases of worry. There is a model called Random Forest that can correctly understand and learn from the DASS-21 anxiety data's complicated and nonlinear connections. Using an ensemble learning approach like Random Forest is a good way to find the minor complexities that come with the different levels of seriousness of anxiety data, as shown by the model's good results. This ability will greatly benefit the field of clinical screening where there is a strong need for accurately assigning severity levels to anxiety. While Support Vector Machine (SVM) produced very good results as well, with F1 scores of (0.90, 1.00), it did fall short of producing high Recall Scores in the Extremely Severe and Severe levels of anxiety. These results indicate that SVM is quite robust and versatile, although it is slightly less sensitive with the most severe cases of anxiety as compared with the ensemble models, Random Forest. All findings affirm that when class imbalance is sufficiently tackled, ensemble-based methods are ideal for capturing severity variations in anxiety data. K-Nearest Neighbors (k-NN) and Naïve Bayes, however, as instance-based and probabilistic models, respectively, show rather moderate performance, especially for the Extremely Severe and Moderate anxiety categories.

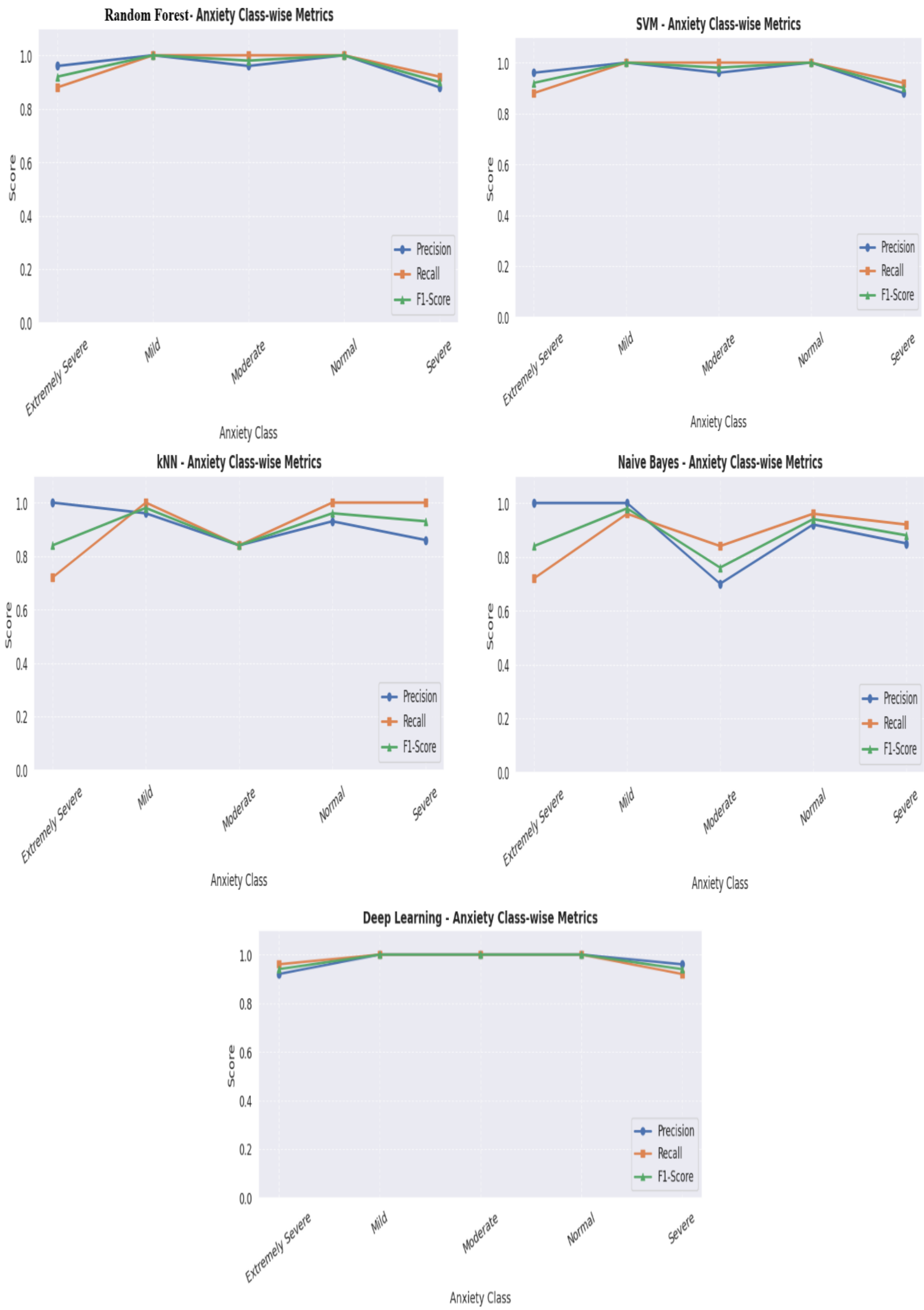
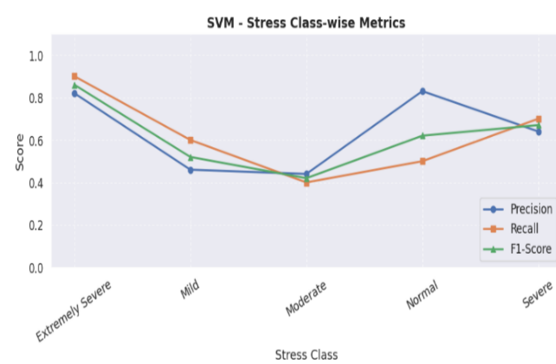


Fig. 3. Category: Anxiety

Models such as Naïve Bayes and k-NN demonstrate a substantial level of steady precision. However, their recalls, which can sink as low as 0.72, suggest that they are potentially missing a few critical situations of a higher degree of severity. In clinical situations where most are needed because they can present a serious risk, this becomes a bottleneck of critical importance. The performance of the Deep Learning model from its F1-score is nearly perfect and is from 0.94 to 1.00, meaning that it also had the best ability to deal with the intricacies of various complex interactions that responses to the DASS-21. The slight difference in performance of this model compared to some of the ensemble models indicates that for this type of relatively low dimensional structure in the questionnaire data, ensemble models of some learning type may be used with a lower level of reliability and lower level of computational resources. Even with all its power, the findings show that these types of models can yield more reliable results, particularly for multi-class classification in mental health assessments that are more concerned about the randomness. In this regard, the results most designate the Random Forest model as the caring and reliable model in regard to the classification of the severity of anxiety, particularly with the classification of the Extremely Severe and Severe categories. The performance of SVM, with respect to these cases, is better than that of the Random Forest, although this is also the case with respect to the other high-severity cases. K-NN and Naïve Bayes are meanwhile also limited in this regard, although the Deep Learning model noted still demonstrates more high-severity cases. In general, Random Forest and comparable methods in ensemble learning are more pragmatic, with greater precision and less in terms of computation, providing scalable, clinically pertinent evaluations.

4.3 Category: Stress

This study looked at different machine learning and deep learning models and found that the results for classifying stress show how well (or poorly) they worked. The Random Forest model, which was the best model overall, had the most predictive power for detecting different levels of stress. With F1 scores of 1.00 for Extremely Severe, Moderate, and Severe stress, and F1 = 0.95 for Mild and Normal stress, it managed to make almost perfect forecasts. After SMOTE-based (Synthetic Minority Over-sampling Technique) balancing was used to fix the class mismatch in the DASS-21 dataset, this result shows whether the model can tell the difference between different levels of stress. It's clear that Random Forest does a great job with multi-class psychological data because it got perfect scores in accuracy, memory, and F1 (0.99) for all five stress intensity groups. Finding and recording the non-linear links between the data traits is also easy for the model. Random Forest is the most reliable model for figuring out how bad stress is because of all of these reasons. KNN (k nearest neighbour) and SVM (support vector machine) don't work very well with the Mild and Moderate stress levels. Moderate stress got an F1 score of 0.42 from SVM, and Severity stress got an F1 score of 0.15 from kNN. Support vector machines and k nearest neighbour models have trouble with this kind of data because they use distance and margin to make sense of it. This is especially true when the symptoms are the same across different amounts of stress. These classifiers struggle to draw the line between multiple classes in mental health data and in this case, the complexity lies in the discrimination of stress levels which is iterative between moderate and mild.



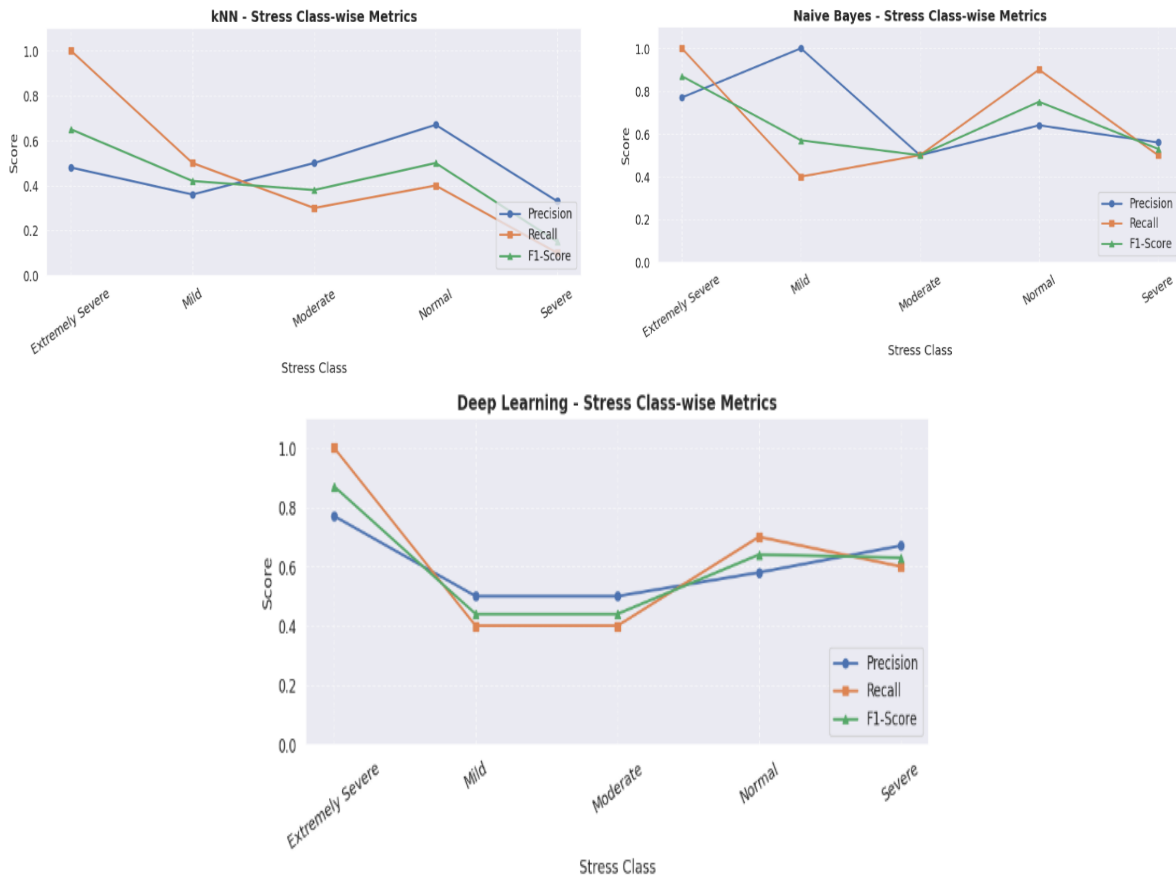


Fig. 4. Category: Stress

The Naive Bayes and Deep Learning Models show adequate sensitivity for higher severity stress levels, however, their performance across all severity categories is not very consistent. Naive Bayes shows adequate performance on the Extremely Severe stress class (F1 = 0.87) and Normal stress (F1 = 0.75), but iit has lower performance on Mild stress, with a recall of 0.40. This shows that Naive Bayes does a poor job on classifying lower severity cases. Deep learning models also show a similar pattern.

Table 7. Model Performance Comparison for Severity Classification of Mental Health

Model	Accuracy (S1)	Accuracy (S2)	Accuracy (S3)	F1-score (S1)	F1-score (S2)	F1-score (S3)
Random Forest	0.983	0.990	0.980	0.983	0.990	0.980
SVM	0.683	0.960	0.620	0.681	0.960	0.618
k-NN	0.567	0.911	0.460	0.556	0.909	0.418
Naive Bayes	0.650	0.879	0.660	0.619	0.881	0.643
Deep Learning	0.683	0.976	0.620	0.676	0.976	0.605

S1 = Severity Level 1 (Depression), S2 = Severity Level 2 (Anxiety), S3 = Severity Level 3 (Stress)

Deep learning models show high recall for Extremely Severe stress (1.00), but they also show a poor performance on the Mild and Moderate categories with both having an F1 = 0.44 and even having a recall of 0.00. This shows that the deep learning models also struggle with the dataset at hand. These results show that on the stress severity classification task, ensemble learning models are the best, particularly random forests. These models are the best on the stress classification task. These models are able to provide high and consistent accuracy across all stress levels which is important in their use in fully automated stress screening systems. Identifying moderate to extreme stress levels is important to allow for early intervention with at-risk individuals and to provide better mental health decision support in the populations of interest (which are the academic and clinical). Fig.5 shows how well five different models Random Forest, SVM, k-NN, Naive Bayes, and Deep Learning classified how bad depression, anxiety, and stress incidents were.

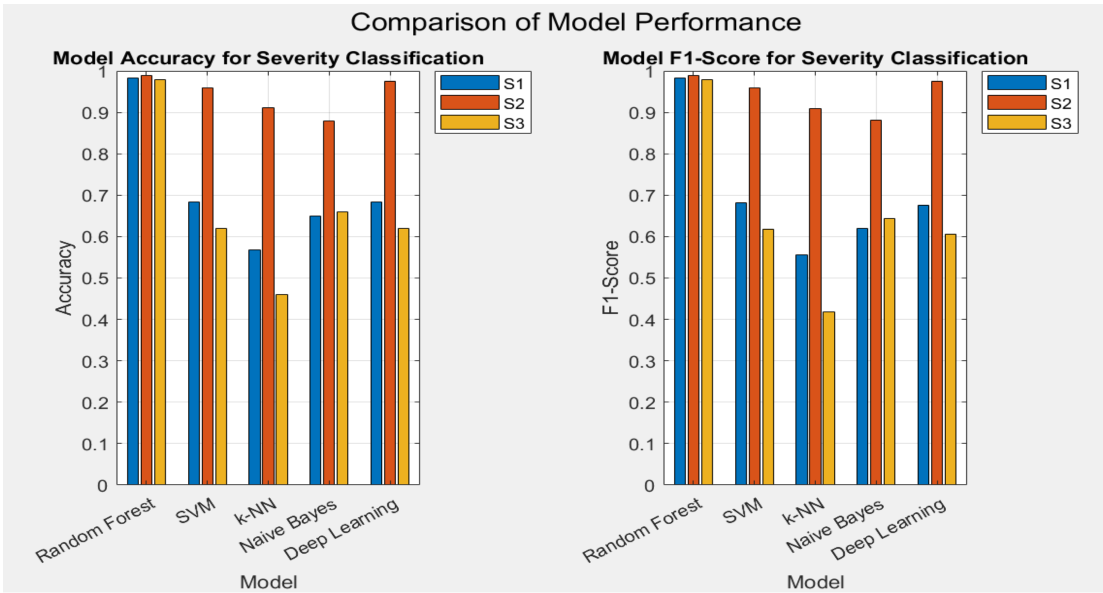


Fig. 5. Model Performance Comparison for Severity Classification of Mental Health

Random Forest regularly does better than the others, reporting the best accuracy and F1-scores for all intensity levels, including S1 (depression), S2 (anxiety), and S3 (stress). As for the model's F1 scores, they are 0.983 for S1, 0.990 for S2, and 0.980 for S3. This means that the model is accurate in all three cases. Our results show that Random Forests is the most accurate machine learning algorithm we tried. It can accurately classify types better than SVM and k-NN, which had the lowest accuracy scores for S1 (SVM=0.683 and k-NN=0.567). Therefore, the study shows that the Random Forest model is the most accurate and reliable one for using machine learning to classify how bad mental health problems are.

4.4 Correlation Heatmaps for Mental Health Severity

The Random Forest model sparked development of the correlation heatmap to assist in unearthing the DASS-21 questionnaire items and their corresponding relationships to depression and the severity of the breadth of associated mental health concerns: anxiety and stress. Examining the depression heatmap, notable positive correlations are a result of the depression label and the above mentioned key DASS-21 responses, which include: q10 (d) (r = 0.70), q21 (d) (r = 0.67), and q17 (d) (r = 0.66) -- all of which serve as depression severity predictors. Further, the Multiply_by_2 variable, which captures the score which is weighted according to the DASS-21, is consistently a strong predictor of severity (r = 0.70). These, combined with the correlations to stress items q16 (d) (r = 0.45) and q13 (d) (r = 0.57), show considerable overlap of depression and stress. Overall, this supports the need for overlap in model components for identifying the ascendant threshold level of concern for depression and the broader concerned mental health spectrum associated with the phenomenon.

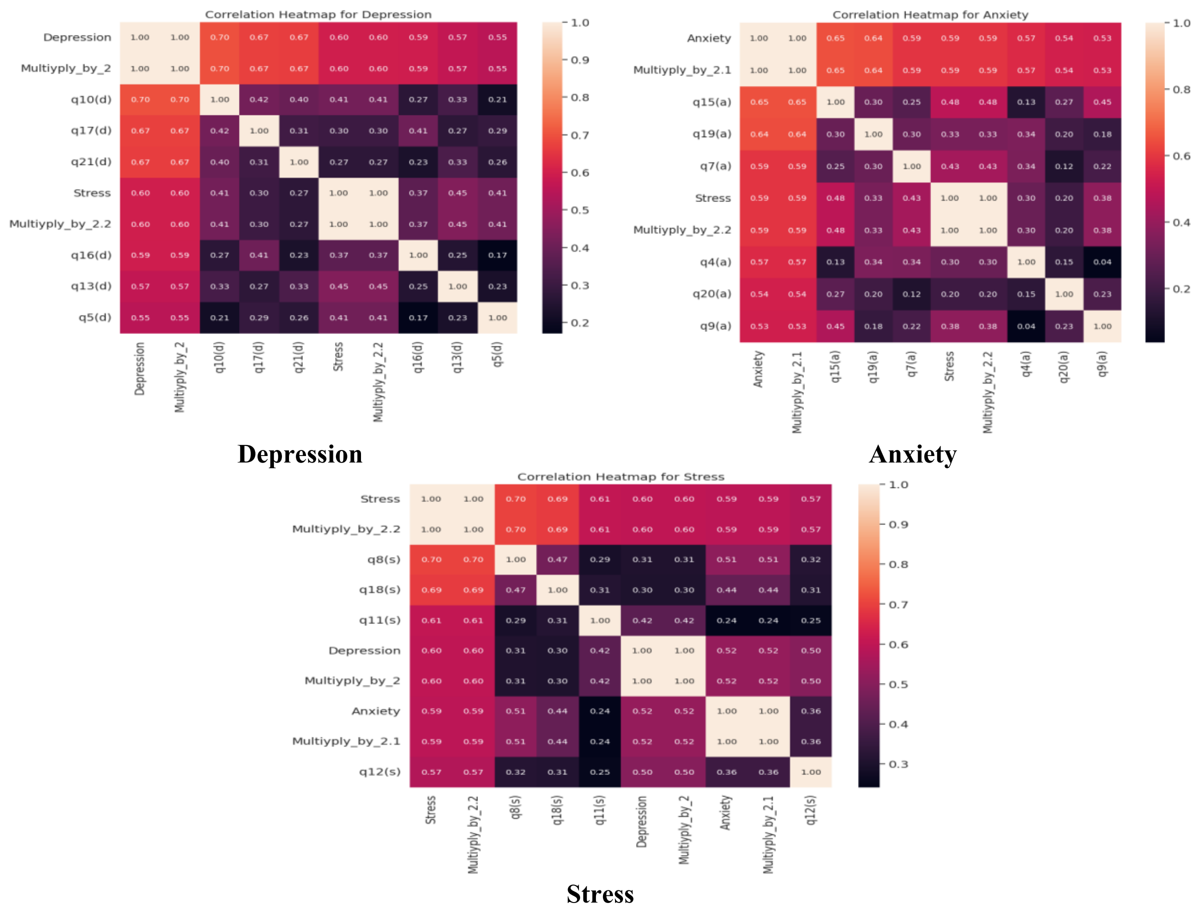


Fig. 6. Correlation Heatmaps for Mental Health Severity Predictions (Random Forest)

In the case of anxiety, the heatmap refers to a perfect correlation with the Multiply_by_2_1 variable (r = 1.00) meaning that the raw scores are possibly scaled correctly. Correlations with q15(a) (r = 0.65), q9(a) (r = 0.63), and q7(a) (r = 0.59) are a good enough reason to give these features priority in focusing on the detection of the level of anxiety. Furthermore, anxiety is moderately correlated with stress (r = 0.59), which is a good enough reason to explain the condition's comorbidity. This most certainly confirms the importance of anxiety and stress's interrelation when a mental health model is being built, and it also adds to the overall interpretability of the model. The different degrees of correlation across features implies that the model is sophisticated enough to pinpoint particular questionnaire responses that are most contributory to the prediction of the severity of anxiety. The heatmap for stress also correlates exactly with the Multiply_by_2_2 variable (r = 1.00), which confirms that the stress scores are also scaled correctly. q8(s) (r = 0.70) and q18(s) (r = 0.69) are the most important correlates of stress and are therefore critical to stress classification. Other moderate correlations with q11(s) (r = 0.62) and q12(s) (r = 0.57) also add to the predictability. Stress displays a moderate correlation with both depression and anxiety (r = 0.64 and r = 0.59, respectively), pointing to the established comorbidity of these conditions. This affirms the model’s capability to identify and categorize severity of stress with concomitant psychosomatic manifestations.

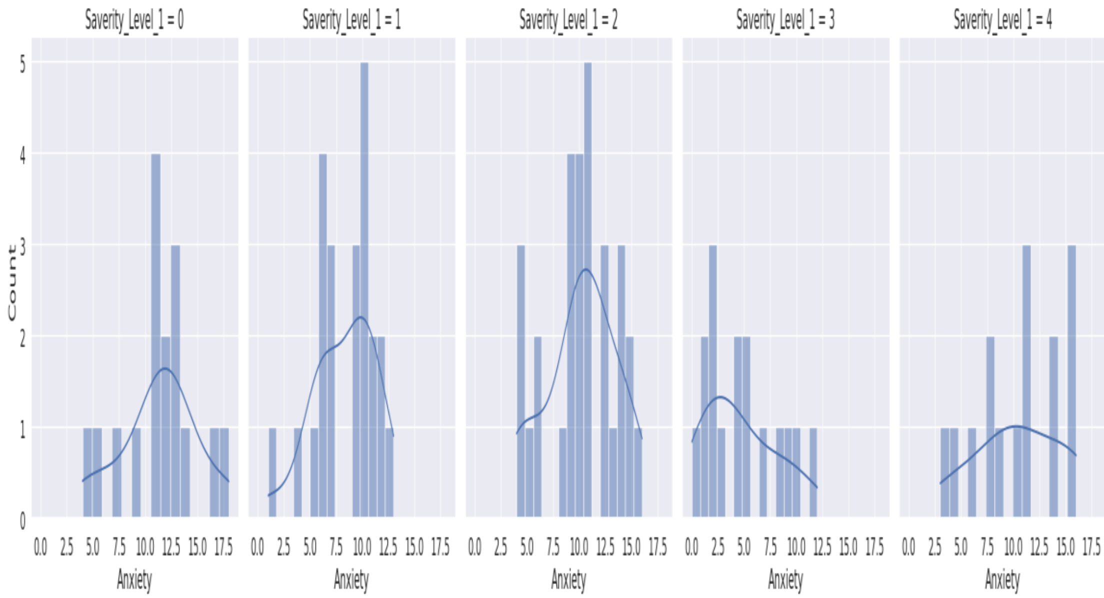
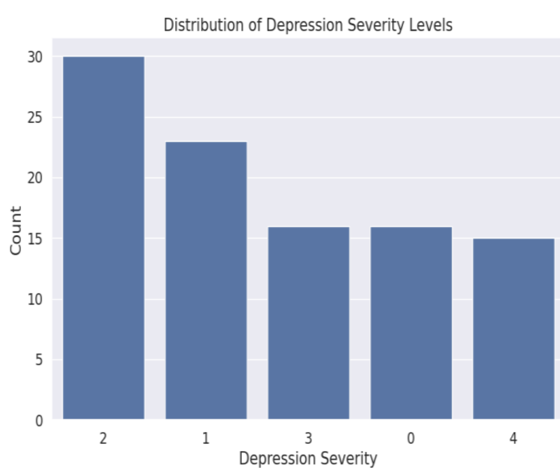


Fig. 7. Distribution of Anxiety Scores across Severity Levels

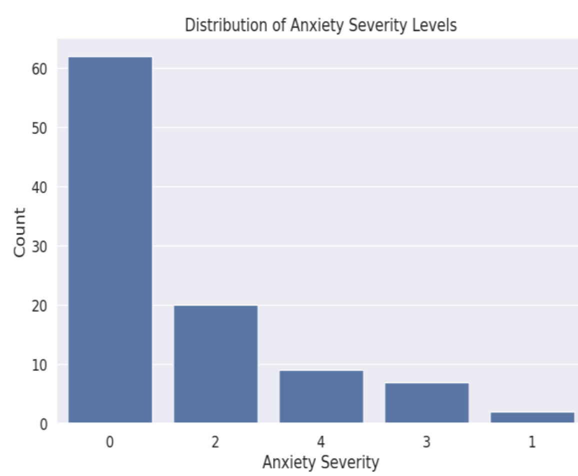
The multi-panel histogram with KDE plots presented in the fig.7 above visually represents the distribution of anxiety scores across five severity levels: Normal (0), Mild (1), Moderate (2), Severe (3), and Extremely Severe (4). For Severity_Level_1 = 0, anxiety scores predominantly fall between 10 and 15, indicating moderate anxiety even in normal cases. For Severity_Level_1 = 1 and Severity_Level_1 = 2, the bell-shaped curves suggest that moderate anxiety scores are common. However, for Severity_Level_1 = 3, the distribution skews towards lower scores, possibly indicating inconsistencies in self-reporting. In Severity_Level_1 = 4, scores peak sharply between 12 and 17, reflecting higher anxiety in extremely severe cases. This visualization aids in classifying and understanding anxiety patterns, particularly distinguishing between moderate and extreme cases, enhancing classification accuracy in mental health assessments.

4.5 Distribution of Mental Health Severity Levels

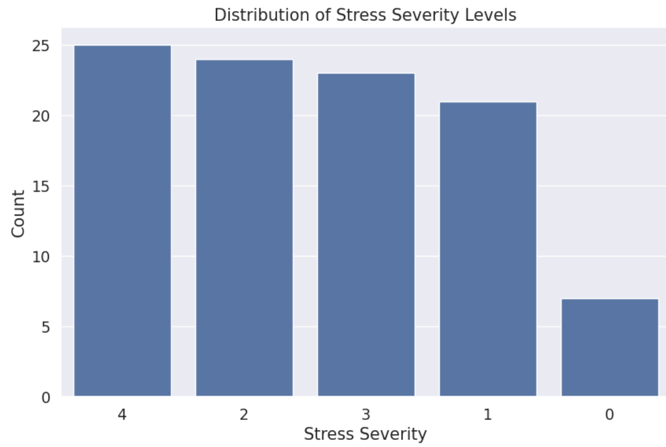
Figure 8 shows how the Mental Health Severity Levels are spread out across sadness, anxiety, and stress. Moderate severity (Severity = 2) is the most common number for depression, with 29 people. Mild severity (Severity = 1) is next, with 23 participants. 15 people are in the Severe (Severity = 3) and Normal (Severity = 0) categories, while only fourteen people are in the Extremely Severe (Severity = 4) group. We need to use SMOTE (Synthetic Minority Over-sampling Technique) to even out the dataset because the lack of serious cases could make it harder for the model to predict critical conditions.



Depression



Anxiety



Stress

Fig. 8. Distribution of Mental Health Severity Levels

For Anxiety, Normal severity (Severity = 0) dominates with more than 60 participants, reflecting a high proportion of individuals without significant anxiety symptoms. The Mild severity group (Severity = 1) has 25 participants, and Moderate (Severity = 2), Severe (Severity = 3), and Extremely Severe (Severity = 4) categories show progressively fewer participants 20, 12, and 10, respectively. The dominance of Normal severity points to the necessity of balancing the dataset, particularly for the underrepresented Severe and Extremely Severe categories, which are crucial for accurate anxiety detection in machine learning models. The distribution for Stress reveals a relatively even spread across Moderate (Severity = 2), Severe (Severity = 3), and Extremely Severe (Severity = 4) levels, each containing approximately 20 to 25 participants. However, the Normal (Severity = 0) and Mild (Severity = 1) categories are underrepresented with 5 and 7 participants, respectively. This distribution emphasizes the presence of moderate to extreme stress levels in the sample, highlighting the need for targeted interventions. Additionally, it underscores the importance of class balancing during model training to ensure fair predictions across all severity levels, especially for Moderate and Severe stress cases.

4.6 Correlation Matrix of Mental Health Severity Levels

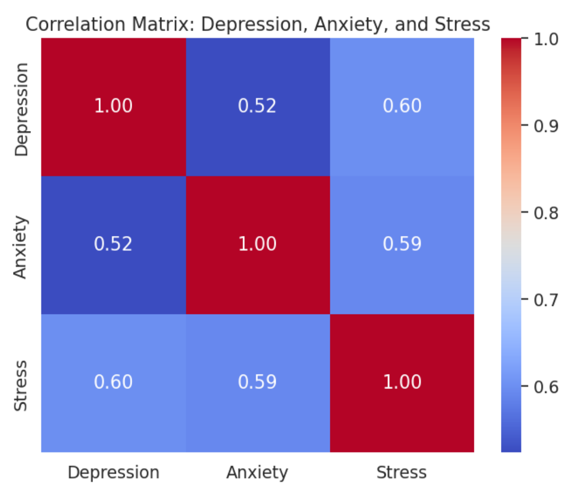


Fig. 9. Correlation Matrix of Mental Health Severity Levels

The DASS-21 answers show how Depression, Anxiety, and Stress are connected in the association matrix (Fig. 9). While 1.00 means there is perfect connection, the other numbers show how strongly these mental health factors are linked. It's possible that people who are depressed are also more likely to have anxiety symptoms, since the link between the two is 0.52. Those who are depressed tend to be more stressed as well, as shown by the slightly stronger link between the two at 0.60. The last thing to note is that anxiety and stress are moderately related. People who have higher levels of anxiety also report high levels of stress. It's important to look at more than one aspect of mental health at the same time in professional exams and treatments, as these results show that these conditions often happen together. The observed correlations emphasize the need for integrated models that account for overlapping symptoms across mental health disorders, which can improve early detection and treatment strategies.

5. Conclusion

The studies are mostly about sorting and understanding the different machine-learning and deep-learning models, as well as how well they do with the DASS-21 assessment for anxiety, depression, and stress. It's hard to do precise class split with mental health information. This is when the Synthetic Minority Over Sampling Technique (SMOTE) is used, and it has been shown that this method makes the model more sensitive. Random Forest (RF) is the winning model, with 98% accuracy over all mental health disorders. The RF ensemble model is said to have a positive performance when trying to detect severe and moderate cases, and is particularly positive when it comes to imbalanced data. In contrast, other models which could be termed traditional, performed at a moderate level, SVM, k-NN, and Naive Bayes models (classifying a set of data) ranges from 56% to 68% (56% to 68%, respectively, in the case of Naive Bayes). Here, the models did not seem to be positive in predicting the minority class. The Deep Learning (DL) model, on the other hand, with an accuracy of between 68 and 97% performed well and is a positive characteristic with respect to the complex model with multi class level severity. The findings of the study confirm the above and prove that RF in particular is the best model to be employed especially in the case of ensemble methods, and Random Forest (RF) particularly stands out from the crowd. The automated mental health severity models and other systems and models are most appropriately applied in a clinical and academic setup. They are models and systems that focus on the early detection of identified high-risk cases from the system, and therefore in a clinical setup, they allow for the necessary and required prompt measures to be taken.

CRediT Authorship Contribution Statement

Neha Rajas: Conceptualization, Methodology, Data curation, Software, Formal analysis, Investigation, Validation, Visualization, Writing – original draft, and Writing – review & editing.

Dr. Balaji Patil: Supervision, Conceptualization, Methodology guidance, Validation, Project administration, Writing – review & editing, and final manuscript approval.

Both authors have read and approved the final version of the manuscript titled **"Severity Detection of Anxiety, Depression, and Stress Using Ensemble Learning and Deep Neural Networks: A Multi-Class Approach."** The study involves DASS-21 based mental health severity classification using machine learning and deep learning models, including SMOTE-based class balancing and model comparison using accuracy, precision, recall, and F1-score.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper titled "Severity Detection of Anxiety, Depression, and Stress Using Ensemble Learning and Deep Neural Networks: A Multi-Class Approach."

References

1. P. Kumar, S. Garg, and A. Garg, "Assessment of Anxiety, Depression and Stress using Machine Learning Models," Procedia Computer Science, vol. 171, pp. 1989–1998, 2020, doi: 10.1016/j.procs.2020.04.213.
2. B. W. Penninx, D. S. Pine, E. A. Holmes, and A. Reif, "Anxiety disorders," The Lancet, vol. 397, no. 10277, pp. 914–927, Mar. 2021, doi: 10.1016/S0140-6736(21)00359-7.
3. I. Marijanović et al., "Use of the Depression, Anxiety and Stress Scale (DASS-21) Questionnaire to Assess Levels of Depression, Anxiety, and Stress in Healthcare and Administrative Staff in 5 Oncology Institutions in Bosnia and Herzegovina During the 2020 COVID-19 Pandemic," Med Sci Monit, vol. 27, Feb. 2021, doi: 10.12659/MSM.930812.
4. A. Chakraborty, A. Jha, A. Banerjee, N. Banerjee, D. Saha, and A. Das, "Depression, Anxiety, and Stress Levels among Medical Students and Interns in a Medical College of Kolkata: A Cross-Sectional Study," JCH, vol. 12, pp. 89–93, Jul. 2024, doi: 10.25259/JCH_18_2024.
5. A. Rahman et al., "Machine learning and deep learning-based approach in smart healthcare: Recent advances, applications, challenges and opportunities," AIMSPH, vol. 11, no. 1, pp. 58–109, 2024, doi: 10.3934/publichealth.2024004.
6. G. W. Lindsay and T. Serre, "Deep Learning Networks and Visual Perception," in Oxford Research Encyclopedia of Psychology, Oxford University Press, 2021. doi: 10.1093/acrefore/9780190236557.013.841.
7. V. Souza, J. Nobre, and K. Becker, "A Deep Learning Ensemble to Classify Anxiety, Depression, and their Comorbidity from Texts of Social Networks," jidm, vol. 12, no. 3, Oct. 2021, doi: 10.5753/jidm.2021.1901.

8. R. Li and Z. Liu, “Stress detection using deep neural networks,” BMC Med Inform Decis Mak, vol. 20, no. S11, p. 285, Dec. 2020, doi: 10.1186/s12911-020-01299-4.
9. T. Nijhawan, G. Attigeri, and T. Ananthakrishna, “Stress detection using natural language processing and machine learning over social interactions,” J Big Data, vol. 9, no. 1, p. 33, Dec. 2022, doi: 10.1186/s40537-022-00575-6.
10. G. Husain et al., “SMOTE vs. SMOTEENN: A Study on the Performance of Resampling Algorithms for Addressing Class Imbalance in Regression Models,” Algorithms, vol. 18, no. 1, p. 37, Jan. 2025, doi: 10.3390/a18010037.
11. S. Graham et al., “Artificial Intelligence for Mental Health and Mental Illnesses: an Overview,” Curr Psychiatry Rep, vol. 21, no. 11, p. 116, Nov. 2019, doi: 10.1007/s11920-019-1094-0.
12. H. Hairani, T. Widiyaningtyas, and D. Dwi Prasetya, “Addressing Class Imbalance of Health Data: A Systematic Literature Review on Modified Synthetic Minority Oversampling Technique (SMOTE) Strategies,” JOIV : Int. J. Inform. Visualization, vol. 8, no. 3, p. 1310, Sep. 2024, doi: 10.62527/joiv.8.3.2283.
13. L. Barker et al., “Community-based mental health interventions for reducing youth suicidal thoughts and behaviours: a systematic review and meta-analysis,” BMC Psychiatry, vol. 25, no. 1, p. 1041, Oct. 2025, doi: 10.1186/s12888-025-07527-3.
14. S. Rashid, T. Badrul, L. Rafid, J. H. Setu, N. Halder, and A. Islam, “Ensemble and Deep Learning Approaches for Automated Screening of Anxiety, Depression, and Burnout in Medical Student Populations,” in 2024 Advances in Science and Engineering Technology International Conferences (ASET), Abu Dhabi, United Arab Emirates: IEEE, Jun. 2024, pp. 01–10. doi: 10.1109/ASET60340.2024.10708680.
15. V. Borba De Souza, J. Campos Nobre, and K. Becker, “DAC Stacking: A Deep Learning Ensemble to Classify Anxiety, Depression, and Their Comorbidity From Reddit Texts,” IEEE J. Biomed. Health Inform., vol. 26, no. 7, pp. 3303–3311, Jul. 2022, doi: 10.1109/JBHI.2022.3151589.
16. A. S. P. J and S. K. K, “Early Detection of Anxiety, Depression and Stress among Potential Patients using machine learning and deep learning models,” in 2023 2nd International Conference on Computational Systems and Communication (ICCSC), Thiruvananthapuram, India: IEEE, Mar. 2023, pp. 1–7. doi: 10.1109/ICCSC56913.2023.10143026.
17. M. S. Rahman, A. Ashik, and Md. S. Rahman, “Comprehensive Analysis of Stress Levels Using Ensemble Learning and Neural Networks for Insightful Understanding,” in 2024 6th International Conference on Electrical Engineering and Information & Communication Technology (ICEEICT), Dhaka, Bangladesh: IEEE, May 2024, pp. 1344–1349. doi: 10.1109/ICEEICT62016.2024.10534534.
18. S. Liu et al., “An objective quantitative diagnosis of depression using a local-to-global multimodal fusion graph neural network,” Patterns, vol. 5, no. 12, p. 101081, Dec. 2024, doi: 10.1016/j.patter.2024.101081.
19. R. Li and Z. Liu, “Stress detection using deep neural networks,” BMC Med Inform Decis Mak, vol. 20, no. S11, p. 285, Dec. 2020, doi: 10.1186/s12911-020-01299-4.
20. S. Zhang, M. Tu, Y. Yan, Y. Zhuang, L. Ge, and G. Wei, “Detecting Depression, Anxiety and Mental Stress in One Sequential Model with Multi-task Learning,” in HCI International 2022 – Late Breaking Papers: HCI for Health, Well-being, Universal Access and Healthy Aging, vol. 13521, V. G. Duffy, Q. Gao, J. Zhou, M. Antona, and C. Stephanidis, Eds., in Lecture Notes in Computer Science, vol. 13521. , Cham: Springer Nature Switzerland, 2022, pp. 191–200. doi: 10.1007/978-3-031-17902-0_14.
21. R. V. Anand, A. Q. Md, S. Urooj, S. Mohan, M. A. Alawad, and A. C., “Enhancing Diagnostic Decision-Making: Ensemble Learning Techniques for Reliable Stress Level Classification,” Diagnostics, vol. 13, no. 22, p. 3455, Nov. 2023, doi: 10.3390/diagnostics13223455.
22. B. Saylam and Ö. D. İncel, “Multitask Learning for Mental Health: Depression, Anxiety, Stress (DAS) Using Wearables,” Diagnostics, vol. 14, no. 5, p. 501, Feb. 2024, doi: 10.3390/diagnostics14050501.
23. V. Singh, S. K. Sahana, and V. Bhattacharjee, “A novel CNN-GRU-LSTM based deep learning model for accurate traffic prediction,” Discov Computing, vol. 28, no. 1, p. 38, Apr. 2025, doi: 10.1007/s10791-025-09526-0.