INTI JOURNAL | Vol.2025, Issue 4, No.2
eISSN:2600-7320

Classification of Mental Health Level of Students Using SMOTE and Soft Voting Ensemble Classifier and the DASS-21 Profile

Muhammad Imron Rosadi[1*], Khoirun Nisa[1], Nanik Kholifah[2]

[1]Informatics Engineering, Universitas Yudharta Pasuruan, Pasuruan 67162, Indonesia
[2]Psychology, Universitas Yudharta Pasuruan, Pasuruan 67162, Indonesia

[*]**Email:** imron.rosadi@yudharta.ac.id

Abstract

This study proposes a comprehensive approach to address the rise in mental health problems among college students. It leverages the Synthetic Minority Over-sampling Technique (SMOTE) to address the class imbalance in the dataset and employs a Voting Ensemble with soft voting to combine several base algorithms (Logistic Regression, Random Forest, Gradient Boosting, and XGBoost/SVM) for accurate prediction of mental health levels (normal, mild, moderate, severe, very severe). Feature importance-based feature selection using Random Forest is utilized to eliminate less relevant attributes. The model evaluation includes accuracy, precision, recall, F1-score, and confusion matrix analysis. The results demonstrate that the ensemble approach improves stability and accuracy compared to individual models. Notably, the application of SMOTE led to significant performance improvements, with classification accuracies reaching up to 100% for the Random Forest model. These findings support the use of ensemble learning and SMOTE for developing effective college student mental health screening systems, ultimately enabling timely intervention and support.

Keywords

Classification, Feature Importance, Mental Health, SMOTE, Ensemble Classifier, Soft Voting

Introduction

Stress is the body's natural response to certain situations, but chronic stress can cause headaches, sleep issues, and physiological harm (Alić et al., 2016). Student mental health is critical for academic success and quality of life, with over 40% experiencing significant stress and depression symptoms (Battista et al., 2022). making it a key concern in developing countries (Duangchaemkarn et al., 2024).

The DASS-21, developed by Lovibond et al. (1995), is a reliable tool for measuring depression, anxiety, and stress and is widely used in cross-national studies (Ahmed et al., 2022). Machine learning applications on DASS profiles show high accuracy, with SVM models achieving

Submission: 10 July 2025; **Acceptance:** 1 October 2025; **Available online:** October 2025

up to 91% in detecting depression and anxiety (Duangchaemkarn et al., 2024); (Singh & Kumar, 2021).

Class imbalance is a major ML challenge that causes biased models, often addressed with sampling methods such as SMOTE to balance data distribution (Matharaarachchi et al., 2024); (Thabtah et al., 2020). While SMOTE has the advantage of expanding the decision space for minority classes, it has some drawbacks that need to be considered. Therefore, it is important to combine SMOTE with other techniques, such as reducing the number of majority examples, to achieve better balance in the dataset (Chawla et al., 2002). Ensemble learning improves prediction accuracy by combining multiple classifiers, reducing errors from individual models (Rokach, 2010). A voting ensemble has shown better stress detection performance, reaching 78% accuracy in ML and IoT-based systems (Hadhri et al., 2024).

Nevertheless, there is currently little research on creating machine learning-based models from DASS-21 characteristics. The goal of this work is to find cutting-edge machine learning-based models for categorizing students' emotional states of stress, anxiety, and sadness. This study employs an ensemble voting classifier to increase score accuracy and suggests an oversampling technique to handle data imbalance. The study's conclusions will help medical practitioners apply these models in practical settings, enhancing clinical judgment when diagnosing and evaluating mental health issues.

Methodology

In the paper, we employed 8 machine learning classification algorithms to identify patients' stress levels, enabling early intervention to prevent significant harm to their lives before it escalates. Figure 1 illustrates the schematic flow diagram of the stress detection methodology..

A. Data collection and Profilling

215 data were obtained from Yudharta University Pasuruan who had finished the online DASS-21 questionnaire in 2025 in order to create the DASS-21 profile for the machine learning-based prediction model. The DASS-21 questionnaire can download from website https://maic.qld.gov.au. In compliance with the Personal Data Protection Act, the dataset was deidentified from the University to safeguard the privacy of personal information.

B. Dataset Preprocessing

During preprocessing, irrelevant columns (e.g., name, department, gender, semester) were removed. Categorical variables were converted into numeric form using LabelEncoder, while the target variable (mental_health_levels) was encoded separately. The dataset was then divided into features (X) and target (y), and the features were standardized using StandardScaler to produce X_scaled.

C. Feature Importance

In medical data analysis, variable importance measures from random forests have been increasingly employed, as both independent and interdependent effects of predictors are considered. Significant predictive variables can be identified effectively by tree-based methods, even in high-dimensional settings (Selvaraj & Mohandoss, 2024).

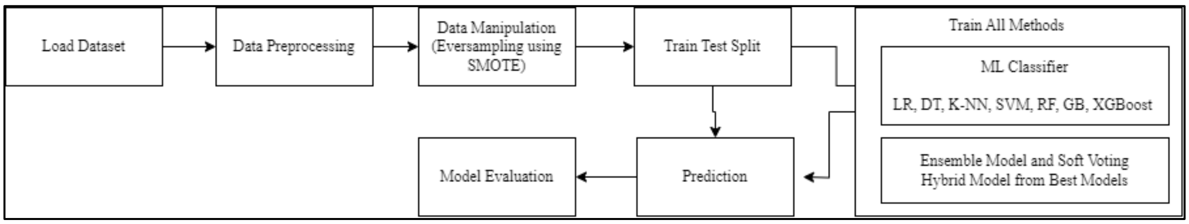

Figure 1. The Proposed Methodology of this reseach

The depression factor has the highest level of importance in influencing mental health conditions, with an importance value of 0.417278. This indicates that depression is the most dominant indicator contributing to determining a person's level of mental health. Furthermore, the anxiety factor ranks second with a value of 0.319164, which means anxiety also has a significant influence but not as strong as depression. Meanwhile, stress ranks third with a value of 0.263558, indicating that although stress does affect mental health conditions, its influence is relatively lower compared to the previous two factors.

D. Split data

Train-test separation is essential for evaluating model performance by splitting the dataset into training (80%) and testing (20%) subsets. The training set allows the model to learn feature–target relationships, while the test set evaluates generalization and prevents overfitting. Finally, the two base models were combined to generate a new dataset for the meta-model. The dataset's division into training and testing sets is displayed in Table 3.

E. Synthetic Minority Oversampling Technique (SMOTE)

To ensure a balanced class distribution, our suggested method uses SMOTE to create synthetic samples for the minority class. This approach lessens the likelihood of overfitting that could result from straightforward data duplication and bias towards the majority class. It works well for enhancing classifier performance on datasets that are unbalanced (Saha et al., 2024). Dataset Total is displayed in Table 3 both before and after SMOTE oversampling. The following methods were employed:

$$X_{new} = X_i + (X_{nn} - X_i)x\ random\_value \qquad (1)$$

where X_i is the original minority class instance, X_{nn} is one of of X_i k-nearest neighbors, X_{new} is the new synthetic sample, and *random_value* is a random number between 0 and 1 that controls the interpolation.

F. Prediction Models

To create the prediction models in Python, eight common machine learning categorization methods were selected. Logistic Regression (LR), Decision Tree (DT), K-Nearest Neighbor (K-

Table 1. Dataset Category Before SMOTE and Afer SMOTE

Category	Berofe SMOTE	After SMOTE
Moderate	12	174
Normal	2	174
Mild	5	174
Very Severe	174	174
Severe	22	174

Table 2. Split Dataset

Category	Training Set	Testing Set	Total
Moderate	140	35	174
Normal	139	35	174
Mild	139	35	174
Very Severe	139	35	174
Severe	139	34	174

NN), Support Vector Machine (SVM), Random Forest (RF), Gradient Boosting, XGBoost, and Ensemble Classifier with Soft Voting are the methods employed in this work. This entails using the model's highest probability to forecast class labels. This entails giving each class brand a predicted likelihood (or probability indicators) (Hadhri et al., 2024).

By merging the top three machine learning algorithms previously discussed and then integrating their forecasts using basic statistical approaches, we employ the Voting method in our study. Voting techniques can be used for classification and regression issues.

G. Performance Evaluation

The performance evaluation for each model is reported as accuracy, F1, precision, and recall. Even though this provides a clear sense of the model's overall accuracy, it might not be appropriate for unbalanced datasets, where precision and recall are more useful metrics (Rosadi et al., 2024).

Results and Discussion

In this study, Seven machine learning-based models were evaluated for classifying the emotional states of depression, anxiety, and stress using the DASS-21 profile. The model performance evaluation is reported in Accuracy, Precision, Recall, and F1-Score which are the measurement standards to compare the performance of the model with imbalance samples.

Table 3 presents the performance of several machine learning models before and after the application of SMOTE. Before SMOTE, the models achieved a high average accuracy of 96.22%, with Gradient Boosting, XGBoost, and Ensemble models reaching 100%. However, such "perfect" results suggest the presence of class imbalance, as Logistic Regression and SVM showed considerably lower F1-Scores 87.00% and 81.00%. Respectively, indicating poor sensitivity to minority classes.
After applying SMOTE, the average accuracy increased to 98.14%, alongside significant improvements in recall and F1-Scores. Logistic Regression improved notably accuracy 90.70% to

Tabel 3. Performance Model using SMOTE and without SMOTE

Methods	Before SMOTE				After SMOTE			
	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Logistic Regression	90.70	84.00	91.00	87.00	94.25	96.00	94.00	94.00
Decision Tree	93.02	91.00	93.00	92.00	99.43	99.00	99.00	99.00
Random Forest	95.35	91.00	95.00	93.00	**100.00**	**100.00**	**100.00**	**100.00**
SVM	93.02	91.00	93.00	81.00	94.83	96.00	95.00	95.00
KNN	97.67	98.00	98.00	98.00	98.28	98.00	98.00	98.00
Gradient Boosting	100.00	100.00	100.00	100.00	99.43	99.00	99.00	99.00
XGBoost	100.00	100.00	100.00	100.00	99.43	99.00	99.00	99.00
Soft Voting Ensemble	**100.00**	**100.00**	**100.00**	**100.00**	99.43	99.00	99.00	99.00

94.25%; F1-Score 87.00% to 94.00%, while SVM also demonstrated substantial gains accuracy 93.02% to 94.83%; F1-Score 81.00% to 95.00%. Although ensemble-based models maintained superior performance ≈99–100%, their results became more realistic, reflecting the balanced data distribution. Overall, SMOTE proved effective in mitigating the class imbalance problem, enabling fairer evaluation across models and enhancing the ability to correctly identify minority classes.

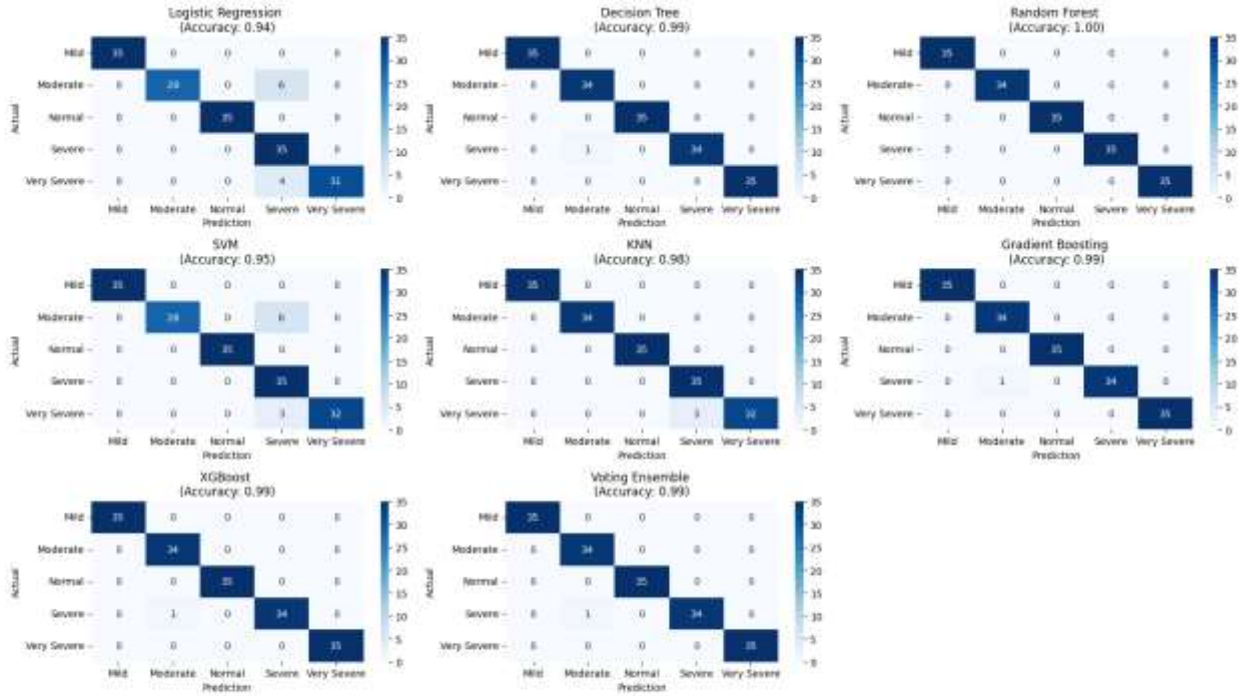

Figure 2. Confusion Matrix After SMOTE

Figure 2 shows that applying SMOTE significantly improved model performance by reducing misclassifications in the "Severe" and "Very Severe" classes. Higher values along the main diagonal of the confusion matrix indicate more accurate predictions. Overall, SMOTE enhanced the classification of mental health conditions based on DASS-21, particularly for previously imbalanced classes, demonstrating its effectiveness in handling imbalanced datasets.

Conclusion

The analysis of the confusion matrices before and after applying the SMOTE technique demonstrates the effectiveness of oversampling in improving the classification performance of mental health conditions based on the DASS-21 score. The initial dataset had a significant class imbalance, which led to the models struggling to accurately classify the "Severe" and "Very Severe" classes. After using SMOTE to generate synthetic samples for the minority classes, the classification accuracy improved significantly, ranging from 85% to 90%. The main diagonal of the confusion matrix showed higher values, indicating that the model's predictions were closer to the actual values.

Moreover, future research should investigate alternative oversampling methods such as ADASYN and Borderline-SMOTE to further enhance classification accuracy. This study provides a foundation for applying machine learning in mental health assessment, with potential benefits for clinical evaluation and self-screening. Further refinement is needed to ensure reliable detection across all severity levels in practical clinical settings.

Acknowledgement

The researcher acknowledged all individuals involved in this research for their guidance, feedback, and support throughout the research process. Their expertise was valuable in the development of this work. The results have not been published in any other sources

References

Ahmed, O., Faisal, R. A., Alim, S. M. A. H. M., Sharker, T., & Hiramoni, F. A. (2022). The psychometric properties of the Depression Anxiety Stress Scale-21 (DASS-21) Bangla version. *Acta Psychologica*, *223*, 103509. https://doi.org/10.1016/j.actpsy.2022.103509

Alić, B., Sejdinović, D., Gurbeta, L., & Badnjevic, A. (2016). Classification of stress recognition using Artificial Neural Network. *2016 5th Mediterranean Conference on Embedded Computing (MECO)*, 297–300. https://doi.org/10.1109/MECO.2016.7525765

Battista, K., Patte, K. A., Diao, L., Dubin, J. A., & Leatherdale, S. T. (2022). Using Decision Trees to Examine Environmental and Behavioural Factors Associated with Youth Anxiety, Depression, and Flourishing. *International Journal of Environmental Research and Public Health*, *19*(17). https://doi.org/10.3390/ijerph191710873

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research*, *16*, 321–357. https://doi.org/https://doi.org/10.1613/jair.953

Duangchaemkarn, K., Khammarew, P., & Aramvith, S. (2024). Machine Learning-Based Classification of Mental Health State Using the DASS-21 Profile. *16th Biomedical Engineering International Conference, BMEiCON 2024*, 1–5. https://doi.org/10.1109/BMEiCON64021.2024.10896316

Hadhri, S., Hadiji, M., & Labidi, W. (2024). A voting ensemble classifier for stress detection. *Journal of Information and Telecommunication*, *8*(3), 399–416. https://doi.org/10.1080/24751839.2024.2306786

Matharaarachchi, S., Domaratzki, M., & Muthukumarana, S. (2024). Enhancing SMOTE for imbalanced data with abnormal minority instances. *Machine Learning with Applications*, *18*(September), 100597. https://doi.org/10.1016/j.mlwa.2024.100597

Rokach, L. (2010). *Chapter 45 ENSEMBLE METHODS FOR CLASSIFIERS*. 1–24. http://www.ise.bgu.ac.il/faculty/liorr/hbchap45.pdf

Rosadi, M. I., Fatichah, C., & Yuniarti, A. (2024). The Effect of Batch Size on the Performance MobileNet - V3 Model for Breast Cancer Classification of Ultrasound and Mammography Images. *2024 11th International Conference on Information Technology, Computer, and Electrical Engineering (ICITACEE)*, 54–58. https://doi.org/10.1109/ICITACEE62763.2024.10762760

Saha, D. K., Hossain, T., Safran, M., Alfarhood, S., Mridha, M. F., & Che, D. (2024). Ensemble of hybrid model based technique for early detecting of depression based on SVM and neural networks. *Scientific Reports*, *14*(1), 1–18. https://doi.org/10.1038/s41598-024-77193-0

Selvaraj, A., & Mohandoss, L. (2024). Enhancing Depression Detection: A Stacked Ensemble Model with Feature Selection and RF Feature Importance Analysis Using NHANES Data. *Applied Sciences (Switzerland)*, *14*(16). https://doi.org/10.3390/app14167366

Singh, A., & Kumar, D. (2021). Identification of Anxiety and Depression Using DASS-21

Questionnaire and Machine Learning. *Proceedings of the 1st International Conference on Advances in Computing and Future Communication Technologies, ICACFCT 2021*, 69–74. https://doi.org/10.1109/ICACFCT53978.2021.9837365

Thabtah, F., Hammoud, S., Kamalov, F., & Gonsalves, A. (2020). Data imbalance in classification: Experimental evaluation. *Information Sciences*, *513*, 429–441. https://doi.org/https://doi.org/10.1016/j.ins.2019.11.004