HANDLING IMBALANCED DATA ON MULTILEVEL DEPRESSION CLASSIFICATION: CHALLENGES AND SOLUTIONS

Mohd Shahrul Nizam Mohd Danuri[1*], Atiqah Miza Ahmad Tarmizie[1], and Rohizah Abd Rahman[2]

[1]Faculty of Computer Science and Information Technology, Universiti Malaya,
50603 Kuala Lumpur, Malaysia

[2]Faculty of Technology and Information Science, Universiti Kebangsaan Malaysia,
43650 Bandar Baru Bangi, Selangor, Malaysia

Emails: msnizam@um.edu.my[1*], u2000906@siswa.um.edu.my[1], rohizah@ukm.edu.my[2]

ABSTRACT

This study addresses the challenges posed by imbalanced data in multilevel depression classification by leveraging the Adaptive Synthetic (ADASYN) technique. Subject Matter Experts (SMEs) annotate data collected from X into four categories: None, Mild, Moderate, and Severe. The imbalanced distribution, particularly with a larger group for the None category, prompts the application of ADASYN for effective data augmentation. The research framework encompasses Data Collection, Expert Data Annotation, Text Preprocessing, and Text Representation and Classification. Evaluation metrics, including Recall and F1 score, gauge the model's effectiveness in multilevel depression classification. Results showcase the efficacy of the ADASYN-enhanced model, specifically with XGBoost, demonstrating improved classification accuracy, especially for minority classes. This study contributes valuable insights to the field of multilevel depression classification, emphasizing the effectiveness of ADASYN in managing imbalanced data scenarios and showcasing the applicability of XGBoost in enhancing model performance.

Keywords: Imbalanced Data; Multilevel Depression Classification; ADASYN; Online Social Network.

1.0 INTRODUCTION

The 2023 National Health and Morbidity Survey (NHMS) in Malaysia reported that 4.6% of adults experienced depression, which doubled from the year 2019 [1]. Subsequent findings from the 2023 NHMS highlighted a concerning 16.5% of children grappling with mental health problems from age 5 to 15 years old. This underscores the urgent need for effective methods to identify, categorize, and address varying levels of depression, emphasizing the importance of tailored interventions and support mechanisms.

Recognizing the evolving landscape of mental health discussions, individuals in Malaysia, like many worldwide, increasingly turn to Online Social Networks (OSN), particularly X, to advocate for mental health and share personal experiences [2]. The platform serves as a crucial outlet for venting about mental health issues and seeking solidarity within a community that understands and empathizes with the challenges of mental well-being.

In response to this pressing societal need, this study adapts traditional psychological assessments, such as the Depression, Anxiety, and Stress Scale (DASS) [3], into a data science framework. Rather than directly integrating the DASS test, the researcher leverages its principles to guide the exploration. The primary aim is to harness the vast data available on OSN, specifically X, to detect and classify depression levels. This research contributes to the field by pioneering novel methodologies for depression multi-classification, utilizing OSN data as a valuable source for individual insights into mental well-being.

This study is committed to developing an effective model for classifying depression into distinct categories: None, Mild, Moderate, and Severe. These categories serve as critical benchmarks for tailoring interventions and support mechanisms, emphasizing the urgency and significance of addressing mental health challenges in Malaysia and beyond.

2.0 BACKGROUND OF THE STUDY

This section outlines this study's current works, data source, and scope.

2.1 Current Works

Depression, characterized by persistent feelings of sadness and loss of interest, significantly impacts an individual's emotions, thoughts, and behavior, leading to various emotional and physical challenges. Often diagnosed as major depressive disorder or clinical depression, this mood disorder affects the ability to perform day-to-day activities, sometimes resulting in a perception that life is not worth living. By recognition as more than a fleeting episode of sadness, depression necessitates long-term treatment, with medication, psychotherapy, or a combination of both being common avenues for recovery [4].

Depression has prompted extensive research in automated prediction and detection. However, many datasets in these studies exhibit class imbalance, where the dominant class overshadows the minority class targeted for detection. Simisani Ndaba [5] employs the PRISMA methodology to review various class imbalance handling techniques in Depression prediction and detection research. The study identifies a research gap, emphasizing the scarcity of under-sampling methods, and proposes considering regression modeling for future exploration.

The Depression, Anxiety, and Stress Scale (DASS-21) is a commonly used psychometric instrument that evaluates the emotional conditions of depression, anxiety, and stress. Originated by the work of Sydney Lovibond and Peter Lovibond, the tripartite structure facilitates a nuanced understanding of mental health, rendering it especially valuable across diverse populations, such as students and healthcare workers [6]. The DASS-21 has undergone validation in various cultural contexts, confirming its reliability and applicability across different settings [7], [8], [9].

Research demonstrates that the DASS-21 accurately reflects the psychological effects of stressors, especially in high-pressure contexts like medical education. A study identified a significant predictive relationship between perceived stressors and DASS-21 scores in medical students, demonstrating the correlation between academic stress and burnout with elevated levels of depression, anxiety, and stress [10]. This relationship highlights the significance of employing the DASS-21 as a screening instrument in educational contexts to detect students vulnerable to mental health challenges.

The DASS-21 demonstrates robust psychometric properties, encompassing both reliability and validity. The instrument has undergone translation into multiple languages and has been validated across diverse demographic groups, thereby increasing its applicability in cross-cultural research [7], [11], [12]. A validation study conducted among Vietnamese adolescents confirmed the factor structure and convergent validity of the DASS-21, thereby reinforcing its effectiveness in measuring mental health outcomes across diverse populations [9]. The scale has been utilized to evaluate mental health in particular contexts, including postpartum women in Malawi, where it showed sufficient reliability in screening for prevalent mental disorders [13].

The COVID-19 pandemic has highlighted the significance of the DASS-21 in evaluating mental health. Research conducted during this period indicated increased levels of depression, anxiety, and stress among diverse populations, including university students and healthcare workers [14], [15], [16]. The DASS-21 served as a standardized tool for assessing the psychological effects of the pandemic. A study comparing mental health outcomes during the pandemic across various countries employed the DASS-21 to demonstrate significant differences in mental health status, underscoring its importance in global mental health research [17], [18]. This has shown that the DASS-21 is an essential instrument for evaluating mental health in various populations and settings. The instrument's strong psychometric properties and capacity to encapsulate the complexities of emotional distress render it essential for researchers and practitioners focused on effectively addressing mental health issues.

SMOTE is a widely utilized data-level approach often employed independently or in conjunction with oversampling and under-sampling strategies. It is particularly effective in addressing class imbalances, as seen by its popularity in the literature [19], [20], [21]. The study emphasizes the continuous requirement for various strategies, namely under-sampling techniques and regression modeling, to improve the effectiveness of automated systems for detecting depression. The Adaptive Synthetic (ADASYN) approach creates synthetic samples by considering the density distribution of the minority class. It targets explicitly difficult-to-learn examples to effectively manage unbalanced data, surpassing the performance of the SMOTE technique. ADASYN focuses on generating synthetic samples that pose a more significant challenge for the classifier, enhancing the model's performance [22], [23]. This is particularly beneficial in intricate datasets where

distinguishing the minority class is vital. Therefore, ADASYN proves to be more successful than SMOTE in enhancing the classification process's accuracy and resilience.

Mahendran and Vincent [24] address Major Depressive Disorder (MDD) as a severe threat to psychophysiology and emphasize the need for an effective predictive model. They focus on reducing feature dimensionality among IT professionals using the Random Forest-based Recursive Feature Elimination technique. Evaluating the model with Naïve Bayes, Support Vector Machines, and Decision Tree classifiers, the study shows a considerable increase in prediction accuracy after applying the feature selection technique. The findings underscore the importance of feature selection in enhancing the performance of classification algorithms, offering valuable insights for the diagnosis of MDD in the IT industry.

Automatic Depression Detection (ADD) has emerged as a valuable tool to provide an objective, efficient, and convenient technique for diagnosing depression in mental health detection. However, existing ADD methods face challenges, such as binary detection limitations in addressing early detection tasks and inadequate learnable representations due to small dataset sizes and improper organization of interview texts. In response, Zhang and Guo [25] propose a novel approach, Multilevel Depression Status Detection, based on Fine-Grained Prompt Learning (MDSD- FGPL). This approach introduces multiple sets of prompts ranging from coarse-grained to fine-grained, training a language model with features extracted at different depressive levels. By reorganizing interview texts at the question-response level and leveraging attention mechanisms, the MDSD-FGPL method exhibits promising results in fine-grained depression detection, achieving a state-of-the-art F1-score of 0.8276 in coarse-grained binary classification [25]. The study highlights the importance of addressing the limitations of existing ADD methods to enhance the accuracy and granularity of depression detection.

Fang et al. [26] propose the Multimodal Fusion Model with Multilevel Attention Mechanism (MFM-Att) for comprehensive depression feature extraction. Utilizing audio, visual, and text data, the model employs two LSTMs and a Bidirectional LSTM in the first stage to learn diverse features. The second stage merges these features using an attention fusion network, capitalizing on modality diversity. The multilevel attention mechanism enhances performance by extracting valuable features within and between modalities. Evaluated on DAIC-WOZ, MFM-Att outperforms state-of-the-art models, attested by superior RMSE metrics. This study underscores the significance of multimodal approaches and attention mechanisms in advancing depression detection models.

2.2 Data Source

The data collected from OSN, namely X (formerly referred to as Twitter), reveals a fundamental disparity in the prevalence of depression. The disparity in the distribution of these levels is a significant barrier to developing an accurate and unbiased model for categorizing depression. This disparity is not only a statistical obstacle but a crucial one that has the potential to undermine the accuracy and reliability of the model. The skewed portrayal of minority groups, namely those who are grappling with severe depression, can result in their inadequate representation and, therefore, a decline in the general accuracy of predictions.

It is crucial to prioritize the resolution of this data mismatch with the growing recognition and approval of mental health conversations on online social networks (OSN). This research addresses the difficulties presented by imbalanced datasets in mental health data while recognizing the possible bias and negative impact on model accuracy. The primary objective is to create a robust categorization model for depression that utilizes OSN data, providing a more effective and digital substitute for conventional evaluations such as the Depression, Anxiety, and Stress Scale (DASS). The study aims to address the imbalance problem and offer a beneficial tool for individuals to quickly identify potential mental health disorders and obtain the required support and intervention.

2.3 Scope of Study

The scope of this project is designed to address a critical need for mental health awareness and support by providing an accessible and digital solution for individuals to identify and understand their depression levels. The project collects data from X, specifically tweets related to mental health, to form the dataset. The following steps encompass training a model for multilevel depression classification and incorporating methods to address the inherent imbalance in the collected data.

This initiative is tailored for civilians and aims to be a valuable resource for anyone seeking to understand their mental well-being. The digital nature of the project allows for widespread accessibility, ensuring that individuals from various backgrounds can benefit from the service. Ultimately, the project strives to contribute to mental health awareness, offering a tool that empowers users to identify their conditions and take proactive steps toward seeking help.

3.0 METHODOLOGY

This section introduces the methodology for studying Imbalance Data on Multilevel Depression Classification. Fig. 1 explains the overall architecture, including data collection and preprocessing, the utilization of advanced techniques for depression classification, evaluation metrics, and the deployment of the model through a user-friendly web application. The business objectives were to develop a model for detecting and classifying depression levels and develop a deployable application incorporating the trained model.

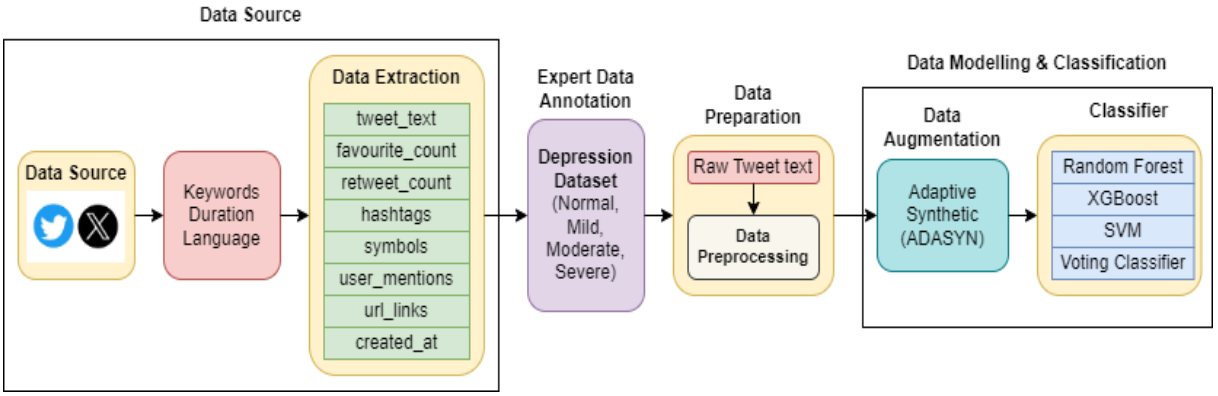

Fig. 1: A Model for Depression Multilevel Detection

3.1 Data Source and Annotation

This phase is to collect as much knowledge as possible about the dataset. The insight gained from understanding the data will be helpful later during the Data Preparation phase. This study collects data from Online Social Network (OSN) sources, mainly X. Tweets were crawled and collected using the X streaming API service and stored in a CSV file. The initial number of tweets gathered was 19,744. The data preparation was greatly aided by Subject Matter Experts (SMEs) in mental health, who manually annotated the data before preprocessing, representation, and classification. This study focused only on the depression category with four negative emotional states: Normal, Mild, Moderate, and Severe.

3.2 Data Preparation and Preprocessing

The data preparation phase involves cleaning and structuring the collected X data for effective use in this study. Python packages are employed for efficient data cleaning, enhancing the dataset's quality for subsequent analysis and model training. Various transformations enhance the X text data for effective model training. The process commences with using LabelEncoder to convert string labels into numerical classes, aiding in model interpretation.

Subsequently, text standardization is implemented by converting all text to lowercase, ensuring uniformity in the dataset. Special characters and punctuation are removed to streamline tokenization, where sentences are broken down into individual words or tokens. The removal of common English stopwords focuses the analysis on meaningful words, while stemming reduces words to their root form for consistency. URLs are eliminated to enhance text clarity, and lemmatization transforms words to their base or dictionary form. Collectively, these preprocessing steps refine and standardize the X text data, creating a more meaningful and streamlined input for subsequent stages.

3.3 Data Augmentation

In the data augmentation step, the researcher addresses the imbalance in the dataset to enhance the model's ability to classify depression levels across all categories accurately. Various methods, including SMOTE, were tested. However, based on the evaluations, ADASYN emerged as the most effective technique for the specific dataset.

Adaptive Synthetic Sampling (ADASYN) excels in generating synthetic instances for the minority classes, which, in this case, study, are the less-represented depression levels (Severe category). This technique focuses on areas where the dataset is sparser, creating additional synthetic data points resembling minority instances. By introducing these synthetic instances, ADASYN aims to balance the class distribution, ensuring that the model

receives sufficient information to classify even the underrepresented depression levels accurately. This augmentation step is crucial for improving the model's performance, especially when dealing with imbalanced datasets, and allows for more robust predictions across all severity levels of depression.

3.4 Data Modelling and Evaluation

In the data modeling and evaluation phase, multiple machine learning models, namely Random Forest, XGBoost, SVM, and a Voting Classifier combining the three, were employed for depression classification. The selection criterion for the best-performing model centered on achieving the highest Recall and F1 score, given the importance of correctly identifying cases of depression, especially in severe instances. Random Forest, XGBoost, and SVM each bring distinct strengths to the classification task. The Voting Classifier leverages the combined insights from all three models, potentially enhancing overall predictive performance.

For evaluation, metrics like accuracy, Recall, precision, and F1 score were calculated. While accuracy provides an overall measure of correctness, Recall, and F1 score assume greater significance in this context. Recall emphasizes the model's ability to capture instances of depression, prioritizing sensitivity to avoid false negatives. The F1 score, which considers both precision and Recall, offers a balanced assessment of the model's precision and ability to avoid false positives and negatives.

Given the nature of depression classification, where the identification of severe cases is particularly crucial, the emphasis on Recall and F1 score ensures that the chosen model correctly identifies and classifies depression across all severity levels. The XGBoost model, identified as the best performer based on these metrics, is subsequently chosen for deployment in the project's next phase.

4.0 FINDING AND DISCUSSION

This section outlines the results and findings of this study.

4.1 Analysis and Finding

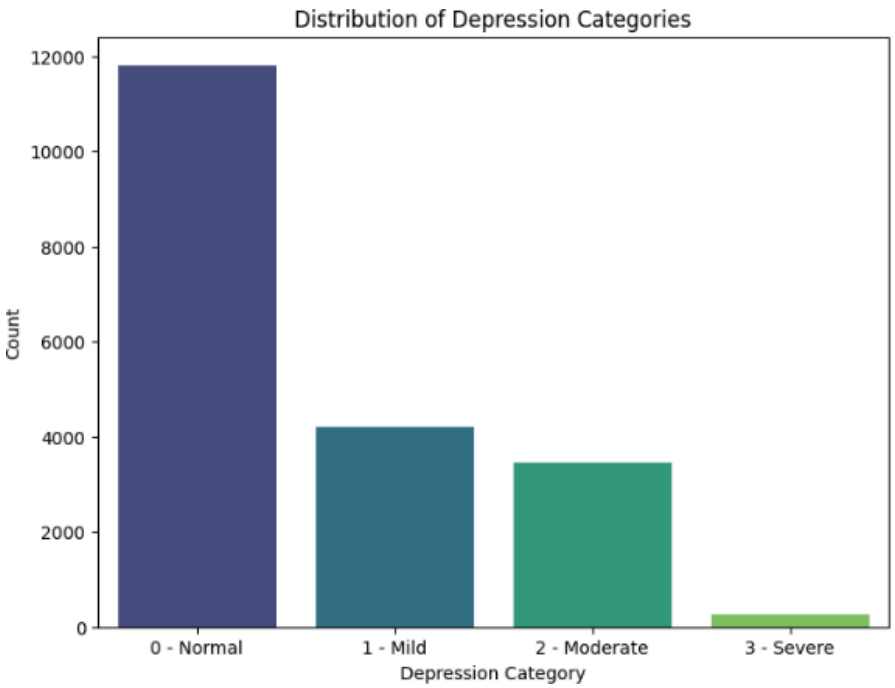

Fig. 2: A Class distribution before ADASYN

Fig. 2 shows the bar plot showing the distribution of depression categories across the dataset, visually representing the significant class imbalance. Specifically, the "normal" category stands out with a substantial 11,816 instances, followed by "mild" with 4,203 instances, "moderate" with 3,467 instances, and the least frequent "severe" category with only 258 instances. This data distribution discrepancy, where "none" dominates

and "severe" is comparatively rare, poses a notable challenge for the model, potentially introducing bias and impacting its overall effectiveness.

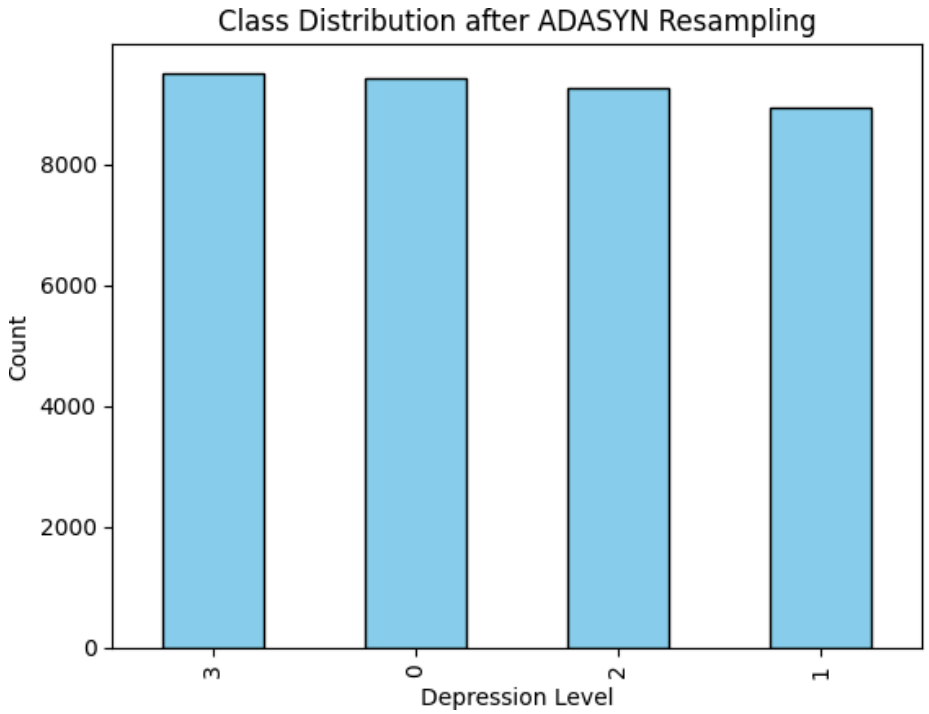

Fig. 3: A Class distribution after ADASYN

Fig. 3 shows a bar plot that reflects the transformed distribution of depression categories after applying the ADASYN technique to address the initial severe class imbalance. This resampling approach has notably balanced the dataset, resulting in a more equitable representation of each category. The "severe" category has experienced a substantial increase, now comprising 9,503 instances, compared to the original count of 258. Similarly, the "normal" category, initially dominant with 11,816 instances, has been adjusted to 9,428. The "moderate" and "mild" categories have also seen adjustments, with "moderate" now at 9,244 instances (up from 3,467) and "mild" at 8,933 instances (up from 4,203). This rebalancing enhances the dataset's diversity, mitigating potential biases and fostering improved model training across all depression categories.

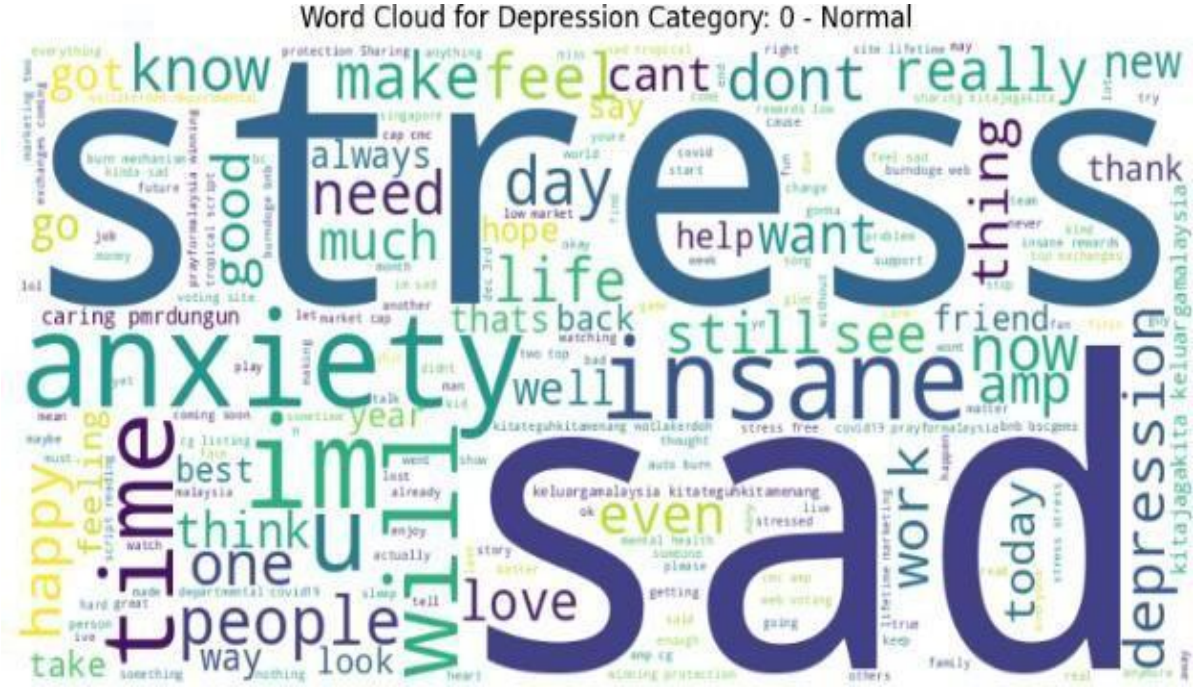

Fig. 4: Word Cloud for "Normal" category

The word cloud for the "Normal" category of depression highlights key terms such as "stress," "sad," "anxiety," and "insane," indicating that individuals without clinically significant depression often employ language linked to negative emotions. The prevalence of terms such as "stress" and "sad" in the visual indicates that conversations about mental health are frequent in everyday discourse, irrespective of the presence of actual depressive conditions. Furthermore, terms such as "life," "time," "need," and "feel" represent broad concerns and expressions that, while not explicitly indicative of depression, are integral to regular considerations of personal well-being. The occurrence of terms like "friend," "love," and "happy" suggests positive or supportive dimensions within these individuals' social environments, potentially serving as protective factors. This word cloud indicates that individuals in the "Normal" category may discuss stress and anxiety, but such discussions likely do not meet the criteria for clinical depression. This underscores the necessity of differentiating between general expressions of stress and symptoms of depression.

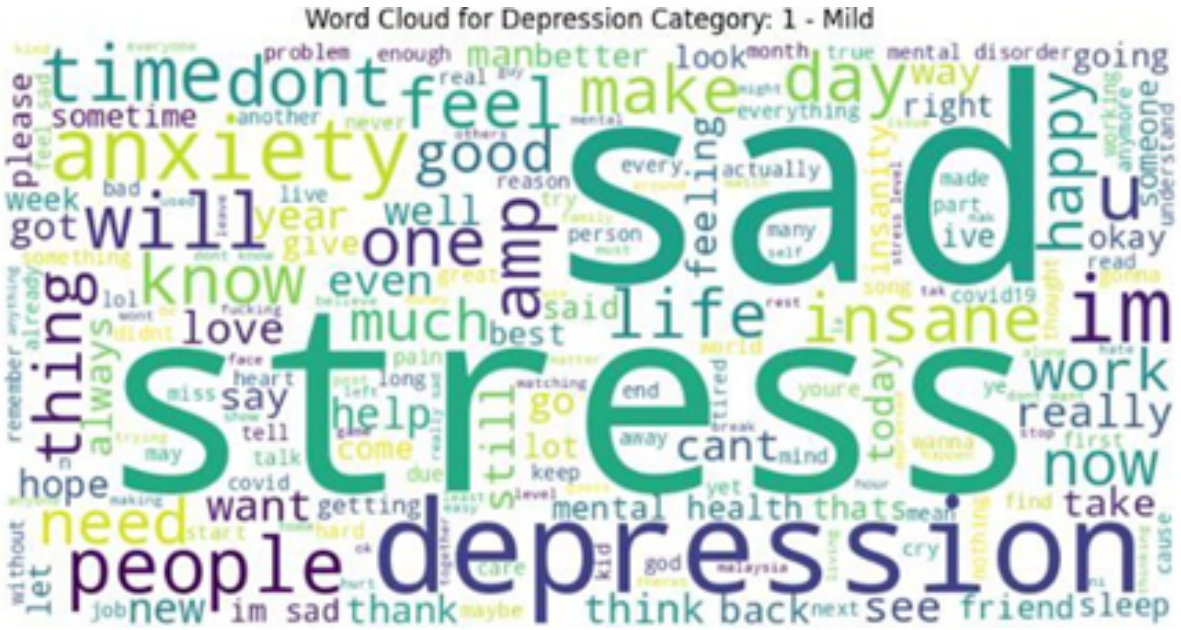

Fig. 5: Word Cloud for "Mild" category

The word cloud for the "Mild" depression category emphasizes terms like "sad," "stress," "depression," and "anxiety," indicating a stronger correlation with negative emotions and mental health issues relative to the "Normal" category. The prevalence of terms such as "depression" indicates a heightened self-identification or recognition of depressive symptoms, regardless of their severity. Terms such as "feel," "time," "make," and "need" reflect an emphasis on emotional conditions and the necessity for transformation or assistance. The presence of terms such as "people," "life," and "help" suggests a need for connection and support, reflecting the initial phases of seeking assistance or comprehension regarding mental health issues. Furthermore, terms like "work," "friends," and "happy" indicate the significance of social relationships and daily activities in the management of mild depressive symptoms. The word cloud indicates that individuals classified as "Mild" are increasingly acknowledging and articulating their depressive symptoms, frequently alongside expressions of stress, anxiety, and a nuanced request for assistance.

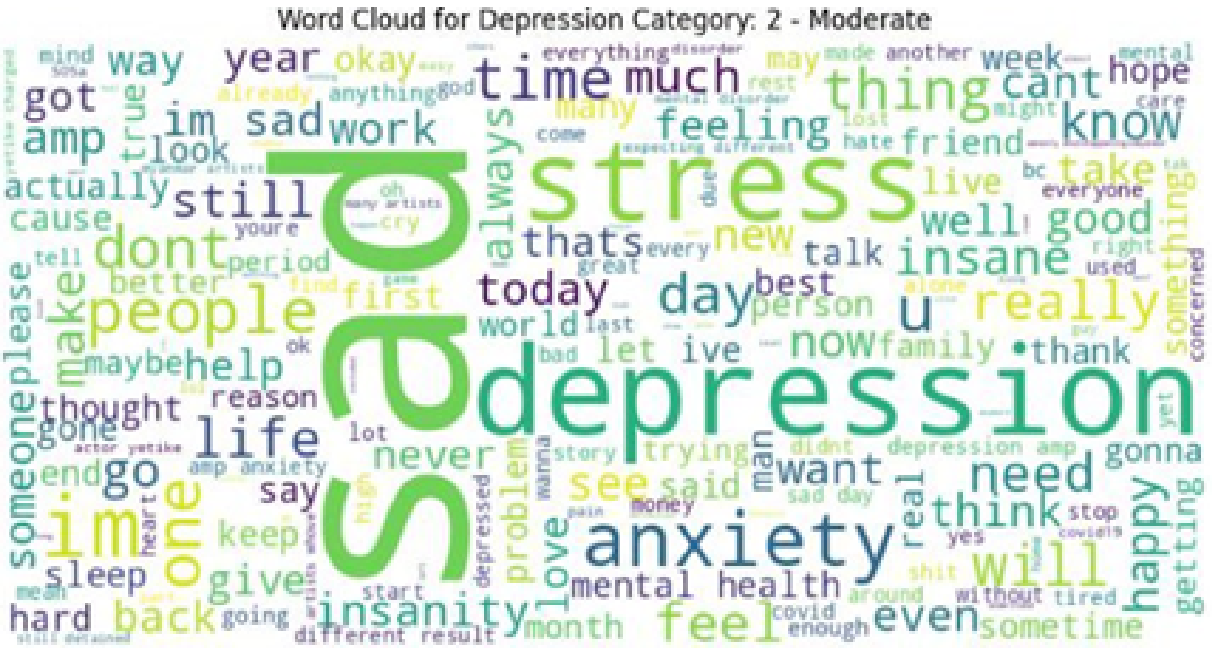

Fig. 6: Word Cloud for "Moderate" category

The word cloud for the "Moderate" depression category prominently includes terms such as "depression," "sad," "stress," and "anxiety," indicating a more profound and sustained engagement with depressive and anxious emotions relative to the "Mild" category. The use of terms like "can't," "hard," "need," and "help" indicates a struggle and a heightened awareness of the influence of these emotions on everyday life. Terms such as "mental health," "feeling," and "tired" suggest that individuals in this group possess a heightened awareness of their mental health issues and may be undergoing emotional fatigue. Furthermore, terms such as "people," "life," and "work" reflect the impact of external social and environmental factors on mental states. The employment of more intense terms, such as "insanity" and "lost," indicates an increased intensity in their emotional experiences, consistent with the "Moderate" classification. This word cloud indicates that individuals in this category often report challenges in coping and may require external support to address their mental health issues.

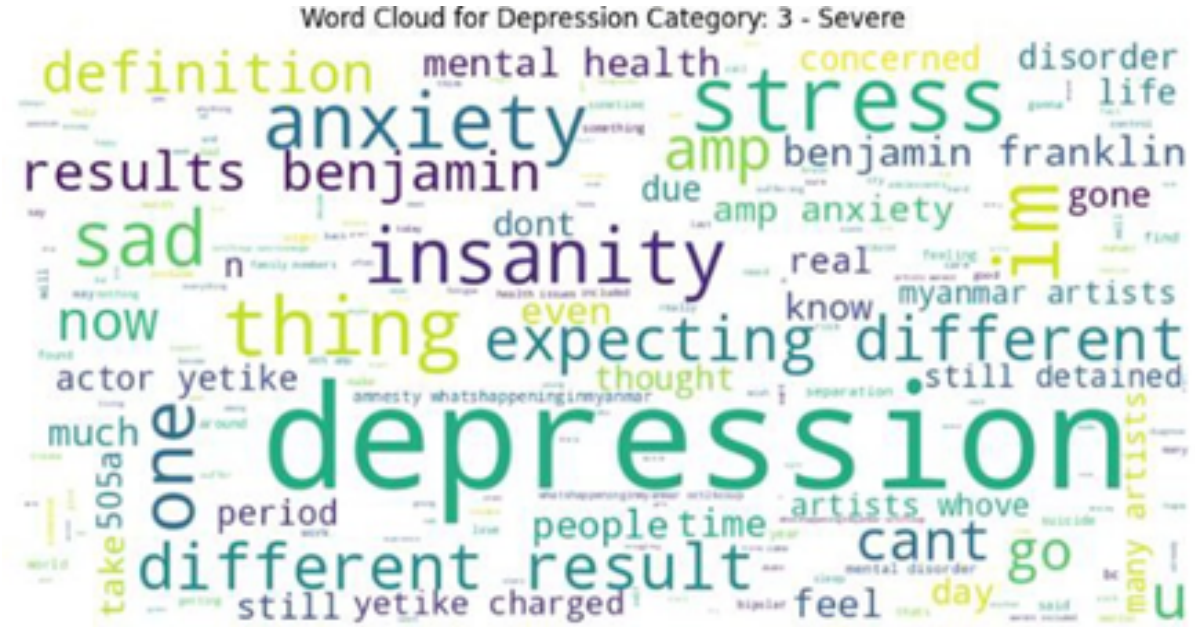

Fig. 6: Word Cloud for "Severe" category

The word cloud for the "Severe" depression category features terms like "depression," "insanity," "stress," "anxiety," and "sad," indicating a significant focus on adverse emotions and mental health challenges. The use of

phrases such as "expecting different" and "different result" indicates a potential frustration or hopelessness, likely associated with unmet expectations or an aspiration for change perceived as unattainable. The term "insanity" denotes an elevated state of mental distress, potentially accompanied by a sense of being overwhelmed. Terms like "can't" and "gone" express a feeling of helplessness or loss, corresponding with the severe classification in which individuals may experience a sense of entrapment or inability to cope. Furthermore, terms such as "mental health," "concerned," and "definition" suggest a significant awareness and potential critical reflection on one's mental state, highlighting the necessity for intervention and support. This word cloud illustrates the emotional turmoil faced by individuals categorized as "Severe," who often experience profound despair, frustration, and a significant need for support in managing their mental health.

4.2 Model Performance

The researcher applied multiple machine learning models, namely Random Forest, XGBoost, SVM, and a Voting Classifier, combining the three methods employed for depression classification. The emphasis lies in evaluating the Recall and F1 score metrics before and after the implementation of ADASYN to address imbalanced data. This comparative analysis identifies and selects the most effective model for further consideration.

4.2.1 Random Forest

Tables 1 and 2 explain that the evaluation of the Random Forest model revealed notable changes in performance metrics before and after the implementation of ADASYN. Before ADASYN, the overall model showed a Recall of 0.58 and an F1-Score of 0.55. Delving into individual categories, distinct Recall and F1-Scores were observed. For the "None" category, the Recall was 0.84 with an F1-Score of 0.76, while the "Mild" category showed a Recall of 0.16 and an F1-Score of 0.2. The "Moderate" category exhibited a Recall of 0.25 and an F1-Score of 0.27; the "Severe" category had a Recall of 0.15 and an F1-Score of 0.15.

After ADASYN implementation, the model's performance demonstrated adjustments, with improved Recall and F1-Scores across all categories. The "None" category saw a Recall of 0.52 and an F1-Score of 0.59, "Mild" showed a Recall of 0.28 and an F1-Score of 0.25, "Moderate" had a Recall of 0.32 and an F1-Score of 0.26, and "Severe" displayed a Recall of 0.44 and an F1-Score of 0.27.

Table 1: Random Forest (Before and After using ADASYN)

ADASYN	Category	Precision	Recall	F1-Score	Support
Before	0-Normal	0.69	0.84	0.76	2388
	1-Mild	0.30	0.16	0.20	808
	2-Moderate	0.30	0.25	0.27	698
	3-Severe	0.16	0.15	0.15	55
After	0-Normal	0.68	0.52	0.59	2388
	1-Mild	0.23	0.28	0.25	808
	2-Moderate	0.22	0.32	0.26	698
	3-Severe	0.20	0.44	0.27	55

Table 2: Accuracy of Random Forest (before and After using ADASYN)

Category	Before	After
Accuracy	0.58	0.43
Precision	0.53	0.50
Recall	0.58	0.43
F1-Score	0.55	0.46

Table 2 presents the overall accuracy of the Random Forest model before and after ADASYN implementation. The model exhibits an initial accuracy of 0.58, which decreases to 0.43 following the application of ADASYN, as the rebalancing process prioritizes improved detection of minority classes at the expense of overall accuracy. This underscores ADASYN's efficacy in enhancing Recall for underrepresented categories, such as "Severe". The reduction in accuracy underscores the trade-off achieved for enhanced Recall and F1-scores, especially within the "Severe" depression category. This trade-off is essential, as the objective is to improve the detection of minority classes, even at the expense of overall accuracy, thereby illustrating ADASYN's effectiveness in increasing sensitivity to underrepresented classes.

Table 3: Random Forest (Before and After using SMOTE)

SMOTE	Category	Precision	Recall	F1-Score	Support
Before	0-Normal	0.67	0.89	0.77	2388
	1-Mild	0.32	0.15	0.21	808
	2-Moderate	0.34	0.18	0.24	698
	3-Severe	0.19	0.13	0.15	55
After	0-Normal	0.66	0.61	0.63	2388
	1-Mild	0.22	0.27	0.24	808
	2-Moderate	0.25	0.23	0.24	698
	3-Severe	0.19	0.36	0.25	55

Table 4: Accuracy of Random Forest (before and After using SMOTE)

Category	Before	After
Accuracy	0.60	0.47
Precision	0.54	0.49
Recall	0.60	0.47
F1-Score	0.55	0.48

Tables 3 and 4 compare the Random Forest model's performance utilizing SMOTE oversampling. Table 3 indicates that the Recall for the "Severe" category increased from 0.13 prior to SMOTE to 0.36, subsequently accompanied by a rise in the balanced F1-score from 0.15 to 0.25. The overall accuracy presented in Table 4 decreased from 0.60 to 0.47 following the application of SMOTE. This indicates that, although SMOTE enhances Recall for the minority class in balancing the dataset, especially "Severe" depression, it does not consistently improve overall model performance compared to ADASYN, as evidenced by lower F1-scores in other categories.

4.2.2 XGBoost

The XGBoost model exhibited robust performance in depression classification, with noteworthy results before and after the implementation of ADASYN. Before ADASYN, the model displayed a Recall of 0.6255 and an F1-Score of 0.5541, showcasing its initial effectiveness. After applying ADASYN to address imbalanced data, the model's performance remained strong, with a Recall of 0.6121 and an F1-Score of 0.5391. Delving into individual categories, the model demonstrated remarkable Recall and F1-Scores across all severity levels.

Before ADASYN, the "None" category exhibited a Recall of 0.94 and an F1-Score of 0.78, while the "Mild" category showed a Recall of 0.16 and an F1-Score of 0.22. The "Moderate" category displayed a Recall of 0.13 and an F1-Score of 0.19; the "Severe" category had a Recall of 0.04 and an F1-Score of 0.06. Post-ADASYN, the "None" category maintained a Recall of 0.93 and an F1-Score of 0.78; the "Mild" category showed a Recall of 0.08 and an F1-Score of 0.13. The "Moderate" category demonstrated a Recall of 0.14 and an F1-Score of 0.21, while the "Severe" category notably improved with a Recall of 0.45 and an F1-Score of 0.17.

Table 5: XGBoost (Before and After using ADASYN)

ADASYN	Category	Precision	Recall	F1-Score	Support
Before	0-Normal	0.67	0.94	0.78	2388
	1-Mild	0.35	0.16	0.22	808
	2-Moderate	0.40	0.13	0.19	698
	3-Severe	0.29	0.04	0.06	55
After	0-Normal	0.67	0.93	0.78	2388
	1-Mild	0.34	0.08	0.13	808
	2-Moderate	0.42	0.14	0.21	698
	3-Severe	0.13	0.45	0.17	55

Table 6: Accuracy of XGBoost (before and After using ADASYN)

Category	Before	After
Accuracy	0.63	0.61
Precision	0.55	0.55
Recall	0.63	0.61
F1-Score	0.55	0.54

Tables 5 and 6 illustrate the performance of the XGBoost model, highlighting significant results prior to and following the application of ADASYN. Table 5 indicates a significant improvement in Recall for the "Severe" category, rising from 0.04 to 0.45, demonstrating an enhanced ability of the model to identify cases within this critical classification. The overall accuracy presented in Table 6 slightly decreased from 0.63 to 0.61. However, the enhancement in Recall for both the “Moderate” and “Severe” categories resulted in a balanced F1-score of 0.54 following the application of ADASYN, indicating that the model effectively addresses imbalanced data while maintaining overall accuracy.

Table 7: XGBoost (Before and After using SMOTE)

SMOTE	Category	Precision	Recall	F1-Score	Support
Before	0-Normal	0.67	0.94	0.78	2388
	1-Mild	0.36	0.15	0.22	808
	2-Moderate	0.43	0.15	0.22	698
	3-Severe	0.25	0.05	0.09	55
After	0-Normal	0.67	0.93	0.78	2388
	1-Mild	0.31	0.08	0.12	808
	2-Moderate	0.36	0.16	0.22	698
	3-Severe	0.1	0.27	0.14	55

Table 8: Accuracy of XGBoost (before and After using SMOTE)

Category	Before	After
Accuracy	0.63	0.61
Precision	0.56	0.54
Recall	0.63	0.61
F1-Score	0.56	0.51

Tables 7 and 8 indicate moderate improvements across categories for the XGBoost model utilizing SMOTE. Table 7 indicates that Recall for the “Severe” category increased from 0.05 to 0.27, while the “Moderate” category increased from 0.15 to 0.22. Table 8 displays the results of XGBoost following the application of SMOTE, indicating that accuracy remains stable at 0.61. While SMOTE improves Recall for minority classes, particularly the "Severe" category, its effect is less significant than ADASYN. The findings indicate that SMOTE is more appropriate for scenarios necessitating moderate rebalancing, as it preserves accuracy without significant alterations. However, it does not improve Recall to the same extent as ADASYN.

4.2.3 Support Vector Machine

The Support Vector Machine (SVM) model demonstrated notable performance in depression classification, with discernible results before and after the implementation of ADASYN. Before ADASYN, the model showcased a Recall of 0.63 and an F1-Score of 0.55, indicating its initial proficiency. Following the integration of ADASYN to address data imbalance, the model's performance saw adjustments, resulting in a Recall of 0.51 and an F1-Score of 0.49.

The SVM model exhibited distinctive Recall and F1-Scores across all severity levels by examining individual categories. Pre-ADASYN, the "None" category displayed a Recall of 0.95 and an F1-Score of 0.78, while the "Mild" category exhibited a Recall of 0.13 and an F1-Score of 0.19. The "Moderate" category demonstrated a Recall of 0.13 and an F1-Score of 0.20; the "Severe" category showed a Recall of 0.02 and an F1-Score of 0.03. Post-ADASYN, the "None" category retained a Recall of 0.71 and an F1-Score of 0.68, while the "Mild" category revealed a Recall of 0.13 and an F1-Score of 0.17. The "Moderate" category indicated a Recall of 0.26 and an F1-Score of 0.23; the "Severe" category portrayed a Recall of 0.25 and an F1-Score of 0.21.

Table 9: Support Vector Machine (Before and After using ADASYN)

ADASYN	Category	Precision	Recall	F1-Score	Support
Before	0-Normal	0.66	0.95	0.78	2388
	1-Mild	0.36	0.13	0.19	808
	2-Moderate	0.41	0.13	0.2	698
	3-Severe	0.33	0.02	0.03	55
After	0-Normal	0.65	0.71	0.68	2388
	1-Mild	0.26	0.13	0.17	808
	2-Moderate	0.21	0.26	0.23	698
	3-Severe	0.18	0.25	0.21	55

Table 10: Accuracy of Support Vector Machine (before and After using ADASYN)

Category	Before	After
Accuracy	0.63	0.51
Precision	0.55	0.49
Recall	0.63	0.51
F1-Score	0.55	0.49

Tables 9 and 10 present the performance metrics of the SVM model before and after the application of ADASYN. Table 9 demonstrates that Recall for the "Severe" category increased from 0.02 to 0.25 following the application of ADASYN, thereby significantly improving the model's capacity to identify minority cases. Table 10 indicates a decrease in the SVM model's accuracy from 0.63 to 0.51 following the application of ADASYN. This trend is consistent with the results observed in the Random Forest model, as SVM emphasizes enhanced detection of underrepresented classes. This demonstrates that ADASYN effectively redistributes model focus across all categories, enhancing SVM's sensitivity to the "Severe" category and improving Recall and F1-scores.

Table 11: Support Vector Machine (Before and After using SMOTE)

SMOTE	Category	Precision	Recall	F1-Score	Support
Before	0-Normal	0.66	0.95	0.78	2388
	1-Mild	0.36	0.14	0.2	808
	2-Moderate	0.43	0.11	0.18	698
	3-Severe	0	0	0	55
After	0-Normal	0.66	0.75	0.7	2388
	1-Mild	0.23	0.2	0.21	808
	2-Moderate	0.31	0.21	0.25	698
	3-Severe	0.14	0.24	0.18	55

Table 12: Accuracy of Support Vector Machine (before and After using SMOTE)

Category	Before	After
Accuracy	0.62	0.53
Precision	0.55	0.51
Recall	0.62	0.53
F1-Score	0.54	0.52

The application of SMOTE to the SVM model yields a slight enhancement in Recall for the “Severe” category, increasing from 0.0 to 0.24, and for the “Moderate” category, rising from 0.11 to 0.21, as illustrated in Tables 11. Table 12 presents the SVM model utilizing SMOTE, demonstrating a reduced accuracy decline from 0.62 to 0.53. This indicates that SMOTE is advantageous for models such as SVM, where accuracy preservation is critical, while ADASYN demonstrates greater effectiveness in managing extreme imbalance. SMOTE's moderate improvement in Recall suggests it may be advantageous when a balance between accuracy and Recall is desired, though it may be less effective in highly imbalanced situations.

4.2.4 Voting Classifier

The Voting Classifier, an ensemble model combining Random Forest (RF), Support Vector Machine (SVM), and XGBoost, exhibited promising results in depression classification before and after the incorporation of ADASYN. Before ADASYN, the model showcased a Recall of 0.6257 and an F1-Score of 0.5499, indicating a robust initial performance. Following the integration of ADASYN to mitigate data imbalance, the model's performance demonstrated adjustments, resulting in a Recall of 0.5424 and an F1-Score of 0.5135.

Delving into the individual severity categories, the Voting Classifier presented distinctive Recall and F1-Scores across all levels. Pre-ADASYN, the "None" category displayed a Recall of 0.95 and an F1-Score of 0.78, while the "Mild" category exhibited a Recall of 0.14 and an F1-Score of 0.20. The "Moderate" category demonstrated a Recall of 0.13 and an F1-Score of 0.20; the "Severe" category showed a Recall of 0.04 and an F1-Score of 0.07. Post-ADASYN, the "None" category retained a Recall of 0.78 and an F1-Score of 0.71, while the "Mild" category revealed a Recall of 0.13 and an F1-Score of 0.18. The "Moderate" category indicated a Recall of 0.23 and an F1-Score of 0.23; the "Severe" category portrayed a Recall of 0.35 and an F1-Score of 0.24. These findings underscore the adaptability of the Voting Classifier, with ADASYN integration contributing to enhanced depression classification across diverse severity levels.

Table 13: Voting Classifier (Before and After using ADASYN)

ADASYN	Category	Precision	Recall	F1-Score	Support
Before	0-Normal	0.67	0.95	0.78	2388
	1-Mild	0.35	0.14	0.2	808
	2-Moderate	0.4	0.13	0.2	698
	3-Severe	0.33	0.04	0.07	55
After	0-Normal	0.66	0.78	0.71	2388
	1-Mild	0.28	0.13	0.18	808
	2-Moderate	0.24	0.23	0.23	698
	3-Severe	0.18	0.35	0.24	55

Table 14: Accuracy of Voting Classifier (before and After using ADASYN)

Category	Before	After
Accuracy	0.63	0.54
Precision	0.55	0.50
Recall	0.63	0.54
F1-Score	0.55	0.51

Tables 13 and 14 present the performance metrics of the Voting Classifier, an ensemble model comprising Random Forest, SVM, and XGBoost, both prior to and following the application of ADASYN. Table 13 indicates that Recall for the “Severe” category increased from 0.04 to 0.35, while F1-scores showed improvement across all categories. Table 14 analyzes the Voting Classifier ensemble model, indicating that the application of ADASYN decreases accuracy from 0.63 to 0.54. The ensemble's enhanced Recall for the "Severe" category accounts for this reduction, as improved identification of severe depression cases is essential in clinical settings. This trade-off illustrates ADASYN's effectiveness in enhancing minority class recall while maintaining overall model performance across categories.

Table 15: Voting Classifier (Before and After using SMOTE)

SMOTE	Category	Precision	Recall	F1-Score	Support
Before	0-Normal	0.66	0.95	0.78	2388
	1-Mild	0.36	0.14	0.2	808
	2-Moderate	0.42	0.11	0.18	698
	3-Severe	0.3	0.05	0.09	55
After	0-Normal	0.66	0.80	0.72	2388
	1-Mild	0.23	0.16	0.19	808
	2-Moderate	0.33	0.18	0.23	698
	3-Severe	0.14	0.25	0.18	55

Table 16: Accuracy of Voting Classifier (before and After using SMOTE)

Category	Before	After
Accuracy	0.62	0.55
Precision	0.55	0.51
Recall	0.62	0.55
F1-Score	0.55	0.52

Tables 15 and 16 assess the performance of the Voting Classifier utilizing SMOTE. Table 15 presents the Recall enhancements for the "Moderate" category, increasing from 0.11 to 0.18, and the "Severe" category, rising from 0.05 to 0.25. Table 16 presents the Voting Classifier utilizing SMOTE, indicating a slight reduction in accuracy from 0.62 to 0.55. The findings demonstrate that SMOTE provides consistent accuracy while improving Recall to a lesser extent than ADASYN. This comparative analysis emphasizes the appropriateness of SMOTE in contexts where maintaining accuracy is prioritized over significantly increasing Recall.

All tables highlight the performance differences between ADASYN and SMOTE across different classifiers. ADASYN effectively enhances Recall for underrepresented classes, especially in Random Forest, XGBoost, and SVM models. In contrast, SMOTE offers a moderate enhancement in Recall with negligible loss in accuracy, rendering it advantageous for tasks that necessitate balanced performance between majority and minority classes. The findings indicate that ADASYN is more suitable for addressing severe class imbalances, while SMOTE may be advantageous in situations necessitating less significant rebalancing.

In conclusion, evaluating four distinct machine learning models, Random Forest, XGBoost, SVM, and a Voting Classifier, illuminated the impact of addressing data imbalance through ADASYN in depression classification. While each model showcased commendable performance, XGBoost emerged as the standout choice, exhibiting consistent and superior Recall and F1-Scores across all severity categories. The robustness of XGBoost, even post-ADASYN, underscores its efficacy in addressing the complexities of imbalanced mental health data. Moving forward, deploying the XGBoost model in the depression classification application holds promise for providing accurate and reliable predictions, contributing to a more nuanced understanding of individuals' mental well-being.

4.3 Limitations and Future Recommendations

Similarly to any data science initiative, the project faces certain limitations. One primary challenge lies in the innate imbalance of mental health data, which can lead to biased model outcomes. Despite employing techniques such as ADASYN for data augmentation, it is crucial to acknowledge that the quality and representativeness of the data may still constrain the model's effectiveness. Additionally, relying on X data introduces potential biases associated with users' demographics and behavior on this platform. These biases may influence the project's generalizability to a broader population.

The present investigation of unbalanced data in multilevel depression classification may gain from further methodological improvements. Initially, although X functions as a principal data source, augmenting the dataset with additional social media platforms, such as Reddit or other appropriate and credible sources, might improve the variety and representativeness of the gathered data. The mental health forums on Reddit might offer more nuanced information that reflects diverse language expressions and demographic traits, enhancing the study. Additionally, feature engineering may be employed to identify behavioral patterns, such as user engagement frequency and interactions with certain subjects. Including this new dimension may enhance the model by integrating a behavioral layer, potentially increasing its classification accuracy.

While ADASYN has demonstrated encouraging outcomes in mitigating data imbalance, a comparative evaluation with more sophisticated oversampling methods, such as Generative Adversarial Networks (GANs) or Variational Autoencoders (VAEs), may be advantageous. Contemporary methodologies aim to synthesize data that more appropriately represents minority groups, possibly enhancing performance in managing intricate, unbalanced datasets. Furthermore, integrating explainability approaches, such as SHapley Additive exPlanations (SHAP), may yield significant insights into the model's decision-making process. SHAP may elucidate the particular words or attributes that predominantly affect the categorization, enhancing transparency and interpretability in the model's predictions.

Analyzing temporal trends in the depression-related data might further improve the contextual comprehension of the model. Examining temporal variations in data patterns may reveal seasonal or event-driven surges that expose fundamental tendencies, enhancing the model's contextual flexibility and classification precision. It is recommended that the evaluation measures be expanded. However, Recall and F1-Score are essential. Metrics like the Matthews Correlation Coefficient (MCC) or any other statistical analysis may provide a more comprehensive assessment of model performance, particularly in unbalanced data.

A hybrid modeling technique that combines text-based classifiers with sentiment analysis or emotion detection frameworks may enhance the system's ability to discern nuanced variations in depression severity more effectively. This integration may enhance the model's ability to identify subtle emotional expressions, therefore improving the accuracy of depression categorization across different severity levels. Adopting these recommendations will establish a more robust, interpretable, and generalizable framework and methodology of the research, improving the model's efficacy in accurately categorizing depression levels.

Alongside the suggested methodological improvements, further study may investigate the possibility of multimodal modeling to categorize depression better. A multimodal method that incorporates textual, audio, and visual data may provide a more comprehensive picture of depressed signs, as individuals frequently convey mental health issues through several modalities beyond text. For example, using vocal intonations and facial expressions from platforms that allow audio and video exchanges could add more meaning to the textual analysis that text alone might miss. These further modalities would facilitate a more comprehensive and nuanced comprehension of depressed symptoms, resulting in possibly enhanced classification accuracy and increased generalizability.

In executing multimodal modeling, approaches like fusion networks may integrate knowledge across modalities efficiently. Recent improvements, shown by the Multimodal Fusion Model with Multilevel Attention Mechanism (MFM-Att), illustrate the use of attention mechanisms in evaluating and synthesizing diverse input kinds, highlighting each modality's most salient features. This methodology may be especially beneficial in the categorization of depression, as nuanced visual signals, such as facial expressions, or auditory signals, such as tonal variations, might offer further signs of emotional distress that are not readily discernible through text alone.

Moreover, multimodal models may be enhanced by including pre-trained deep learning frameworks tailored for specific data types (e.g., BERT for text, CNNs for pictures, and RNNs or LSTMs for audio). This method would enable the model to discern complex associations between variables across modalities, enhancing the depression classification job. Furthermore, employing multimodal data augmentation methods, including cross-modal data synthesis, would rectify possible imbalances within each modality and establish a fairer foundation for model training.

Future research should investigate multimodal interpretability strategies to guarantee that insights from each modality significantly enhance the categorization result. Multimodal SHAP or Layer-wise Relevance Propagation (LRP) can clarify how certain features such as words, voice intonations, or visual cues collectively influence model predictions. The advancements in multimodal modeling might significantly boost the interpretability, flexibility, and resilience of depression classification systems, hence facilitating a more holistic approach to recognizing and comprehending mental health issues. Integrating these sophisticated multimodal approaches will enhance the existing framework, enabling it to tackle the intricacies of real-world mental health data more effectively.

Other than that, future research may also concentrate on classifying depression into two categories only, either “depression” or “non-depression” to avoid the issue of the enormous imbalance in data sources, especially the “non-depression”. The researcher may also exclusively classify depression into the categories of "Mild," "Moderate," and "Severe," omitting the "None" (non-depressed) category to yield more precise and clinically pertinent outcomes. Excluding the "None" category, often overrepresented in datasets and contributing to class imbalance, allows researchers to allocate model resources more effectively toward understanding and differentiating between various levels of depression severity. This approach may simplify the classification process and enable the model to focus on the nuances of depression rather than being distracted by non-depressed cases. Models specifically tailored for depression levels may offer more nuanced insights into variations among depressive states, supplying detailed data for healthcare practitioners and potentially enhancing the efficacy of targeted interventions. Prioritizing these three categories of depression may enable future research to create a specialized classification framework that better aligns with clinical needs, thereby facilitating more accurate assessments and interventions for individuals experiencing different intensities of depressive symptoms.

5.0 CONCLUSION

This study has demonstrated the effectiveness of a comprehensive approach in addressing the challenges of classifying depression levels using online social network (OSN) data. The model was trained on a balanced dataset through meticulous preprocessing, including the implementation of ADASYN for data augmentation, mitigating the impact of imbalances in depression severity. The evaluation criteria used four distinct models strategically focused on Recall and F1-Score, which were crucial for accurately identifying varying levels of depression. XGBoost emerged as the most robust performer, achieving an impressive F1-Score, showcasing its proficiency in classifying depression across multiple categories, enhancing the dataset size, exploring alternative OSN platforms such as Reddit, testing additional classifiers or delving into deep learning methods, and experimenting with further data augmentation techniques stand out as promising avenues for refining and expanding the scope of depression classification models.

REFERENCES

[1] Institute for Public Health, “National Health and Morbidity Survey (NHMS) 2023: Non-communicable Diseases and Healthcare Demand - Key Findings,” 2024, *Institute for Public Health, Shah Alam*.

[2] R. Abd Rahman, K. Omar, S. A. Mohd Noah, and M. S. N. Mohd Danuri, “A Survey on Mental Health Detection in Online Social Network,” *Int J Adv Sci Eng Inf Technol*, vol. 8, no. 4–2, pp. 1431–1436, Sep. 2018, doi: 10.18517/ijaseit.8.4-2.6830.

[3] L. Parkitny and J. McAuley, “The Depression Anxiety Stress Scale (DASS),” *J Physiother*, vol. 56, no. 3, p. 204, 2010, doi: 10.1016/S1836-9553(10)70030-8.

[4] L. Soleimani, K. A. B. Lapidus, and D. V. Iosifescu, “Diagnosis and Treatment of Major Depressive Disorder,” Feb. 2011. doi: 10.1016/j.ncl.2010.10.010.

[5] S. Ndaba, “Class Imbalance Handling Techniques used in Depression Prediction and Detection,” *International Journal of Data Mining & Knowledge Management Process*, vol. 13, no. 1/2, pp. 17–33, Mar. 2023, doi: 10.5121/ijdkp.2023.13202.

[6] S. H. Lovibond and P. F. Lovibond, *Manual for the Depression Anxiety Stress Scales*, 2nd ed. Sydney: Psychology Foundation of Australia, 1995.

[7] X. Li, D. T. L. Shek, and E. Y. W. Shek, “Psychological Morbidity Among University Students in Hong Kong (2014–2018): Psychometric Properties of the Depression Anxiety Stress Scales (DASS) and Related Correlates,” *Int J Environ Res Public Health*, vol. 18, no. 16, p. 8305, 2021, doi: 10.3390/ijerph18168305.

[8] K. Marfoh, E. Okyere, P. Kushigbor, and F. Acheampong, “Validation of Depression, Anxiety and Stress Scale (DASS-21) Among Healthcare Workers During the Outbreak of Delta Variant of SARS-CoV-2 in Ghana,” *F1000Res*, vol. 12, p. 229, 2023, doi: 10.12688/f1000research.130447.1.

[9] M. L. P. Le, T. Tran, S. Holton, N. T. Huong, R. Wolfe, and J. Fisher, "Reliability, Convergent Validity and Factor Structure of the DASS-21 in a Sample of Vietnamese Adolescents," *PLoS One*, vol. 12, no. 7, p. e0180557, 2017, doi: 10.1371/journal.pone.0180557.

[10] I. Mansoor, M. A. Khan, and K.- Kubra, "Predictive Relationship of Perceived Stressors and Mental Health of Medical Students," *Life and Science*, vol. 3, no. 3, p. 7, 2022, doi: 10.37185/lns.1.1.218.

[11] L. T. Lam, "The Roles of Parent-and-Child Mental Health and Parental Internet Addiction in Adolescent Internet Addiction: Does a Parent-and-Child Gender Match Matter?," *Front Public Health*, vol. 8, 2020, doi: 10.3389/fpubh.2020.00142.

[12] A. Thiyagarajan, T. G. James, and R. R. Marzo, "Psychometric Properties of the 21-Item Depression, Anxiety, and Stress Scale (DASS-21) Among Malaysians During COVID-19: A Methodological Study," *Humanit Soc Sci Commun*, vol. 9, no. 1, 2022, doi: 10.1057/s41599-022-01229-x.

[13] E. Moya, L. M. Larson, R. C. Stewart, J. Fisher, M. N. Mwangi, and K. S. Phiri, "Reliability and Validity of Depression Anxiety Stress Scale (DASS)-21 in Screening for Common Mental Disorders Among Postpartum Women in Malawi," *BMC Psychiatry*, vol. 22, no. 1, 2022, doi: 10.1186/s12888-022-03994-0.

[14] S. Islam, Md. Reza-A-Rabby, and G. K. Chabra, "Immediate Psychological Responses & Associated Demographic Factors During the Lockdown Period of the COVID- 19 Outbreak Among the Bangladeshi University Students," *Mind and Society*, vol. 11, no. 04, pp. 68–74, 2023, doi: 10.56011/mind-mri-114-20228.

[15] N. Mohamad, "Assessing Mental Health Outcomes in Quarantine Centres: A Cross-Sectional Study During COVID-19 in Malaysia," *Healthcare*, vol. 11, no. 16, p. 2339, 2023, doi: 10.3390/healthcare11162339.

[16] M. Ali, S. S. Hasan, R. Iftikhar, M. U. Fayyaz, and F. A. Anjum, "Mental and Physical Health Correlates of the Psychological Impact of the First Wave of COVID-19 Among General Population of Pakistan," *Front Psychol*, vol. 13, 2022, doi: 10.3389/fpsyg.2022.942108.

[17] C. Wang *et al.*, "The Impact of the COVID-19 Pandemic on Physical and Mental Health in the Two Largest Economies in the World: A Comparison Between the United States and China," *J Behav Med*, vol. 44, no. 6, pp. 741–759, 2021, doi: 10.1007/s10865-021-00237-7.

[18] C. Wang *et al.*, "The Impact of COVID-19 Pandemic on Physical and Mental Health of Asians: A Study of Seven Middle-Income Countries in Asia," *PLoS One*, vol. 16, no. 2, p. e0246824, 2021, doi: 10.1371/journal.pone.0246824.

[19] H. Kaur, H. S. Pannu, and A. K. Malhi, "A Systematic Review on Imbalanced Data Challenges in Machine Learning," *ACM Comput Surv*, vol. 52, no. 4, pp. 1–36, Jul. 2020, doi: 10.1145/3343440.

[20] D. Elreedy and A. F. Atiya, "A Comprehensive Analysis of Synthetic Minority Oversampling Technique (SMOTE) for handling class imbalance," *Inf Sci (N Y)*, vol. 505, pp. 32–64, Dec. 2019, doi: 10.1016/j.ins.2019.07.070.

[21] J. Brownlee, "Imbalanced Classification with Python: Choose Better Metrics, Balance Skewed Classes, and Apply Cost-Sensitive Learning," 2020.

[22] I. Dey and V. Pratap, "A Comparative Study of SMOTE, Borderline-SMOTE, and ADASYN Oversampling Techniques using Different Classifiers," in *Proceedings - 2023 3rd International Conference on Smart Data Intelligence, ICSMDI 2023*, Institute of Electrical and Electronics Engineers Inc., 2023, pp. 294–302. doi: 10.1109/ICSMDI57622.2023.00060.

[23] J. Brandt and E. Lanzén, "A Comparative Review of SMOTE and ADASYN in Imbalanced Data Classification," Uppsala, 2020.

[24] N. Mahendran and D. R. Vincent, "Effective Classification of Major Depressive Disorder Patients Using Machine Learning Techniques," *Recent Patents on Computer Science*, vol. 12, no. 1, pp. 41–48, Jan. 2019, doi: 10.2174/2213275911666181016160920.

[25] J. Zhang and Y. Guo, “Multilevel depression status detection based on fine-grained prompt learning,” *Pattern Recognit Lett*, vol. 178, pp. 167–173, Feb. 2024, doi: 10.1016/j.patrec.2024.01.005.

[26] M. Fang, S. Peng, Y. Liang, C.-C. Hung, and S. Liu, “A multimodal fusion model with multi-level attention mechanism for depression detection,” *Biomed Signal Process Control*, vol. 82, p. 104561, Apr. 2023, doi: 10.1016/j.bspc.2022.104561.