

## Accurate Classification of Human Mental States Using a Stacked Ensemble Machine Learning Framework

Yifan Zhao<sup>1\*</sup>, Liang Chen<sup>1</sup>, Ahmad Fauzi<sup>2</sup>, Nor Hayati<sup>3</sup>

<sup>1</sup> Department of Psychiatry and Brain-State Classification, Faculty of Medicine, China Agricultural University, Beijing, China.

<sup>2</sup> Department of Ensemble Learning and Mental State Prediction, Faculty of Medicine, University of Malaya, Kuala Lumpur, Malaysia.

<sup>3</sup> Department of Mental Health Informatics, Faculty of Medicine, Universiti Putra Malaysia, Serdang, Malaysia.

\*E-mail ✉ yifan.zhao@cau.edu.cn

### Abstract

Psychological disorders such as stress, depression, and anxiety remain highly prevalent and exert considerable negative effects on personal well-being. Effective recognition and analysis of these conditions are vital for minimizing their detrimental consequences. This investigation applies machine learning methodologies to advance mental health evaluation. The developed “Stacking-Classifer-Ensemble-Learning” methodology aggregates multiple weak learners to strengthen overall predictive capacity. Base learners comprised neural networks, decision trees, random forests, Naive Bayes, support vector machines (SVM), K-nearest neighbors (KNN), linear discriminant analysis, and AdaBoost, with SVM functioning as the meta-learner. Class imbalance prevalent in mental health data was mitigated through the Synthetic Minority Over-sampling Technique (SMOTE), while model robustness was confirmed through cross-validation. Given the difficulties of obtaining high-quality mental health data, this work relied on a publicly accessible dataset from [www.kaggle.com](http://www.kaggle.com), which contained known imbalances and incomplete entries. Empirical results showed that ensemble stacking classifiers reliably outperformed individual models, reaching an accuracy approaching 98%. Ultimately, this research seeks to refine mental health assessment practices and treatment approaches by harnessing advanced machine learning techniques to deepen understanding of these disorders.

**Keywords:** Machine learning, Ensemble learning, Mental health, Stress, Anxiety, Depression

### Introduction

In today’s rapidly evolving world, individuals from all walks of life—such as students, parents, workers, and organizational leaders—face significant pressure to remain competitive. Those immersed in demanding lifestyles often feel greater pressure to excel, requiring them to manage obstacles, setbacks, and expectations. Within this high-pressure environment, anxiety disorders have increasingly compromised mental health. Stress is defined as any stimulus provoking physiological,

emotional, or psychological tension. It constitutes the body’s intrinsic response to demands that require adaptive behavior. Stress affects people universally, though to varying degrees. Depression typically involves feelings of irritation, anger, and reduced engagement in pleasurable activities. It frequently results in multiple psychological and somatic complications, along with diminished capacity for effective daily and personal functioning. Nowadays, stress ranks among the most widespread psychological issues. Although often overlooked in societal discourse, this condition continues to influence large segments of the global population during the twenty-first century. Stress is frequently associated with stimulus-rich settings in contemporary life; nevertheless, it also profoundly affects structured environments such as educational institutions [1].

The human body evaluates stress by activating behavioral and physiological mechanisms in response to

Access this article online

<https://smerpub.com/>

Received: 17 February 2026; Accepted: 24 May 2026

Copyright CC BY-NC-SA 4.0

**How to cite this article:** Zhao Y, Chen L, Fauzi A, Hayati N. Accurate Classification of Human Mental States Using a Stacked Ensemble Machine Learning Framework. *J Med Sci Interdiscip Res.* 2026;6(1):200-18. <https://doi.org/10.51847/uEXVG57Ru>

emotions such as joy, fear, aversion, astonishment, repulsion, and other affective experiences [2]. Anxiety represents a psychological state capable of generating sensations of dread, apprehension, and unease. It can trigger agitation and increased heart rate, sometimes serving as an adaptive response to stressors. An anxiety disorder involves prolonged and intensifying worry that disrupts key areas of life, encompassing workplace efficiency, scholastic achievement, and social connections. Affected individuals commonly strive to circumvent anxiety-provoking stimuli. Insufficient availability of suitable mental health resources may aggravate these difficulties, especially within younger demographics [3]. Social embodiment anxiety constitutes a common form of social psychological disorder that exhibits higher prevalence among women than men [4]. Individuals with panic disorder (PD) may encounter abrupt episodes of trembling, perspiration, or disorientation, occasionally accompanied by breathing difficulties [5]. Agoraphobia develops when people harbor fears regarding entrapment in situations where escape could become problematic [6]. Scholarly sources present a variety of assessment instruments for evaluating distinct anxiety disorders and for informing related clinical treatments.

Social networking sites have emerged as effective instruments for health monitoring and disease forecasting. For example, Ref. [7] derived subject-specific and personality-related features from Live Journal publications and integrated them into a joint modeling system to detect mental health concerns, particularly depression. This paired modeling structure exceeded earlier single-task learning (STL) and multitask learning (MLT) standards, showing that online community discussions extend well beyond depression symptomatology. Investigators have also used social media data for early detection of depression, drawing on the CLEF eRisk collection. Assessment of five alternative models showed that fusing data-centric and knowledge-acquisition strategies delivered superior performance. Audio-textual methods have also been applied to simulate depressive patterns by gathering information from individuals experiencing depression. One study [8] used an expanded long short-term memory (LSTM) model to detect depressive states.

This work surveys existing literature on machine learning applications for mental health assessment (Section 2). Section 3 elaborates the proposed framework, which incorporates the Synthetic Minority Oversampling

Technique (SMOTE) to counteract class imbalance alongside a stacking classifier ensemble method. This includes processes for data gathering, preprocessing, annotation, and examination of feature significance. Model efficacy is subsequently evaluated employing standard performance indicators (Section 4). The study concludes with a synthesis of outcomes, articulation of contributions, and discussion of broader implications (Section 5). Finally, Section 6 outlines recommendations for subsequent investigations and opportunities for refinement.

#### *Related work*

Numerous investigations have employed data derived from cross-sectional surveys and clinical studies to identify emotional disorders, including depression, through machine learning algorithms applied to various populations such as adolescents, university students, older adults engaged in domestic or professional activities, users of social media platforms, and additional demographic groups [9-12]. Reference [13] emphasizes the importance of feature analysis in such studies. This process often takes considerable time to derive meaningful insights using random forest (RF) models and achieve notable area under the curve (AUC) values. In another study [14], a Random Forest (RF) classifier attained an improvement of nearly 0.92 across 100 iterations when detecting bipolar disorder, depression, schizophrenia, and obsessive-compulsive disorder, despite the presence of substantial class imbalance and a high volume of missing data within a clinical mental health dataset. A Decision Tree model applied to Facebook social media data achieved 0.73 accuracy in identifying sadness [9]. By comparison, Naive Bayes demonstrated superior performance with 85% accuracy when trained on responses from 348 participants. According to Ref. [15], stress detection accuracy among Bangladeshi university students reached 89%. Machine learning techniques utilizing a limited set of features successfully identified indicators of common mental health conditions in university student populations [16]. Most studies on psychosocial prediction have applied diverse machine learning classifiers, including RF, LR, SVM, NB, KNN, DT, ANN, instance-based learning, and ensemble approaches [9, 10, 12, 13, 15, 17-19].

Electroencephalography (EEG) is a robust tool for detecting mental health disorders. This technique records brain activity by positioning electrodes on the scalp. Although alternative neuroimaging methods, such as

computed tomography and functional magnetic resonance imaging (fMRI), are used in diagnostic contexts, EEG offers distinct benefits by capturing dynamic temporal patterns of brain activity at a lower cost than other modalities [20, 21]. Research on the application of EEG for diagnosing social anxiety disorder (SAD) remains relatively scarce. Distinguishing EEG signal variations between control and treatment groups in SAD cases proves challenging, as patterns often appear similar. Consequently, clinicians increasingly rely on automated machine learning-based detection systems. Such approaches facilitate more accurate diagnoses and represent an initial step toward enhanced connectivity analyses, predictive modeling, and evaluation of therapeutic efficacy in SAD [22]. Although other models frequently outperform neural networks, exceptions exist in specific contexts. Contemporary researchers employ neural networks extensively across disciplines, including mental health applications. These networks excel at identifying cancer cells in MRI and CT scan images more effectively than human experts. Experimental identification of SAD can be achieved by analyzing EEG signal patterns. Moreover, investigators have

incorporated the spatial arrangement of EEG electrodes into classification tasks for the first time, resulting in a 7% improvement in convolutional neural network (CNN) accuracy when channel positioning is considered compared to configurations that disregard it [20]. Deep learning has demonstrated remarkable success in areas such as strategic gameplay, speech recognition, medical imaging, and health informatics, owing to its ability to discern complex patterns [23].

Nevertheless, studies exploring deep learning for mental health assessment remain limited. CNN-LSTM combinations have been widely adopted in medical domains to analyze physiological signals, supporting diagnosis, detection, and continuous monitoring [24, 25]. The LSTM method belongs to the family of recurrent neural networks (RNN) and is particularly effective for time-series signal classification. In contrast, CNNs are recognized for their ability to extract salient features from input data. A novel deep learning architecture was developed to systematically characterize mental health indicators, serving as a real-time supportive tool for psychiatric evaluation [26]. **Table 1** summarizes selected prior studies along with their respective limitations.

**Table 1.** Summary of selected related studies in the domain of mental disorders.

| Authors                       | Applied method                     | Performance metric                  | Study limitation               | Reference |
|-------------------------------|------------------------------------|-------------------------------------|--------------------------------|-----------|
| <b>Islam <i>et al.</i></b>    | Decision Tree, Naive Bayes         | Accuracy: 73% and 85%, respectively | Limited dataset size (n = 348) | [9]       |
| <b>Wade <i>et al.</i></b>     | Random Forest                      | AUC                                 | High computational time        | [13]      |
| <b>Rois <i>et al.</i></b>     | Random Forest                      | Accuracy: 92%                       | Small sample size (n = 355)    | [15]      |
| <b>Mokatren <i>et al.</i></b> | Convolutional Neural Network (CNN) | Accuracy: 87%                       | Applicable only to SAD         | [20]      |

Ensemble methods constitute a prominent category within machine learning that enhances predictive performance by aggregating outputs from multiple base models. Although the potential number of ensemble configurations for any given predictive task is virtually unlimited, three principal strategies prevail in ensemble learning research. These approaches have spawned numerous specialized variants rather than entirely distinct algorithms. The fundamental categories of ensemble learning are bagging, stacking, and boosting. A comprehensive understanding of each is essential before applying them to computational modeling tasks.

- Bagging involves constructing multiple decision trees using random subsets of the same dataset and deriving the final prediction through averaging the individual outcomes.

- Stacking entails training several heterogeneous model architectures on the identical dataset and employing a meta-learner to combine their predictions optimally.
- Boosting sequentially incorporates additional ensemble members to rectify errors from prior predictions, ultimately producing a weighted aggregation of forecasts.

#### *Proposed model*

This study introduces an ensemble-based machine learning framework designed to enhance classification accuracy regarding the severity levels of mental health conditions, specifically stress, anxiety, and depression. **Figure 1** illustrates the overall structure of the proposed approach.

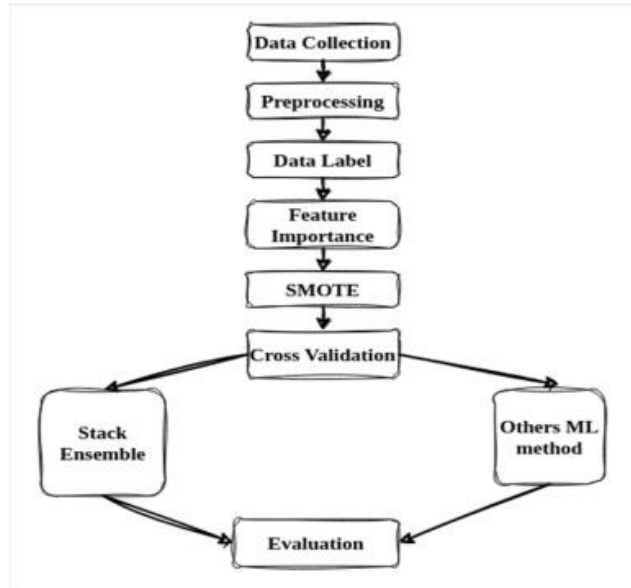
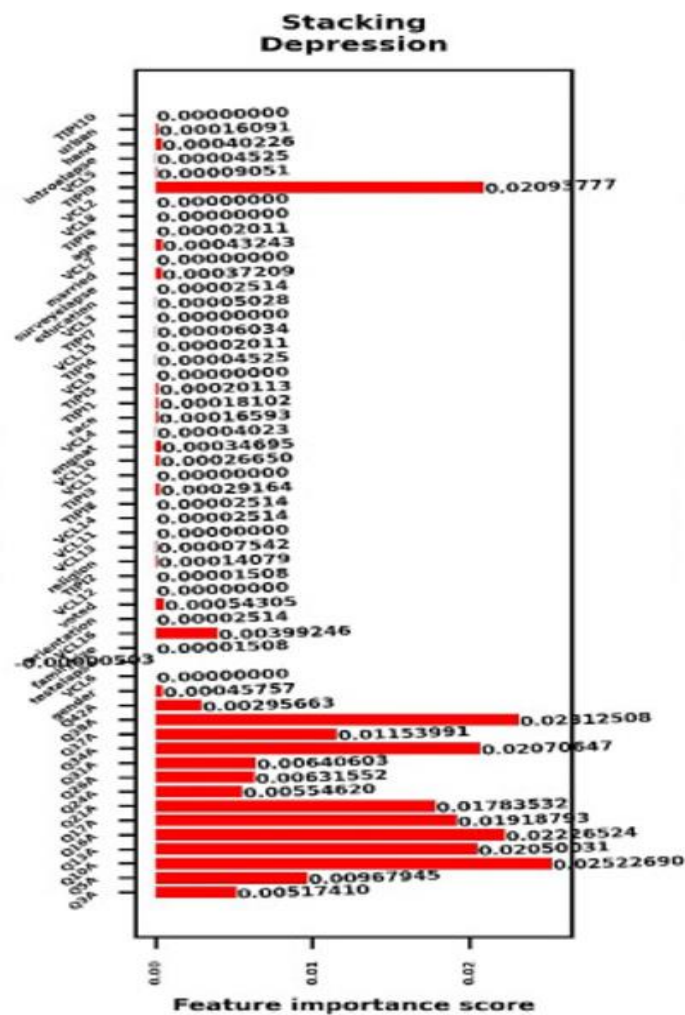
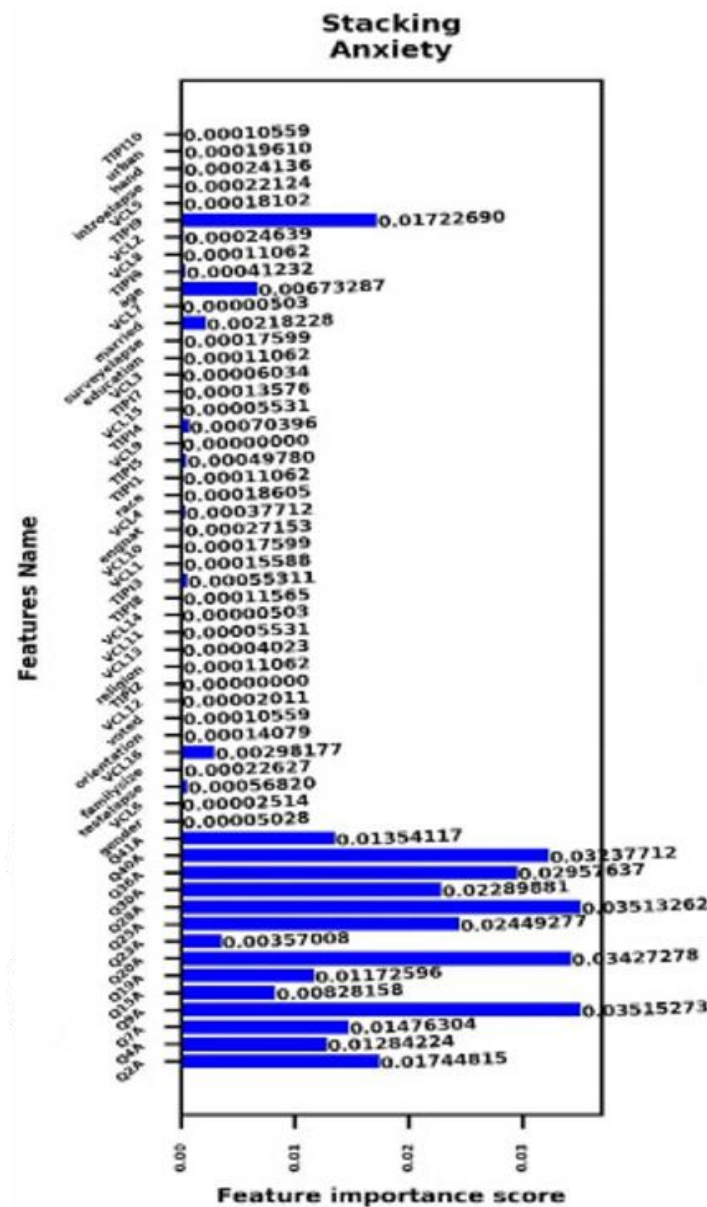


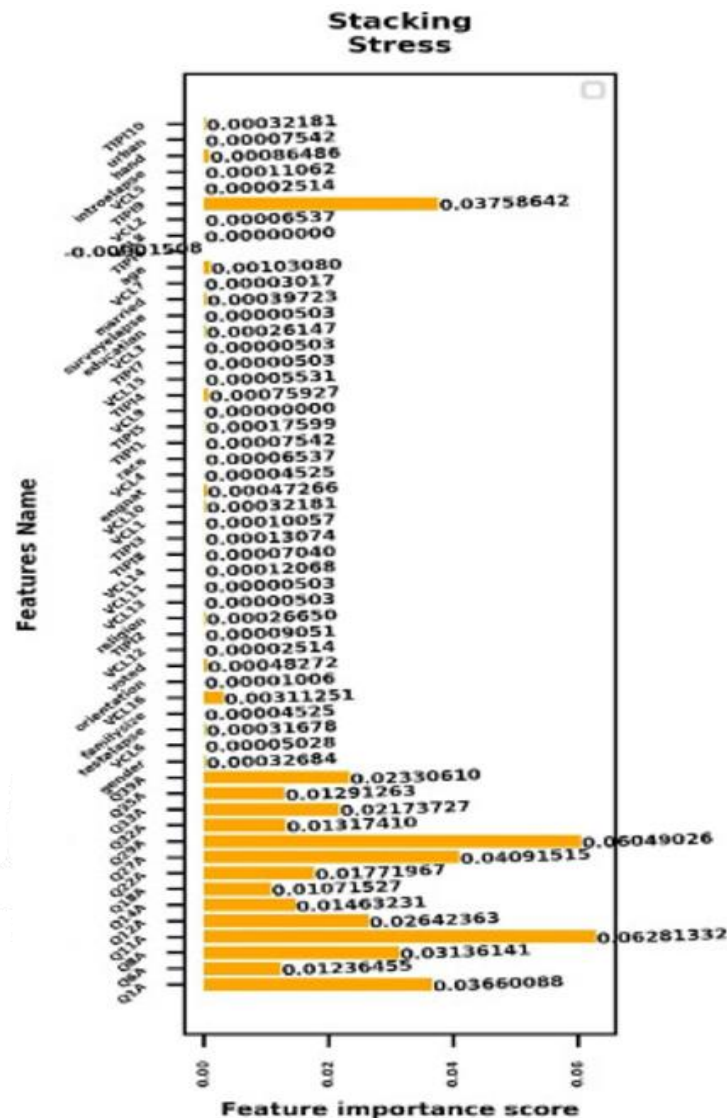
Figure 1. Overview of the proposed methodology.



(a)



(b)



(c)

**Figure 2.** Feature importance scores derived from the stacking classifier model.

### Data collection

Data acquisition constitutes a fundamental phase in any empirical investigation. Gathering information related to mental disorders presents particular difficulties. The mental healthcare domain suffers from limited data availability, further complicated by stringent ethical considerations surrounding patient confidentiality. For this research, data were obtained through an online questionnaire and the established DASS dataset [27]. The resulting dataset is extensive, possessing dimensions of (39775, 175), corresponding to 39,775 individual observations, each characterized by 175 distinct attributes. The survey was accessible to all interested participants, who were invited to contribute in exchange

for personalized feedback. After completion, respondents were allowed to complete a concise follow-up questionnaire. The dataset includes only those individuals who consented to the study and affirmed the honesty of their responses when queried at the conclusion: “Have you supplied honest answers? It may be used for studies?” Data collection occurred between 2017 and 2019. The Depression Anxiety Stress Scale (DASS) is a 42-item self-report instrument developed to evaluate three interrelated negative emotional states: depression, anxiety, and tension.

### Preprocessing

During this phase, the acquired dataset underwent systematic preprocessing. Missing values were eliminated, and categorical variables were transformed into numerical formats suitable for model input. Preprocessing represents a critical component of machine learning workflows. Poorly prepared data containing missing entries or biases can hinder effective model learning and compromise predictive performance. Given the sensitive nature of the information, preprocessing procedures were executed with particular care and strict adherence to ethical standards. The dataset was largely preprocessed before acquisition; therefore, the primary additional step involved applying L2 normalization. L2 normalization adjusts each data point to possess a Euclidean norm of 1. This is accomplished by dividing every element of a vector  $X$  by the square root of the sum of its squared components, as expressed in Eq. (1). This method preserves the directional properties of data points in multidimensional space while standardizing their magnitudes. Consequently, it proves highly beneficial in

machine learning and data analysis for maintaining features on a comparable scale.

$$L_2 = \sqrt{(x_1^2 + x_2^2 + x_3^2 + x_4^2 \dots x_n^2)} \quad (1)$$

#### Data labeling

The DASS-42 instrument adopts a dimensional rather than categorical perspective on psychological disorders. It was constructed upon the premise—supported by empirical evidence—that differences in depression, anxiety, and stress between non-clinical and clinical populations reflect variations in intensity rather than distinct categories. Accordingly, the DASS-42 does not directly align with discrete diagnostic classifications outlined in systems such as the DSM or ICD. Instead, it generates total scores along with subscale scores for each of the three domains. Percentile rankings for the subscales were established based on a normative community sample [28]. Each condition was categorized into five severity levels, with the complete labeling scheme presented in **Table 2**.

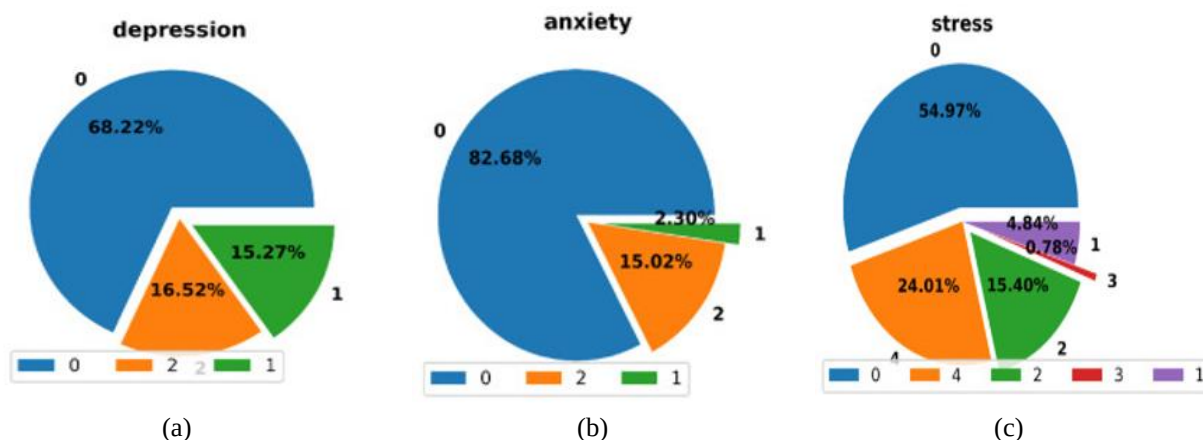
**Table 2.** Labeling scheme for the DASS dataset.

| Severity level          | Stress score | Anxiety score | Depression score |
|-------------------------|--------------|---------------|------------------|
| <b>Normal</b>           | 0–14         | 0–7           | 0–9              |
| <b>Mild</b>             | 15–18        | 8–9           | 10–13            |
| <b>Moderate</b>         | 19–25        | 10–14         | 14–20            |
| <b>Severe</b>           | 26–33        | 15–19         | 21–27            |
| <b>Extremely Severe</b> | 34+          | 20+           | 28+              |

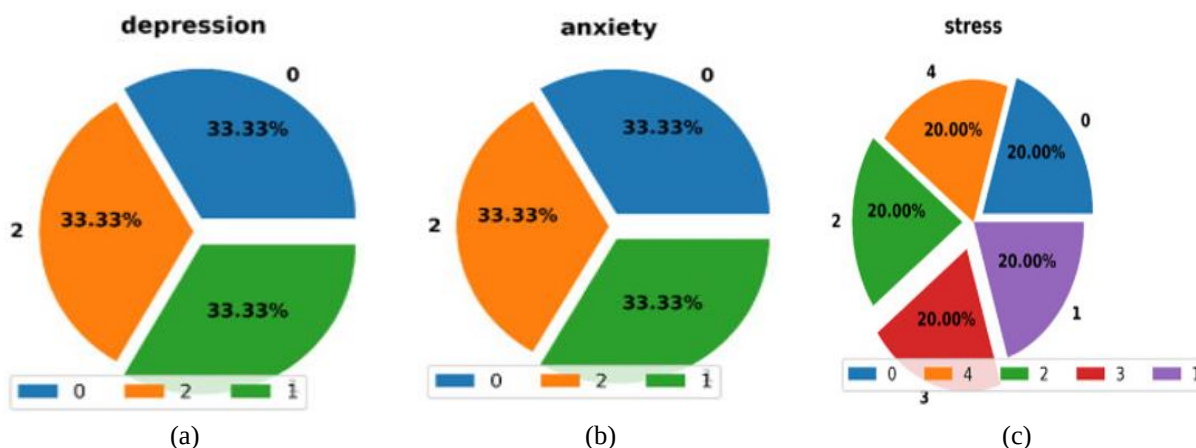
#### SMOTE technique

Imbalanced classification involves developing predictive models for datasets exhibiting substantial disparity between class distributions. Such imbalances often lead conventional machine learning algorithms to overlook or underperform on the underrepresented class. Nevertheless, the minority class frequently holds the greatest analytical importance. One effective strategy for addressing class imbalance is synthetically generating additional minority-class instances. A basic approach is to duplicate existing minority samples; however, this method adds no new information to the learning process. More sophisticated techniques generate new samples by interpolating between existing ones. The Synthetic Minority Oversampling Technique (SMOTE) is a widely adopted data augmentation method that specifically targets the minority class.

SMOTE operates by identifying nearby instances within the feature space, computing the vector difference between a selected sample and its neighbors in high-dimensional space, and creating a new synthetic instance along the connecting line segment. The procedure begins by randomly selecting a minority-class instance. It then identifies its nearest neighbors (typically  $k = 5$ ). One neighbor is chosen at random, and a synthetic example is generated at a random point along the line segment joining the two instances in feature space. **Figure 3** illustrates the dataset's class distribution before applying SMOTE. **Figure 4** shows the class distribution after SMOTE. Notably, for depression and anxiety disorders, the normal and mild severity categories contain no observations, which may substantially impair machine learning model performance.



**Figure 3.** Class distribution of the original dataset before SMOTE application.



**Figure 4.** Class distribution of the dataset following SMOTE application.

#### Feature importance

During statistical model construction, feature selection reduces the dimensionality of input variables. Reducing the number of features helps alleviate computational demands and, in some cases, improves overall model effectiveness. Various techniques can identify the most relevant features in machine learning tasks. One common approach relies on feature importance metrics derived from mathematical computations. These metrics assign scores to input variables based on their contribution to predicting the target outcome. Examples include correlation coefficients, regression coefficients, decision

tree-based measures, and permutation importance scores. In this study, we used permutation importance. Unlike several alternative methods that are model-specific, permutation importance applies broadly across nearly all machine learning algorithms.

**Tables 3-5** present the descriptive statistics for the response variable distributions and the primary independent variables associated with depression, anxiety, and stress, respectively. Although the complete dataset contains over 170 features, only the most influential variables are displayed here.

**Table 3.** Descriptive statistics of key independent variables for the depression label.

|       | Q3A   | Q5A   | Q10A  | Q13A  | Q16A  | Q17A  | Q21A  | Q24A  | Q26A  | Q31A  | Q34A  | Q37A  | Q38A  | Q42A  |
|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| count | 39775 | 39775 | 39775 | 39775 | 39775 | 39775 | 39775 | 39775 | 39775 | 39775 | 39775 | 39775 | 39775 | 39775 |
| mean  | 2.226 | 2.521 | 2.447 | 2.785 | 2.520 | 2.659 | 2.350 | 2.437 | 2.659 | 2.377 | 2.634 | 2.374 | 2.393 | 2.680 |

|      |       |       |       |       |       |       |       |       |       |       |       |       |       |       |
|------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| std  | 1.039 | 1.070 | 1.139 | 1.074 | 1.111 | 1.157 | 1.166 | 1.051 | 1.067 | 1.044 | 1.151 | 1.140 | 1.187 | 1.033 |
| min  | 1     | 1     | 1     | 1     | 1     | 1     | 1     | 1     | 1     | 1     | 1     | 1     | 1     | 1     |
| 25 % | 1     | 2     | 1     | 2     | 2     | 2     | 1     | 2     | 2     | 2     | 2     | 1     | 1     | 2     |
| 50 % | 2     | 2     | 2     | 3     | 2     | 3     | 2     | 2     | 3     | 2     | 3     | 2     | 2     | 3     |
| 75 % | 3     | 3     | 4     | 4     | 4     | 4     | 3     | 3     | 4     | 3     | 4     | 3     | 4     | 4     |
| max  | 4     | 4     | 4     | 4     | 4     | 4     | 4     | 4     | 4     | 4     | 4     | 4     | 4     | 4     |

**Table 4.** Descriptive statistics of key independent variables for the anxiety label.

|       | Q2A   | Q4A   | Q7A   | Q9A   | Q15A  | Q19A  | Q20A  | Q23A  | Q25A  | Q28A  | Q30A  | Q36A  | Q40A  | Q41A  |
|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| count | 39775 | 39775 | 39775 | 39775 | 39775 | 39775 | 39775 | 39775 | 39775 | 39775 | 39775 | 39775 | 39775 | 39775 |
| mean  | 2.172 | 1.950 | 1.925 | 2.670 | 1.827 | 1.946 | 2.323 | 1.562 | 2.184 | 2.217 | 2.392 | 2.268 | 2.651 | 1.966 |
| std   | 1.112 | 1.042 | 1.034 | 1.068 | 0.986 | 1.072 | 1.116 | 0.862 | 1.080 | 1.074 | 1.078 | 1.108 | 1.108 | 1.046 |
| min   | 1     | 1     | 1     | 1     | 1     | 1     | 1     | 1     | 1     | 1     | 1     | 1     | 1     | 1     |
| 25 %  | 1     | 1     | 1     | 2     | 1     | 1     | 1     | 1     | 1     | 1     | 1     | 1     | 2     | 1     |
| 50 %  | 2     | 2     | 2     | 3     | 2     | 2     | 2     | 1     | 2     | 2     | 2     | 2     | 3     | 2     |
| 75 %  | 3     | 3     | 3     | 4     | 2     | 3     | 3     | 2     | 3     | 3     | 3     | 3     | 4     | 3     |
| max   | 4     | 4     | 4     | 4     | 4     | 4     | 4     | 4     | 4     | 4     | 4     | 4     | 4     | 4     |

**Table 5.** Descriptive statistics of key independent variables for the stress label.

|       | Q1A   | Q6A   | Q8A   | Q11A  | Q12A  | Q14A  | Q18A  | Q22A  | Q27A  | Q29A  | Q32A  | Q33A  | Q35A  | Q39A  |
|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| count | 39775 | 39775 | 39775 | 39775 | 39775 | 39775 | 39775 | 39775 | 39775 | 39775 | 39775 | 39775 | 39775 | 39775 |
| mean  | 2.619 | 2.540 | 2.480 | 2.803 | 2.426 | 2.580 | 2.478 | 2.344 | 2.612 | 2.653 | 2.446 | 2.414 | 2.303 | 2.454 |
| std   | 1.032 | 1.050 | 1.052 | 1.049 | 1.066 | 1.080 | 1.068 | 1.030 | 1.050 | 1.058 | 1.018 | 1.051 | 0.998 | 1.019 |
| min   | 1     | 1     | 1     | 1     | 1     | 1     | 1     | 1     | 1     | 1     | 1     | 1     | 1     | 1     |
| 25 %  | 2     | 2     | 2     | 2     | 2     | 2     | 2     | 2     | 2     | 2     | 2     | 2     | 2     | 2     |
| 50 %  | 3     | 2     | 2     | 3     | 2     | 2     | 2     | 2     | 3     | 3     | 2     | 2     | 2     | 2     |
| 75 %  | 4     | 3     | 3     | 4     | 3     | 4     | 3     | 3     | 4     | 4     | 3     | 3     | 3     | 3     |
| max   | 4     | 4     | 4     | 4     | 4     | 4     | 4     | 4     | 4     | 4     | 4     | 4     | 4     | 4     |

**SMOTE technique**

Imbalanced classification refers to developing predictive models for datasets with considerable disparity in class representation. This imbalance poses a significant challenge, as most machine learning algorithms tend to ignore or perform poorly on the underrepresented minority class, despite its often critical importance. A common strategy to address such imbalances is the synthetic interpolation of minority-class instances. The most rudimentary method replicates existing minority samples; however, this provides no additional information for the model. More advanced approaches generate novel instances by combining features of

current samples. The Synthetic Minority Oversampling Technique (SMOTE) constitutes a specialized form of data augmentation focused on the minority class.

SMOTE proceeds as follows:

- Selecting sufficiently proximate samples within the feature space,
- Computing the vector difference between these examples in higher-dimensional space, and
- Generating a new synthetic instance along the interpolated line segment.

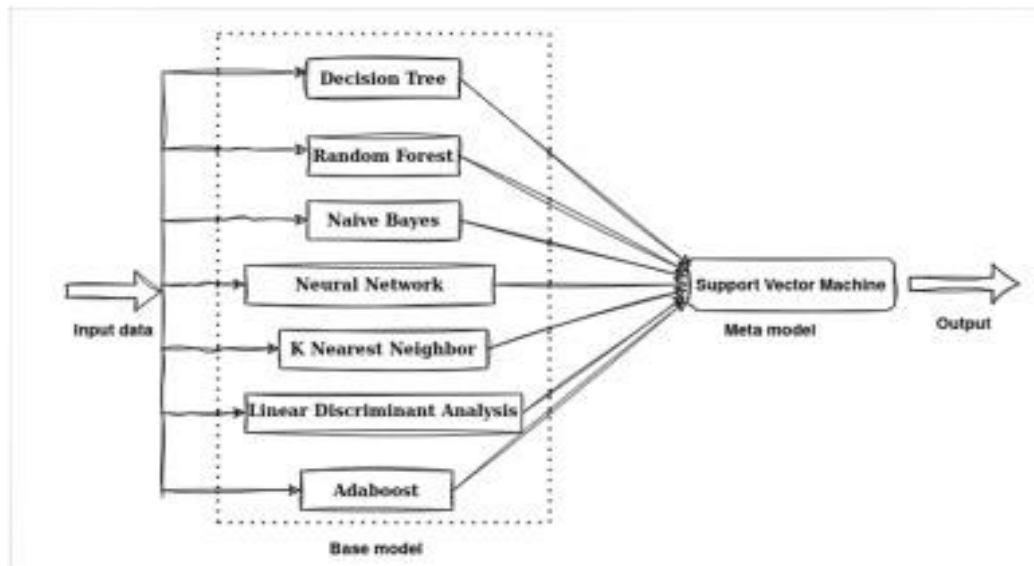
The process initiates with the random selection of a minority class instance. Its nearest neighbors are

subsequently identified (typically  $k = 5$ ). A neighbor is then chosen at random, and a synthetic sample is created at a random location along the line connecting the two points in feature space. **Figure 3** depicts the dataset's class distribution before SMOTE. **Figure 4** illustrates the distribution following SMOTE implementation. It is noteworthy that the normal and mild severity categories for both depression and anxiety disorders contain no observations, a circumstance that may substantially undermine machine learning model performance.

#### Stacking classifier ensemble

Given the high dimensionality often present in datasets, relying upon a single classifier to achieve high predictive accuracy can prove overly optimistic, particularly when the available data volume is limited. A stacking classifier is an ensemble learning strategy that uses predictions from multiple base classifiers as input features for a meta-classifier, which then produces the final output. Various models can serve as base learners. For instance, in multi-class problems addressed through a one-versus-rest framework, some classes may be linearly separable from the remaining data points, while others are not. Consequently, base learners must be selected or adapted to accommodate these varying conditions. This ensemble approach enhances flexibility and effectiveness in handling complex multi-class classification tasks. All

base learners are trained on the complete dataset. Their collective predictions are then aggregated to form the training input for the meta-classifier, with the original target labels serving as the corresponding output variable. The meta-classifier can be any chosen algorithm tasked with deriving the ultimate prediction from the base model outputs. In this study, the meta-classifier is trained using the predictions generated by the base learners. **Figure 5** presents the architecture of the proposed stacking classifier. Within this framework, the base models comprise Decision Tree, Random Forest, Naive Bayes, SVM, KNN, Linear Discriminant Analysis, and AdaBoost, while a Neural Network serves as the meta-model. Cross-validation was applied to both the base learners and the meta-model. The meta-model receives as input the outputs produced by the base models. Although the meta-model generates the final classifications, the original data labels are utilized differently during training of the base and meta levels. Stacking seeks to enhance predictive performance; however, success is not guaranteed in all scenarios. Performance gains depend on the inherent complexity of the problem, the adequacy of the training data representation, and the potential for complementary learning through combined predictions. Outcomes also hinge on carefully selecting diverse base models whose predictions are sufficiently accurate and whose errors are minimally correlated.



**Figure 5.** Architecture of the proposed stacking classifier.

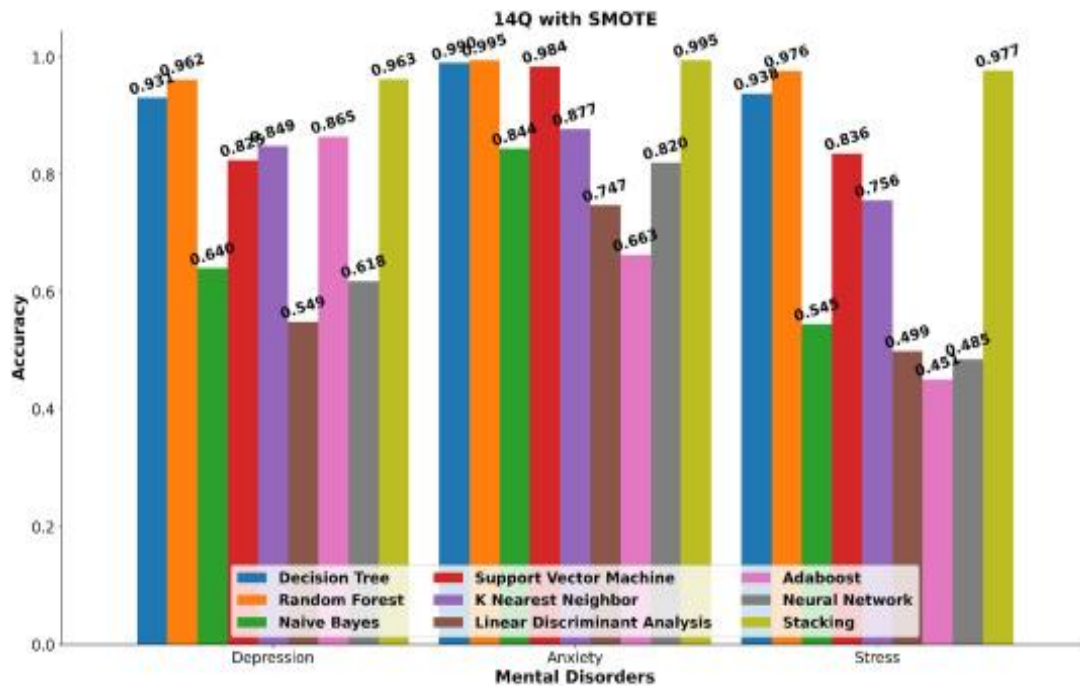
#### Results analysis

##### Experiment using only 14Q features

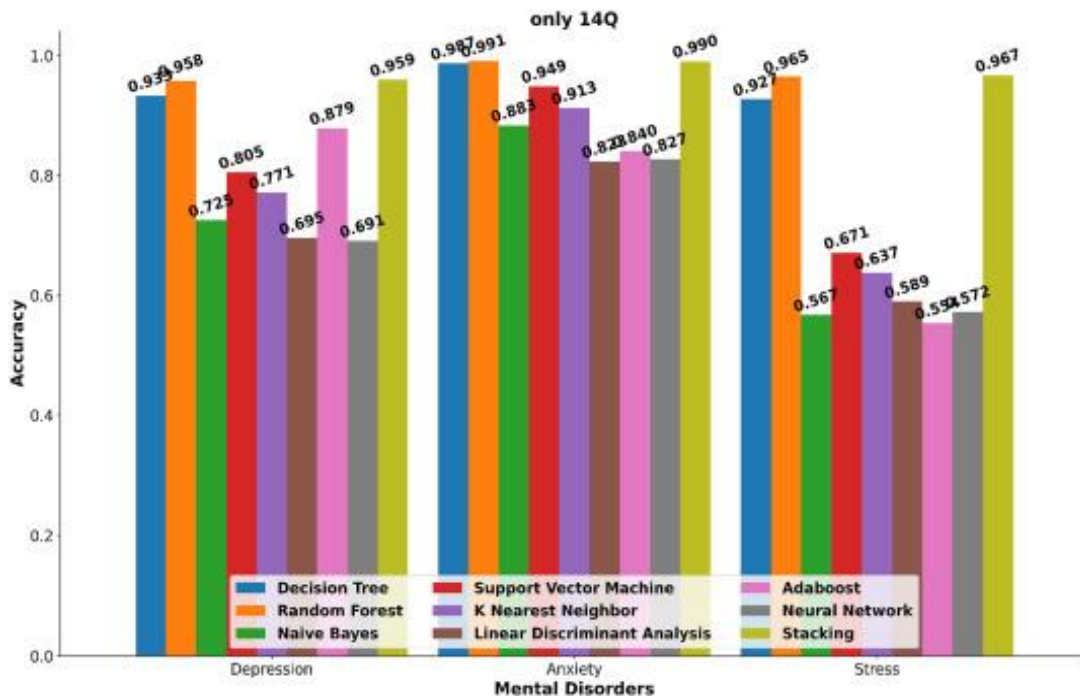
The DASS dataset includes 14 questions dedicated to assessing each psychological condition—namely depression, anxiety, and stress. To evaluate the proposed

model's efficacy, we used only these 14 questions as input features. As illustrated in **Figure 6**, the stacking classifier achieved the highest accuracy when restricted to the 14Q features, recording 0.959% for depression, 0.990% for anxiety, and 0.966% for stress. Random Forest and Decision Tree models performed similarly to

the proposed ensemble, whereas several other classifiers showed comparatively lower results on the DASS dataset. **Figure 7** presents the accuracy scores of the proposed model alongside those of other established algorithms. **Table 6** details additional evaluation metrics.



**Figure 6.** Accuracy scores obtained using the 14Q features with SMOTE.



**Figure 7.** Accuracy scores obtained using only the 14Q features.

**Table 6.** Training and testing performance metrics (accuracy, macro-averaged precision, recall, and F1-score) for the 14Q features.

| Classifier                         | Mental health condition | Training F1-score (Macro) | Testing F1-score (Macro) | Training recall (Macro) | Testing recall (Macro) | Training precision (Macro) | Testing precision (Macro) | Training accuracy | Testing accuracy |
|------------------------------------|-------------------------|---------------------------|--------------------------|-------------------------|------------------------|----------------------------|---------------------------|-------------------|------------------|
| Logistic regression                | Depression              | 0.4150                    | 0.4140                   | 0.4174                  | 0.4168                 | 0.5363                     | 0.5337                    | 0.6968            | 0.6960           |
|                                    | Anxiety                 | 0.3100                    | 0.3102                   | 0.3362                  | 0.3362                 | 0.3836                     | 0.3852                    | 0.8245            | 0.8244           |
|                                    | Stress                  | 0.2602                    | 0.2584                   | 0.2672                  | 0.2659                 | 0.2983                     | 0.2959                    | 0.5932            | 0.5917           |
| Decision tree                      | Depression              | 0.9616                    | 0.9109                   | 0.9443                  | 0.8959                 | 0.9824                     | 0.9286                    | 0.9723            | 0.9340           |
|                                    | Anxiety                 | 0.9530                    | 0.9451                   | 0.9966                  | 0.9911                 | 0.9199                     | 0.9104                    | 0.9922            | 0.9868           |
|                                    | Stress                  | 0.7905                    | 0.7377                   | 0.7922                  | 0.7421                 | 0.7890                     | 0.7338                    | 0.9831            | 0.9267           |
| Random forest                      | Depression              | 0.9616                    | 0.9423                   | 0.9457                  | 0.9249                 | 0.9808                     | 0.9633                    | 0.9723            | 0.9581           |
|                                    | Anxiety                 | 0.9530                    | 0.9516                   | 0.9969                  | 0.9955                 | 0.9197                     | 0.9184                    | 0.9922            | 0.9910           |
|                                    | Stress                  | 0.7905                    | 0.7733                   | 0.7937                  | 0.7747                 | 0.7875                     | 0.7720                    | 0.9831            | 0.9645           |
| Naïve bayes                        | Depression              | 0.6180                    | 0.6160                   | 0.6092                  | 0.6074                 | 0.6469                     | 0.6461                    | 0.7258            | 0.7247           |
|                                    | Anxiety                 | 0.7735                    | 0.7724                   | 0.7884                  | 0.7875                 | 0.8097                     | 0.8083                    | 0.8840            | 0.8833           |
|                                    | Stress                  | 0.3809                    | 0.3791                   | 0.4877                  | 0.4858                 | 0.4540                     | 0.4529                    | 0.5691            | 0.5670           |
| Support vector machine (SVM)       | Depression              | 0.7450                    | 0.6802                   | 0.6900                  | 0.6352                 | 0.8764                     | 0.7934                    | 0.8400            | 0.8050           |
|                                    | Anxiety                 | 0.8894                    | 0.8781                   | 0.9300                  | 0.9175                 | 0.8721                     | 0.8641                    | 0.9559            | 0.9486           |
|                                    | Stress                  | 0.5445                    | 0.4637                   | 0.5173                  | 0.4468                 | 0.6164                     | 0.5164                    | 0.7347            | 0.6709           |
| K-Nearest neighbor (KNN)           | Depression              | 0.7431                    | 0.6488                   | 0.7227                  | 0.6389                 | 0.7897                     | 0.6879                    | 0.8294            | 0.7713           |
|                                    | Anxiety                 | 0.8880                    | 0.8590                   | 0.9524                  | 0.9301                 | 0.8387                     | 0.8072                    | 0.9390            | 0.9126           |
|                                    | Stress                  | 0.5827                    | 0.4670                   | 0.6483                  | 0.5459                 | 0.6009                     | 0.4823                    | 0.7350            | 0.6374           |
| Linear discriminant analysis (LDA) | Depression              | 0.4305                    | 0.4291                   | 0.4331                  | 0.4321                 | 0.5396                     | 0.5398                    | 0.6965            | 0.6954           |
|                                    | Anxiety                 | 0.3138                    | 0.3138                   | 0.3375                  | 0.3375                 | 0.3793                     | 0.3808                    | 0.8230            | 0.8231           |
|                                    | Stress                  | 0.2642                    | 0.2631                   | 0.2743                  | 0.2735                 | 0.2913                     | 0.2895                    | 0.5905            | 0.5893           |
| AdaBoost                           | Depression              | 0.8361                    | 0.8330                   | 0.8162                  | 0.8126                 | 0.8712                     | 0.8701                    | 0.8805            | 0.8787           |
|                                    | Anxiety                 | 0.7490                    | 0.7482                   | 0.8616                  | 0.8613                 | 0.6883                     | 0.6879                    | 0.8412            | 0.8401           |
|                                    | Stress                  | 0.1877                    | 0.1845                   | 0.2191                  | 0.2175                 | 0.2648                     | 0.2585                    | 0.5537            | 0.5536           |
| Neural network                     | Depression              | 0.3381                    | 0.3389                   | 0.3689                  | 0.3699                 | 0.4012                     | 0.3992                    | 0.6915            | 0.6917           |
|                                    | Anxiety                 | 0.3017                    | 0.3017                   | 0.3333                  | 0.3333                 | 0.2756                     | 0.2756                    | 0.8268            | 0.8268           |
|                                    | Stress                  | 0.1853                    | 0.1837                   | 0.2218                  | 0.2207                 | 0.3441                     | 0.3365                    | 0.5713            | 0.5699           |
| Stacking ensemble (Proposed)       | Depression              | 0.9616                    | 0.9444                   | 0.9456                  | 0.9286                 | 0.9810                     | 0.9636                    | 0.9723            | 0.9594           |
|                                    | Anxiety                 | 0.9530                    | 0.9504                   | 0.9967                  | 0.9939                 | 0.9198                     | 0.9176                    | 0.9922            | 0.9901           |
|                                    | Stress                  | 0.7904                    | 0.7741                   | 0.7924                  | 0.7750                 | 0.7885                     | 0.7734                    | 0.9830            | 0.9661           |

**Table 6** reports key performance indicators commonly employed in machine learning evaluation—accuracy, precision, recall, and F1-score—computed using macro

averaging for both training and testing phases. Across all metrics, the proposed stacking model consistently delivered superior results.

### Experiment using 14Q features with SMOTE

The DASS dataset comprises 14 questions for the identification of depression, anxiety, and stress. These 14 questions, combined with SMOTE for class balancing, were supplied to the models to assess the proposed framework's effectiveness. As shown in **Figure 6**, the stacking classifier attained the highest accuracies of 0.963% for depression, 0.995% for anxiety, and 0.977%

for stress when the 14Q features were augmented with SMOTE. SMOTE improved outcomes overall, demonstrating its capacity to enhance model effectiveness. Random Forest and Decision Tree classifiers again approached the ensemble's performance, while other models produced markedly inferior results. **Table 7** presents detailed evaluation metrics.

**Table 7.** Training and testing performance metrics (accuracy, macro-averaged precision, recall, and F1-score) for the 14Q features with SMOTE.

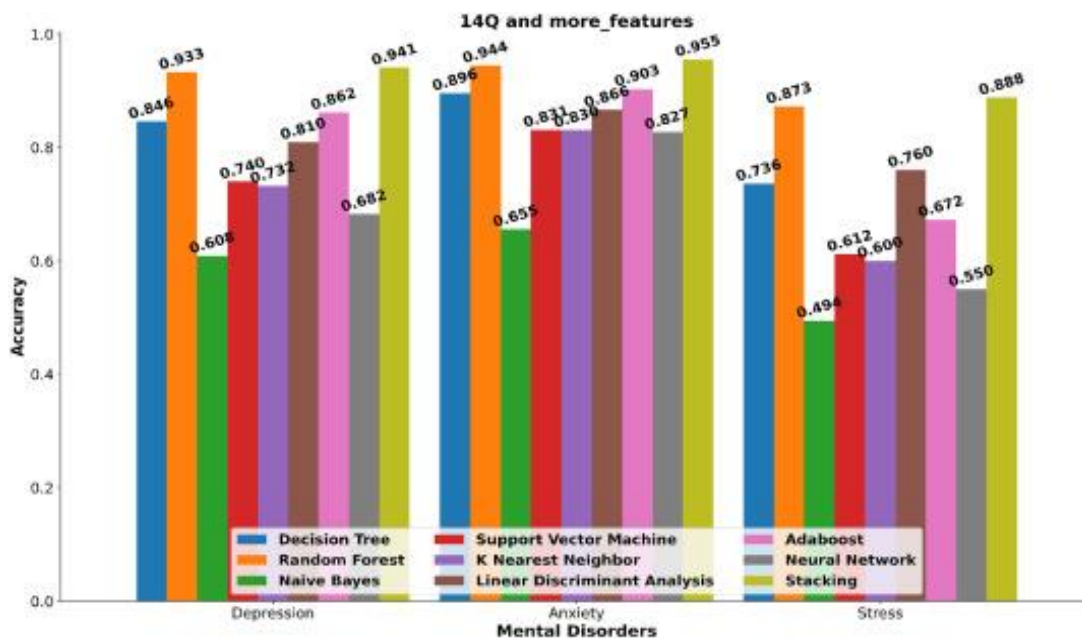
| Algorithm                    | Condition  | Macro F1-score (train) | Macro F1-score (test) | Macro recall (train) | Macro recall (test) | Macro precision (train) | Macro precision (Test) | Training accuracy | Testing accuracy |
|------------------------------|------------|------------------------|-----------------------|----------------------|---------------------|-------------------------|------------------------|-------------------|------------------|
| Logistic regression          | Depression | 0.5441                 | 0.5429                | 0.5505               | 0.5494              | 0.5470                  | 0.5459                 | 0.5505            | 0.5494           |
|                              | Anxiety    | 0.7798                 | 0.7792                | 0.7840               | 0.7835              | 0.7775                  | 0.7769                 | 0.7840            | 0.7835           |
|                              | Stress     | 0.4845                 | 0.4829                | 0.5068               | 0.5053              | 0.4727                  | 0.4714                 | 0.5068            | 0.5053           |
| Decision tree                | Depression | 0.9760                 | 0.9331                | 0.9761               | 0.9336              | 0.9772                  | 0.9338                 | 0.9761            | 0.9336           |
|                              | Anxiety    | 0.9969                 | 0.9894                | 0.9969               | 0.9894              | 0.9969                  | 0.9894                 | 0.9969            | 0.9894           |
|                              | Stress     | 0.9889                 | 0.9404                | 0.9889               | 0.9407              | 0.9892                  | 0.9404                 | 0.9889            | 0.9407           |
| Random forest                | Depression | 0.9761                 | 0.9610                | 0.9762               | 0.9613              | 0.9774                  | 0.9623                 | 0.9762            | 0.9613           |
|                              | Anxiety    | 0.9969                 | 0.9951                | 0.9969               | 0.9951              | 0.9969                  | 0.9951                 | 0.9969            | 0.9951           |
|                              | Stress     | 0.9890                 | 0.9762                | 0.9890               | 0.9764              | 0.9893                  | 0.9766                 | 0.9890            | 0.9764           |
| Naïve bayes                  | Depression | 0.6438                 | 0.6425                | 0.6426               | 0.6414              | 0.6575                  | 0.6567                 | 0.6426            | 0.6414           |
|                              | Anxiety    | 0.8413                 | 0.8408                | 0.8417               | 0.8412              | 0.8412                  | 0.8408                 | 0.8417            | 0.8412           |
|                              | Stress     | 0.5344                 | 0.5329                | 0.5456               | 0.5444              | 0.5292                  | 0.5284                 | 0.5456            | 0.5444           |
| Support vector machine       | Depression | 0.8458                 | 0.8249                | 0.8451               | 0.8243              | 0.8503                  | 0.8289                 | 0.8451            | 0.8243           |
|                              | Anxiety    | 0.9878                 | 0.9836                | 0.9878               | 0.9837              | 0.9880                  | 0.9839                 | 0.9878            | 0.9837           |
|                              | Stress     | 0.8607                 | 0.8320                | 0.8634               | 0.8361              | 0.8621                  | 0.8336                 | 0.8634            | 0.8361           |
| K-nearest neighbors          | Depression | 0.8770                 | 0.8475                | 0.8779               | 0.8494              | 0.8783                  | 0.8493                 | 0.8779            | 0.8494           |
|                              | Anxiety    | 0.8695                 | 0.8555                | 0.8910               | 0.8769              | 0.8607                  | 0.8488                 | 0.8910            | 0.8769           |
|                              | Stress     | 0.8068                 | 0.7694                | 0.8327               | 0.7947              | 0.7938                  | 0.7619                 | 0.8327            | 0.7947           |
| Linear discriminant analysis | Depression | 0.5402                 | 0.5398                | 0.5480               | 0.5475              | 0.5446                  | 0.5443                 | 0.5480            | 0.5475           |
|                              | Anxiety    | 0.7391                 | 0.7381                | 0.7484               | 0.7475              | 0.7411                  | 0.7401                 | 0.7484            | 0.7475           |
|                              | Stress     | 0.4807                 | 0.4793                | 0.5014               | 0.5000              | 0.4697                  | 0.4686                 | 0.5014            | 0.5000           |
| AdaBoost                     | Depression | 0.8754                 | 0.8731                | 0.8748               | 0.8727              | 0.8804                  | 0.8782                 | 0.8748            | 0.8727           |
|                              | Anxiety    | 0.5520                 | 0.5520                | 0.6629               | 0.6629              | 0.4953                  | 0.4953                 | 0.6629            | 0.6629           |

|                              |            |        |        |        |        |        |        |        |        |
|------------------------------|------------|--------|--------|--------|--------|--------|--------|--------|--------|
|                              | Stress     | 0.3719 | 0.3711 | 0.4512 | 0.4504 | 0.4940 | 0.5004 | 0.4512 | 0.4504 |
| Neural network               | Depression | 0.6068 | 0.6060 | 0.6078 | 0.6069 | 0.6250 | 0.6239 | 0.6078 | 0.6069 |
|                              | Anxiety    | 0.8168 | 0.8161 | 0.8221 | 0.8215 | 0.8233 | 0.8225 | 0.8221 | 0.8215 |
| Stacking ensemble (proposed) | Stress     | 0.4484 | 0.4467 | 0.4859 | 0.4839 | 0.4470 | 0.4457 | 0.4859 | 0.4839 |
|                              | Depression | 0.9757 | 0.9628 | 0.9758 | 0.9630 | 0.9769 | 0.9640 | 0.9758 | 0.9630 |
|                              | Anxiety    | 0.9969 | 0.9948 | 0.9969 | 0.9948 | 0.9969 | 0.9949 | 0.9969 | 0.9948 |
|                              | Stress     | 0.9888 | 0.9774 | 0.9888 | 0.9775 | 0.9891 | 0.9777 | 0.9888 | 0.9775 |

#### Experiment using 14q features combined with additional variables

The DASS dataset contains 14 core questions for evaluating depression, anxiety, and stress. These questions were supplemented with further variables, such as age, educational attainment, religious affiliation, marital status, and others highlighted in **Figure 2**, to examine the proposed model's performance. As shown in **Figure 8**, incorporating these extra features alongside the

14Q set reduced effectiveness compared with the 14Q-only configuration. The stacking classifier achieved accuracies of 0.942 % for depression, 0.957 % for anxiety, and 0.897 for stress. Integrating SMOTE again produced improved results, underscoring its utility in enhancing model performance. Random Forest and Decision Tree models remained the closest competitors, whereas other algorithms performed substantially worse. **Table 8** provides comprehensive evaluation metrics.



**Figure 8.** Accuracy scores obtained when combining 14Q features with additional variables.

**Table 8.** Training and testing performance metrics (accuracy, macro-averaged precision, recall, and F1-score) for the 14Q features combined with additional variables.

| Classification method | Mental health condition | Training macro F1-score | Testing macro F1-score | Training macro recall | Testing macro recall | Training macro precision | Testing macro precision | Training accuracy | Testing accuracy |
|-----------------------|-------------------------|-------------------------|------------------------|-----------------------|----------------------|--------------------------|-------------------------|-------------------|------------------|
| Logistic regression   | Depression              | 0.2783                  | 0.2776                 | 0.2840                | 0.2815               | 0.3084                   | 0.3067                  | 0.5826            | 0.5812           |

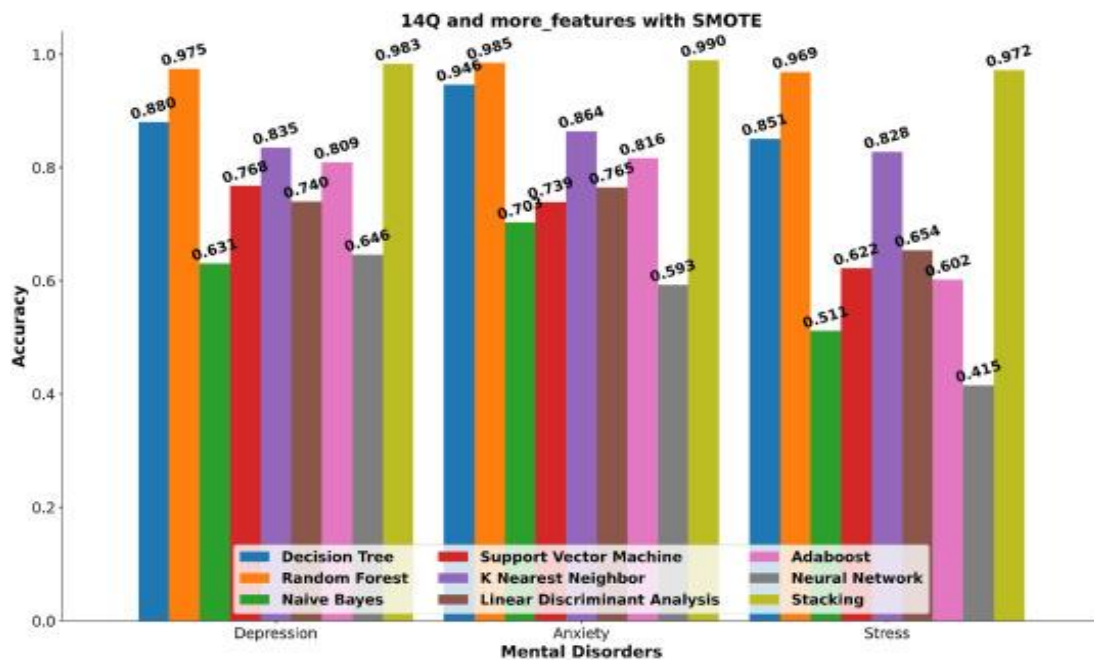
|                                     |            |        |        |        |        |        |        |        |        |
|-------------------------------------|------------|--------|--------|--------|--------|--------|--------|--------|--------|
|                                     | Anxiety    | 0.4148 | 0.4084 | 0.6838 | 0.6822 | 0.3918 | 0.3863 | 0.7165 | 0.7156 |
|                                     | Stress     | 0.3130 | 0.3114 | 0.6141 | 0.6106 | 0.3196 | 0.3184 | 0.6338 | 0.6305 |
| <b>Decision tree</b>                | Depression | 1.0000 | 0.7594 | 1.0000 | 0.7596 | 1.0000 | 0.7596 | 1.0000 | 0.8464 |
|                                     | Anxiety    | 1.0000 | 0.7844 | 1.0000 | 0.7896 | 1.0000 | 0.7797 | 1.0000 | 0.8946 |
|                                     | Stress     | 1.0000 | 0.6511 | 1.0000 | 0.6514 | 1.0000 | 0.6516 | 1.0000 | 0.7361 |
| <b>Random forest</b>                | Depression | 1.0000 | 0.8889 | 1.0000 | 0.8723 | 1.0000 | 0.9085 | 1.0000 | 0.9317 |
|                                     | Anxiety    | 1.0000 | 0.8480 | 1.0000 | 0.7841 | 1.0000 | 0.9438 | 1.0000 | 0.9450 |
|                                     | Stress     | 1.0000 | 0.7252 | 1.0000 | 0.6761 | 1.0000 | 0.8599 | 1.0000 | 0.8698 |
| <b>Naïve bayes</b>                  | Depression | 0.5520 | 0.5497 | 0.6268 | 0.6240 | 0.5498 | 0.5487 | 0.6099 | 0.6084 |
|                                     | Anxiety    | 0.5691 | 0.5633 | 0.6784 | 0.6697 | 0.5792 | 0.5794 | 0.6565 | 0.6552 |
|                                     | Stress     | 0.4446 | 0.4415 | 0.5203 | 0.5167 | 0.4253 | 0.4256 | 0.4946 | 0.4940 |
| <b>Support vector machine</b>       | Depression | 0.4570 | 0.4564 | 0.4615 | 0.4613 | 0.5296 | 0.5387 | 0.7400 | 0.7397 |
|                                     | Anxiety    | 0.3229 | 0.3227 | 0.3437 | 0.3437 | 0.4511 | 0.4338 | 0.8312 | 0.8311 |
|                                     | Stress     | 0.2393 | 0.2379 | 0.2601 | 0.2591 | 0.2675 | 0.2619 | 0.6133 | 0.6117 |
| <b>K-nearest neighbors</b>          | Depression | 0.6712 | 0.5163 | 0.6316 | 0.4987 | 0.7647 | 0.5804 | 0.8102 | 0.7324 |
|                                     | Anxiety    | 0.5644 | 0.4284 | 0.5083 | 0.4081 | 0.7524 | 0.5417 | 0.8730 | 0.8305 |
|                                     | Stress     | 0.5241 | 0.3416 | 0.4837 | 0.3279 | 0.6866 | 0.4430 | 0.7285 | 0.5999 |
| <b>Linear discriminant analysis</b> | Depression | 0.6629 | 0.6589 | 0.6414 | 0.6382 | 0.7060 | 0.7017 | 0.8118 | 0.8097 |
|                                     | Anxiety    | 0.5708 | 0.5595 | 0.5353 | 0.5273 | 0.6360 | 0.6312 | 0.8688 | 0.8661 |
|                                     | Stress     | 0.5305 | 0.5136 | 0.5156 | 0.4996 | 0.5625 | 0.5491 | 0.7644 | 0.7601 |
| <b>AdaBoost</b>                     | Depression | 0.7986 | 0.7884 | 0.7848 | 0.7754 | 0.8233 | 0.8121 | 0.8692 | 0.8621 |
|                                     | Anxiety    | 0.6516 | 0.6393 | 0.6194 | 0.6104 | 0.8033 | 0.8008 | 0.9058 | 0.9030 |
|                                     | Stress     | 0.5218 | 0.5143 | 0.5122 | 0.5056 | 0.6247 | 0.6335 | 0.6768 | 0.6721 |
| <b>Artificial neural network</b>    | Depression | 0.2735 | 0.2728 | 0.3348 | 0.3345 | 0.4965 | 0.4513 | 0.6828 | 0.6826 |
|                                     | Anxiety    | 0.3017 | 0.3017 | 0.3333 | 0.3333 | 0.2756 | 0.2756 | 0.8268 | 0.8268 |
|                                     | Stress     | 0.1428 | 0.1427 | 0.2002 | 0.2002 | 0.1208 | 0.1212 | 0.5499 | 0.5498 |
| <b>Proposed stacking ensemble</b>   | Depression | 1.0000 | 0.9081 | 1.0000 | 0.9042 | 1.0000 | 0.9125 | 1.0000 | 0.9422 |
|                                     | Anxiety    | 1.0000 | 0.8867 | 1.0000 | 0.8567 | 1.0000 | 0.9247 | 1.0000 | 0.9575 |
|                                     | Stress     | 0.9776 | 0.7965 | 0.9641 | 0.7612 | 0.9951 | 0.8595 | 0.9986 | 0.8902 |

#### Experiment using 14Q features with additional variables and SMOTE

The DASS dataset contains 14 questions designed to identify each psychological condition, specifically depression, anxiety, and stress. These 14 questions, supplemented by additional variables such as age, educational attainment, religious affiliation, marital status, and others referenced in **Figure 2**, were provided

as input to evaluate the proposed model's performance. As shown in **Figure 9**, the stacking classifier achieved the highest accuracies when the 14Q features were combined with extra variables and SMOTE, yielding 0.982% for depression, 0.990% for anxiety, and 0.972% for stress. Random Forest and Decision Tree models again produced results close to the proposed ensemble, while other classifiers performed considerably worse on

the DASS dataset. Even modest improvements in mental health classification can yield substantial practical benefits.



**Figure 9.** Accuracy scores obtained when applying the 14Q features with additional variables and SMOTE.

**Table 9** presents standard machine learning evaluation metrics, including accuracy, macro-averaged precision, recall, and F1-score, for both training and testing phases.

The proposed stacking model consistently outperformed alternatives across all evaluated parameters.

**Table 9.** Training and testing performance metrics (accuracy, macro-averaged precision, recall, and F1-score) for the 14Q features combined with additional variables and SMOTE.

| Learning algorithm  | Target disorder | Training macro F1-score | Test macro F1-score | Training macro recall | Test macro recall | Training macro precision | Test macro precision | Training accuracy | Test accuracy |
|---------------------|-----------------|-------------------------|---------------------|-----------------------|-------------------|--------------------------|----------------------|-------------------|---------------|
| Logistic regression | Depression      | 0.4159                  | 0.4162              | 0.4169                | 0.4173            | 0.5300                   | 0.5281               | 0.6976            | 0.6968        |
|                     | Anxiety         | 0.3912                  | 0.3903              | 0.3913                | 0.3902            | 0.4276                   | 0.4270               | 0.8379            | 0.8364        |
|                     | Stress          | 0.3329                  | 0.3329              | 0.6420                | 0.6420            | 0.3310                   | 0.3310               | 0.6606            | 0.6608        |
| Decision tree       | Depression      | 1.0000                  | 0.8818              | 1.0000                | 0.8817            | 1.0000                   | 0.8826               | 1.0000            | 0.8817        |
|                     | Anxiety         | 1.0000                  | 0.9455              | 1.0000                | 0.9456            | 1.0000                   | 0.9460               | 1.0000            | 0.9456        |
|                     | Stress          | 1.0000                  | 0.8521              | 1.0000                | 0.8523            | 1.0000                   | 0.8524               | 1.0000            | 0.8523        |
| Random forest       | Depression      | 1.0000                  | 0.9740              | 1.0000                | 0.9740            | 1.0000                   | 0.9749               | 1.0000            | 0.9740        |
|                     | Anxiety         | 1.0000                  | 0.9856              | 1.0000                | 0.9856            | 1.0000                   | 0.9859               | 1.0000            | 0.9856        |
|                     | Stress          | 1.0000                  | 0.9691              | 1.0000                | 0.9692            | 1.0000                   | 0.9696               | 1.0000            | 0.9692        |
| Naïve bayes         | Depression      | 0.6322                  | 0.6304              | 0.6362                | 0.6345            | 0.6413                   | 0.6404               | 0.6362            | 0.6345        |
|                     | Anxiety         | 0.7109                  | 0.7087              | 0.7046                | 0.7024            | 0.7466                   | 0.7448               | 0.7046            | 0.7024        |
|                     | Stress          | 0.5112                  | 0.5080              | 0.5145                | 0.5114            | 0.5182                   | 0.5158               | 0.5145            | 0.5114        |

|                              |            |        |        |        |        |        |        |        |        |
|------------------------------|------------|--------|--------|--------|--------|--------|--------|--------|--------|
| Support vector machine       | Depression | 0.7758 | 0.7725 | 0.7701 | 0.7667 | 0.8066 | 0.8044 | 0.7701 | 0.7667 |
|                              | Anxiety    | 0.7444 | 0.7422 | 0.7393 | 0.7369 | 0.7571 | 0.7556 | 0.7393 | 0.7369 |
|                              | Stress     | 0.6349 | 0.6300 | 0.6295 | 0.6245 | 0.6510 | 0.6469 | 0.6295 | 0.6245 |
| K-nearest neighbors          | Depression | 0.8875 | 0.8283 | 0.8903 | 0.8347 | 0.9004 | 0.8524 | 0.8903 | 0.8347 |
|                              | Anxiety    | 0.9005 | 0.8583 | 0.9036 | 0.8649 | 0.9157 | 0.8851 | 0.9036 | 0.8649 |
|                              | Stress     | 0.8881 | 0.8208 | 0.8910 | 0.8279 | 0.8943 | 0.8323 | 0.8910 | 0.8279 |
| Linear discriminant analysis | Depression | 0.7496 | 0.7471 | 0.7418 | 0.7392 | 0.7912 | 0.7899 | 0.7418 | 0.7392 |
|                              | Anxiety    | 0.7715 | 0.7697 | 0.7658 | 0.7640 | 0.7959 | 0.7948 | 0.7658 | 0.7640 |
|                              | Stress     | 0.6601 | 0.6571 | 0.6582 | 0.6553 | 0.6745 | 0.6720 | 0.6582 | 0.6553 |
| AdaBoost                     | Depression | 0.8198 | 0.8138 | 0.8166 | 0.8105 | 0.8348 | 0.8300 | 0.8166 | 0.8105 |
|                              | Anxiety    | 0.8264 | 0.8213 | 0.8235 | 0.8183 | 0.8379 | 0.8335 | 0.8235 | 0.8183 |
|                              | Stress     | 0.6080 | 0.6047 | 0.6094 | 0.6059 | 0.6656 | 0.6633 | 0.6094 | 0.6059 |
| Artificial neural network    | Depression | 0.6421 | 0.6381 | 0.6549 | 0.6508 | 0.6548 | 0.6515 | 0.6549 | 0.6508 |
|                              | Anxiety    | 0.5711 | 0.5688 | 0.5956 | 0.5928 | 0.5805 | 0.5774 | 0.5956 | 0.5928 |
|                              | Stress     | 0.4002 | 0.3956 | 0.4333 | 0.4283 | 0.4105 | 0.4048 | 0.4333 | 0.4283 |
| Stacking ensemble (proposed) | Depression | 1.0000 | 0.9820 | 1.0000 | 0.9820 | 1.0000 | 0.9823 | 1.0000 | 0.9820 |
|                              | Anxiety    | 1.0000 | 0.9900 | 1.0000 | 0.9900 | 1.0000 | 0.9902 | 1.0000 | 0.9900 |
|                              | Stress     | 1.0000 | 0.9717 | 1.0000 | 0.9717 | 1.0000 | 0.9722 | 1.0000 | 0.9717 |

## Conclusion

This study proposed a stacking-based ensemble machine learning framework capable of classifying varying severity levels of depression, stress, and anxiety. Data were sourced from [www.kaggle.com](http://www.kaggle.com), and the application of the Synthetic Minority Over-sampling Technique (SMOTE) substantially improved performance on the DASS dataset. Experiments incorporating the 14 core questions alongside supplementary variables produced state-of-the-art results. The findings underscore SMOTE's critical role in managing class imbalance in mental health datasets. Additionally, the analysis revealed a notably low proportion of individuals exhibiting normal mental states, potentially reflecting the increasing pressures and complexities of contemporary life on psychological well-being. Based on these outcomes, the following recommendations are offered: 1) Adopt SMOTE or comparable resampling strategies to mitigate class imbalance in mental health data, as evidenced by enhanced classification performance. 2) Promote continued research in mental health assessment

through interdisciplinary collaboration with domain experts to refine predictive models. 3) Emphasize model interpretability and clinical validation to foster confidence among healthcare practitioners. 4) Examine societal factors contributing to the underrepresentation of normal mental health states in modern populations, with a focus on initiatives that support psychological well-being.

## Future work

Ensemble machine learning approaches can be combined with emerging technologies, such as wearable devices and electronic health records, to generate more comprehensive and accurate assessments of an individual's mental state. Enhanced management of complex data: Future investigations may concentrate on developing sophisticated ensemble models capable of effectively processing high-dimensional and intricate datasets, including brain imaging information, to improve diagnostic precision for mental disorders. Transfer learning: Ensemble frameworks could be applied within transfer learning paradigms to leverage

knowledge across different domains. This strategy holds particular promise for mental health research, where demographic or geographic limitations often constrain data availability.

**Acknowledgments:** None

**Conflict of Interest:** None

**Financial Support:** None

**Ethics Statement:** None

## References

1. Bastías Arriagada EM, Stiepovich Bertoni J. Una revisión de los estilos de vida de estudiantes universitarios iberoamericanos. *Cienc Enferm.* 2014;20:93-101.
2. Burghardt GM. A place for emotions in behaviour systems research. *Behav Process.* 2019;166:103881.
3. Coles ME, Ravid A, Gibb B, George-Denn D, Bronstein LR, McLeod S. Adolescent mental health literacy: young people's knowledge of depression and social anxiety disorder. *J Adolesc Health.* 2016;58(1):57-62.
4. Portman RM, Bradbury J, Lewis K. Social physique anxiety and physical activity behaviour of male and female exercisers. *Eur J Sport Sci.* 2018;18(2):257-65.
5. Shin C, Choi KY, Kim YK. Risk factors and preventions in panic disorder. In: Nova Science Publishers, Inc; 2018. p. 91-107.
6. Bandelow B, Domschke K, Baldwin D. Panic disorder and agoraphobia. Oxford University Press; 2013.
7. Saha B, Nguyen T, Dinh P, Venkatesh S. A framework for classifying online mental health-related communities with an interest in depression. *IEEE J Biomed Health Inform.* 2016;20(4):1008-15.
8. Al Hanai T, Ghassemi MM, Glass JR. Detecting depression with audio/text sequence modeling of interviews. In: Proceedings of Interspeech 2018; Hyderabad, India; 2018. p. 1716-20. doi:10.21437/Interspeech.2018-2522
9. Islam MR, Kabir MA, Ahmed A, Kamal ARM, Wang H, Ulhaq A, et al. Depression detection from social network data using machine learning techniques. *Health Inf Sci Syst.* 2018;6(1).
10. Cacheda F, Fernandez D, Novoa FJ, Carneiro V. Early detection of depression: social network analysis and random forest techniques. *J Med Internet Res.* 2019;21(6):e12554.
11. Nguyen MH, Nguyen QYT, Ho MT, Vuong QH. A dataset of students' mental health and help-seeking behaviours in a multicultural environment. *Data.* 2019;4(3):124.
12. Priya A, Garg S, Tigga NP. Predicting anxiety, depression and stress in modern life using machine learning algorithms. *Procedia Comput Sci.* 2020;167:1258-67.
13. Wade BSC, Joshi SH, Pirnia T, Leaver AM, Woods RP, Thompson PM, et al. Random forest classification of depression status based on subcortical brain morphometry following electroconvulsive therapy. In: IEEE 12th International Symposium on Biomedical Imaging (ISBI). IEEE; 2015.
14. Abou-Warda H, Belal NA, El-Sonbaty Y, Darwish S. A random forest model for mental disorders diagnostic systems. In: Advances in Intelligent Systems and Computing. Springer International Publishing; 2016. p. 670-80.
15. Rois R, Ray M, Rahman A, Roy SK. Prevalence and predicting factors of perceived stress among Bangladeshi university students using machine learning algorithms. *J Health Popul Nutr.* 2021;40(1).
16. Srividya M, Mohanavalli S, Bhalaji N. Behavioural modelling for mental health using machine learning algorithms. *J Med Syst.* 2018;42(5).
17. Nithya B, Ilango V. Predictive analytics in healthcare using machine learning tools and techniques. In: International Conference on Intelligent Computing and Control Systems (ICICCS). IEEE; 2017.
18. Byeon H. Is the random forest algorithm suitable for predicting Parkinson's disease with mild cognitive impairment out of Parkinson's disease with normal cognition? *Int J Environ Res Public Health.* 2020;17(7):2594.
19. Feres M, Louzoun Y, Haber S, Faveri M, Figueiredo LC, Levin L. Support vector machine-based differentiation between aggressive and chronic periodontitis using microbial profiles. *Int Dent J.* 2018;68(1):39-46.
20. Mokatren LS, Ansari R, Cetin AE, Leow AD, Ajilore O, Klumpp H, et al. EEG classification based

- on image configuration in social anxiety disorder. In: 9th International IEEE/EMBS Conference on Neural Engineering (NER). IEEE; 2019.
21. Sanei S, Chambers JA. EEG signal processing. John Wiley and Sons; 2013.
  22. Smith EE, Reznik SJ, Stewart JL, Allen JJB. Assessing and conceptualising frontal EEG asymmetry: an updated primer on recording, processing, analyzing, and interpreting frontal alpha asymmetry. *Int J Psychophysiol.* 2017;111:98-114.
  23. Kaufman LD, Pratt J, Levine B, Black SE. Antisaccades: a probe into the dorsolateral prefrontal cortex in Alzheimer's disease. A critical review. *J Alzheimers Dis.* 2010;19(3):781-93.
  24. Tan JH, Hagiwara Y, Pang W, Lim I, Oh LS, Adam M, et al. Application of stacked convolutional and long short-term memory network for accurate identification of CAD ECG signals. *Comput Biol Med.* 2018;94:19-26.
  25. Yıldırım Ö, Pławiak P, Tan RS, Acharya UR. Arrhythmia detection using deep convolutional neural networks with long duration ECG signals. *Comput Biol Med.* 2018;102:411-20.
  26. Shafiei SB, Lone Z, Elsayed AS, Hussein AA, Guru KA. Identifying mental health status using deep neural networks trained by visual metrics. *Transl Psychiatry.* 2020;10(1).
  27. Depression Anxiety Stress Scales - DASS [Internet]. Sydney (AU): University of New South Wales; [cited 2023 Oct 6]. Available from: <http://www2.psy.unsw.edu.au/dass/>
  28. Henry JD, Crawford JR. The short-form version of the Depression Anxiety Stress Scales (DASS-21): construct validity and normative data in a large non-clinical sample. *Br J Clin Psychol.* 2005;44(2):227-39.