Discover Artificial Intelligence

https://doi.org/10.1007/s44163-025-00708-9

Article in Press

HyOPTEnsemble: custom-weighted soft voting hyperparameter optimization ensemble model, explainable-AI for predicting mental state among university students

Received: 25 June 2025

Accepted: 21 November 2025

Published online: 21 January 2026

Cite this article as: Ahmed R., Fahad N., Miah M.S.U. *et al.* HyOPTEnsemble: custom-weighted soft voting hyperparameter optimization ensemble model, explainable-AI for predicting mental state among university students. *Discov Artif Intell* (2025). https://doi.org/10.1007/s44163-025-00708-9

Rasel Ahmed, Nafiz Fahad, Md. Saef Ullah Miah, Md. Jakir Hossen & Kanta Bhattacharjee

HyOPTEnsemble: Custom-Weighted Soft Voting Hyperparameter Optimization Ensemble Model, Explainable-AI for Predicting Mental State among University Students

Rasel Ahmed[b], Nafiz Fahad[c], Md Saef Ullah Miah[c], Md. Jakir Hossen[d,*], Kanta Bhattacharjee[a]

[a] *Faculty of Science and Technology, American International University-Bangladesh, Dhaka, Bangladesh*
[b] *Faculty of Artificial Intelligence and Engineering, Multimedia University, Cyberjaya, Malaysia*
[c] *Faculty of Information Science and Technology, Multimedia University, Melaka, Malaysia*
[d] *Faculty of Engineering and Technology, Multimedia University, Melaka, Malaysia*

Abstract

The emergence of mental health issues among university students has a detrimental impact on their academic and job performance, leading to increased levels of anxiety, dropout mentality, and even suicide. Male students are at higher risk compared to their female counterparts. Early detection of mental health issues is crucial for effective treatment. While previous research has focused primarily on specific departments or universities, few studies have examined data from multiple institutions. This study proposes a novel hypertuned ensemble machine learning model that demonstrates its effectiveness in predicting the mental state of university students based on their past two weeks' experiences. The dataset was collected through an online survey conducted over a period of one month, involving 50 universities, with a total of 400 responses from undergraduate and graduate students. Of these, 67.1% were male, 31.7% were female, and 1.2% declined to disclose their gender identities. To address the data imbalance issue, a hybrid synthetic minority oversampling method, SMOTE-ENN, is employed. Initially, single machine learning models were utilized, including Decision Tree (DT), Naive Bayes (NB), Ada-Boost, Bagging, Random Forest (RF), and Modified Random

*Corresponding author

Email addresses: raselahmed1337@gmail.com (Rasel Ahmed), fahadnafiz1@gmail.com (Nafiz Fahad), saef@aiub.edu (Md Saef Ullah Miah), jakir.hossen@mmu.edu.my (Md. Jakir Hossen), 20-43542-1@student.aiub.edu (Kanta Bhattacharjee)

Forest. In addition, a grid search was performed for hyperparameter optimization. However, the ensemble model with 5-fold cross-validation achieved the highest accuracy of 0.977, along with outstanding performance metrics of precision, recall, and f1 score, at 0.983, 0.977, and 0.979, respectively. Furthermore, Cohen's kappa coefficient and Matthews correlation coefficient were utilized to measure reliability. The proposed model demonstrates the ability to predict the mental state of university students.

Keywords: Mental State Prediction, Chi-square Test, SMOTE-ENN, OPTUNA, University Students, Explainable AI

1. Introduction

According to World Health Organization (WHO), 1 in every 8 people or estimated 970 million people worldwide suffers from a mental disorder with anxiety and depression which encompasses significant disruptions in thinking, emotional regulation, or behavior (WHO, 2025a). The increasing global prevalence of mental illnesses, impacting approximately 280 million people living with depression worldwide, has led to an intensified focus on mental health prevention and treatment (Solmi et al., 2022; Breslau et al., 2023; Collaborators et al., 2022; WHO, 2023). Among them, the mental health of university students is an issue that has received huge public attention since the COVID-19 pandemic (Pérez et al., 2023). The pandemic has exacerbated this issue and has significantly impacted students' studies, jobs, and social life (universitiesuk, 2024). Research has shown that the mental health of university students has deteriorated significantly, with the proportion of those experiencing mental health problems tripling from 6% in 2016-17 to 16% in 2022-23 (Hall, 2023). Many students live alone and away from their families, which can increase financial expenses and contribute to mental health problems, causing some to drop out of university (Hall, 2023). Feelings of loneliness and lack of financial stability can lead to anxiety and depression among these students. In some cases, this can lead to students giving up their studies and work and even resorting to suicide (universitiesuk, 2024; Valencia-Arias et al., 2023) According to a report by the World Health Organization (WHO), depression is one of the leading causes of disability and suicide is the fourth leading cause of death worldwide (WHO, 2025b). However, Predicting the state of mental stress at early stage can significantly boost treatment outcomes and quality of life of university students.

Mental health disorders among university students have become a global public health concern, with studies reporting high prevalence rates of depression, anxiety, and stress (WHO, 2025a). In Bangladesh, the situation is particularly severe due to cultural stigma, limited mental health resources, and a lack of institutional support. Recent studies indicate that 45–65% of Bangladeshi university students experience moderate to severe mental health issues (Hossain et al., 2020; Islam et al., 2020). Despite these alarming statistics, there is a paucity of empirical research utilizing survey data to identify predictors of mental health outcomes in this population. This study aims to address this gap by analyzing survey data to develop predictive models for mental health outcomes among Bangladeshi university students.

This study endeavors to provide an innovative and improved model for predicting the mental state of university students with greater accuracy and precision. To achieve this, we aim to refine evaluation metrics such as accuracy, precision, recall, and f1-score, Cohens Kappa, Matthews Correlation Coefficient(MCC) by utilizing the data collected from 50 universities in Bangladesh. One of the primary challenges of this research is to conduct the survey across various institutions, rather than restricting the study to specific universities or departments to enhance the diversity, which expands upon prior studies. Another challenge is the imbalanced dataset resulting from the survey. This research proposes a unique technique to address this issue. We aim to bridge this gap by employing categorical data collected from an online survey administered across multiple higher education institutions. This study aims to propose an exceptionally efficient hyperparameter optimization method with a custom-weighted soft voting classifier for the proposed ensemble (HyOPTEnsemble) model that can predict university students' mental state to overcome mental health issues, including dropout mentality, depression, anxiety, and suicide, by accurately predicting mental states based on mental health questionnaires from the past two weeks. The questionnaires fall under both psychology and clinical psychology, as it is a subfield that draws from both general psychology and the specific branch of clinical psychology. Psychology provides a broader framework, while clinical psychology is directly involved in the assessment and treatment of mental health issues that can arise in students. Psychology is the general study of the mind and behavior (how they think, act, react and interact), and it includes areas like educational psychology and health psychology that are relevant to student life (Kendra Cherry, 2024). Furthermore, this research seeks to improve the interpretability of the proposed model by integrating SHapley Additive exPlanations (SHAP). This approach will provide valuable insights into the factors that can contribute to mental state prediction, addressing a gap in previous studies.

There have been many studies on mental health disorders such as depression, anxiety, stress during the period of COVID although not enough studies, especially after the pandemic among university students. Most of the prior research tried to address the effective solution to mitigate mental health disorders problems. Moreover, there are still some limitations in previous studies that can be addressed. Prior research worked with data from a specific department or university and collecting and analyzing less data may enlarge the chances of biases in the prediction result (Sahlan et al., 2021; Bhatnagar et al., 2023). The collection of data from a university, more precisely from the engineering department in India but not from any other departments from the same university limits their study (Bhatnagar et al., 2023). A study worked with a small number of records or small datasets which could be the reason for over-fitting and under-fitting for the machine learning model. Although, a large dataset could improve the performance of predicting mental health (Mutalib et al., 2021; Sahlan et al., 2021; Vaishnavi et al., 2022). Sahlan et al. (2021) worked with an old data set called Mental Health Tech Survey: Data set, collected from Kaggle before COVID and surveying between 2014 to 2016 instead of considering a new survey dataset, and missing values were not removed during pre-processing time. No further pre-processing was mentioned to enhance predictive performance. The data was collected only from one university and the condition of 1994 students were not revealed who did not provide answers in 2020 and 2021 (Baba and Bunji, 2023). Abdul Rahman et al. (2023) mentioned the improvement in the precision score for the prediction model, and their model was not primarily based on social and cultural context. Mantas et al. (2023) did not use data augmentation approaches and multi-class anxiety predictions. In addition, (Sahu and Debbarma, 2023; Ogunseye et al., 2022) suggested that ensemble models can be used to improve the evaluation metrics such as precision, recall, and f1 score.

The proposed method enhances the prediction accuracy of mental state to provide earlier assistance to prospective students instead of risking their lives. The notable contribution of this research includes the following:

- Addressing the limitation of previous works, this study offers a wide diverse survey dataset from 50 universities to analyze the mental health state.
- Worrying too much, being depressed, hopeless, feeling tired, and restless students are highly associated with mental stress.
- A hyperparameter optimization method is employed for the ensemble model with a custom-weighted soft voting classifier to predict mental state.
- Our proposed hyperparameter optimization technique (Optuna) outperformed the traditional hyperparameter Grid Search technique.
- A distinct way to enrich the prediction of evaluation metrics including Cohen's kappa and Matthews Correlation Coefficient(MCC) of mental health state is through a custom-weighted soft-voting optimal Hyper-Tuned ensemble model for multiclass classification.
- Highlighting the explainable-ai for feature extraction and interpreting output to overcome the previous study gap as prior research focused on black box Machine Learning models.

The proposed predictive model aims to assist student counsellors and other stakeholders in anticipating and addressing students' mental health disorders, potentially averting catastrophic outcomes. This study may enable individual university students to identify mental health disorders through using the proposed predictive model, thereby enhancing their academic and professional success.

The following sections comprise the remainder of this paper: Section 2 presents a comprehensive review of prior research relevant to the current topic of investigation.

Section 3 explains the research methodology undertaken to achieve the study objectives. Section 4 offers an in-depth analysis and discussion of the findings, culminating in the conclusion, which is followed by potential areas for further exploration in Section 5.

2. Literature Review

The global concern for the mental health disorders of university students is evident. Utilizing student survey items, including demographics and self-rated health, the authors conducted a study that aimed to predict anxiety symptoms among 329 respondents (Mantas et al., 2023). In order to improve the precision of their model, the authors applied feature selection algorithms (Sahu and Debbarma, 2023). The ultimate objective was to develop a machine learning (ML) model that could predict mental health disorders in students within one year through a health survey (Baba and Bunji, 2023). Machine learning algorithms were employed to classify students into different categories of mental health disorders, including stress, depression (Deng et al., 2024), and anxiety (Mutalib et al., 2021; Kim et al., 2020; Chatterjee et al., 2023; Fatima et al., 2021; Aleem et al., 2022). Early detection of mental disorders in university students based on features can help to prevent negative psychological well-being states and mitigate the consequences of mental health issues (Sahlan et al., 2021; Abdul Rahman et al., 2023). ML models have been utilized to predict mental health treatment outcomes (Ogunseye et al., 2022). Five machine learning techniques were identified and evaluated for their accuracy in identifying mental health issues using various criteria (Vaishnavi et al., 2022). Bhatnagar et al. aimed to identify the extent of anxiety, as well as its effects, in Indian university students (Bhatnagar et al., 2023).

There are several machine learning algorithms that was approached by the previous research to provide treatments and predict mental health disorders. These ML algorithms include Logistic Regression, Decision Trees, Random Forests, KNN (K-Nearest Neighbors) Classifiers, and Neural Networks. However, their accuracies were compared on different measures (Sahu and Debbarma, 2023; Sahlan et al., 2021). Baba and Bunji (2023); Mutalib et al. (2021); Deng et al. (2024) employed other algorithms such as SVM, XGBoost, and LightGBM to improve the evaluation metrics. Ogunseye et al. (2022) developed an algorithm called Ada-Boost that outperformed other single machine learning algorithms by comparing their accuracy. Mantas et al. (2023) employed five machine learning algorithms such as Logistic Regression, KNN Classifier, Decision Tree Classifier, Random Forest, and also stacking ensemble model as well as a multilayer perceptron (MLP) neural network to analyze and Bhatnagar et al. (2023) predict with the small dataset to find the best accuracy. Abdul Rahman et al. (2023) utilized several machine-learning models, including generalized linear models, KNN, naive bayes, neural networks, random forests, recursive partitioning, bagging, and boosting.

Therefore, based on the previous studies in Table 1, more data from different disciplines and various countries can help to enlarge the performance of mental health (Mutalib et al., 2021; Vaishnavi et al., 2022; Abdul Rahman et al., 2023; Bhatnagar et al., 2023), and working with new datasets can be more efficient (Ogunseye et al., 2022). However, surveying data from various continents of the world could increase ethnic diversity, which may lead to better prediction of mental health (Rahman et al., 2020). In addition, collecting data from social media such as Facebook and Twitter can be effective in analyzing sentiment and mental health more accurately (Kim et al., 2020). Due to the sensitivity of depression data, it may not be possible to collect more instead high accuracy can be accomplished through limited data (Aleem et al., 2022). Data analysis over sentiment can be applied to track changes in a clear expression of depression and stress over history (Fatima et al., 2021). Missing values (Ogunseye et al., 2022) and data augmentation approaches and multiclass features can be effective for better prediction (Mantas et al., 2023). Implementing an ensemble model can provide better accuracy which mitigates the overfitting and underfitting (Sahu and Debbarma, 2023). Furthermore, research on various mental health issues and their impacts such as depression and suicide can be vital for extended studies (Chatterjee et al., 2023).

Table 1: Related works in the field of mental health disorder

Reference	Scope	Method Used	Limitation
Bhatnagar et al. (2023)	The study focused on understanding the factors affecting the mental health and well-being of Indian engineering students. It involves developing a questionnaire, collecting data, and applying machine learning algorithms to classify anxiety levels.	Naive Bayes, Decision Tree, Random Forest, and Support Vector Machine (SVM).	Their constraints they used secondary data, which may not have captured all relevant variables influencing mental health treatment access. Small sample size (n=127). Lacked the application of Explainable AI for model output interpretation.
Chatterjee et al. (2023)	The study focused on monitoring depression in real-time by analyzing social media interactions, specifically Twitter posts.	Two-level Depression Profile Detection framework (classification, feature extraction), tokenizing, and sentiment Analysis	Lacked the application of explainable AI for model output interpretation.
Abdul Rahman et al. (2023)	The study aimed to early detection and prediction of negative psychological mental states.	Generalized linear models, k-nearest neighbors, naïve Bayes, neural networks, random forest, and recursive partitioning (RPART).	Their limitation is that they did not remove overfitting and depended on machine learning models, and they also lacked the application of Explainable AI for model output interpretation.
Iyortsuun et al. (2023)	The study conducted a systematic review focusing on machine learning (ML) and deep learning (DL) methodologies for diagnosing various mental health conditions	Systematic Reviews and Meta-Analyses (PRISMA)	Lacked the application of Explainable AI for model output interpretation.
Mantas et al. (2023)	This study predicted anxiety among university students in Lebanon by using machine learning.	Development and comparison of various machine learning algorithms for anxiety prediction, and identification of key predictive factors	Lacked the application of Explainable AI for model output interpretation.
Ogunseye et al. (2022)	The research developed machine learning models to predict mental health issues among students.	AdaBoost, LR, KNN, NN, RF, bagging, and stacking.	Their limitation is model overfitting, where the machine learning model may too closely fit the specific sample data used, possibly not generalizing well to other similar datasets. They also lacked the application of Explainable AI for model output interpretation.
(Baba and Bunji, 2023)	The research developed machine learning models to predict mental health issues among students.	Logistic regression, elastic net, random forest, XGBoost, and LightGBM	Their limitation is model overfitting, where the machine learning model may too closely fit the specific sample data used, possibly not generalizing well to other similar datasets. They also lacked the application of Explainable AI for model output interpretation
Fatima et al. (2021)	The study aimed to extract depression, anxiety, and stress levels from text using semi-supervised machine learning models.	semi-supervised learning regression model named DASentimental.	They lacked the application of Explainable AI for model output interpretation.
Sahlan et al. (2021)	The research employed machine learning techniques to forecast mental health problems among students by utilizing data collected through surveys.	KNN, SVM, DT	Their used machine learning models depend heavily on the quality and breadth of the data available, which may not encompass all relevant variables or nuances of mental health conditions. Small sample size (n=219). Lacked the application of Explainable AI for model output interpretation.
Mutalib et al. (2021)	They researched about mental health prediction models in higher education using machine learning.	Decision Trees, Neural Networks Support Vector Machines (SVM), Naïve Bayes, Logistic Regression.	Lacked the application of Explainable AI for model output interpretation.
Kim et al. (2020)	The study developed a deep learning model to identify a user's mental state based on their posting information on Reddit, focusing on specific mental disorders including depression, anxiety, bipolar disorder, borderline personality disorder, schizophrenia, and autism.	XGBoost and convolutional neural networks (CNN)	Their limitation is that they focused on specific mental states for classification, which did not allow for accurate assessment of comorbid mental health conditions, leaving room for future research to address these complexities and they also lacked the application of Explainable AI for model output interpretation.
Liu and Wang (2025)	The study analyzed LLM based model for psychological counseling, focusing on mental health issues.	LLM(ChatGPT)	The study lacks in addressing the university student mental health specifically and the model may struggle to understand the psychology in diverse scenarios
Linlin et al. (2023)	They focused on the impact of short videos on individual college students behaviors and their psychological disorders	Grounded theory	The study limited to analyzing impact of short videos on students psychological behaviors but did not discuss on academic, social aspects of students mental health

3. Methodology

University students' mental health state prediction task is considered a multiclass classification experiment. As shown in Fig.1, this research involves several steps, such as data collection and pre-processing, optimal feature extraction, and selection using chi-square (X^2) test (p<=0.05). This chi-square test is used to see the significance of each feature. The p-value is less than 0.05, which means the improvement is statistically significant. In addition, Standard Scaling is applied to ensure the features are on the same scale. Furthermore, a Synthetic Minority Oversampling Technique and Edited Nearest Neighbours (SMOTE-ENN) hybrid technique is employed to balance the data. Then, the data is partitioned, and seven individual models are selected using the grid-search hyperparameter tuning method. An ensemble model with a custom weighted soft voting classifier and grid-search hyper-parameter tuning method is evaluated. To enhance the evaluation metrics of the previous grid-search-based ensemble model, an optimization hyper-parameter tuning method (optuna) is employed. The best model is determined based on the evaluation metrics. Finally, an explainable-ai technique is also applied to interpret the output of this study.

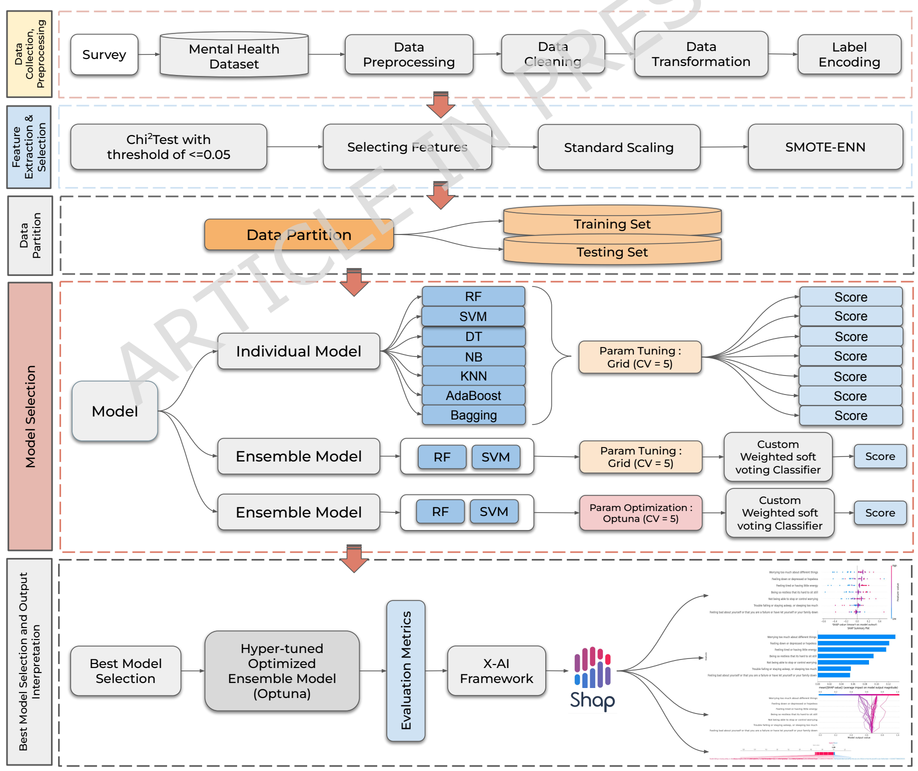

Fig. 1: Overview of the methodology employed in this study

3.1. Dataset Description

3.1.1. Dataset Collection and Survey Procedure

Data Collection is a crucial step in any research. Students' mental disorder-based data collection is one of the challenging tasks. The dataset for this study was collected through a structured or closed-ended survey. We started surveying data through a Google form from August 2023 to September 2023, where all the participants were from undergraduate and graduate levels. All the survey participants have given written informed consent and none of the participants were minors. As evidence of the diversity of our dataset, we collected information from 400 students at 50 different universities in Bangladesh, both public and private universities. All features of our dataset contained categorical values (mentioned in the appendix section) and a total number of 28 features in (Fig.2) is used to analyze this study.

3.1.2. Dataset Validation

The questionnaires for the survey of this study were inspected several times by the independent researchers. At first, the survey form was distributed among a few researchers to test before being shared among university students. All the questionnaires were revised by the experts, which may create confusion for the participants to answer. However, the questions were designed in a structured way. Meanwhile, all the features and variables were written in English (Fig.2). The socio-demographic and academic characteristics and the mental stress of university students based on the last two weeks were analyzed based on the survey taking time. The categorical covariance that was adjusted for all multi-variable models were gender, marital status, having a source of income, feeling alone and lonely, and how frequently bothered by difficulty focusing on things such as reading a newspaper or watching television, feeling down, depressed, or hopeless, thoughts that one would be better off dead, or hurting oneself in some way, based on perception over the past two weeks.

Outcome	Overall Sample (n=400)
University Name	A total of 50 Universities
Gender	Male (67.1%), Female (31.7%), Prefer not to say (1.2%)
Marital status	Married (7.7%), Unmarried (92.3%)
Studying in	Undergraduate level (85.4%), Graduate level (14.6%)
Have own income source	No (73.3%), Yes (26.7%)
Travel distance from university	Walking Distance (45.5%),30 min vehicle travel distance (17.6%), more than 30 min vehicle travel distance (36.9%)
Stay in/with	Hostel/mess (56.9%), Family (43.1%)
Family income per month (BDT)	Up to 40,000 (56.2%), 41,000-100,000 (33.9%), Above 100,000 (9.9%)
Family type	Nuclear family (76.7%), Joint family (23.3%)
How often do you feel left out	Often (30%), Some of the time (54.5%), Hardly ever (15.6%)
How often do you feel isolated from others	Often (34.7%), Some of the time (46.8%), Hardly ever (18.6%)
Feeling nervous, anxious, or on edge	Not at all (23.3%), Several days (54.2%), Nearly every day (16.6%), More than half the days (5.9%)
Not being able to stop or control worrying	Not at all (27%), Several days (45%), Nearly every day (21.8%), More than half the days (6.2%)
Worrying too much about different things	Not at all (23.8%), Several days (53.2%), Nearly every day (17.3%), More than half the days (5.7%)
Trouble relaxing	Not at all (32.2%), Several days (45.3%), Nearly every day (18.3%), More than half the days (4.2%)
Being so restless that it's hard to sit still	Not at all (32.9%), Several days (43.6%), Nearly every day (18.8%), More than half the days (4.7%)
Becoming easily annoyed or irritable	Not at all (32.2%), Several days (46.8%), Nearly every day (16.1%), More than half the days (5%)
Feeling afraid as if something awful might happen	Not at all (36.9%), Several days (38.6%), Nearly every day (17.3%), More than half the days (7.2%)
Little interest or pleasure in doing things	Not at all (28.5%), Several days (54.2%), Nearly every day (14.6%), More than half the days (2.7%)
Feeling down, depressed, or hopeless	Not at all (25.7%), Several days (40.8%), Nearly every day (25.5%), More than half the days (7.9%)
Trouble falling or staying asleep, or sleeping too much	Not at all (28%), Several days (37.6%), Nearly every day (27%), More than half the days (7.4%)
Feeling tired or having little energy	Not at all (20.3%), Several days (44.1%), Nearly every day (26.5%), More than half the days (9.2%)
Poor appetite or overeating	Not at all (38.4%), Several days (41.1%), Nearly every day (15.8%), More than half the days (4.7%)
Feeling bad about yourself or that you are a failure or have let yourself or your family down	Not at all (29%), Several days (37.6%), Nearly every day (27.7%), More than half the days (5.7%)
Trouble concentrating on things, such as reading the newspaper or watching television	Not at all (40.8%), Several days (37.9%), Nearly every day (17.1%), More than half the days (4.2%)
Moving or speaking so slowly that other people could have noticed? Or the opposite being so fidgety or restless that you have been moving around a lot more than usual	Not at all (44.3%), Several days (39.4%), Nearly every day (11.1%), More than half the days (5.2%)
Thoughts that you would be better off dead or of hurting yourself in some way	Not at all (54.5%), Several days (26%), Nearly every day (14.4%), More than half the days (5.2%)
Feel that you lack companionship	Often (32.4%), Some of the time (46.4%), Hardly ever (21%)
Stress Level	Low stress (26.2%), Moderate stress (45.3%), High stress (28.3%)

Fig. 2: Survey responses overview (sample size, n = 400)

3.2. Dataset Pre-Processing:

3.2.1. Data Cleaning

Data cleaning includes identifying and correcting errors within the data set. Thus, the Python library Pandas (Pandas) data frame *data.dropna*() method is used to eliminate the rows that contain $NULL$ values. A total number of three features were dropped that were not important to analyze for this study. However, the Pandas data-frame *data.isnull*().*sum*() method is used to check the missing values for individual features in the dataset, which ensures the dataset has zero null values.

3.2.2. Data Transformation

Data transformation is a process of transforming data from one format or structure to another without changing the content of the datasets. There are several methods such as label encoding (Hancock and Khoshgoftaar, 2020). Label encoding helps to simplify the data. However, all the features in the dataset of this study are categorical data. We employed the label encoding method to transform the categorical features into numerical ones. The features variable is transformed as 0, 1, 2, 3, . . respectively to analyze the data.

3.3. Feature Selection

Feature selection (FS) is used to reduce data dimensionality and leads to enrich the performance of any model. This technique is one of the vital parts of a machine learning model as it improves the performance of the original dataset. In the pre-processing phase, dimension reduction is used to improve the accuracy of learning features and reduce training time by removing irrelevant data, noise, and duplicated features (Zebari et al., 2020). The efficiency of data processing and storage can be enriched by Feature extraction (FE) that can handle the challenge of selecting the most distinct and highly correlated collection of features. Initially, we selected the target feature from the original dataset and Chi^2 Test method is also employed to select highly correlated features (heatmap, Fig.3) to this study or the target feature with the threshold of p values less than equal to 0.05 ($p \leq 0.05$). The method selects the features that satisfy the given condition. Lastly, a total of 7 features (Table 2) were selected as highly correlated from 27 features.

Fig. 3: Heatmap generated from the dataset to show the relationship among the selected optimal features

Table 2: A Total of 7 Optimal Selected Features

Attributes	**Data Type**	**Transformed Values**
Not being able to stop or control worrying	Categorical	0, 1, 2, 3
Worrying too much about different things	Categorical	0, 1, 2, 3
Feeling down or depressed or hopeless	Categorical	0, 1, 2, 3
Being so restless that its hard to sit still	Categorical	0,1,2,3
Trouble falling or staying asleep, or sleeping too much	Categorical	0, 1, 2, 3
Feeling tired or having little energy	Categorical	0, 1, 2, 3
Feeling bad about yourself or that you are a failure or have let yourself or your family down	Categorical	0, 1, 2, 3

3.4. Data Standardization

Standardization is a technique for scaling data that converts the statistical distribution of the data into a mean of 0 (zero) and a standard deviation of 1. This method results in the entire dataset scaling with a zero mean and unit variance. The Python library sklearn provides the $StandardScaler()$ function, which can standardize data values into a standard format. To use this function, we first create an object of the $StandardScaler()$ function. Then, we apply the $fit\_transform()$ method along with the assigned object to transform and standardize the data.

$$z = \frac{x - \mu}{\sigma} \tag{1}$$

where, z = scaled data, x = data to be scaled, μ = Mean, σ = Standard deviation

3.5. Data Augmentation for Label Imbalance

Data augmentation is used to enhance the robustness and accuracy of the machine-learning model. The sample size of university students in the survey was small (sample size = 400). Therefore, the dataset of this experiment indicates the existence of label imbalance. Consequently, we employed the Synthetic Minority Oversampling Technique - Edited Nearest Neighbours (SMOTE-ENN) algorithm (a hybrid method to balance data) to augment the data due to class imbalance and improve generalization performance. The basic oversampling technique uses a normal strategy to reproduce a sample of a target class.

3.5.1. SMOTE-ENN

The Synthetic Minority Oversampling Technique with Edited Nearest Neighbors (SMOTE-ENN) is a hybrid algorithm designed to balance a dataset (Lamari et al., 2021). SMOTE synthesizes new samples in the minority class through linear interpolation, while ENN reduces the number of samples in the majority class by removing noise samples (as depicted in Fig.4).

3.6. Data Partition (Train Test)

Data partition also known as train-test-split technique is a commonly employed method for assessing the performance of machine learning algorithms designed to

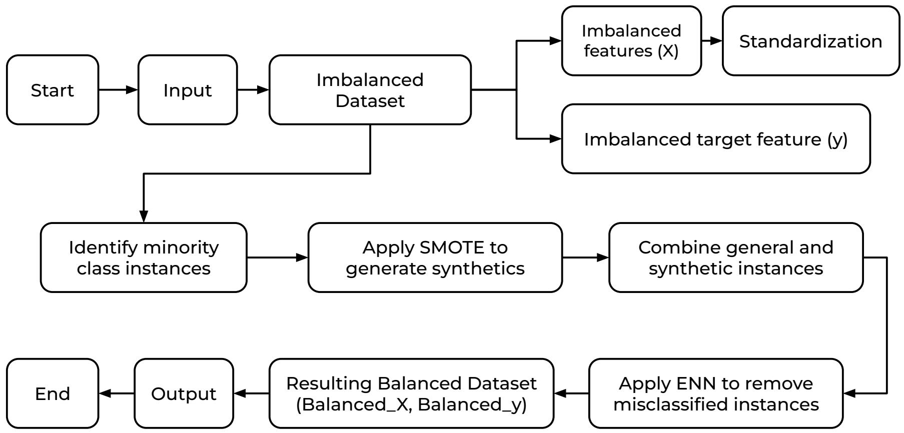

Fig. 4: The workflow of SMOTE-ENN employed in this study

make predictions on unseen data. In this study, train_test_split function is utilized, which shuffles the dataset before dividing it into two segments. The proportion of the original dataset allocated to the test split is determined by the test_size parameter. In this particular case, the test_size has been set at 0.2 or 20%, which indicates that 20% of the data will be used for the test set, while the remaining 80% will be used for the training set after performing the data balancing technique. However, after partitioning the train test data, the shape of the dataset such as X_train shape (165, 7), X_test shape (42, 7), y_train shape(165, 1), and y_test shape (42, 1) (Fig. 5). The target feature of this experiment dataset is the student 'Stress Level' of university students. The standard scaling method is used to balance the data and enhance the performance metrics of this study.

3.7. Model Selection

3.7.1. Individual Models

In the previous studies, the researchers proposed single models to illustrate the best performance. Following the prior studies here in this research, various single classifier models is also employed at the onset of this study to analyze the performance utilizing models such as Decision Tree (DT), Naive Bayes (NB), Ada-Boost, Bagging, Random

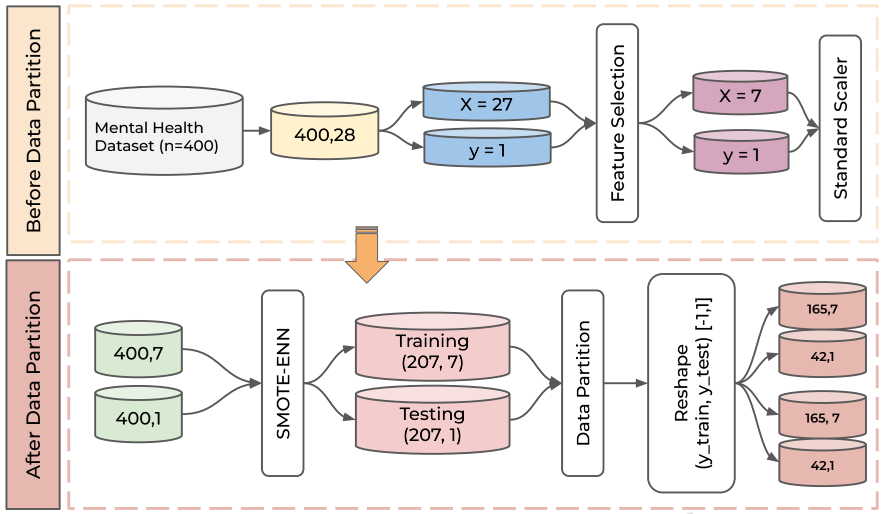

Fig. 5: Graphical view of dataset shape before and after the partition

Forest (RF), Support vector machine(SVM), k-nearest neighbors(KNN). Hyperparameter tuning is used to get the best performance and optimal results. Consequently, GridSearchCV is utilized to obtain the best parameters for the model. This research utilizes a range of classification models, each possessing unique capabilities to address multiclass classification problems, which is essential given the categorical nature of our data.

Random Forest(RF). : RF is an ensemble learning method that constructs a diverse array of decision trees during training and selects the class that receives the highest vote from the individual trees. This makes it well-suited for multiclass classification problems(Random-Forest). The prediction function of a trained Random Forest model can be represented as follows:

$$\hat{y} = \frac{1}{B}\sum_{b=1}^{B} T_b(x) \tag{2}$$

Where $\hat{y}$ is the predicted output, B is the number of trees in the forest, $T_b(x)$ is the prediction of the b-th Decision Tree

SVM: Support Vector Machines (SVMs) have gained significant popularity as a machine learning tool for tasks such as classification. They have demonstrated strong generalization capabilities on a variety of real-world problems, and their theoretical basis is well-established. One of the key advantages of SVMs is that they have a relatively small number of adjustable parameters, and the learning machine's architecture does not require experimental discovery (Yue et al., 2003).

K-Nearest Neighbors (KNN): The k-nearest neighbors (KNN) algorithm is a widely used, non-parametric, supervised learning classifier that operates by determining the proximity of an individual data point to make classifications or predictions about its grouping. With its simplicity and popularity, it remains a prevalent classification and regression tool in the field of machine learning (KNN).

Naive Bayes (NB): NB is a probabilistic classifier that relies on Bayes' theorem with the assumption of independence between the features. It predicts the probabilities of all classes and selects the class with the highest probability, allowing it to effectively handle multiclass classification problems(Webb, 2010).

$$\hat{y} = \arg\max_{y} P(y) \prod_{i=1}^{n} P(x_i \mid y) \tag{3}$$

where $P(y)$ is the prior probability of class y and $P(xi||y)$ is the likelihood of feature xi given class y.

Decision Tree (DT). : DT is a diagrammatic representation in the form of a flowchart, where each internal node signifies a test conducted on an attribute, each branch depicts the outcome of the test, and leaf nodes represent the classes or the distribution of classes. This structure allows it to efficiently handle multiclass classification problems (Szczerbicki, 2001). the prediction function of a trained Decision Tree model can be represented as follows:

$$Entropy(S) = \sum_{i=1}^{n} -p_i * \log_2(p_i) \tag{4}$$

$$Gain(S, A) = Entropy(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} Entropy(S_v) \tag{5}$$

Gain(S,A) is the information gain of dataset S on attribute A, Values(A) is the set of all possible values of attribute A, Sv is the subset of S for which attribute A has value v, $|S_v|$ and $|S|$ are the number of instances in S_v and S, respectively.

AdaBoost. : AdaBoost is an ensemble technique that employs a series of decision trees to enhance the accuracy of classification tasks. This method utilizes boosting, which involves connecting a group of weak classifiers in sequence, whereby each weak classifier strives to rectify the errors made by the previous classifier on misclassified samples (Chatterjee et al., 2019). The AdaBoost algorithm updates the weights of the training instances and the weak learners.

$$H(x) = sign\left(\sum_{t=1}^{T} \alpha_t h_t(x)\right) \tag{6}$$

Bagging. : Bagging is an ensemble meta-algorithm designed to enhance the stability and accuracy of machine learning algorithms. It operates by generating multiple subsets of the original data, training a model (typically a decision tree) on each, and making predictions by aggregating predictions from all models (BaggingClassifier). The prediction function of a trained Bagging classifier can be represented as follows: For a new input vector $x = (x1, x2, ..., xp)$, the predicted output $\hat{y}$ is given by

$$\hat{y} = \frac{1}{B} \sum_{b=1}^{B} C_b(x) \tag{7}$$

where B is the number of base classifiers in the ensemble, and $Cb(x)$ is the prediction of the b-th base classifier

3.7.2. Ensemble Model

Ensemble modeling is a method that involves constructing a collection of diverse models to predict an outcome, either by employing numerous modeling algorithms or utilizing distinct training data sets. The ensemble model subsequently combines the individual predictions of each base model and generates a single final prediction for unseen data (Hancock Jr, 2015).

In this study, we employed an ensemble model (Fig.6) after utilizing the above individual models to enhance the performance, comprising two individual models such as a random forest and a support vector machine classifier. A custom-weighted soft voting classifier technique is used to construct the ensemble model. Soft voting classifier considers the probability scores of each class predicted by individual models and averages them to produce a more refined final prediction. The decision to employ a custom weighted soft voting technique of [0.1, 0.9] for the Random Forest and SVM models, respectively, was a deliberate choice based on our internal evaluation metrics. The custom designation highlights that this was not a standard, equally-weighted ensemble but a strategic one. The empirical analysis showed that the optimized SVM model consistently demonstrated higher predictive performance on our validation set. Therefore, assigning a higher weight of 0.9 to the SVM allowed this study to more heavily leverage its superior predictive capabilities within the ensemble. The Random Forest was included with a smaller weight of 0.1 to provide a supplementary, diversifying perspective, which can help smooth out potential errors while primarily trusting the more robust SVM output. This technique is effective in enhancing the accuracy of prediction by considering the output of multiple models and averaging their scores. Grid-search, a hyper-parameter tuning method is used for this current ensemble model to find the best parameters with 5-fold cross-validation. Consequently, it produces more reliable results compared to individual models.

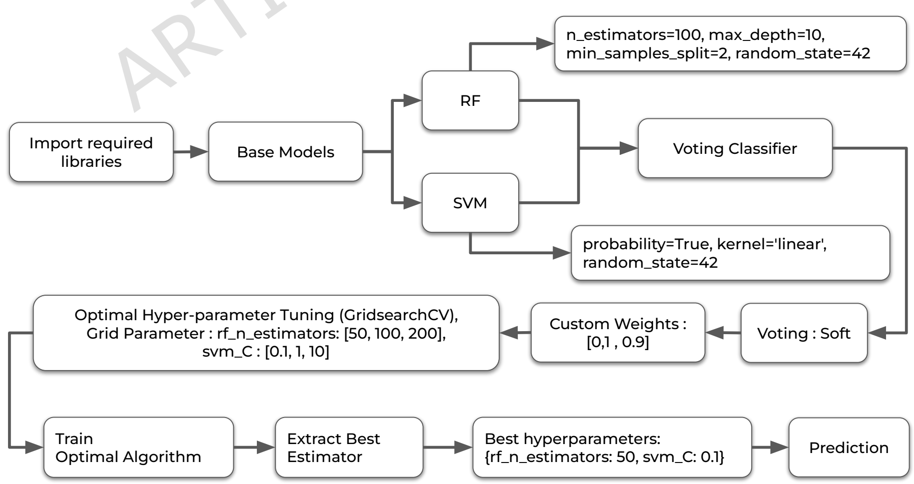

Fig. 6: The Ensemble Model with grid search in details

3.8. Proposed Methods

3.8.1. Proposed HyOPTEnsemble Model

We proposed and developed a hyper-parameter optimized ensemble (HyOPTEnsemble) model comprising two base models in Fig.7 such as, random forest (RF) and support vector machine (SVM) to enrich the performance of previously employed models in this study. Initially, we employed OPTUNA, a hyper-parameter optimized framework instead of grid-search. However, after hyperparameter optimization using OPTUNA. The method determined optimal hyperparameters and trained them to find the optimized models. The optimized models are combined utilizing a custom-weighted soft voting classifier. Custom weights are assigned to emphasize each model's contribution. Consequently, our proposed HyOPTEnsemble model outperformed the previous models and achieved significant evaluation metrics on the test dataset, demonstrating the effectiveness of the proposed model.

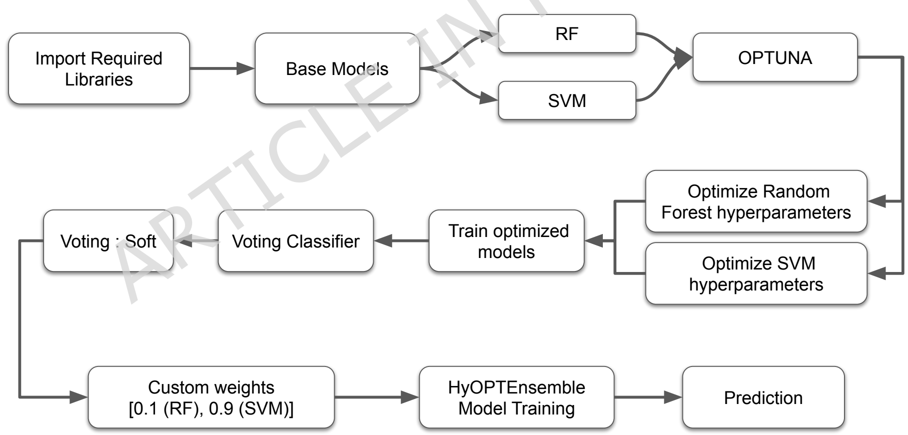

Fig. 7: The Proposed Ensemble Model with OPTUNA (HyOPTEnsemble)

3.8.2. OPTUNA

Optuna is a software framework specifically designed for the purpose of hyperparameter optimization in machine learning (Akiba et al., 2019). Its primary objective is to identify the optimal set of hyperparameter values through a series of repeated trials. In this context, the terms ”study” and ”trials” are commonly employed. A study refers to an optimization process centered on an objective function, while a trial refers to a single instance of executing the objective function Fig.8. However, the direction for optimization is determined to be maximum for each base model.

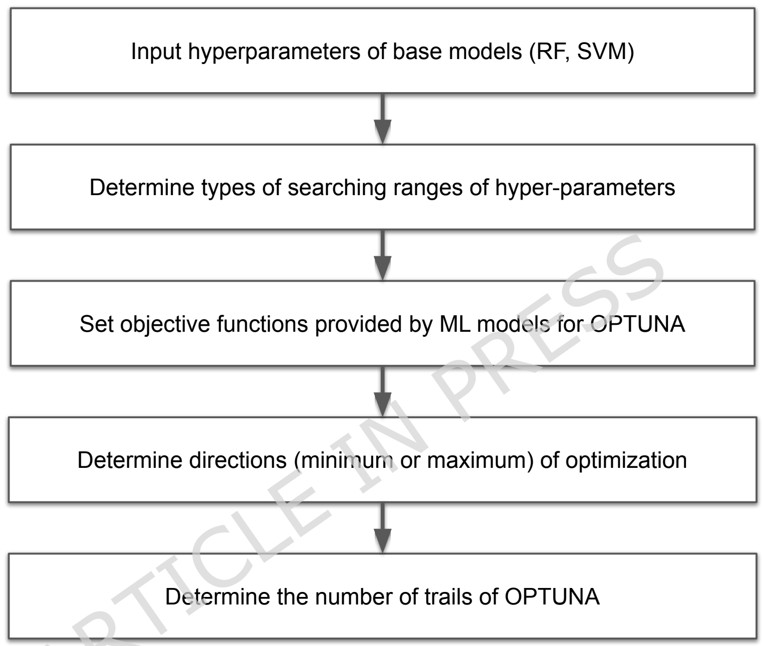

Fig. 8: Workflow of OPTUNA employed in this study

3.8.3. Explainable AI (XAI)

Explainable artificial intelligence (XAI) encompasses a variety of procedures and techniques that enable human users to grasp and believe in the outcomes and findings generated by machine learning algorithms (Explainable-AI).

For this purpose, SHAP Shapley Additive Explanations (SHAP) have been introduced. SHAP is a technique employed in machine learning models to interpret the output of any model. It was first utilized in a game theoretic approach. Specifically, Shapley values were derived from game theory, to quantify the contribution of each feature (or variable) to the prediction for every individual feature (Alomari and Andó, 2024). In simpler terms, SHAP values enable us to comprehend how each feature in our data affects the model's prediction. Features with positive SHAP values have a positive impact on the prediction (Fig.9), whereas those with negative values exhibit a negative impact. The magnitude of these values signifies the strength of the effect. SHAP delivers a consistent and objective explanation of how each feature affects the model's prediction, thereby enhancing the transparency and interpretability of machine learning models.

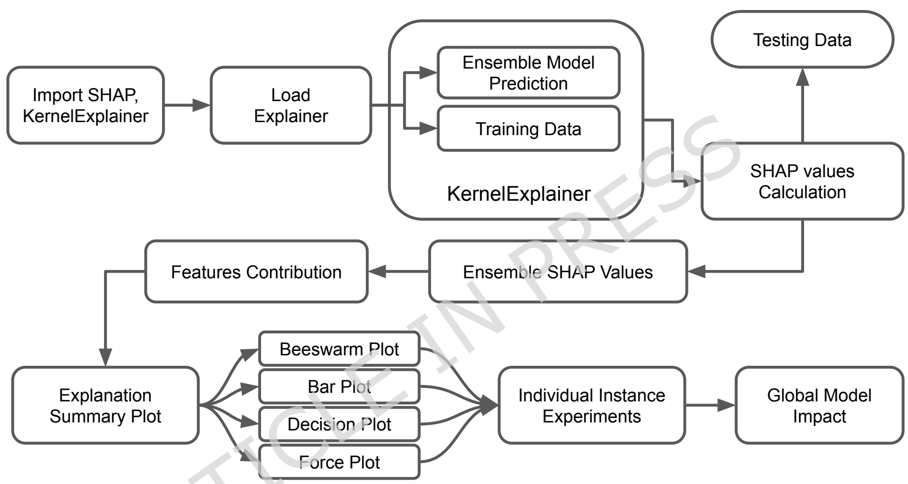

Fig. 9: Steps of implementing SHAP in this study in details

We proposed a HyOPTEnsemble model or algorithm, integrating the strengths of two base models, Random Forest and Support Vector Machine. To provide interpretable explanations for individual predictions, we computed SHAP (Shapely Additive explanations) values, which offered insights into the decision-making process of the ensemble model (SHAP).

3.9. Evaluation Metrics

3.9.1. Accuracy

Accuracy is a fundamental metric used in various fields to evaluate the performance of a predictive model or measurement system (Serrano-Guerrero et al., 2024).

It quantifies the degree to which the model's predictions or measurements align with the true or observed values. In the context of classification tasks, accuracy measures the proportion of correctly classified instances out of the total number of instances. Mathematically, accuracy is defined as (Rainio et al., 2024; Ahmed et al., 2024): **added citation (Ahmed 2024)**

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}} \tag{8}$$

Where, TP (True Positive), TN (True Negative), FP (False Positive), FN (False Negative).

3.9.2. Precision

Precision is a fundamental metric in the fields of statistics and machine learning, commonly used to assess the accuracy of a classification or prediction model. It quantifies the ability of a model to correctly identify positive instances out of the total instances it predicts as positive. Precision is particularly crucial in scenarios where false positives are costly or undesirable, such as medical diagnoses or fraud detection. Precision is calculated using the following equation (Pellegrino et al., 2021):

$$\text{Precision} = \frac{\text{TP}}{\text{TP + FP}} \tag{9}$$

Where, TP (True Positive), FP (False Positive).

3.9.3. Recall

Recall is a fundamental performance metric used in machine learning and information retrieval to assess the effectiveness of a model or system in correctly identifying relevant instances from a data set (Hossain et al., 2024). It is particularly important in scenarios where the goal is to minimize false negatives, such as in medical diagnoses or search engines. Recall denoted as R, is defined as the ratio of true positives (TP) to the sum of true positives and false negatives (FN). Recall is calculated using the following equation (Mahmud et al., 2024):

$$\text{Recall} = \frac{\text{TP}}{\text{TP + FN}} \tag{10}$$

Where, TP (True Positive), (False Negative).

3.9.4. F1-Score

The F1-Score, also known as the F-Measure, is a widely used metric in machine learning and information retrieval to evaluate the performance of binary classification models. It combines both precision and recall into a single measure and is particularly useful when dealing with imbalanced datasets where one class significantly outweighs the other. The F1-Score is defined as the harmonic mean of precision (P) and recall (R):

$$F1\text{-score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{11}$$

According to the confusion matrix of this study in Fig.10, the confusion matrix for the testing data is of paramount importance. Particularly noteworthy are the True Positive (TP) values, which reflect the instances where the model accurately predicted each class. For classes 0, 1, and 2, the TP values were 20, 1, and 21, respectively. These values demonstrate the model's capacity to accurately identify positive instances. Additionally, False Positives (FP), which are instances where the model incorrectly predicts a positive class, are also of great significance. In this instance, there was one instance from Class 2 that was incorrectly predicted as Class 0. Equally important are False Negatives (FN), which are instances where the model fails to predict a positive class. Fortunately, there were no FN instances, indicating that the model was able to correctly identify all positive instances.

3.9.5. Cohen's Kappa Coefficient

Cohen's kappa statistic is a measure of agreement between two raters and is mostly used to test inter-rater reliability and most correlation statistics, kappa can range from (-1) to (+1) (cohens-kappa). The calculation was considered based on the confusion matrix in this study. The equation for the Cohen's kappa coefficient is as follows (Tan et al., 2024):

$$\begin{aligned} p_o &= \frac{\text{number of agreements}}{\text{total number of instances}} \\ &= \frac{12 + 25 + 3}{12 + 25 + 3 + 1 + 1} \\ &= \frac{40}{42} \end{aligned} \tag{12}$$

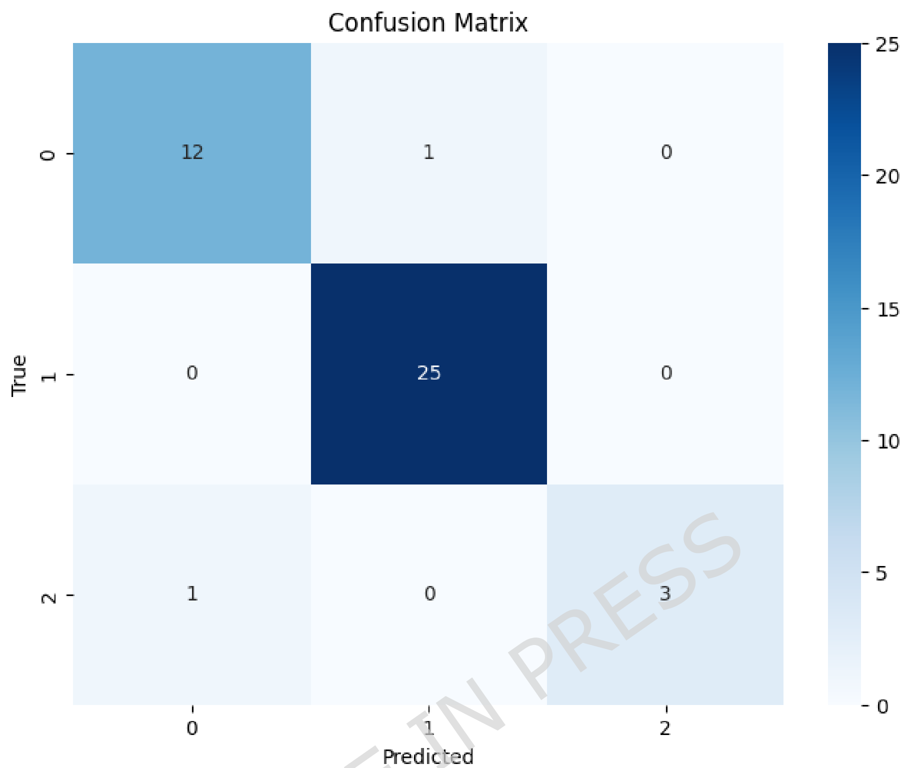

Fig. 10: The confusion matrix of the Ensemble Model of this study

$$\approx 0.95238$$

$$p_e = \sum_{i=1}^{k} p_{i,\text{actual}} \times p_{i,\text{predicted}} \tag{13}$$

$$= \frac{(12+1)\times(12+1)+(1+25)\times(1+25)+(3+1)\times(3+1)}{42^2}$$

$$= \frac{13\times 13+26\times 26+4\times 4}{42^2}$$

$$= \frac{169+676+16}{1764}$$

$$= \frac{861}{1764}$$

$$\approx 0.4881$$

$$\begin{aligned}\kappa &= \frac{p_o - p_e}{1 - p_e} \\ &= \frac{0.95238 - 0.4881}{1 - 0.4881} \\ &\approx 0.9048\end{aligned}$$

Where, (< 0) = Poor, (0.01 - 0.20) = Slight, (0.21-0.40) = Fair, (0.41-0.60) = Moderate, (0.61-0.80) = Substantial, (0.81 - 1.00) = Almost perfect. However, Cohen's kappa score of this study indicates almost perfect agreement, which is approximately 0.9048.

3.10. Matthews Correlation Coefficient

The Matthews correlation coefficient is widely employed in machine learning as a measure of multiclass classification quality (matthews-corr-coef). This measure considers both true and false positives and negatives, rendering it a balanced and suitable choice regardless of class size disparities. In this research, Matthews correlation coefficient (MCC) score of 0.911198, represents the model's predictive precision and reliability, and testament to its exceptional performance. This metric, which ranges from (-1) to (+1), demonstrates a strong positive correlation between the model's predictions and the actual labels. The MCC score of 0.911198 is particularly noteworthy as it is significantly higher than the value of 0.5, which indicates the model's ability to make accurate classifications. The results of this experiment provide compelling evidence of the model's effectiveness. Furthermore, Matthews' correlation coefficient (MCC), mathematically presents as (Rainio et al., 2024).

$$\text{MCC} = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP + FP) \cdot (TP + FN) \cdot (TN + FP) \cdot (TN + FN)}} \in [-1, 1] \tag{14}$$

Where, TP (True Positive), TN (True Negative), FP (False Positive), FN (False Negative).

Furthermore, the algorithm of the proposed model is mentioned in algorithms part 1 and part 2.

Algorithm 1 Pseudocode Algorithm for HyOPTEnsemble Model (Part 1)

```
procedure OBJECTIVE_RF(trial)
    rf_model ← RandomForestClassifier(
        n_estimators = trial.suggest_int(’n_estimators’, 50, 200 , 300),
        max_depth = trial.suggest_int(’max_depth’, 5, 20) )
    rf_model.fit(X_train, y_train)
    rf_predictions ← rf_model.predict(X_test)
    accuracy ← accuracy_score(y_test, rf_predictions)
    return accuracy
end procedure
procedure OBJECTIVE_SVM(trial)
    svm_model ← SVC(
        C = trial.suggest_loguniform(’C’, 0.001, 1000),
        kernel = trial.suggest_categorical(’kernel’, [’linear’, ’rbf’]) )
    svm_model.fit(X_train, y_train)
    svm_predictions = svm_model.predict(X_test)
    accuracy ← accuracy_score(y_test, svm_predictions)
    return accuracy
end procedure
Optimize Random Forest hyperparameters:
study_rf = optuna.create_study(direction=’maximize’)
study_rf.optimize(objective_rf, n_trials=100)
best_params_rf ← study_rf.best_params
Optimize SVM hyperparameters:
study_svm = optuna.create_study(direction=’maximize’)
study_svm.optimize(objective_svm, n_trials=100)
best_params_svm ← study_svm.best_params
Train optimized models:
optimized_rf_model ← RandomForestClassifier(**best_params_rf)
optimized_svm_model ← SVC(probability=True, **best_params_svm)
optimized_rf_model.fit(X_train, y_train)
optimized_svm_model.fit(X_train, y_train)
custom_weights = [0.1 for RF, 0.9 for SVM]
```

Algorithm 2 Pseudocode Algorithm for HyOPTEnsemble Model (Part 2)

```
Create an ensemble model using the optimized models:
ensemble_model = VotingClassifier(
    estimators=[('rf', optimized_rf_model),
    ('svm', optimized_svm_model)],
    voting='soft',
    weights=custom_weights)
ensemble_model.fit(X_train, y_train)
ensemble_predictions ← ensemble_model.predict(X_test)
ensemble_accuracy ← accuracy_score(y_test, ensemble_predictions)
Result:
result ← classification_report(y_test, ensemble_predictions)
print(result)
```


3.11. Experimental Setup

The following applications were used for this study:

- Google Form: An application of Google is utilized to survey any events (google-form).
- Google Colab: Colab is supported by Google for the research community. Python is the only programming language used in this application through a web browser (google-colab).

In addition to the above application, the following Python libraries were used:

- NumPy: An effective Python package for manipulating matrices and multidimensional arrays (numpy).
- Matplotlib: This cross-platform library is typically used in conjunction with NumPy to facilitate data visualization and graph plotting (matplotlib).
- Pandas: A software library was used to process and analyze the data in tabular form (Pandas).

- Scikit-learn: This powerful software library provides a large range of tools for applying machine learning algorithms, including algorithms for regression and classification (scikit-learn).

4. Results and Discussion

The study population comprised 400 students, including 67.1% (268) male, 31.7% (127) female, and only 1.2% (5) students (Fig.11) who preferred not to reveal their identity from 50 universities in Bangladesh (Fig.12). However, The highest number of 123 (31%) male students have moderate stress in comparison to female students 57(14%), and also male students have higher stress compared to female students (Fig.13). The collected dataset for this study was imbalanced which causes model overfitting and underfitting. Initially this study applied SMOTE-ENN following after data partition for training and testing.Although the result remains overfitted. After that, this study employed SMOTE-ENN initially to balance the dataset before train test and which interestingly helps to model to fit the model and improve the performance.

Our investigation involved employing various machine learning (ML) classification models, including decision tree, Naïve Bayes, KNN, SVM, AdaBoost, Bagging, Random Forest, grid hyper-tuned ensemble model, a proposed HyOPTEnsemble model to predict the mental state. Table 3 presents the outcomes of our investigation, which indicate that the proposed HyOPTEnsemble model demonstrated superior performance across several metrics, including accuracy, precision, recall, F1 score, CV score (cv=5), Cohen's Kappa score, and MCC, surpassing the other models. Furthermore, Table 4 illustrates the results of prior studies focused on mental health prediction, revealing that our proposed model achieved higher evaluation metrics compared to previous studies.

The five-fold cross-validation is a procedure in which data is randomly partitioned into five folds, with the model being trained on the four folds and the remaining fold reserved for testing. This process is known as k-fold cross-validation, where k represents the number of folds. This method is commonly used to evaluate the performance of machine learning models. The results of our study showed that the mean scores of proposed model performance metrics were remarkably higher (cv score = 0.952) than individual models of this study and previous studies (Fig.14).

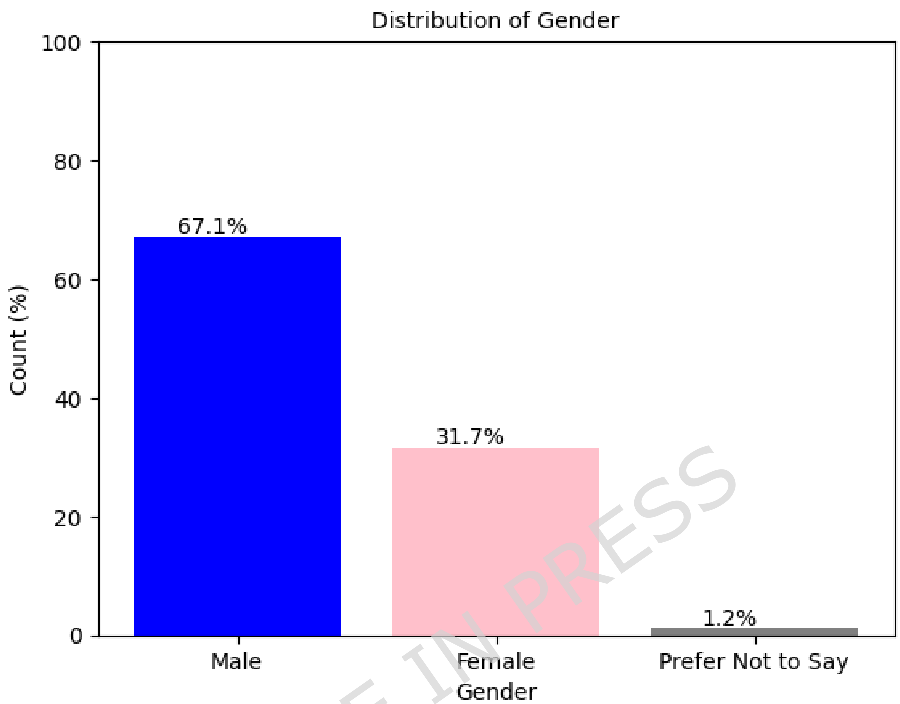

Fig. 11: Gender ratio of male and female students

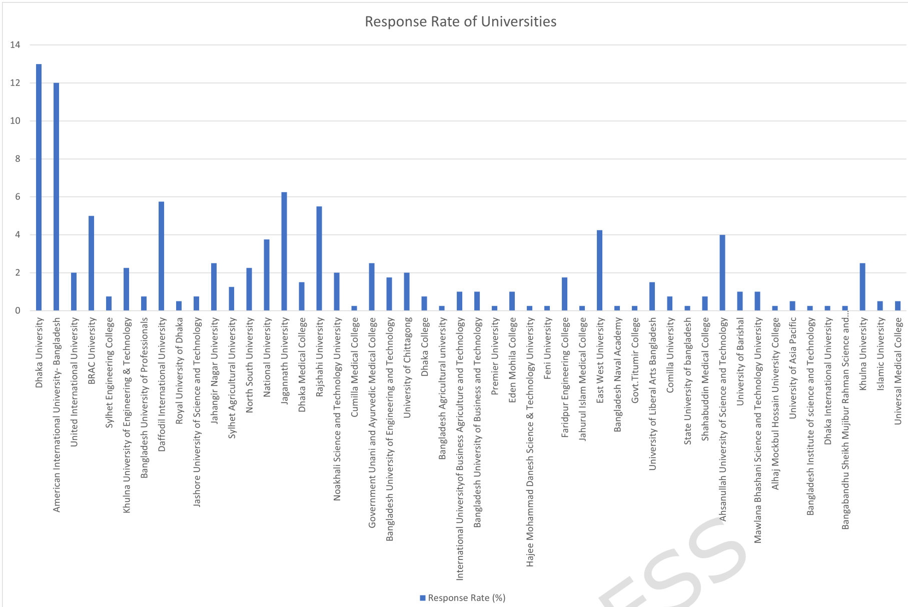

Fig. 12: Response ratios of all university students

Table 3: Performance Comparison of Proposed Hyper-tuned Ensemble with individual ML model of this study

Model (sample=400)	Test Accuracy	Precision	Recall	F1-score	CV Score	kappa	MCC
DT	0.833	0.823	0.833	0.827	0.909	0.689	0.689
KNN	0.881	0.799	0.881	0.838	0.915	0.765	0.778
SVM	0.905	0.912	0.905	0.885	0.951	0.811	0.822
NB	0.857	0.777	0.857	0.815	0.897	0.718	0.730
AdaBoost	0.810	0.734	0.810	0.770	0.836	0.624	0.635
Bagging	0.881	0.876	0.881	0.877	0.915	0.778	0.779
Random Forest	0.857	0.869	0.857	0.840	0.939	0.725	0.731
Ensemble (RF+SVM) with Grid	0.857	0.869	0.857	0.840	0.934	0.724	0.731
Proposed HyOPTEnsemble Model	**0.952**	**0.953**	**0.952**	**0.951**	**0.952**	**0.909**	**0.911**

The proposed model's performance was further validated through 5-fold cross-validation, which confirmed the robustness of the algorithm or model. The cross-validation scores, ranging from 0.909 to 1.0, reflect the robust performance of the ensemble model across diverse folds (Fig.14). Notably, the mean cross-validation score of 0.952 underscores the overall effectiveness of the model in generalizing to unseen

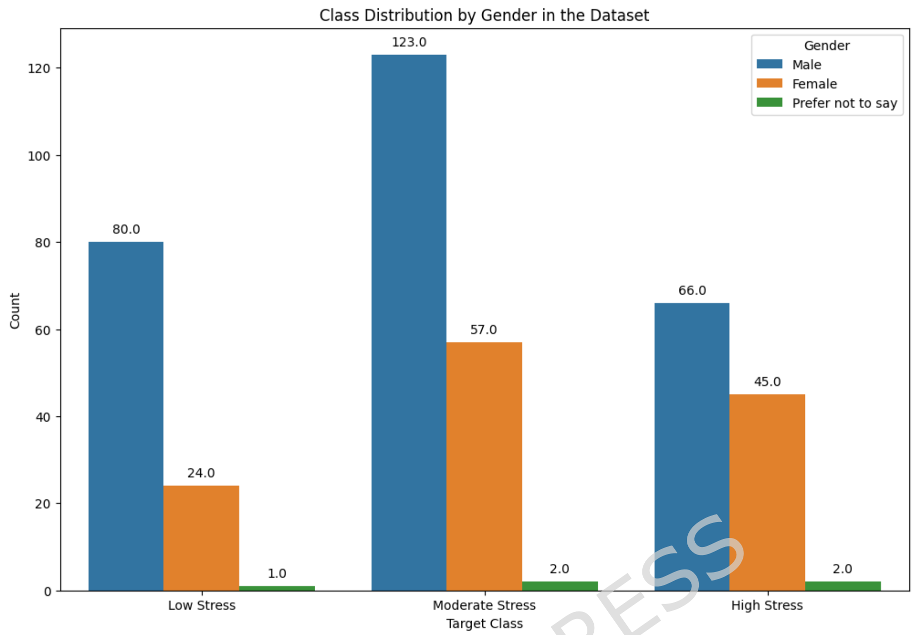

Fig. 13: Stress levels of gender via analyzing the dataset(n=400)

Table 4: Performance Comparison of Previous Studies

Study/Work	Model	Sample Size	Test Accuracy	Precision	Recall	F1-score
(Sahu and Debbarma, 2023)	SVM	-	0.755	-	-	-
(Mutalib et al., 2021)	Linear Regression, Neural Networks	-	0.68, 0.88	-	-	-
(Sahlan et al., 2021)	Decision Tree	219	0.64	-	-	0.61
(Ogunseye et al., 2022)	AdaBoost	-	0.8175	-	-	-
(Vaishnavi et al., 2022)	Stacking	-	0.8175	-	-	-
(Abdul Rahman et al., 2023)	Random Forest	-	0.921	-	-	-
(Bhatnagar et al., 2023)	Random Forest	127	0.789	-	-	-
(Rahman et al., 2020)	Random Forest	-	0.96	-	-	-
(Kim et al., 2020)	CNN	-	0.9696	-	-	-
(Chatterjee et al., 2023)	SVM	672	0.89	-	-	0.88
(Iyortsuun et al., 2023)	Random Forest	-	0.88	-	-	-

data. The consistently high individual scores affirm the model's proficiency in capturing varied patterns. This ensemble approach demonstrates reliability and consistency, showcasing its potential for real-world applications. The results validate the ensemble model's capacity to achieve notable predictive accuracy and underscore its suitability for tasks requiring robust generalization.

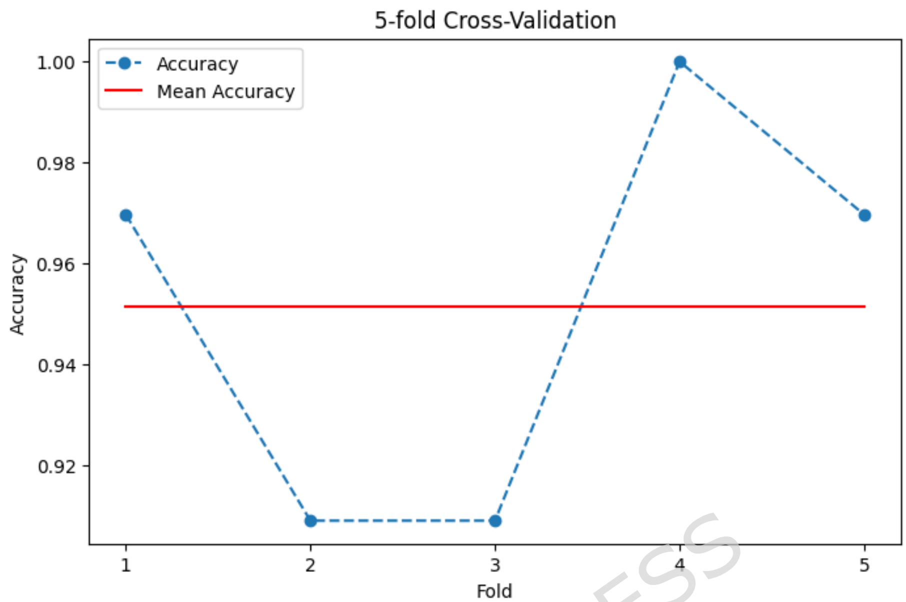

Fig. 14: 5-Fold Cross Validation

The proposed HyOPTEnsemble model has showcased impressive performance on both the training and testing data. During the training phase shown in Fig.15, the model achieved perfect accuracy, correctly identifying all instances across all classes. This demonstrates the model's ability to effectively capture underlying patterns. On unseen data, the model maintained strong performance, achieving high precision for classes 0 and 1. However, the minor misclassification from class 2 to class 0 warrants further investigation. Overall, our approach exhibits promising generalization capabilities for predicting university students' mental states

The proposed model was thoroughly examined across 50 epochs, revealing a consistent and outstanding performance pattern (Fig.16). Notably, the training accuracy consistently maintained a pristine 100%, while the testing accuracy score fluctuated from 0.88 to 0.95.8. Surprisingly, this study achieved an impressive accuracy score of 95.2%, throughout the entire epoch span. This robust and persistent accuracy profile underscores the model's efficacy and generalization capability, confirming its reliability in various scenarios.

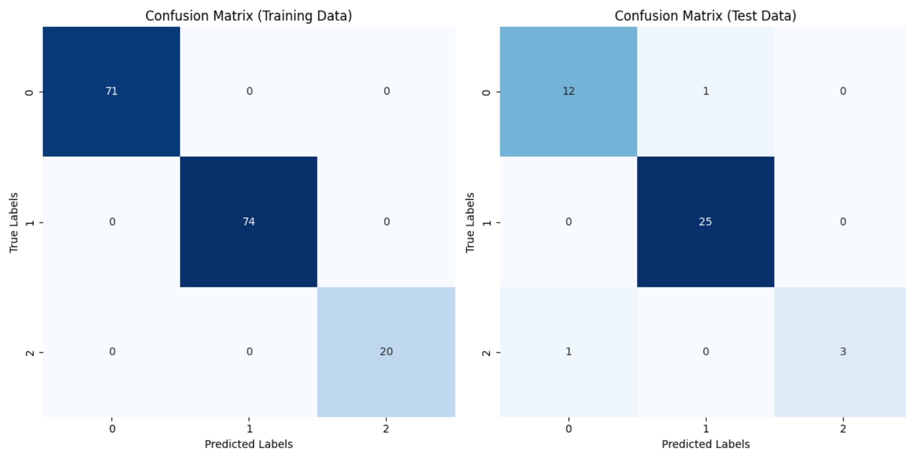

Fig. 15: Confusion Matrix of train test data

Across the 50 meticulously conducted epochs, the proposed HyOPTEnsemble model displayed remarkable stability and proficiency (Fig.17). Training loss consistently maintained an exemplary 0.05% with only a small fluctuation, attesting to the model's aptitude for capturing intricate patterns within the training data. Moreover, the testing loss remained steadfast at a commendable 0.25%, exhibiting the model's robust generalization capacity. This consistent performance highlights the proposed model's ability to navigate complexity, achieving a balanced approach between the training and testing phases. The minimal deviation between the training and testing losses throughout the epochs underscores the resilience of the model, instilling a sense of confidence in its dependability.

The graph (Fig.18) depicting the learning curve reveals the HyOPTEnsemble model's competence and adaptability. The Training scores, ranging from 0.92 to 1.0, indicate robust learning from the dataset, while the Cross-Validation scores, ranging from 0.52 to 0.96, showcase the model's ability to generalize effectively. The convergence of these curves at high scores suggests a well-balanced model that is capable of capturing complexity without overfitting. This cohesive performance underscores the proposed model's suitability for practical applications, demonstrating a commendable balance between training and generalization across varying complexities.

Furthermore, the performance of the model was evaluated using the area under

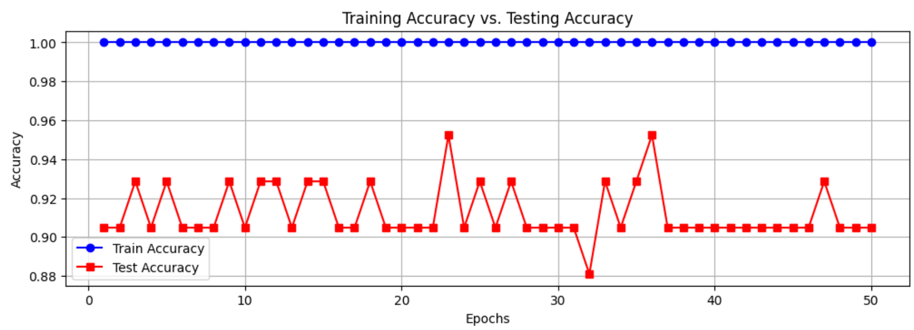

Fig. 16: Train Test Accuracy

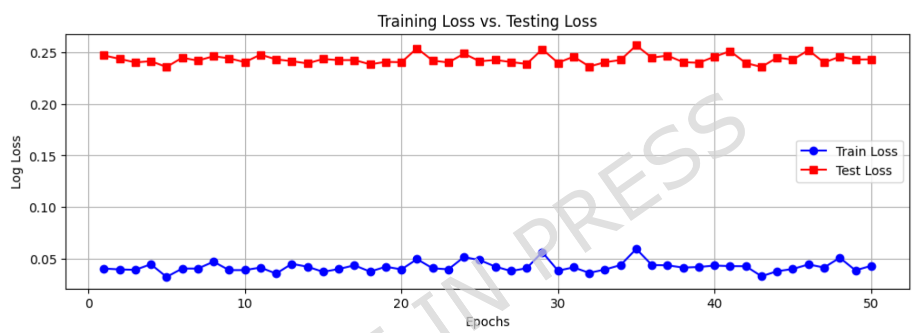

Fig. 17: Train Test Loss

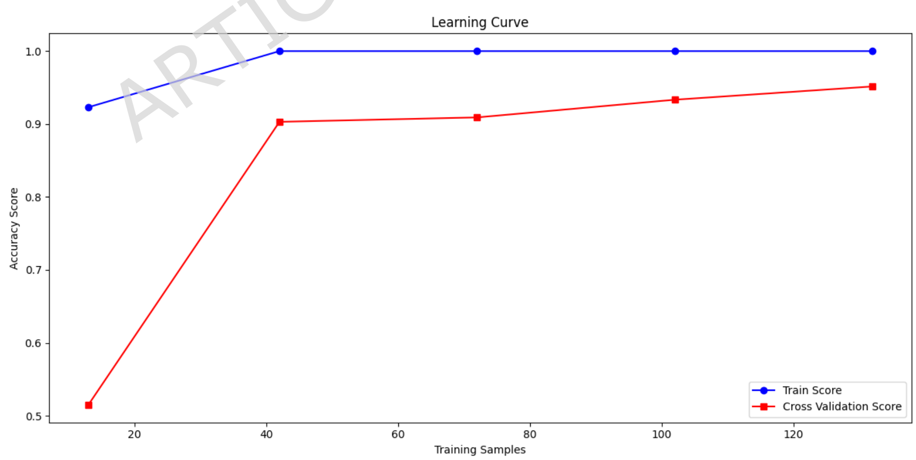

Fig. 18: Learning Curve

the receiver operating characteristic (ROC) curve shown in (Fig.19), which measures the model's ability to discriminate between positive and negative instances. Class 0 achieved an AUC of 0.98, indicating strong discrimination. Class 1 demonstrated an AUC of 0.99, indicating high discriminatory power. Class 2 exhibited an excellent AUC of 0.97, demonstrating the ability of the model to distinguish between positive and negative instances.

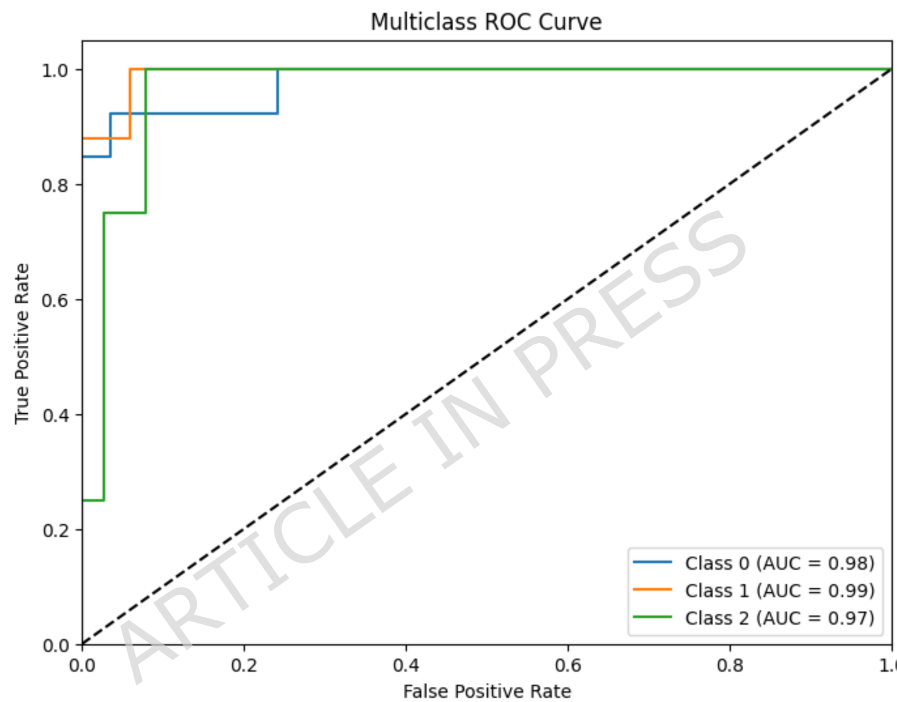

Fig. 19: ROC curve graphically presented for multiclass classification of this study.

The combination of two base models, presented as the ensemble model with feature selection and hyperparameter optimization method, has exhibited exceptional performance. It has attained a notable level of accuracy and has proven to be dependable in a diverse range of circumstances, as evidenced in Table 4. In order to comprehend the rationale behind the predictions generated by this ensemble model, we employed interpretable artificial intelligence frameworks referred to as SHAP (SHAP). SHAP values aid in our comprehension of the extent to which each feature (or variable) contributes to the prediction for each individual instance. SHAP summary and Decision plot, as depicted in Fig.20 and Fig.21, furnish us with a comprehensive perspective on the contribution of each feature Fig.22 towards the predictions. This contributes significantly towards enhancing our comprehension of the inner workings of the ensemble model. According to SHAP values in this study, worrying too much about different things, being depressed, feeling down or hopeless, feeling tired or having little energy, and being restless students are affected by high stress.

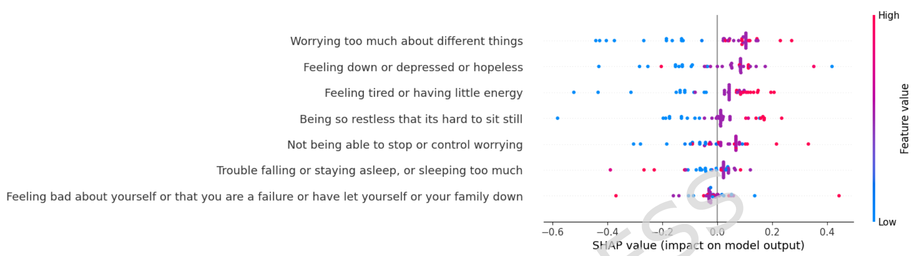

Fig. 20: SHAP summary plots for interpreting the proposed ensemble model

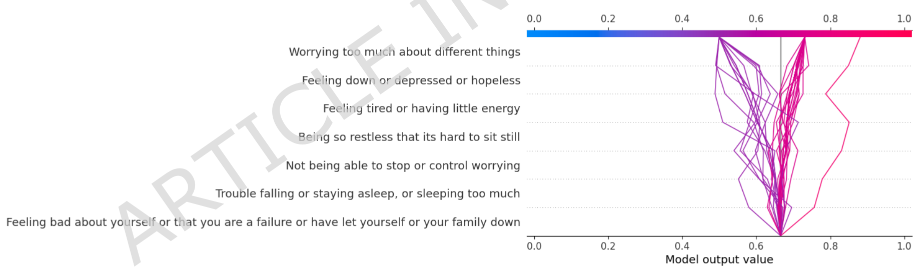

Fig. 21: Decision Plot to interpret the prediction result

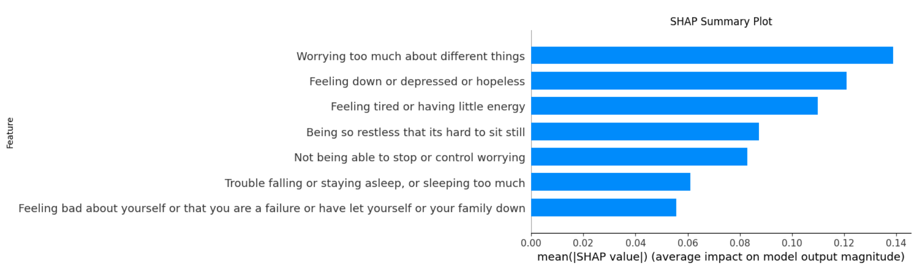

Fig. 22: SHAP summary bar plots for interpreting the proposed ensemble model

In Fig.23, this plot permits the user to visually evaluate the individual contributions of each feature towards the predicted outcome for a specific instance. The level of detail provides offers a comprehensive understanding of a particular case. Red indicators represent features that contributed to a higher model score, while blue indicators represent features that contributed to a lower score. The size of the arrow indicates the extent of the feature's impact on the output. The x-axis displays the degree of increase or decrease in the impacts.

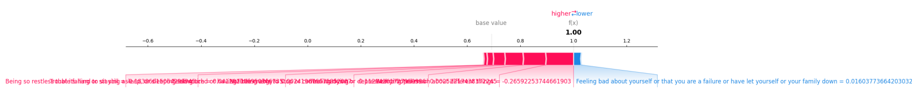

Fig. 23: Force Plot to interpret the prediction result

By overcoming the challenges of imbalanced data, the performance of the proposed HyOPTEnsemble model, which integrates Random Forest (RF) and Support Vector Machine (SVM) algorithms, is notably high. The combination of these two techniques with optimal hyperparameter optimization and the custom-weighted soft voting process allows the proposed model to capitalize on the strengths of each, leading to enhanced predictive capabilities. The use of SHapley Additive exPlanations (SHAP) provided invaluable information about the model's predictions. The SHAP values revealed that 'Worrying too much', 'Feeling down, depressed, hopeless', 'Feeling tired, having little energy', 'Being restless' had the most significant influence on predicting mental states, followed by 'Not being able to stop or control worrying', 'Trouble falling or staying asleep, sleeping too much' and 'Feeling bad about Failure, let yourself or family down'. These findings suggest that these features may serve as valuable indicators of mental states and could be targeted in early intervention strategies. Furthermore, This study highlights the potential of machine learning, particularly the HyOPTEnsemble model, in predicting states. The implications of these findings could be substantial for the early prediction and treatment of mental disorders.

5. Conclusions

This study proposed and demonstrated the usefulness of a custom-weighted soft voting hyperparameter optimized ensemble machine learning algorithm for predicting the mental state of university students by utilizing categorical online survey data. The proposed model was developed, which combined random forest (RF) and support vector machine (SVM), to predict the earlier mental state. We considered a custom-weighted soft-voting technique hyperparameter optimized ensemble machine model to predict mental state, which achieved higher prediction results than individual Machine Learning classifier models and exhibited exceptional performance as measured by the Matthews Correlation Coefficient (MCC) and Cohens Kappa Score. One of the primary obstacles in this research was the initial disparity in the collected data. To confront this issue, we utilized the Synthetic Minority Over-sampling Technique - Edited Nearest Neighbours (SMOTE-ENN) method, which not only rectified the imbalance in the data but also improved the model's generalizability by preventing it from being biased towards the majority class. The HyOPTEnsemble model, when enhanced with SHapley Additive exPlanations (SHAP) values, has demonstrated not only robust predictive performance but also valuable insights into the complex dynamics of feature contributions. The feature contribution helped to determine and extract the best optimal features for this study. The utilization of SHAP provided critical insights into the model's predictions, thereby enhancing the model's interpretability. This study emphasizes the potential of machine learning, particularly ensemble learning, in healthcare applications and highlights the importance of model interpretability. This study found that 'depressed, hopeless', and 'worrying too much', students are highly associated with mental stress. This work also contributes to the growing body of research that highlights the synergy between ensemble models and advanced interpretability techniques. Although this research is limited to 400 samples for predicting the university student mental health. Despite the limited sample size (n=400), the proposed model outperformed the existing metrics on university student mental health prediction especially on Bangladeshi university students. Moreover, The experimental performance revealed that our proposed model remarkably predicted university students' mental state which will help the student counselors to predict earlier and assist the students before any unsatisfactory outcome occurs. Although we have shown the performance of the proposed ensemble model with rich experiments, there are still ways to improve in this research. Firstly, future research could focus to enhance the dataset size and secondly, examine additional machine learning models and compare their results with the ensemble model such as XGBoost, LightGBM, and deep learning, transfer learning as well as incorporate more extensive data sources for comprehensive analysis. The findings of this study have significant implications for the early identification and treatment of mental health disorders, ultimately contributing to improved healthcare outcomes.

6. Declaration

· Ethics approval and consent to participate: NA

· Consent for publication: NA

· Availability of data and material: Data Will be available on request.

· Competing interests: No Competing Interest

· Funding: NA

· Authors' contributions:Rasel Ahmed: Writing Original Draft, Writing , reviewing and editing final draft, data curation, methodology design, experiment, visualization; Nafiz Fahad: Writing Original Draft, Writing , reviewing and editing final draft, data curation; Md Saef Ullah Miah: Supervision, Writing , reviewing and editing final draft, data curation;Md. Jakir Hossen: Supervision, Writing Original Draft, Writing , reviewing and editing final draft, data curation; Kanta Bhattacharjee: Writing , reviewing and editing final draft.

· Acknowledgements:The authors want to thank Multimedia University, Jalan Ayer Keroh Lama 75450, Bukit Beruang, Melaka

References

Abdul Rahman, H., Kwicklis, M., Ottom, M., Amornsriwatanakul, A., H. Abdul-Mumin, K., Rosenberg, M., Dinov, I.D., 2023. Machine Learning-Based Prediction of Mental Well-Being Using Health Behavior Data from University Students. Bioengineering 10, 575. URL: https://www.mdpi.com/2306-5354/10/5/575, doi:10.3390/bioengineering10050575.

Ahmed, R., Fahad, N., Miah, M.S.U., Hossen, M.J., Morol, M.K., Mahmud, M., Rahman, M.M., 2024. A novel integrated logistic regression model enhanced with recursive feature elimination and explainable artificial intelligence for dementia prediction. Healthcare Analytics 6, 100362.

Akiba, T., Sano, S., Yanase, T., Ohta, T., Koyama, M., 2019. Optuna: A next-generation hyperparameter optimization framework, in: Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining, pp. 2623–2631.

Aleem, S., Huda, N.u., Amin, R., Khalid, S., Alshamrani, S.S., Alshehri, A., 2022. Machine learning algorithms for depression: diagnosis, insights, and research directions. Electronics 11, 1111.

Alomari, Y., Andó, M., 2024. Shap-based insights for aerospace phm: Temporal feature importance, dependencies, robustness, and interaction analysis. Results in Engineering 21, 101834.

Baba, A., Bunji, K., 2023. Prediction of mental health problem using annual student health survey: Machine learning approach. JMIR Mental Health 10, e42420.

BaggingClassifier, . sklearn.ensemble.BaggingClassifier. URL: https://scikit-learn/stable/modules/generated/sklearn.ensemble.BaggingClassifier.html.

Bhatnagar, S., Agarwal, J., Sharma, O.R., 2023. Detection and classification of anxiety in university students through the application of machine learning. Procedia Computer Science 218, 1542–1550.

Breslau, J., Roth, E.A., Baird, M.D., Carman, K.G., Collins, R.L., 2023. A longitudinal study of predictors of serious psychological distress during covid-19 pandemic. Psychological medicine 53, 2418–2426.

Chatterjee, M., Kumar, P., Sarkar, D., 2023. Generating a mental health curve for monitoring depression in real time by incorporating multimodal feature analysis through social media interactions. International Journal of Intelligent Information Technologies (IJIIT) 19, 1–25.

Chatterjee, R., Datta, A., Sanyal, D.K., 2019. Ensemble learning approach to motor imagery eeg signal classification, in: Machine Learning in Bio-Signal Analysis and Diagnostic Imaging. Elsevier, pp. 183–208.

cohens-kappa, 2023. Kappa Statistics - an overview | ScienceDirect Topics. URL: https://www.sciencedirect.com/topics/medicine-and-dentistry/kappa-statistics.

Collaborators, G..M.D., et al., 2022. Global, regional, and national burden of 12 mental disorders in 204 countries and territories, 1990–2019: a systematic analysis for the global burden of disease study 2019. The Lancet Psychiatry 9, 137–150.

Deng, S., Wang, F., Cai, Y., Wang, H., Wang, Z., Qian, Q., Ding, W., 2024. Prediction and analysis of multiple causes of mental health problems based on machine learning, in: International Conference on Information, Springer. pp. 150–160.

Explainable-AI, . What is Explainable AI (XAI)? | IBM. URL: https://www.ibm.com/topics/explainable-ai.

Fatima, A., Li, Y., Hills, T.T., Stella, M., 2021. Dasentimental: Detecting depression, anxiety, and stress in texts via emotional recall, cognitive networks, and machine learning. Big Data and Cognitive Computing 5, 77.

google-colab, 2023. Google Colaboratory. URL: https://colab.research.google.com/.

google-form, 2023. Google Forms. URL: https://docs.google.com/.

Hall, R., 2023. University students more at risk of depression than non-students – study. The Guardian URL: https://www.theguardian.com/education/2023

/sep/29/university-students-more-at-risk-of-depression-than-non-students-study.

Hancock, J.T., Khoshgoftaar, T.M., 2020. Survey on categorical data for neural networks. Journal of big data 7, 28.

Hancock Jr, M.F., 2015. Data mining: Process. Encyclopedia of Information Systems and Technology-Two Volume Set , 396.

Hossain, M.M., Tasnim, S., Sultana, A., Faizah, F., Mazumder, H., Zou, L., McKyer, E.L.J., Ahmed, H.U., Ma, P., 2020. Epidemiology of mental health problems in covid-19: a review. F1000Research 9.

Hossain, M.N., Fahad, N., Ahmed, R., Sen, A., Al Huda, M.S., Hossen, M.I., 2024. Preventing student's mental health problems with the help of data mining. International Journal of Computing 23, 101–108. URL: https://www.computingonline.net/computing/article/view/3441, doi:10.47839/ijc.23.1.3441.

Islam, M.A., Barna, S.D., Raihan, H., Khan, M.N.A., Hossain, M.T., 2020. Depression and anxiety among university students during the covid-19 pandemic in bangladesh: A web-based cross-sectional survey. PloS one 15, e0238162.

Iyortsuun, N.K., Kim, S.H., Jhon, M., Yang, H.J., Pant, S., 2023. A review of machine learning and deep learning approaches on mental health diagnosis, in: Healthcare, MDPI. p. 285.

Kendra Cherry, M., 2024. How different branches of psychology study the brain and behavior. URL: https://www.verywellmind.com/major-branches-of-psychology-4139786.

Kim, J., Lee, J., Park, E., Han, J., 2020. A deep learning model for detecting mental illness from user content on social media. Scientific reports 10, 11846.

KNN, 2024. What is the k-nearest neighbors algorithm? | IBM. URL: https://www.ibm.com/topics/knn.

Lamari, M., Azizi, N., Hammami, N.E., Boukhamla, A., Cheriguene, S., Dendani, N., Benzebouchi, N.E., 2021. Smote–enn-based data sampling and improved dynamic ensemble selection for imbalanced medical data classification, in: Advances on Smart and Soft Computing: Proceedings of ICACIn 2020, Springer. pp. 37–49.

Linlin, W., Wanyu, H., Yuting, L., Huimin, Q., Zhi, L., Qinchen, J., Tingting, W., Fan, W., Minghao, P., Wei, Z., 2023. Research on the mechanism of short video information interaction behavior of college students with psychological disorders based on grounded theory. BMC Public Health 23, 2256.

Liu, Y., Wang, F., 2025. Investigating the interpretability of chatgpt in mental health counseling: An analysis of ai-generated content differentiation. Computer Methods and Programs in Biomedicine , 108864.

Mahmud, S.H., Goh, K.O.M., Hosen, M.F., Nandi, D., Shoombuatong, W., 2024. Deep-wet: a deep learning-based approach for predicting dna-binding proteins using word embedding techniques with weighted features. Scientific reports 14, 2961.

Mantas, J., et al., 2023. A machine learning study to predict anxiety on campuses in lebanon. Healthcare Transformation with Informatics and Artificial Intelligence 305, 85.

matplotlib, 2023. Matplotlib — Visualization with Python. URL: https://matplotlib.org/.

matthews-corr-coef, . sklearn.metrics.matthews_corrcoef. URL: https://scikit-learn/stable/modules/generated/sklearn.metrics.matthews_corrcoef.html.

Mutalib, S., et al., 2021. Mental health prediction models using machine learning in higher education institution. Turkish Journal of Computer and Mathematics Education (TURCOMAT) 12, 1782–1792.

numpy, 2023. NumPy. URL: https://numpy.org/.

Ogunseye, E.O., Adenusi, C.A., Nwanakwaugwu, A.C., Ajagbe, S.A., Akinola, S.O., 2022. Predictive analysis of mental health conditions using adaboost algorithm. ParadigmPlus 3, 11–26.

Pandas, 2023. pandas - Python Data Analysis Library. URL: https://pandas.pydata.org/.

Pellegrino, E., Jacques, C., Beaufils, N., Nanni, I., Carlioz, A., Metellus, P., Ouafik, L., 2021. Machine learning random forest for predicting oncosomatic variant ngs analysis. Scientific reports 11, 21820.

Pérez, T., Pardo, M.C., Cabellos, Y., Peressini, M., Ureña-Vacas, I., Serrano, D.R., González-Burgos, E., 2023. Mental health and drug use in college students: Should we take action? Journal of Affective Disorders .

Rahman, R.A., Omar, K., Mohd Noah, S.A., Danuri, M.S.N.M., Al-Garadi, M.A., 2020. Application of Machine Learning Methods in Mental Health Detection: A Systematic Review. IEEE Access 8, 183952–183964. URL: https://ieeexplore.ieee.org/document/9214815/, doi:10.1109/ACCESS.2020.3029154.

Rainio, O., Teuho, J., Klén, R., 2024. Evaluation metrics and statistical tests for machine learning. Scientific Reports 14, 6086. URL: https://www.nature.com/articles/s41598-024-56706-x, doi:10.1038/s41598-024-56706-x.

Random-Forest, 2024. sklearn.ensemble.RandomForestClassifier. URL: https://scikit-learn/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html.

Sahlan, F., Hamidi, F., Misrat, M.Z., Adli, M.H., Wani, S., Gulzar, Y., 2021. Prediction of mental health among university students. International Journal on Perceptive and Cognitive Computing 7, 85–91.

Sahu, S., Debbarma, T., 2023. Mental health prediction among students using machine learning techniques, in: Evolution in Computational Intelligence, Springer Nature Singapore, Singapore. pp. 529–541. doi:10.1007/978-981-19-7513-4_46.

scikit-learn, 2023. scikit-learn: machine learning in Python. URL: https://scikit-learn.org/stable/.

Serrano-Guerrero, J., Alshouha, B., Bani-Doumi, M., Chiclana, F., Romero, F.P., Olivas, J.A., 2024. Combining machine learning algorithms for personality trait prediction. Egyptian Informatics Journal 25, 100439.

SHAP, . Welcome to the SHAP documentation — SHAP latest documentation. URL: https://shap.readthedocs.io/en/latest/.

Solmi, M., Radua, J., Olivola, M., Croce, E., Soardo, L., Salazar de Pablo, G., Il Shin, J., Kirkbride, J.B., Jones, P., Kim, J.H., et al., 2022. Age at onset of mental disorders worldwide: large-scale meta-analysis of 192 epidemiological studies. Molecular psychiatry 27, 281–295.

Szczerbicki, E., 2001. Management of complexity and information flow. Agile manufacturing: The 21st century competitive strategy , 247–263.

Tan, K.S., Yeh, Y.C., Adusumilli, P.S., Travis, W.D., 2024. Quantifying interrater agreement and reliability between thoracic pathologists: Paradoxical behavior of cohen's kappa in the presence of a high prevalence of the histopathologic feature in lung cancer. JTO Clinical and Research Reports 5, 100618. URL: https://www.sciencedirect.com/science/article/pii/S2666364323001613, doi:https://doi.org/10.1016/j.jtocrr.2023.100618.

universitiesuk, 2024. Why mental health should be a priority for universities. URL: https://www.universitiesuk.ac.uk/what-we-do/policy-and-research/publications/features/stepchange-mentally-healthy-universities/why-mental-health-should-be-priority.

Vaishnavi, K., Nikhitha Kamath, U., Ashwath Rao, B., Subba Reddy, N.V., 2022. Predicting Mental Health Illness using Machine Learning Algorithms. Journal of Physics: Conference Series 2161, 012021. URL: https://iopscience.iop.org/article/10.1088/1742-6596/2161/1/012021, doi:10.1088/1742-6596/2161/1/012021.

Valencia-Arias, A., Chalela, S., Cadavid-Orrego, M., Gallegos, A., Benjumea-Arias, M., Rodríguez-Salazar, D.Y., 2023. University Dropout Model for Developing Countries: A Colombian Context Approach. Behavioral Sciences 13, 382. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10215226/, doi:10.3390/bs13050382.

Webb, G.I., 2010. Naïve Bayes. Springer US, Boston, MA. pp. 713–714. URL: https://doi.org/10.1007/978-0-387-30164-8_576, doi:10.1007/978-0-387-30164-8_576.

WHO, 2023. Depression. URL: https://www.who.int/health-topics/depression.

WHO, 2025a. Mental disorder. URL: https://www.who.int/news-room/fact-sheets/detail/mental-disorders.

WHO, 2025b. Mental health. URL: https://www.who.int/health-topics/mental-health.

Yue, S., Li, P., Hao, P., 2003. SVM classification:Its contents and challenges. Applied Mathematics-A Journal of Chinese Universities 18, 332–342. URL: http://link.springer.com/10.1007/s11766-003-0059-5, doi:10.1007/s11766-003-0059-5.

Zebari, R., Abdulazeez, A., Zeebaree, D., Zebari, D., Saeed, J., 2020. A Comprehensive Review of Dimensionality Reduction Techniques for Feature Selection and Feature Extraction. Journal of Applied Science and Technology Trends 1, 56–70. URL: https://www.jastt.org/index.php/jasttpath/article/view/24, doi:10.38094/jastt1224.