Developing Smart Process for Predicting Mental illness like Depression, Anxiety, and Stress using Machine Learning Methods

By

Rabia Riaz

(Registration No: 00000401810)

Department of Sciences

School of Interdisciplinary Engineering & Sciences National University of Sciences & Technology (NUST)

Islamabad, Pakistan

(August 2024)

Developing Smart Process for Predicting Mental illness like Depression, Anxiety, and Stress using Machine Learning Methods

By

Rabia Riaz

(Registration No: 00000401810)

A thesis submitted to the National University of Sciences and Technology, Islamabad, in partial fulfillment of the requirements for the degree of

Masters of Science in
Bioinformatics

Supervisor: Dr. Zamir Hussain

School of Interdisciplinary Engineering & Sciences

National University of Sciences & Technology (NUST)

Islamabad, Pakistan

(August 2024)

Thesis Acceptance Certificate

Annex A to NUST Letter No.
0972/102/Exams/Thesis-Cert
dated 23 Dec 16.

THESIS ACCEPTANCE CERTIFICATE

Certified that final copy of MS/MPhil thesis written by **Ms. Rabia Riaz** Registration No. **00000401810** of **SINES** has been vetted by undersigned, found complete in all aspects as per NUST Statutes/Regulations, is free of plagiarism, errors, and mistakes and is accepted as partial fulfillment for award of MS/MPhil degree. It is further certified that necessary amendments as pointed out by GEC members of the scholar have also been incorporated in the said thesis.

Signature with stamp: ________ Associate Professor SINES - NUST, Sector H-12 Islamabad

Name of Supervisor: Dr. Zamir Hussain

Date: Aug 16, 2024

For HoD Sci

Signature of HoD with stamp: ________ Dr. Mian Ilyas Ahmad HoD Engineering Professor SINES - NUST, Sector H-12 Islamabad

Date: 20/08/2024

Countersign by

Signature (Dean/Principal): ________ Principal SINES

Date: 20/08/2024

AUTHOR'S DECLARATION

I Rabia Riaz hereby state that my MS thesis titled "Developing Smart Process for Predicting Mental illness like Depression, Anxiety, and Stress using Machine Learning Methods" is my own work and has not been submitted previously by me for taking any degree from National University of Sciences and Technology, Islamabad or anywhere else in the country/ world. At any time if my statement is found to be incorrect even after I graduate, the university has the right to withdraw my MS degree.

Name of Student: Rabia Riaz

Date: 18-07-2024

PLAGIARISM UNDERTAKING

I solemnly declare that research work presented in the thesis titled "Developing Smart Process for Predicting Mental illness like Depression, Anxiety, and Stress using Machine Learning Methods" is solely my research work with no significant contribution from any other person. Small contribution/ help wherever taken has been duly acknowledged and that complete thesis has been written by me.

I understand the zero-tolerance policy of the HEC and National University of Sciences and Technology (NUST), Islamabad towards plagiarism. Therefore, I as an author of the above titled thesis declare that no portion of my thesis has been plagiarized and any material used as reference is properly referred/cited.

I undertake that if I am found guilty of any formal plagiarism in the above titled thesis even after award of MS degree, the University reserves the rights to withdraw/revoke my MS degree and that HEC and NUST, Islamabad has the right to publish my name on the HEC/University website on which names of students are placed who submitted plagiarized thesis.

Student Signature: Rabia
Name: Rabia Riaz

DEDICATION

I dedicate this thesis to my beloved parents. To my late father, whose memory continues to inspire and guide me, and to my mother, whose strength and sacrifices have been my foundation, I am eternally grateful.

ACKNOWLEDGEMENTS

My profound gratitude is directed towards Almighty Allah for the divine blessings that sustained me throughout the demanding yet fulfilling pursuit of this research. His unwavering support empowered me to navigate challenges and ultimately achieve my desired goals, resulting in the successful completion of this project.

I extend my deepest gratefulness to my esteemed supervisor, Dr. Zamir Hussain, whose support, insightful guidance, and profound knowledge were helpful in the successful completion of this research. His constant motivation, enthusiasm, and patience were indispensable throughout the entire process. Undoubtedly, his dedication to research serves as a perpetual source of inspiration for me.

Furthermore, I acknowledge the pivotal role played by the School of Interdisciplinary Engineering and Sciences (SINES), NUST, in providing the necessary research infrastructure and enabling environment for this project. I am deeply indebted to the members of my Guidance and Examination Committee (GEC), Dr. Zartasha Mustansar, Dr. Masood Ur Rehman Kayani, and Dr. Qurrat Ulain Hamdan (External), whose expert guidance, critiques, and support proved invaluable throughout the research process.

Table of Contents

LIST OF TABLES

Page No.

LIST OF FIGURES

Page No.

LIST OF SYMBOLS, ABBREVIATIONS AND ACRONYMS

AI	Artificial Intelligence
ML	Machine Learning
DASS21	Depression, Anxiety and Stress Scale 21
χ^2	Chi-square Statistic
C	Contingency Coefficient
RFE	Recursive Feature Elimination
DT	Decision Tree
RF	Random Forest
SVM	Support Vector Machines
NB	Naive Bayes
KNN	k-Nearest Neighbors

ABSTRACT

Depression, anxiety, and stress are growing problems worldwide with 3.8% of people experiencing depression and 4.4% dealing with anxiety. These concerns are particularly widespread in Pakistan, affecting 20 million individuals in the country. The treatment of mental illness is hindered by stigma, a scarcity of mental healthcare resources, prolonged consultations, and the expensive fees associated with consulting psychologists and psychiatrists. The objective of this study is to develop machine learning (ML) prediction models for mental illness issues like depression, anxiety and stress among Pakistani students. The Depression Anxiety Stress Scale 21 (DASS21) is utilized to gather data from 115 students. A study has introduced a novel process of data generation considering the utilization of multinomial probability distribution with correlation. The use of chi-square test and Recursive Feature Elimination (RFE) with random forest reveals that all 21 features of DASS21 (depression, anxiety, and stress) exhibit statistical significance. The prediction models are developed using five machine learning algorithms i.e., random forest, decision tree, support vector Machines, naive bayes, and k-nearest neighbors. A comparison reveals SVM outperforms other models with an F1 Score of 0.97, 0.96 and 0.97 for depression, anxiety and stress. The proposed models could be used in a practical framework to facilitate the quick screening decision support system for depression, anxiety, and stress.

Keywords: Depression, Anxiety, Stress Scale 21 (DASS21); Support Vector Machines (SVM); Feature Selection; Wrapper Method; Filter Method

1. INTRODUCTION

Mental illness issues like depression, anxiety, and stress are becoming an important concern in the setting of higher academic pursuits, with consequences that extend far beyond the classroom [1]. A sizable fraction of the student body reports symptoms are associated with these disorders. These symptoms tend to appear in a variety of ways, ranging from persistent despair and hopelessness to extreme stress and tension [2]. Unresolved mental health issues in higher education have far-reaching consequences [3]. They influence academic performance, overall quality of life, and the overall well-being of the campus community early detection and action are essential solutions to these problems. Universities can assist students manage their mental health and academic development by identifying signs and giving support at an early stage [1].

1.1 The Global Impact of Mental Health Issues

The World Health Organization (WHO) estimates that 350 million people worldwide suffer from depression, and that number is increasing [4]. Depression is expected to become the leading worldwide illness burden by 2030, surpassing heart disease and cancer, according to the WHO. It is expected to be the second leading cause of illness by 2020.

1.2 Mental Illness Issues in Pakistan

Mental illness is a significant public health issue in Pakistan affects up to 10% of Pakistan’s population, approximately 20 million people, and is associated with a significant economic burden [5]. According to the WHO, with an estimated prevalence of 16% to 25% among adults and 2% to 10% among children and adolescents [6]. In 2023, Pakistan was ranked ninth out of 177 nations in terms of the worldwide burden of mental diseases .The number of university students facing mental health issues is on the rise. Mental illness is significantly more prevalent among students globally compared to the general population. This is supported by the American College Health Association's findings, which revealed that over 40% of college students have experienced significant depression and over 60% have dealt with overwhelming anxiety in the past year [7]. University students are more likely to experience mental health challenges compared to the general population. This vulnerability stems from the unique pressures they face during this pivotal transition period.

1.2.1 *Challenges in Mental healthcare*

Pakistan has around 400 certified psychiatrists, most of who work in cities. However, district psychiatrist positions have been created nationwide [5]. These psychiatrists mainly practice alone, while larger centers are moving towards multidisciplinary treatment. The current generation of psychiatrists is advocating for a team-based strategy that includes outreach clinics, community service development, and improving mental health literacy, especially in rural areas. According to the WHO's Mental Health Atlas 2017, Pakistan has four large psychiatric hospitals, 344 residential care facilities, and 654 psychiatric units in general hospitals, totaling 2.1 mental health beds per 100,000 inhabitants. The nation has 3,729 outpatient mental health institutions, with 343.34 patients per 100,000 people. The average number of contacts per user is 9.31. Forty-six percent of outpatient clinics offer community follow-up treatment, while only one percent has mobile mental health teams. There are available community-based psychiatric inpatient facilities, with 1.926 beds per 100,000 people. In conclusion, Pakistan faces challenges in providing adequate healthcare services. There is a need for an effective support system for prediction of state of mental health. Improving mental health infrastructure and using predictive techniques are critical to meeting the population's increasing mental health requirements [8].

1.3 Types of Mental Illness

Mental health is a crucial part of our overall health and happiness. It encompasses our emotional, psychological, and social well-being, shaping our thoughts, feelings, and behaviors [9]. Just like physical health, mental well-being is an ongoing journey, with moments of peak wellness and challenges that may arise throughout life. There are several types of mental health conditions, each with a distinct set of symptoms. Some of the most common ones include:

- Depression

This is a persistent feeling of sadness, hopelessness, and a loss of interest in activities that were once enjoyable [10].

- Anxiety

This is characterized by excessive worry, uneasiness, and physical symptoms such as a rapid heartbeat or sweating [11].

- Stress

This is the body's response to a perceived threat or demand [12].

Unfortunately, these three conditions - depression, anxiety, and stress - are becoming increasingly common in modern life. Often overlooked or disregarded, they can have a significant impact on a person's daily life and relationships.

1.4 Comparison of Mental Illness Assessment Tools

The (Table 1.1) below presents a brief comparison of several mental illness evaluation tools, emphasizing their important features and target demographics.

Table 1.1: Comparisons of Mental Health Assessment Tools

Tool Name	**Format**	**Copyright Status**	**Target Population**	**Mental Illness issues**
GAD7(Generalized Anxiety Disorder Questionnaire)	Self-report questionnaire	Yes	Major country populations	Anxiety
K6 & K10 (Kessler Psychological Distress Scales)	Self-report & interviewer-administered	No	Diverse ethnic groups, validated in 21 languages	Nonspecific psychological distress

WHO5 Wellbeing Index	Self-report questionnaire	No	Age 9+, multiple translations	General mental well-being
GHQ (General Health Questionnaire)	Self-report questionnaire	Yes	Adults, not validated for children	Psychological distress
DASS21 and DASS42 (Depression, Anxiety, Stress)	Self-report questionnaire	No	Age 17+, translated into 50+ languages	Depression, anxiety, and stress

The DASS21 and DASS42 are unique in that they can assess depression, anxiety, and stress simultaneously, providing a comprehensive view of these effects [13]. The GAD7 focuses on anxiety while also screening for related conditions such as PTSD and panic attacks [14] [15]. The K6 and K10 are effective for identifying individuals at high risk for mental health issues, and they have been validated across multiple populations [16]. The WHO-5 places emphasis on overall mental well-being over specific mental health disorders. The GHQ is helpful for evaluating overall psychological distress but does not pinpoint specific conditions [16] [17]. The DASS21 is well-suited for use in university student populations due to its comprehensive assessment capabilities, shorter duration compared to the DASS-42, and free availability without copyright restrictions [13].

1.5 Depression, Anxiety and Stress Scale 21 (DASS21)

The Depression, Anxiety, and Stress Scale (DASS21) are a widely used self-report tool for assessing three emotional states: depression, anxiety, and stress. It was developed in the early 1990s by psychologists Peter Lovibond and Sydney Lovibond at the University of New South Wales, as a simplified version of the original DASS-42 [19]. The DASS21 consists of 21 items divided into three scales, each containing 7 items. The

Depression scale assesses feelings of despair and disinterest, the Anxiety scale measures autonomic arousal and situational anxiety, and the Stress scale evaluates chronic non-specific arousal and difficulties relaxing.

The DASS21 is well-known for its good psychometric qualities, and it is commonly used in clinical and research contexts to measure the intensity of emotional states. All 21 items of the DASS are defined in the (Table 1.2).

Table 1.2: DASS21 Items

Sr. no.	Code	Questions	Scales
1	DASS1	I found it hard to wind down (Found it hard to relax)	S
2	DASS2	I was aware of dryness of my mouth	A
3	DASS3	I couldn't seem to experience any positive feeling at all	D
4	DASS4	I experienced breathing difficulty (eg, excessively rapid breathing, breathlessness in the absence of physical exertion)	A
5	DASS5	I found it difficult to work up the initiative to do things	D
6	DASS6	I tended to over-react to situations	S
7	DASS7	I experienced trembling (eg, in the hands)	A
8	DASS8	I felt that I was using a lot of nervous energy	S
9	DASS9	I was worried about situations in which I might panic and make a fool of myself	A
10	DASS10	I felt that I had nothing to look forward to	D
11	DASS11	I found myself getting agitated	S
12	DASS12	I found it difficult to relax	S
13	DASS13	I felt down-hearted and blue	D
14	DASS14	I was intolerant of anything that kept me from getting on with what I was doing	S
15	DASS15	I felt I was close to panic	A
16	DASS16	I was unable to become enthusiastic about anything	D
17	DASS17	I felt I wasn't worth much as a person	D
18	DASS18	I felt that I was rather touchy	S

19	DASS19	I was aware of the action of my heart in the absence of physical exertion (eg, sense of heart rate increase, heart missing a beat)	A
20	DASS20	I felt scared without any good reason	A
21	DASS21	I felt that life was meaningless	D

In table 1.2, A represents anxiety, D represents depression, and S represents stress.

1.6 Overview of the SINES, NUST

The National University of Sciences & Technology (NUST) in Pakistan is renowned for its extensive and varied academic offerings. NUST has positioned itself as a prominent institution of higher education in the region, with seven campuses in five cities. It offers a diverse student population, including both National and international students. The university's dedication to inclusivity is demonstrated via its numerous scholarships based on financial need and significant partnerships with various industries, guaranteeing a nurturing and rewarding atmosphere for every student. NUST's extensive influence and significance position it as a crucial institution in molding the educational framework of Pakistan.

The School of Interdisciplinary Engineering & Sciences (SINES) is a prestigious research institute that was established in 2007 within this recognised university. SINES is well-known for its cutting-edge supercomputing infrastructure, which previously secured its position on the prestigious list of the top 500 supercomputers globally. This advanced tool offers a substantial benefit to its technical studies, creating a favourable atmosphere for groundbreaking research and invention. SINES employs a multidisciplinary approach that combines the fields of Computational Sciences and Computational Engineering. This integration facilitates a full educational experience, equipping students with the skills and knowledge necessary to address intricate, practical issues. The school's education, training, and research activities are carefully categorized into these two primary areas, guaranteeing a curriculum that is both concentrated and comprehensive. SINES provides a variety of academic programs, encompassing Bachelor's (BS), Master's (MS), and Doctoral (PhD) degrees. The staff and students at SINES are actively involved in research that makes a substantial contribution to the socio-economic development of the country.

SINES is choose as the subject of this research because of its multidisciplinary character, which facilitates collaboration between students and faculty members from both Computational Sciences and Computational Engineering. The combination of these different fields creates a cooperative atmosphere that is perfect for groundbreaking research. Furthermore, my research specifically centers on forecasting mental illness in the adult demographic, and the availability of local data was of utmost importance.

1.7 Problem Statement

The adult mental health remains a serious public health issues due to cultural a stigma, lack of awareness, and limited mental healthcare resources. Current intervention options are often time-consuming and expensive, contributing to the increasing prevalence of early-onset mental health issues. Undiagnosed mental health issues among students may lead to serious consequences such as addiction, suicidal behavior, and personality disorders later in life. Therefore, it is crucial to improve the accessibility, convenience, and effectiveness of mental health interventions. One potential approach is to enhance existing mental health screening methods, particularly for mental health using ML algorithms. These advancements could accelerate the development of intelligent, data-driven decision support systems, enabling healthcare providers to make prompt, informed decisions based on local data. Because of their low cost and adaptability, these technologies may gain widespread adoption and accessibility, thereby improving patient outcomes.

1.8 Objectives

The objectives of the study will be:

- Descriptive analysis of the state of mental illness in Pakistani students with respect to gender, age, semester, academic level and accommodation status.
- Generation of hybrid synthetic data based on the multinomial probability distributional of the DASS21 features to overcome the constraints of small

dataset size.

- Development of predictive models for the state of mental illness using ML methods.

1.9 Relevance to National Need

Mental illness is a major public health concern in Pakistan, impacting around 20 million people; however the true prevalence is likely higher due to underreporting and stigma. WHO estimates a prevalence of 16% to 25% among adults and 2% to 10% among children and adolescents, emphasizing the critical need for mental health treatments.

Counseling, coaching, and career advisory centers (C3A) are essential for meeting Pakistani student's mental health requirements. This center is established in 2008 in National University of Sciences and Technology (NUST), offer a designated area where students can go for support while navigating the challenges associated with mental health. C3A provide a secure and encouraging atmosphere for people to seek treatment and support due to the rising trend of mental illness among students. By providing a variety of services, including as workshops, training programmers, and counseling sessions. C3A centers have made significant contributions, one of which is to lessen the stigma attached to mental diseases.

The predictive model of mental assessment in learners can contribute to the attainment of both SDG3 - Good Health and Well-being and SDG9 - Industry, Innovation, and Infrastructure - in the following ways:

- SDG3 - Good Health and Well-Being

  The predictive model can assist students improve their mental health results. Early identification of students at risk of mental health difficulties allows for interventions and support to address these concerns. This can result in increased general well-being, less stigma surrounding mental health, and improved access to mental health treatments, all of which are essential for attaining SDG3.

- SDG9 - Industry, Innovation, and Infrastructure

  Although the predictive model has no direct link to SDG9, it can indirectly promote the objective through technical innovation and infrastructural development. The creation and application of the prediction model are examples of innovation in the field of mental health evaluation. This breakthrough could open the path for technological breakthroughs in mental health support in educational settings. Furthermore, the predictive model's deployment may need infrastructure enhancements in terms of data collecting, storage, and analysis, contributing to SDG9's goal of resilient infrastructure.

1.10 Thesis Structure

The work is structured according to a comprehensive framework to accomplish the objectives outlined in Section 1.7. In Chapter 2, we conducted a review of the literature on AI-driven prediction of mental illness in young populations and identified potential areas for further research. Chapter 3 outlines the methodology of the study, while Chapter 4 compares the findings with existing literature. Finally, Chapter 5 provides a summary of the research, key findings, identifies limitations, and proposes solutions to address them.

2. LITERATURE REVIEW

Machine Learning (ML) has revolutionized industries such as healthcare, banking, e-commerce, manufacturing, and others, enabling increased efficiency, productivity, and decision-making [20]. ML has helped the development of data-driven prediction in mental health care. This helps mental healthcare professional for risk assessment at early stages [21].

ML is a subset of AI that entails the creation of algorithms and models that allow systems to learn from data. ML techniques are classified into two types: supervised learning and unsupervised learning. Supervised learning entails training a model on a labeled dataset, where the algorithm learns from input data and the labels associated with the output. It is used for classification and regression problems [22]. Unsupervised learning, on the other hand, works with unlabeled data, with the algorithm attempting to understand the underlying structure or patterns within the data. This method is used for applications including as grouping, association, and anomaly detection [23] [24].

The subsequent sections explain current literature that focuses on the identification of mental illness in young population. The literature review is divided into two primary categories: International level studies and local level studies.

2.1 ML for Mental Health Issues

2.1.1 *International Level*

Nayan et al. investigated the use of several ML algorithms to predict mental illness in university students [25]. A systematic questionnaire-based online survey was conducted among 2,121 university students from both private and public institutions in Dhaka, Bangladesh. After providing informed permission, participants completed a web-based survey that included sociodemographic information and behavioral assessments such as the Patient Health Questionnaire (PHQ-9) and the Generalised Anxiety Disorder Assessment-7 (GAD-7). The research used six well-known ML algorithms: logistic regression, random forest (RF), Support Vector Machines (SVM), linear discriminant analysis, K-nearest neighbours (KNN), and Naïve Bayes (NB) to predict mental illness among students. The sample comprised of 45% male and 55% female students, with roughly 76.9% aged 21 to 25 years. The findings suggested that women had a greater incidence of severe depression and anxiety than males. Among the models tested, the RF algorithm had the best accuracy for predicting depression (89%), while the SVM algorithm excelled in predicting anxiety. Based on these results, the research advises using RF and SVM algorithms to predict mental health status among Bangladeshi university students.

Khan et al. used several ML algorithms to predict mental problems in Bangladeshi individuals [26]. The research used a dataset of 466 mental health patients to determine the correlations between diagnosis and qualities. Three ML techniques—RF, SVM, and KNN—were used to diagnose mental health disorders, and their performance was evaluated using different accuracy criteria. The experimental findings showed that RF outperformed the other methods. In the initial dataset, factors such as age, marital status, department unit, gender, and symptoms were deemed independent features, with 29 illness kinds serving as target classes. RF has the greatest accuracy of 0.851 when compared to SVM (0.787), Decision Tree (0.723), and KNN (0.765).

A study conducted in Uzbekistan in 2022 that developed a viable ML based prediction model for perceived stress using real-world data collected from an online survey of 444 university students from various ethnic origins [27]. The prediction models were built using supervised ML algorithms. Principal Component Analysis (PCA) and the chi-squared test were used to reduce the number of features. Grid Search Cross-Validation (GSCV) and the Genetic Algorithm (GA) were used to do hyperparameter optimization (HPO).The study's findings revealed that around 11.26% of respondents had significant levels of social stress. Concerning, over 24.10% of participants were found to be suffering extremely high levels of psychological stress, indicating serious implications for students' mental health. With an accuracy of 80.5%, precision of 1.000, F1 score of 0.890, and recall value of 0.826, the ML models produced encouraging prediction results. When paired with PCA for feature reduction and GSCV for HPO, the Multilayer Perceptron model demonstrated the maximum accuracy.

In 2023, the researchers in Turkey developed a viable ML based prediction model for perceived stress prediction and validate it with real-world data obtained through an online survey of 444 university students of various races [28]. The supervised ML methods were used to build the ML models. As feature reduction strategies, they used PCA and the chi-squared test. In addition, for HPO, they used GSCV and the GA. According to the data, roughly 11.26% of persons had significant levels of social stress, while approximately 24.10% experienced extremely high psychological stress, indicating a serious condition for the students' mental health. Furthermore, the ML models' prediction results revealed the highest accuracy (80.5%), precision (1.000), F1 score (0.890), and recall value (0.826). They discovered that the Multilayer Perceptron model performed best when paired with PCA as a feature reduction strategy and GSCV for HPO.

ML approaches were also used to predict mental well-being in children. For this they obtained a dataset contained 60 instances [29]. For the classification process, they selected various features and attributes. The average one-dependence estimator (AODE) ML technique achieved 71% accuracy. Meanwhile, Multi-Layer Perceptron (MLP) had the highest accuracy (78%). The logical analysis tree (LAT) came in second with 70% accuracy, followed by the multiclass classifier with 58% accuracy. Another ML technique, radial basis function network (RBFN), achieved 57% accuracy.

Srividya et al. investigated how to predict mental health statuses using different ML techniques. This study utilized different ML methods including SVM, decision trees (DT), NB classifiers, KNN classifiers, and logistic regression (LR) to assess mental wellness [21]. Target populations included high school and college students, as well as working professionals. The questionnaire results were first analyzed using unsupervised learning methods. Clustered labels were verified using the Mean Opinion Score. Experiments show that SVM, KNN, and RF perform almost identically. The usage of ensemble classifiers dramatically improved performance Srividya et al. investigated how to predict mental health statuses using different ML techniques.

2.1.2 *Few Local Studies*

Mental health is a major concern among students in Pakistan, with about 33,000 mental health issues reported annually within the student community [30]. This covers certain problems including stress, sadness, and anxiety. According to the study conducted in University of Management and Technology, Lahore, 39% of the students reported experiencing stress, 36% expressed anxiety, and 25% claimed depression. These disorders can significantly impact a student's well-being and academic performance.

A study conducted in Lahore in 2022 by Lahore Garrison University on the prevalence of depression and anxiety among university students using ML methods [31]. They used the KNN method, which is well-known for detecting and analyzing mental distress and anxiety. The researchers collected data of 1366 individuals on numerous psychological factors. Following that, they used the acquired data to apply the KNN algorithm to predict the chances of sadness and anxiety. After running the studies; they discovered that the initial accuracy of the predictions obtained without using PCA was 76.5%. This means that the model correctly recognized depression and anxiety cases in the study population 76.5% of the time. However, by applying PCA, a statistical approach used to reduce data dimensionality, they were able to improve accuracy. The findings suggest that using the KNN algorithm in conjunction with PCA can be effective in identifying and predicting depression and anxiety among university students, potentially providing valuable insights for developing intervention strategies and support systems to address mental health challenges in this population.

2.2 Gap in the Literature

Previous studies on state of mental illness have often been constrained by small datasets due to privacy concerns. People are hesitant to share their personal information, leading to limited sample sizes. This hinders the generalizability and reliability of research findings. To address this gap, this research introduces the generation of hybrid synthetic data. By using statistical distribution such as probability distribution, a comprehensive synthetic dataset is generated that replicates the characteristics of the original data while maintaining privacy. This approach not only overcomes the challenge of data scarcity but also enhances the validity and reliability of the analysis, providing a more comprehensive and accurate representation of the state of mental illness.

3. METHODOLOGY

This research followed a conventional approach to ML encompassing data collection, data preprocessing, feature selection and model training. A primary local dataset is collected using Google form which is then pre-processed and labeled according to DASS21 scoring guidelines. Following this, a through feature selection is conducted incorporating both statistical and computational strategies. Subsequently, the data are employed to train and test five ML algorithms, covering classical and ensemble approaches. The performance metrics are evaluated for each algorithm to identify the best performing model which can be converted to decision support system. The complete workflow is illustrated in the (Figure 3.1).

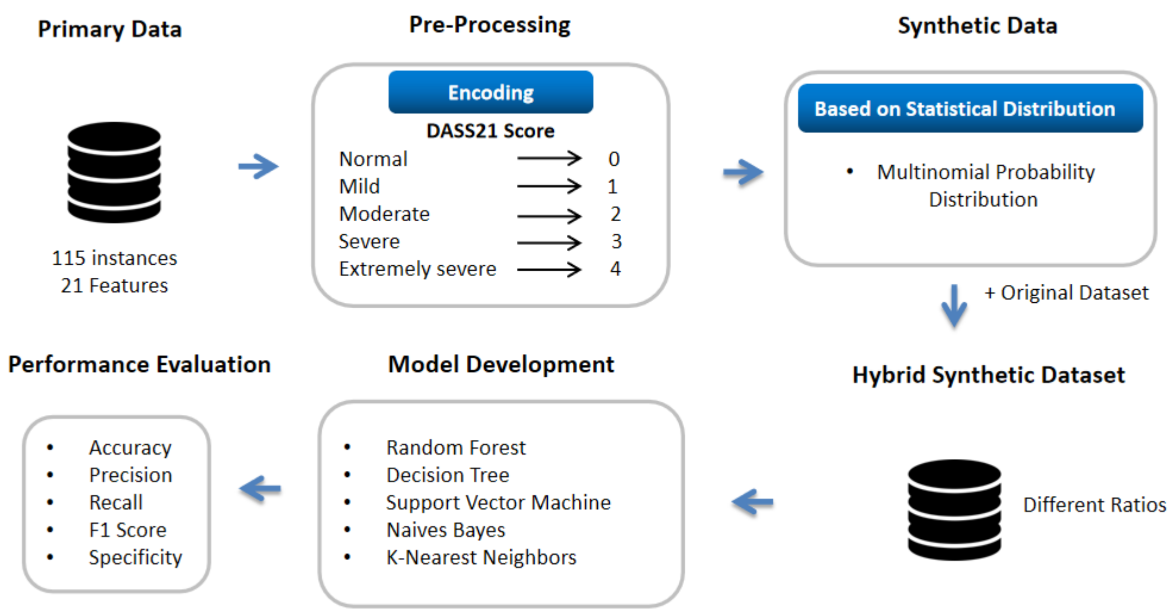


Figure 3.1: Workflow of the Research Methodology

3.1 Data collection

3.1.1 *Ethical Considerations*

The data collection for this research is done with a significant focus on ethical issues. Participants are told that the survey is done by a researcher from the National University of Sciences and Technology's (NUST) School of Interdisciplinary Engineering and Sciences (SINES). Before beginning the survey, participants were given thorough information on the study's aims, voluntary participation, confidentiality, and data security. They were advised that their participation is fully voluntary, and that they could withdraw at any moment without reason.

3.1.2 *Survey Development*

The survey was divided into two parts: a generalized demographic questionnaire and the DASS21 questionnaire. The demographic part attempted to collect information on age, gender, educational level, and accommodation status. The DASS21 questionnaire, a well-known instrument for measuring depression, anxiety, and stress, is included due to its reliability and validity, especially in academic contexts. Participants are made aware that the DASS21 is a self-report instrument and not a replacement for clinical diagnosis. They are recommended to seek professional treatment if they are having serious issues, while also being made aware of the survey's limitations and the significance of specialist mental health services.

3.1.3 Data Collection

Data is collected via a Google Form between February and May 2024. The forms are distributed to SINES, NUST students. Participants who were actively enrolled at NUST are eligible for the research, however only 115 are included in the final analysis. The survey takes around 5-10 minutes to complete, and participants are requested to provide honest and accurate answers. This strategy aims to collect complete data on the prevalence and severity of anxiety, depression, and stress among the target group, using procedures similar to those employed in previous research.

3.2 Data Pre-Processing

All pre-processing was done using SPSS (Version 20) and Python (Version 3.10) [32]. Missing value analysis found that the data included no missing values. The DASS21 questionnaire consists of three scales: anxiety, depression, and stress. These were rated using the DASS21 scoring standards, and the final scores were computed and labeled [33]. There were five potential results for each scale: normal (0), mild (1), moderate (2), severe (3), and Extremely severe (4). This process is done to generate numeric categories or targets in the data for subsequent application of ML algorithms. The resulting data consists of 21 features, with 7 features for each scale. Below is the scoring range for each scale.

Table 3.1: Scoring Range for each Scale of DASS21

Level	Depression	Anxiety	Stress
Normal	0-4	0-3	0-7
Mild	5-6	4-5	8-9
Moderate	7-10	6-7	10-12
Severe	11-13	8-9	13-16
Extremely Severe	≥ 14	≥ 10	≥ 17

The Table 3.2 gives a full detail of the features after pre-processing, including feature names, descriptions, types (qualitative or quantitative), and values. The dataset contains demographic information such as age, gender, academic level, phase of study, and accommodation type. It also includes scores from the DASS21 depression, anxiety and stress scales, which each include seven questions. The severity levels for each scale are classified as normal, mild, moderate, severe, and very severe according to the DASS21 standards. This structured data will be utilized to create numerical categories for ML techniques to be applied later.

Table 3.2: Data Description after pre-processing

Feature	Description	Type	Values
Age	Age range of the participants	Qualitative	18-21, 22-25, 26-30, above 30
Gender	Gender of the participants	Qualitative	Male, Female, Do not want to disclose
Academic Level	Academic level of the participants	Qualitative	MS, BS, PhD
Phase of Study	Current phase of study	Qualitative	Research, Course work, Semester 1-4
Accommodation	Living arrangements of the participants	Qualitative	Hostelite inside campus, Outside campus, Day scholar
Depression	Depression scale item scores	Quantitative	Depression 1: 0-3
Anxiety	Anxiety Scale item scores	Quantitative	Anxiety 1: 0-3
Stress	Stress scale item scores	Quantitative	Stress1: 0-3
Depression Severity Level	Depression severity level based on DASS21 guidelines	Qualitative	Normal (0), Mild (1), Moderate (2), Severe (3), Extremely severe (4)
Anxiety Severity Level	Anxiety severity level based on DASS21 guidelines	Qualitative	Normal (0), Mild (1), Moderate (2), Severe (3), Extremely severe (4)
Stress Severity Level	Stress severity level based on DASS21 guidelines	Qualitative	Normal (0), Mild (1), Moderate (2), Severe (3), Extremely severe (4)

3.3 Data Analysis

3.3.1 *Descriptive Analysis*

To define data characteristics, demographics such as gender, age, academic level (BS, MS, PhD), semester, and of the research population accommodation status are examined. The sample consists of 115 students. Measures of central tendency, such as mean and mode, are determined for subject age, academic level, semester, and kind of accommodation. Furthermore, the gender distribution in the study population is investigated to offer a thorough perspective of the sample demographics. The study tries to find patterns and trends in the data, providing insights into the students' characteristics and living situations.

3.3.2 *Internal Consistency and Reliability*

To the best of our knowledge, the self-report version of the DASS21 has been utilized and validated on the Student Population in Pakistan [34]. This study aims to utilize self-report scale of the DASS21 to assess depression, anxiety, and stress scores among students of SINES, NUST, as well as to examine its internal consistency and reliability. The reliability analysis is carried out in SPSS by calculating Cronbach's alpha and analyzing inter-item correlations.

Cronbach's alpha is a statistical measure used to evaluate the internal consistency, reliability, and stability of surveys, especially in the fields of social, psychological, and biological sciences. It measures the degree of correlation between the elements in a questionnaire. The association is established by computing inter-item correlations among all items and assessing the level of Cronbach's alpha. Typically, magnitudes of these correlations should be within the range of 0.2 to 0.6.Values less than 0.2 show very weak correlation among items, implying that the questions are distinct and are diverging from the main emphasis of the questionnaire [35]. Values larger than 0.6 imply a strong correlation which shows that the items are substantially similar to each other. This renders them repetitious and redundant. Correlations between 0.2 and 0.6 ensure that the items of the questionnaire are broad enough to appropriately encompass the scope of the questionnaire while avoiding repetition. In SPSS, the interitem correlations will be generated as a 21x21 matrix, which will be difficult to interpret. Therefore, the inter-item correlations have been generated as heatmaps in Python using the same statistical approaches to ensure better and easier presentation. The magnitude of Cronbach's alpha runs from 0 to 1, where higher values indicate greater internal consistency, suggesting that the items are assessing the same underlying concept or domain of the questionnaire. Generally, a value between 0.7-0.9 is acceptable [36].

3.3.3 *Hypothesis Testing*

Hypothesis testing is a fundamental statistical concept in inferential analysis. It is used to make estimates about population parameters based on sample data and to determine the significance of an effect or relationship. The process begins with establishing two opposing hypotheses. The null hypothesis (H_0) suggests that there is no effect or relationship, and the alternative hypothesis (H_1 or Ha) asserts that the effect or relationship being studied does exist. The choice of the appropriate hypothesis is based on the evaluation of statistical significance [37].

In order to determine the statistical significance of a study, a statistical test and a level of significance need to be chosen. The choice of the test depends on the characteristics of the data and the type of inferential analysis being conducted. The standard level of significance (α) is typically set at 0.05, indicating a 5% probability of making an error in rejecting the researcher's claims. When statistical tests are performed, they generate a test statistic along with a corresponding p-value, which indicates the statistical significance of the analysis. If the p-value is less than α, the null hypothesis is rejected, supporting the existence of the relationship being investigated. On the other hand, if the p-value is greater than α, the null hypothesis cannot be rejected, thereby disproving the researcher's claim.

In this research, the goodness of fit for the 21 variables of the Depression, Anxiety, and Stress Scales (DASS21) was investigated via hypothesis testing utilizing the Chi-squared goodness-of-fit test. The following null and alternative hypotheses were established for each variable:

H_0: The observed data for the variable follows the specified multinomial distribution with the provided probability.
H_1: The observed data for the variable does not follow the predefined multinomial distribution with supplied probabilities.

The Chi-squared goodness-of-fit test was chosen as the test statistic because it is suited for comparing the observed frequencies of categorical data to the anticipated frequencies under a specific distribution. The significance threshold was set at 0.05, and the observed numbers for each category were compared to the predicted counts derived using the stated probabilities. This technique enables us to examine if the observed data for each variable fits the multinomial distribution with the prescribed probability.

3.4 Feature Selection

Prior to building a ML model, it's important to perform feature selection in order to identify the most significant features. This process is essential for improving model performance and reducing complexity. The dataset in question contains a total of 21 features. Two methods are used for feature selection; filter method viz. Chi-square Test and wrapper method viz. Recursive feature elimination (RFE) [38].

3.4.1 *Filter Method*

This method involves using statistical techniques, such as the Chi-Square test, to assess the relevance of features in relation to the target variable. Features with the highest Chi-Square scores are considered the most relevant [39].

The Chi-Square test is a statistical method used to examine the relationship between categorical variables. It provides several key values: the Chi-Square statistic (X^2), the p-value, and the contingency coefficient (C) [40] [41] .Below is the table indicating range of p value and contingency coefficient and strength of these relationship.

Table 3.3: p Value and Strength of Relationship

P value	Relationship Strength
< 0.05	Statistically Significant
>= 0.05	Not Statistically Significant

Table 3.4: Contingency values and Strength of Relationship

Contingency Values	Strength of Relationship
0.0-0.1	Weak
0.1-0.3	Normal
0.3-0.5	Moderate
0.5-0.6	Moderate to Strong
0.6-0.8	Strong
0.8-1.0	Very Strong

3.4.2 *Wrapper Method*

This method evaluates features by iteratively training models and assessing their performance. It utilizes Recursive Feature Elimination (RFE) to recursively remove the least important features and build models on the remaining features to identify the combination that yields the best performance [42].

3.5 Synthetic Data Generation

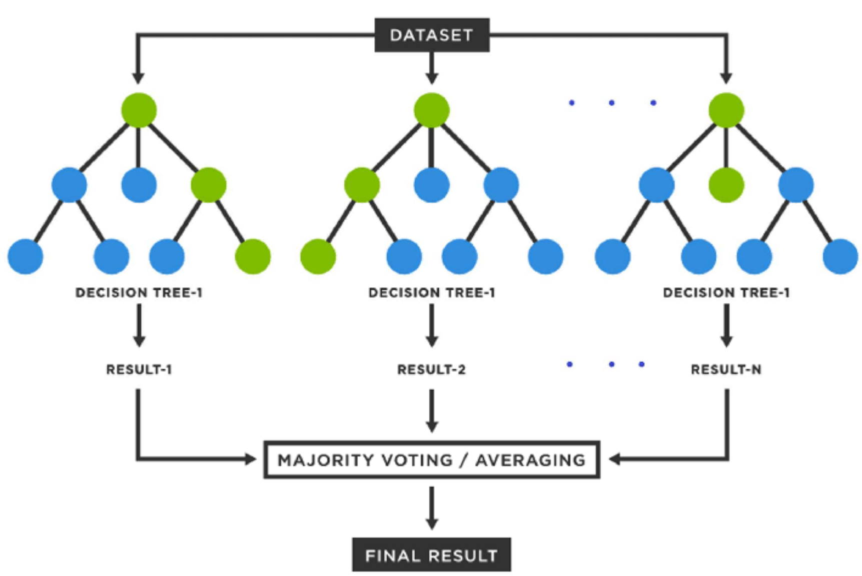
The dataset used in this research is small, with just 115 cases. Due to the difficulties in getting data, such as privacy and ethical considerations, synthetic data is generated to improve the models' resilience and adaptability [43]. Chi-square test is used to determine the quality of fit. We performed hypothesis testing to check goodness of fit. We assumed Null Hypothesis as data follows multinomial probability distribution and Alternative Hypothesis as data do not follow multinomial probability distribution. After checking goodness of fit, synthetic data was generated in R studio. It was generated for the depression, anxiety, and stress scales separately. A correlation between feature for each scale was calculated using pearson correlation, and we reported probabilities values for each outcome (0, 1, 2, 3). A correlation matrix was generated by calculating an average correlation of 0.53, 0.56, and 0.54 for each scale of depression, anxiety, and stress. In order to replicate the relationship observed in real data. A correlation matrix was required to model feature dependency to ensure that synthetic data accurately reflects real data relationships. A multivariate normal distribution was generated using this correlation matrix. Using the cumulative distribution function (CDF), normal variables were converted to uniform variables. These uniform values were converted to discrete outcomes using probabilities to ensure that the synthetic data matched the real data results. This conversion was required to maintain correlation so that discrete data resembled actual data. . The code for synthetic data generation is provided in appendix B.

3.6 Model Selection

In this study, five ML models are utilized-Random Forest(RF), Decision Tree (DT), Support Vector Machiness (SVM), Naïve Bayes (NB) and k-nearest neighbors (KNN) [25]. These models are built using the Scikit-learn library with the default parameters and Python as the programming language [44] .

3.6.1 *Random Forest*

Random Forest (RF) is an ensemble method that aggregates the outputs of many Decision Trees to improve forecast accuracy. RF is flexible, able to handle both classification and regression problems. The RF algorithm's primary hyperparameter are node size, the number of estimators (trees), and the amount of features examined for



splitting at every node. RF typically aggregates the results of 100 Decision Trees by default. This strategy is frequently used in a variety of areas, including business, finance, e-commerce, and healthcare, to enhance decision-making. In healthcare, RF is often used to predict drug reactions, identify biomarkers, and other uses [45].

3.6.2 *Decision Tree*

Decision trees (DT) are supervised ML algorithm. These are applicable for both classification and regression problems. A DT includes a root node, internal nodes, and branches. The algorithm determines the best characteristic for each node to divide into leaf nodes. The optimum feature is the one that accurately meets the criterion at the present node. This process is repeated until no subset of independent traits exists. DTs are simple to understand and adaptable. Gini impurity is the default measure of quality used for splitting. Gini impurity indicates the likelihood of a random sample being erroneously classified. It is an effective measure for creating a divide.

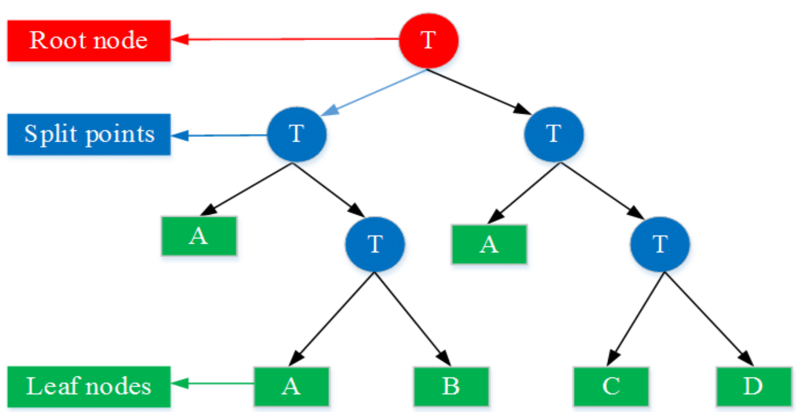


Figure 3.3: A simplified example of Decision Tree Model

3.6.3 *Support Vector Machines*

Support Vector Machiness (SVMs) are a strong ML technique that can do both classification and regression tasks, such as picture categorization and illness prediction.

SVM works by determining the best hyperplane in an N-dimensional space that divides vectors from distinct target classes. The ideal hyperplane is defined as the one that maximizes the distance between vectors from two different target classes. SVM is also useful for analyzing multidimensional and nonlinear connections. This algorithm may be modified to do multiclass classifications using either the one-vs-rest or one-vs-one methods. The one-vs-rest function compares each target class against all other classes. The one-vs-one strategy, on the other hand, requires each target class to be classified separately from each other class.

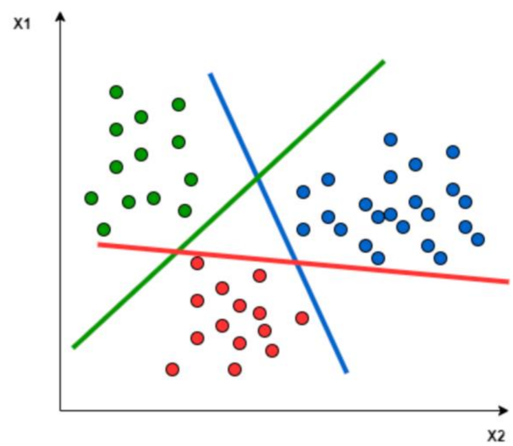
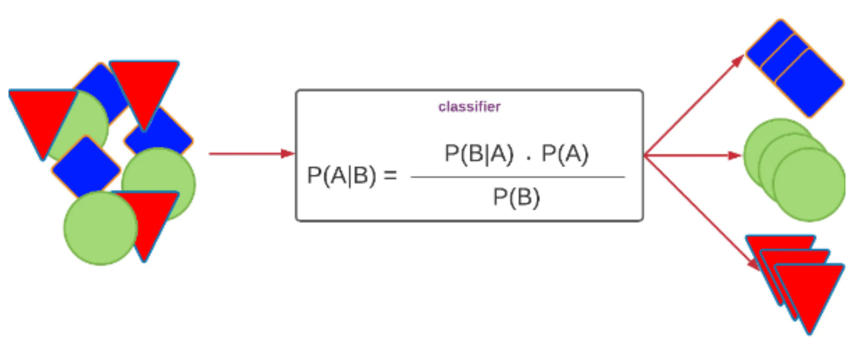


Figure 3.4: Graphical representation of the Support Vector Machines

3.6.4 *Naive Bayes*

Naive Bayes (NB) is a well-known data mining algorithm for classification [46]. It predicts the chance that a new sample belongs to a certain class based on the assumption that all qualities are independent of one another inside the class. This assumption is driven by the necessity to estimate multivariate probabilities using training data. In fact, the majority of attribute value combinations are either absent from the training data or insufficiently represented, rendering direct calculation of each relevant multivariate probability inaccurate. Naive Bayes avoids this problem by assuming



conditional independence [47]. Despite this stringent independence condition, Naive Bayes is a very good classifier in many real-world situations.

3.6.5 *The k-nearest neighbors*

The k-Nearest Neighbors (KNN) method is a simple and effective tool utilized for both classification and regression applications [48]. The algorithm is based on the idea of similarity, where it predicts the output by considering the 'k' closest training instances in the feature space. When producing a prediction, the KNN algorithm computes the distance between the query point and all the points in the training dataset. This is usually done using the Euclidean distance measure; however other distance metrics such as Manhattan or Minkowski can also be employed. The method subsequently chooses the 'k' nearest points and determines an outcome based on their values: for classification, it assigns the class that appears most frequently among the neighbors, and for regression, it calculates the average of the neighbor's values. KNN, being a type of instance-based learning, is easy to build and comprehend. It does not require any training phase other than storing the training data. However, this simplicity comes at the cost of computational expense for large datasets. Although KNN is a simple algorithm, it can be highly effective, particularly in situations when the data has a distinct clustering pattern.

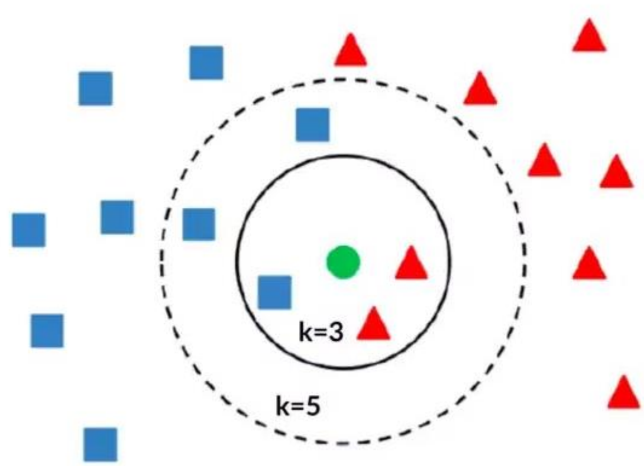


Figure 3.6: Graphical representation of k-Nearest Neighbors

3.7 Performance Evaluation

Performance evaluation is critical for developing reliable classifier. ML is concerned with two primary tasks: regression (predicting continuous values) and classification (sorting data into discrete groups). For classification issues with clear outcomes, performance measures are critical for determining how effectively a classifier works. The performance of the classifiers used in this study is assessed using the following metrics:

- Accuracy
- F1 Score
- Precision
- Recall
- Specificity
- Confusion Matrix

3.7.1 *Confusion Matrix*

A Confusion Matrix is not a performance matrix itself, but it does give valuable assessment insights for a classifier. It is a grid-like table that provides a summary of the performance of a classification model. Each category displays the number of cases that were correctly classified (True Positives and True Negatives) and the number of cases that were incorrectly classified (False Positives and False Negatives). Each Row in the matrix represents the actual labels for that particular class and each column represents the predicted labels (Figure 3.2).

Figure 3.7: Confusion Matrix for the given Muti-Class Classification problem

		Predicted labels				
	Classes	**Normal**	**Mild**	**Moderate**	**Severe**	**Extremely Severe**
Actual labels	**Normal**	**TrueN**	FalseM	FalseM	FalseS	FalseE
	Mild	FalseN	**TrueM**	FalseM	FalseS	FalseE
	Moderate	FalseN	FalseM	**TrueM**	FalseS	FalseE
	Severe	FalseN	FalseM	FalseM	**TrueS**	FalseE
	Extremely Severe	FalseN	FalseM	FalseM	FalseS	**TrueE**

The performance of a classifier in predicting each target class will be evaluated individually, as this study refers to a multi-class classification problem. The evaluation factors for each class are shown in Figure 3.3 with respect to the confusion matrix. Normal, mild, moderate, severe, and extremely severe are represented by the letters N, M, M, S, and E, respectively.

(A)	Normal				
	N	**M**	**M**	**S**	**E**
N	TP	FN	FN	FN	FN
M	FP	TN	TN	TN	TN
M	FP	TN	TN	TN	TN
S	FP	TN	TN	TN	TN
E	FP	TN	TN	TN	TN

(B)	Mild				
	N	**M**	**M**	**S**	**E**
N	TN	FP	TN	TN	TN
M	FN	TP	FN	FN	FN
M	TN	FP	TN	TN	TN
S	TN	FP	TN	TN	TN
E	TN	FP	TN	TN	TN

(C)	Moderate				
	N	**M**	**M**	**S**	**E**
N	TN	TN	FP	TN	TN
M	TN	TN	FP	TN	TN
M	FN	FN	TP	FN	FN
S	TN	TN	FP	TN	TN
E	TN	TN	FP	TN	TN

(D)	Severe				
	N	**M**	**M**	**S**	**E**
N	TN	TN	TN	FP	TN
M	TN	TN	TN	FP	TN
M	TN	TN	TN	FP	TN
S	FN	FN	FN	TP	FN
E	TN	TN	TN	FP	TN

(E)	Extremely Severe				
	N	**M**	**M**	**S**	**E**
N	TN	TN	TN	TN	FP
M	TN	TN	TN	TN	FP
M	TN	TN	TN	TN	FP
S	TN	TN	TN	TN	FP
E	FN	FN	FN	FN	TP

Figure 3.8: Interpretation of Confusion Matrix for different Classes

3.7.2 *Accuracy*

Accuracy is the most fundamental metrics used for evaluating a classification models. It is the measure of the number of correct predictions out of total predictions. The calculation of this can be done by multiplying the ratio of correct predictions to total predictions by 100.

$$\text{Accuracy} = \frac{\text{Number of correct predictions}}{\text{Total number of predictions}} \tag{3.1}$$

3.7.3 *F1 Score*

F1 score is the most widely used evaluation metrics calculates the harmonic mean of recall and precision. It addresses the possible trade-off between recall and precision and offers a fair assessment of both. A high F1-score means the model is good at preventing false positives or negatives and at recognizing true positives.

$$\text{F1 Score} = \frac{2*Precision*Recall}{Precision+Recall} \tag{3.2}$$

3.7.4 *Precision*

Precision is the metrics that measures positive prediction that are actually correct. It focuses on the model's ability to recognize between true positives and false positives.

$$Precision = \frac{True\ Positives\ (TP)}{True\ Positives\ (TP)+False\ Positives\ (FP)} \tag{3.3}$$

For this study, precision is evaluated for all the five target classes individually.

$$Precision\ (Normal) = \frac{True\ Normal}{Total\ Normal\ (Predicted)} \tag{3.4}$$

$$Precision\ (Mild) = \frac{True\ Mild}{Total\ Mild\ (Predicted)} \tag{3.5}$$

$$Precision\ (Moderate) = \frac{True\ Moderate}{Total\ Moderate\ (Predicted)} \tag{3.6}$$

$$Precision(Severe) = \frac{True\ Severe}{Total\ Severe\ (Predicted)} \tag{3.7}$$

$$Precision\ (Extremely\ Severe) = \frac{True\ Severe}{Total\ Extremely\ Severe\ (Predicted)} \quad (3.8)$$

3.7.5 *Recall*

Recall is the metrics that measures the correct positive predictions out of the actual positives. It is computed by dividing a class's true positives by its actual positives, which include both true positives and false negatives.

$$Recall = \frac{True\ Positives\ (TP)}{True\ Positives\ (TP) + False\ Negatives\ (FN)} \quad (3.9)$$

Recall for each class is as follow:

$$Recall\ (Normal) = \frac{True\ Normal}{Total\ Normal\ (Actual)} \quad (3.10)$$

$$Recall\ (Mild) = \frac{True\ Mild}{Total\ Mild\ (Actual)} \quad (3.11)$$

$$Recall\ (Moderate) = \frac{True\ Moderate}{Total\ Moderate\ (Actual)} \quad (3.12)$$

$$Recall\ (Severe) = \frac{True\ Severe}{Total\ Severe\ (Actual)} \quad (3.13)$$

$$Recall\ (Extremely\ Severe) = \frac{True\ Extremely\ Severe}{Total\ Extremely\ Severe\ (Actual)} \quad (3.14)$$

3.7.6 *Specificity*

Specificity is the metrics that measures the proportion of true negative predictions out of the actual negative instances. For a given class, it can be computed by dividing the true negative predictions by the true negative and false positive predictions.

$$Specificity = \frac{True\ Negatives\ (TN)}{True\ Positives\ (TP) + False\ positives\ (FP)} \tag{3.15}$$

Specificity for each target class is as follow:

$$Specificity\ (Normal) = \frac{TN\ (Normal)}{TN + FP\ (Normal)} \tag{3.16}$$

$$Specificity\ (Mild) = \frac{TN\ (Mild)}{TN + FP\ (Mild)} \tag{3.17}$$

$$Specificity\ (Moderate) = \frac{TN\ (Moderate)}{TN + FP\ (Moderate)} \tag{3.18}$$

$$Specificity\ (Severe) = \frac{TN\ (Severe)}{TN + FP\ (Severe)} \tag{3.19}$$

$$Specificity\ (Extremely\ Severe) = \frac{TN\ (Extremely\ Severe)}{TN + FP\ (Extremely\ Severe)} \tag{3.20}$$

4. RESULTS

4.1 Demographic Characteristics

Table 4.1 presents a comprehensive breakdown of the demographic profile of the 115 SINES students who participated in the study. The data analysis reveals a predominantly female student body, with 65.2% identifying as such. The largest age group falls within the range of 22–25 years old, representing 67.0% of the student population. Furthermore, Master's students make up the majority of the group, comprising 76.5%. Notably, the majority of students (60.9%) are in their second semester. In terms of accommodation status, day scholars make up the largest percentage (45.2%), followed by students residing in on-campus accommodation (32.2%).

This insightful analysis depicts a vivid picture of the SINES student body, illustrating a predominantly female cohort of young adults pursuing Master's degrees, and likely residing off-campus or commuting.

Table 4.1: Demographic Data

	Freq.	**%**
Gender		
Male	39	33.9
Female	75	65.2
Academic Level		
BS	23	20
MS	88	76.5
PhD	4	3.5
Age		
18-21	21	18.3
22-25	77	67.0
26-30	14	12.2
Above 30	3	2.6

Semester		
1st	4	3.5
2nd	70	60.9
3rd	8	7.0
4th	28	24.3
Accommodation Status		
Day Scholar	52	45.2
Hostelite (Inside Campus)	37	32.2
Hostelite (Outside Campus)	26	22.6

4.2 Prevalence of Mental Health problem like depression, Anxiety and Stress

Table 4.2 examines the prevalence of mental health issues among the 115 SINES student participants using the DASS21 scale, which categorizes mental health into five levels: normal, mild, moderate, severe, and extremely severe.

A significant number of students experience various levels of depression, anxiety and stress. In terms of depression, 16.1% of students report very severe symptoms, while 11.3% suffer from serious depression. 19.4% of students experience moderate depression, while 12.9% have mild depression. A larger percentage of students (33.1%) report typical levels of depression.

Specifically, 37.1% of students deal with anxiety, with 9.7% experiencing severe anxiety and 14.5% experiencing moderate anxiety. Additionally, 2.4% of students report moderate anxiety and 29.0% have typical anxiety levels.

47.8% of participants reported extremely severe stress, with 12.2% experiencing severe stress. 8.7% and 7.8% reported moderate and mild stress, respectively. Stress levels among students reveal concerning trends. 8.1% of students experience very severe stress, while 12.1% and 15.3% report severe and moderate stress, respectively. 14.5% of students feel mild stress, with 42.7% reporting normal stress levels. This research highlights the significant mental health challenges that many students face. Mental health status of the study participants is highlighted below.

Table 4.2: Mental health status of the study participants: depression, anxiety, and stress

Severity	Number of participant (n=115) (%).		
	Depression	Anxiety	Stress
Normal	41 (33.1)	36 (29)	53 (42.7)
Mild	16 (12.9)	3 (2.4)	18 (14.5)
Moderate	24 (19.4)	18 (14.5)	19 (15.3)
Severe	14 (11.3)	12 (9.7)	15 (12.1)
Extremely Severe	20 (16.1)	46 (37.1)	10 (8.1)

4.3 Internal Consistency and Reliability Analysis

Table summarizes the reliability analysis of the DASS21 and its scales. Cronbach alpha of the DASS21 is 0.941 and for its scales i.e. depression, anxiety and stress it is 0.900, 0.884 and 0.893 correspondingly. These are satisfactory values as they are above the recommended value of 0.7 for statistically valid and reliable questionnaires.

Table 4.3: Cronbach Value for each Scale

Scale	Cronbach value
DASS21	0.941
Depression	0.900
Anxiety	0.884
Stress	0.893

Table 4.3 summarizes the mean inter-item correlation of the DASS21 scales. The average inter-item correlations for scales of the DASS21 are between 0.5 to 0.6 for, indicating moderate associations.

Table 4.4: Average Inter-Item Correlation for each Scale

Scale	Average inter-item correlation
Depression	0.56
Anxiety	0.53
Stress	0.54

Table 4.4 displays the detailed item statistics of the reliability analysis for the DASS21 that was administered to students. This analysis includes the item-total correlation of each question of the DASS21 and the impact of the deletion of any item on the scale mean, scale variance, and Cronbach's alpha. These items are effective indicators of the underlying construct being measured, as evidenced by correlations between 0.4 and 0.7, which indicate that individual items are moderately to strongly relate to the overall scale. Deletion of any item results in negligible variation in the scale mean and variance. Cronbach's alpha decreases by a small magnitude upon deletion of any item, indicating that each item of the DASS21 is significant to maintain its validity.

Table 4.5: Detailed Item Statistics of the Depression items Administered to students

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
DASS1	6.77	25.036	0.643	0.461	0.892
DASS2	6.34	22.419	0.658	0.479	0.890
DASS3	6.61	22.679	0.790	0.639	0.876
DASS4	6.44	22.144	0.748	0.611	0.880
DASS5	6.47	21.234	0.707	0.532	0.885
DASS6	6.68	22.378	0.636	0.455	0.893
DASS7	6.70	22.684	0.764	0.631	0.878

Table 4.6: Detailed Item Statistics of the Anxiety items Administered to students

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
DASS8	6.91	25.220	0.460	0.275	0.892
DASS9	6.83	22.355	0.721	0.552	0.862
DASS10	6.93	22.942	0.663	0.467	0.869
DASS11	6.44	22.144	0.729	0.615	0.861
DASS12	6.63	21.988	0.736	0.637	0.860
DASS13	6.57	22.336	0.740	0.561	0.859
DASS14	6.70	22.564	0.664	0.472	0.869

Table 4.7: Detailed Item Statistics of the Stress items Administered to students

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
DASS15	7.14	21.630	0.728	0.621	0.73
DASS16	7.05	21.630	0.728	0.621	0.873
DASS17	6.98	21.263	0.554	0.318	0.893
DASS18	7.04	21.393	0.720	0.564	0.874
DASS19	7.06	20.952	0.783	0.671	0.866
DASS20	7.06	20.952	0.783	0.671	0.866
DASS21	7.15	21.864	0.711	0.548	0.875

The inter-item correlations among the 21 items of the DASS (Depression Anxiety Stress) Scale indicate a weak to strong relationship, with the preponderance of correlations ranging from 0.1 to 0.7. This range of correlations suggests that the items effectively capture the distinct yet related domains of depression, anxiety, and stress.

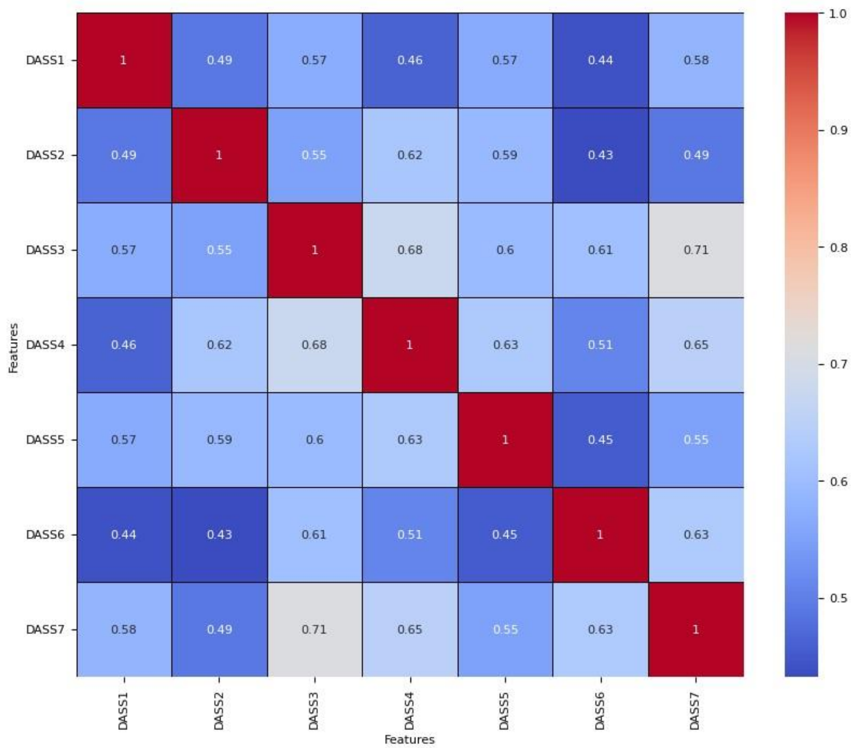


Figure 4.1: Heatmap of the inter-item correlations of the Depression Items

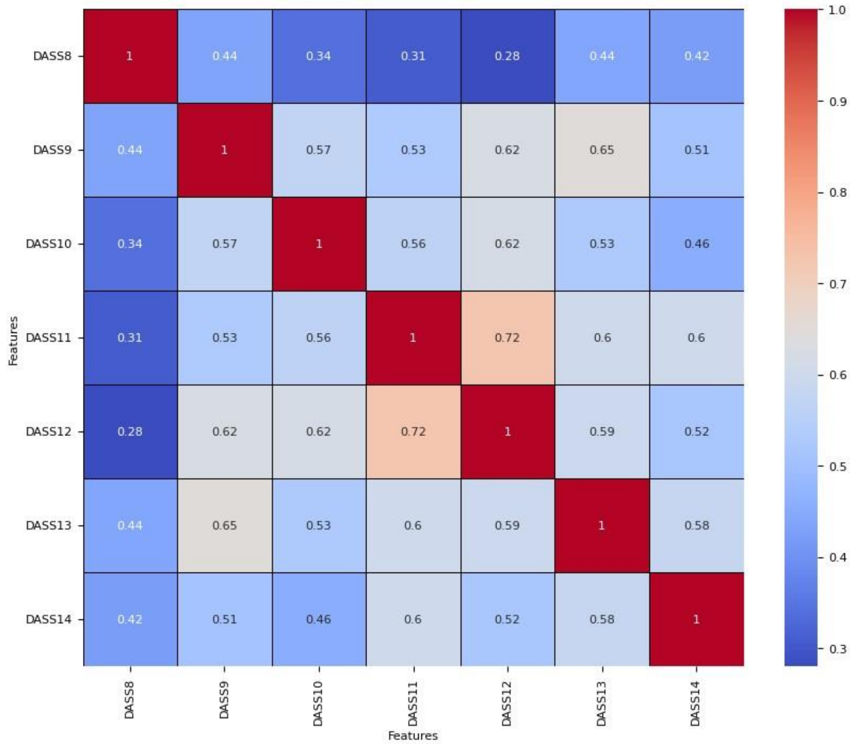


Figure 4.2: Heatmap of the inter-item correlations of the Anxiety Items

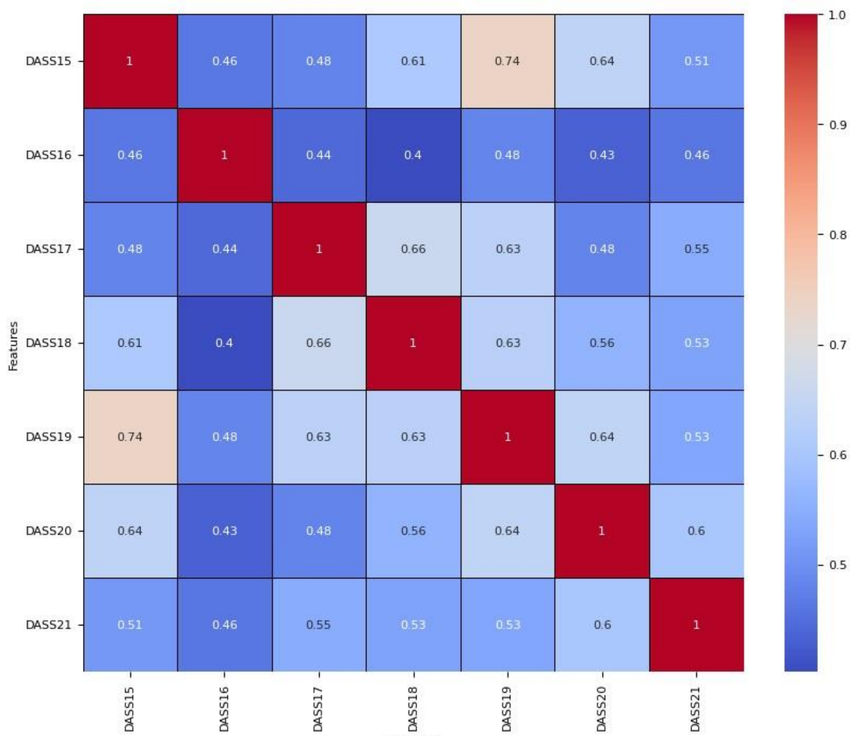


Figure 4.3: Heatmap of the inter-item correlations of the Stress Items

The satisfactory magnitudes of Cronbach's alpha indicate that the DASS is a statistically reliable and consistent tool for screening mental illness in Pakistani students, despite the fact that all but a few of the items in the test have weaker correlations, with a range primarily between 0.2 and 0.6. For the purpose of early mental disease identification and intervention, the suggested predictive models based on DASS21 will therefore also be appropriate as decision support systems.

4.4 Hypothesis Testing

The Chi-squared goodness-of-fit tests performed on the 21 variables of the Depression, Anxiety, and Stress Scales (DASS21) indicated no significant deviation from the specified multinomial distribution with given probabilities at the 0.05 significance level. Therefore, the null hypothesis (H_0) for each variable was not rejected. This implies that there is no significant difference between the observed and expected frequencies for each variable, suggesting that the data fit the multinomial distribution well. Below is the table presenting the Chi-squared values and accompanying p-values for all 21 variables:

Table 4.8: Results of the chi-square and p-value of DASS items

Features	χ^2	p value
DASS1	0.009411	0.9998
DASS2	0.014116	0.9996
DASS3	0.024891	0.999
DASS4	0.014639	0.9995
DASS5	0.022687	0.9991
DASS6	0.0129	0.9996
DASS7	0.039337	0.9979
DASS8	0.014116	0.9996
DASS9	0.027494	0.9988
DASS10	0.03286	0.9984
DASS11	0.019936	0.9993
DASS12	0.023513	0.999

DASS13	0.019259	0.9993
DASS14	0.061422	0.996
DASS15	0.013424	0.9996
DASS16	0.029471	0.9987
DASS17	0.011184	0.9997
DASS18	0.025247	0.9989
DASS19	0.043751	0.9976
DASS20	0.040843	0.9978
DASS21	0.030185	0.9986

4.5 Feature Selection

Based on the fact that all p-values are less than 0.05, we can conclude that there is is a statistically significant relation between target variable and all the features. So we rejected Null Hypothesis which says that there is no significant relation between target variable and features.

Table 4.9: Chi-square tests between 21 items of the DASS and Target Variable

Features	χ^2	p value	C
DASS1	99.90	0.00	0.68
DASS2	122.57	0.00	0.71
DASS3	120.35	0.00	0.71
DASS4	89.74	0.00	0.62
DASS5	115.01	0.00	0.70
DASS6	126.13	0.00	0.70
DASS7	49.18	0.00	0.54
DASS8	81.70	0.00	0.64
DASS9	73.86	0.00	0.62
DASS10	93.31	0.00	0.66
DASS11	89.06	0.00	0.66

DASS12	89.06	0.00	0.66
DASS13	85.38	0.00	0.65
DASS14	49.18	0.00	0.54
DASS15	111.99	0.00	0.70
DASS16	58.98	0.00	0.58
DASS17	103.36	0.00	0.68
DASS18	102.14	0.00	0.68
DASS19	147.63	0.00	0.75
DASS20	147.63	0.00	0.69
DASS21	80.01	0.00	0.64

Both the filter method (the Chi-Square test) and the wrapper method (Recursive Feature Elimination, RFE) show that all 21 features are significant. The p-value for each of the DASS21 items is 0.00, which is less than the generally accepted significance level of 0.05. Thus, we reject the null hypothesis for all variables. This suggests a significant relationship between each DASS21 component and the target variable.

The contingency coefficients for the DASS21 items vary between 0.5 and 0.7. According to the contingency coefficient, these values suggest a moderate to strong association between the DASS21 features and the target variable. These results show that all features are necessary for predicting the desired outcomes. Below is the table given of all the DASS21 items.

The feature importance scores for each scales Depression, Anxiety and Stress calculated utilizing RFE are represented in Bar chart given below.

Figure 4.5: Feature Importance of Depression Items using RFE

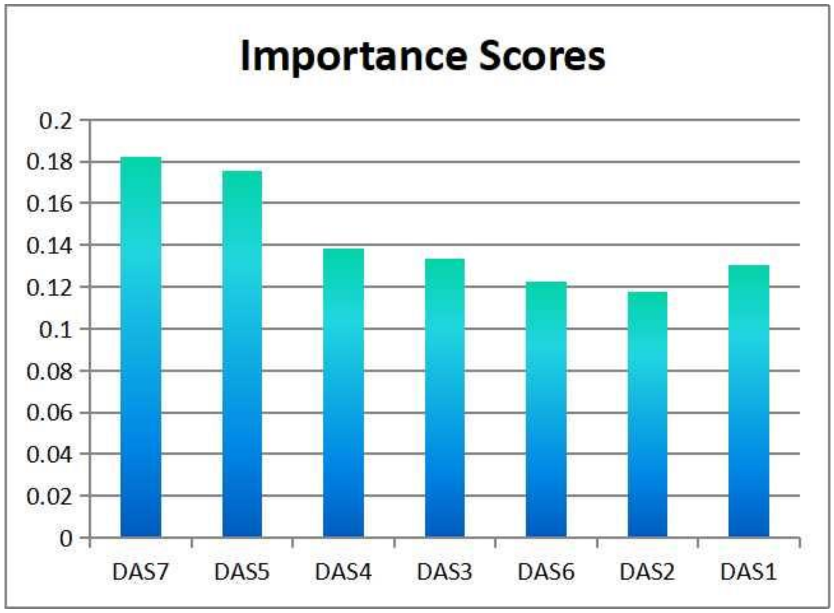


Figure 4.4: Feature Importance of Anxiety Items using RFE

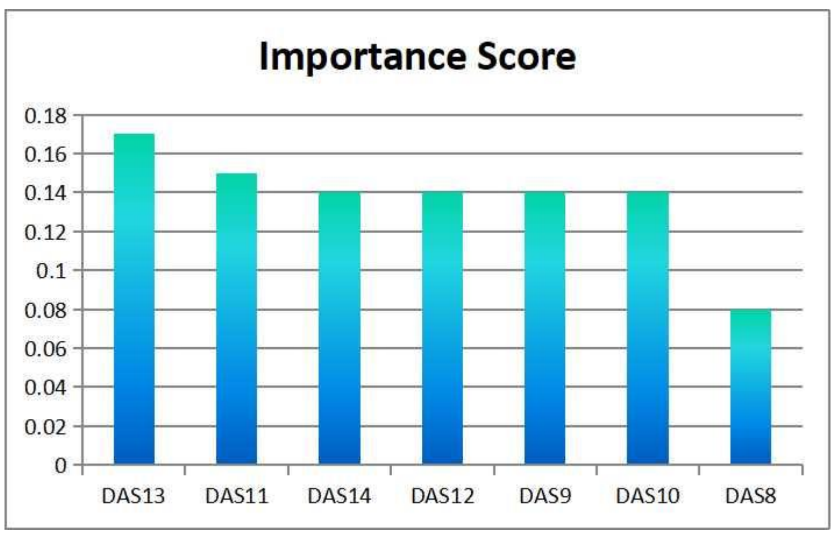


Figure 4.6: Feature Importance of Stress Items using RFE

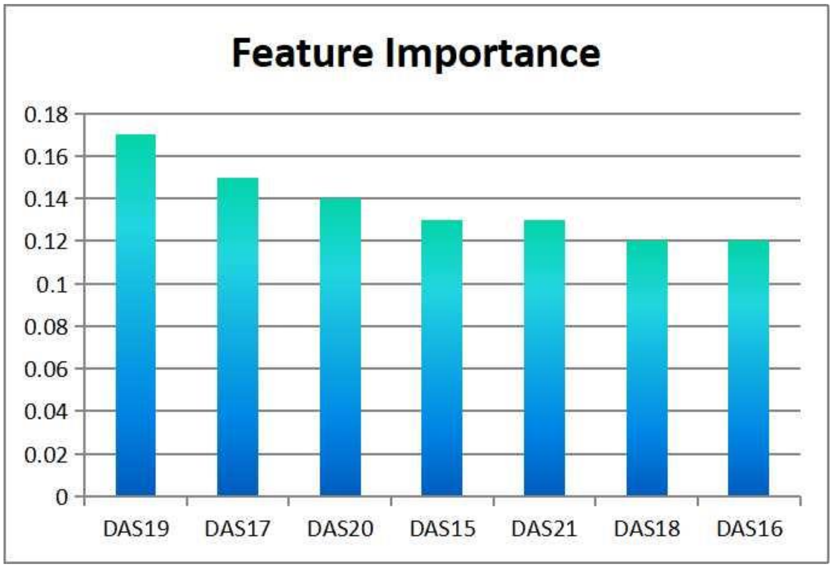


Table shows the number of optimal features that provide the highest RFE accuracies for the anxiety, depression, and stress scales. For depression, RF-RFE chooses all 7 features to train a model with an accuracy of 0.80. Similarly, for anxiety and stress, RF-RFE selects all 7 features for each scale, yielding model accuracies of 0.83 and 0.79, respectively.

Table 4.10: Results of Recursive Feature Elimination

Scale	RF-RFE		
	No. of Features	Eliminated Features	Mean Accuracy
Depression	7	None	0.80
Anxiety	7	None	0.83
Stress	7	None	0.79

4.6 Synthetic data generation

For depression, synthetic data is generated using the correlation values and probabilities for each of the seven variables. This procedure was then replicated for anxiety and stress, each with seven variables. The probability of seven variables of classes 0, 1, 2, and 3 is presented in a single function. The synthetic data were created in odd ratios of 1:3, 1:7, and 1:11 for each scale. As a consequence, three datasets were constructed for each scale, totaling 9 datasets. These synthetic datasets were then integrated with the real data to create a more complete dataset for further analysis and model training.

Table 4.11: Data Generation and Total Instances per Ratio for each Scale

Ratio	Synthetic data	Hybrid Data
1:3	345	460
1:7	805	920
1:11	1265	1380

The table 4.9 displays the number of synthetic instances created and the total number of hybrid instances (original + synthetic) for each ratio, for scales (depression, anxiety and stress).

4.7 Machine Learning Results

The original dataset for depression, anxiety and Stress with 7 features for each and 115 instances each has been utilized for model development.

4.7.1 *Depression*

For dataset for Depression, RF correctly classified 41 out of 41 normal cases, 7 out of 16 mild cases, 16 out of 24 moderate cases, 9 out of 14 severe cases, and 18 out of 20 extremely severe cases. DT correctly classified 36 out of 41 normal cases, 8 out of 16 mild cases, 11 out of 24 moderate cases, 8 out of 14 severe cases, and 11 out of 20 extremely severe cases. SVM correctly classified 39 out of 41 normal cases, 15 out of 16 mild cases, 22 out of 24 moderate cases, 12 out of 14 severe cases, and 17 out of 20 extremely severe cases. NB correctly classified 32 out of 41 normal cases, 14 out of 16 mild cases, 14 out of 24 moderate cases, 9 out of 14 severe cases, and 15 out of 20 extremely severe cases. KNN correctly classified 40 out of 41 normal cases, 10 out of 16 mild cases, 15 out of 24 moderate cases, 7 out of 14 severe cases, and 17 out of 20 extremely severe cases. The best performing model came out to be SVM with an F1 score of 91%. The results of all these five models for Depression and their confusion matrices are given in the Table 4.11 and Figure 4.6 respectively.

Table 4.12: Performance evaluation of the five ML models on the original data for Depression

Model	Accuracy	F1 Score	Classes	Precision	Recall	Specificity
RF	0.79	0.78	Normal	0.82	1.00	0.88
			Mild	0.70	0.44	0.97
			Moderate	0.73	0.67	0.93
			Severe	0.64	0.64	0.95
			Extremely Severe	0.95	0.90	0.99
DT	0.64	0.64	Normal	0.84	0.88	0.91
			Mild	0.50	0.50	0.92
			Moderate	0.55	0.46	0.90
			Severe	0.38	0.57	0.87
			Extremely Severe	0.73	0.55	0.96

SVM	**0.91**	**0.91**	**Normal**	**0.97**	**0.95**	**0.99**
			Mild	**0.79**	**0.94**	**0.96**
			Moderate	**0.92**	**0.92**	**0.98**
			Severe	**0.80**	**0.86**	**0.97**
			Extremely Severe	**1.00**	**0.85**	**1.00**
NB	0.73	0.74	Normal	0.97	0.78	0.99
			Mild	0.54	0.88	0.88
			Moderate	0.58	0.58	0.89
			Severe	0.53	0.64	0.92
			Extremely Severe	0.95	0.89	0.97
KNN	0.77	0.77	Normal	0.93	0.98	0.96
			Mild	0.50	0.62	0.90
			Moderate	0.65	0.62	0.91
			Severe	0.70	0.50	0.97
			Extremely Severe	0.89	0.85	0.98

(A)

	RF				
	N	**M**	**M**	**S**	**E**
N	41	0	0	0	0
M	7	7	2	0	0
M	2	3	16	3	0
S	0	0	4	9	1
E	0	0	0	2	18

(B)

	DT				
	N	**M**	**M**	**S**	**E**
N	36	3	2	0	0
M	5	8	3	0	0
M	2	5	11	6	0
S	0	0	2	8	4
E	0	0	2	7	11

(C)

	SVM				
	N	**M**	**M**	**S**	**E**
N	39	2	0	0	0
M	1	15	0	0	0
M	0	2	22	0	0
S	0	0	2	12	0
E	0	0	0	3	17

(D)

	NB				
	N	**M**	**M**	**S**	**E**
N	32	5	4	0	0
M	1	14	1	0	0
M	0	7	14	3	0
S	0	0	5	9	0
E	0	0	0	5	15

(E)

	KNN				
	N	**M**	**M**	**S**	**E**
N	40	1	0	0	0
M	3	10	3	0	0
M	0	9	15	0	0
S	0	0	5	7	2
E	0	0	0	3	17

Figure 4.7: Confusion matrices of ML models for the original dataset for Depression

4.7.2 *Anxiety*

For dataset for Anxiety, RF correctly classified 36 out of 36 normal cases, 0 out of 3 mild cases, 15 out of 18 moderate cases, 1 out of 12 severe cases, and 43 out of 46 extremely severe cases. DT correctly classified 33 out of 36 normal cases, 0 out of 3 mild cases, 11 out of 18 moderate cases, 3 out of 12 severe cases, and 35 out of 46 extremely severe cases. SVM correctly classified 36 out of 36 normal cases, 0 out of 3 mild cases, 16 out of 18 moderate cases, 10 out of 12 severe cases, and 43 out of 46 extremely severe cases. NB correctly classified 21 out of 36 normal cases, 0 out of 3 mild cases, 13 out of 18 moderate cases, 7 out of 12 severe cases, and 41 out of 46 extremely severe cases. KNN correctly classified 36 out of 36 normal cases, 0 out of 3 mild cases, 14 out of 18 moderate cases, 0 out of 12 severe cases, and 40 out of 46 extremely severe cases. The best performing model came out to be SVM with an F1 score of 91%.However, the recall for class mild was lower due to a lower number of instances, resulting in poor classification for this class. The results of all these five models and their confusion matrices are given in the Table 4.10 and Figure 4.5 respectively.

Table 4.13: Performance evaluation of the five ML models on the original data for Anxiety

Model	Accuracy	F1 Score	Classes	Precision	Recall	Specificity
RF	0.83	0.79	Normal	0.90	1.00	0.95
			Mild	0.00	0.00	1.00
			Moderate	0.71	0.83	0.94
			Severe	0.20	0.08	0.96
			Extremely Severe	0.88	0.93	0.91
DT	0.71	0.72	Normal	0.85	0.92	0.92
			Mild	0.00	0.00	0.97
			Moderate	0.61	0.61	0.93
			Severe	0.20	0.25	0.88
			Extremely Severe	0.88	0.76	0.93
SVM	**0.91**	**0.91**	**Normal**	**0.92**	**1.00**	**0.96**
			Mild	**0.00**	**0.00**	**0.99**
			Moderate	**0.89**	**0.89**	**0.98**
			Severe	**0.71**	**0.83**	**0.96**
			Extremely Severe	**1.00**	**0.93**	**1.00**

NB	0.71	0.76	Normal	1.00	0.58	1.0
			Mild	0.00	0.00	0.85
			Moderate	0.68	0.72	0.94
			Severe	0.47	0.58	0.92
			Extremely Severe	0.95	0.89	0.97
KNN	0.78	0.75	Normal	0.84	1.00	0.91
			Mild	0.00	0.00	0.00
			Moderate	0.48	0.78	0.85
			Severe	1.00	0.87	1.00
			Extremely Severe	1.00	0.87	1.00

(A)

	RF				
	N	**M**	**M**	**S**	**E**
N	36	0	0	0	0
M	2	0	1	0	0
M	2	0	15	1	0
S	0	0	5	1	6
E	0	0	0	3	43

(B)

	DT				
	N	**M**	**M**	**S**	**E**
N	33	0	2	1	0
M	1	0	2	0	0
M	4	2	11	1	0
S	1	1	3	3	5
E	0	1	0	10	35

(C)

	SVM				
	N	**M**	**M**	**S**	**E**
N	36	0	0	0	0
M	3	0	0	0	0
M	0	1	16	1	0
S	0	0	2	10	9
E	0	0	0	3	43

(D)

	NB				
	N	**M**	**M**	**S**	**E**
N	21	12	3	0	0
M	0	0	2	1	0
M	0	3	13	2	0
S	0	2	1	7	2
E	0	0	0	5	41

(E)

	KNN				
	N	**M**	**M**	**S**	**E**
N	36	0	0	0	0
M	3	0	0	0	0
M	4	0	14	0	0
S	0	0	12	0	0
E	0	0	3	3	40

Figure 4.8: Confusion Matrices of ML Models for the Original Dataset for Anxiety

4.7.3 *Stress*

For dataset for Stress, RF correctly classified 52 out of 53 normal cases, 9 out of 18 mild cases, 12 out of 19 moderate cases, 11 out of 15 severe cases, and 7 out of 10 extremely severe cases. DT correctly classified 51 out of 53 normal cases, 8 out of 18 mild cases, 8 out of 19 moderate cases, 7 out of 15 severe cases, and 7 out of 10 extremely severe cases. SVM correctly classified 53 out of 53 normal cases, 12 out of 18 mild cases, 16 out of 19 moderate cases, 14 out of 15 severe cases, and 8 out of 10 extremely severe cases. NB correctly classified 46 out of 53 normal cases, 10 out of 18 mild cases, 12 out of 19 moderate cases, 11 out of 15 severe cases, and 6 out of 10 extremely severe cases. KNN correctly classified 53 out of 53 normal cases, 9 out of 18 mild cases, 11 out of 19 moderate cases, 11 out of 15 severe cases, and 7 out of 10 extremely severe cases. The best performing model came out to be SVM with an F1 score of 89%. The results of all these six models for Depression and their confusion matrices are given in the Table 4.12 and Figure 4.7 respectively (Table 4.12 and Figure 4.7).

Table 4.14: Performance evaluation of the five ML models on the original data for Stress

Model	Accuracy	F1 Score	Classes	Precision	Recall	Specificity
RF	0.79	0.78	Normal	0.88	0.98	0.89
			Mild	0.64	0.50	0.95
			Moderate	0.67	0.63	0.94
			Severe	0.69	0.73	0.95
			Extremely Severe	0.88	0.70	0.99
DT	0.70	0.70	Normal	0.88	0.86	0.89
			Mild	0.44	0.44	0.90
			Moderate	0.42	0.42	0.89
			Severe	0.64	0.47	0.96
			Extremely Severe	0.78	0.78	0.98
SVM	**0.90**	**0.89**	**Normal**	**0.95**	**1.00**	**0.95**
			Mild	**1.00**	**0.67**	**1.00**
			Moderate	**0.84**	**0.84**	**0.97**
			Severe	**0.74**	**0.93**	**0.95**
			Extremely Severe	**0.89**	**0.80**	**0.99**
NB	0.74	0.75	Normal	1.00	0.87	1.00
			Mild	0.42	0.56	0.86
			Moderate	0.52	0.63	0.89
			Severe	0.73	0.73	0.96
			Extremely Severe	0.86	0.60	0.99
KNN	0.79	0.78	Normal	0.85	1.00	0.85
			Mild	0.60	0.50	0.94
			Moderate	0.69	0.58	0.95
			Severe	0.73	0.73	0.96
			Extremely Severe	1.00	0.70	1.00

(A)

	RF				
	N	**M**	**M**	**S**	**E**
N	52	1	0	0	0
M	6	9	3	0	0
M	1	4	12	2	0
S	0	0	3	11	1
E	0	0	0	3	7

(B)

	DT				
	N	**M**	**M**	**S**	**E**
N	51	1	1	0	0
M	6	8	4	0	0
M	1	8	8	2	0
S	0	1	5	7	2
E	0	0	1	2	7

(C)

	SVM				
	N	**M**	**M**	**S**	**E**
N	53	0	0	0	0
M	3	12	3	0	0
M	0	0	16	3	0
S	0	0	0	14	1
E	0	0	0	2	8

(D)

	NB				
	N	**M**	**M**	**S**	**E**
N	46	7	0	0	0
M	0	10	8	0	0
M	0	7	12	0	0
S	0	0	3	11	1
E	0	0	0	4	6

(E)

	KNN				
	N	**M**	**M**	**S**	**E**
N	53	0	0	0	0
M	8	9	1	0	0
M	1	6	11	1	0
S	0	0	4	11	0
E	0	0	0	3	7

Figure 4.9: Confusion matrices of ML models for the original dataset for Stress

The SVM exhibited exceptional performance in evaluating mental health measures, namely in the domains of anxiety, depression, and stress assessments. Regarding anxiety, the SVM demonstrated an F1 score of 91%, whilst for depression; it achieved a rate of 91%. The SVM achieved an F1 Score rate of 89% when assessing stress levels. The results demonstrate the resilience and efficacy of the SVM model in properly forecasting outcomes using the original data. The results are given below in the

Table 4.15.

Table 4.15: F1 Score of the five ML models on the original data for Depression, Anxiety, and Stress

Model	**Scales**		
	Depression	**Anxiety**	**Stress**
RF	78	74	78
DT	64	72	70
SVM	91	91	89
NB	74	76	75
KNN	78	75	78

4.8 Synthetic data Results

4.8.1 *Depression (1:3)*

For synthetic dataset for depression ratio 1:3, RF correctly classified 126 out of 133 normal cases, 53 out of 78 mild cases, 103 out of 120 moderate cases, 36 out of 60 severe cases, and 62 out of 69 extremely severe cases. DT correctly classified 122 out of 133 normal cases, 50 out of 78 mild cases, 87 out of 120 moderate cases, 36 out of 60 severe cases, and 54 out of 69 extremely severe cases. SVM correctly classified 124 out of 133 normal cases, 76 out of 78 mild cases, 116 out of 120 moderate cases, 50 out of 60 severe cases, and 64 out of 69 extremely severe cases. NB correctly classified 111 out of 133 normal cases, 64 out of 78 mild cases, 107 out of 120 moderate cases, 41 out of 60 severe cases, and 56 out of 69 extremely severe cases. KNN correctly classified 126 out of 133 normal cases, 47 out of 78 mild cases, 102 out of 120 moderate cases, 37 out of 60 severe cases, and 57 out of 69 extremely severe cases. The best performing model came out to be SVM with F1 score of 93 % .The results of all these six models for Depression and their confusion matrices are given in the 6 and 10 respectively.

Table 4.16: Performance evaluation of the five ML models on the Hybrid data for Depression (1:3)

Model	Accuracy	F1 Score	Classes	Precision	Recall	Specificity
RF	0.83	0.82	Normal	0.86	0.95	0.94
			Mild	0.72	0.68	0.95
			Moderate	0.79	0.86	0.92
			Severe	0.82	0.60	0.98
			Extremely Severe	0.95	0.90	0.99
DT	0.76	0.76	Normal	0.84	0.92	0.93
			Mild	0.50	0.64	0.96
			Moderate	0.76	0.72	0.92
			Severe	0.58	0.60	0.94
			Extremely Severe	0.87	0.78	0.98
SVM	**0.93**	**0.93**	**Normal**	**1.00**	**0.93**	**0.96**
			Mild	**0.86**	**0.97**	**0.97**
			Moderate	**0.92**	**0.97**	**0.97**
			Severe	**0.89**	**0.83**	**0.98**
			Extremely Severe	**0.97**	**0.93**	**0.99**
NB	0.82	0.83	Normal	1.00	0.83	1.00
			Mild	0.14	0.06	0.98
			Moderate	0.76	0.89	0.90
			Severe	0.68	0.68	0.95
			Extremely Severe	1.00	0.81	1.00
KNN	0.80	0.80	Normal	0.83	0.95	0.92
			Mild	0.70	0.60	0.95
			Moderate	0.77	0.85	0.91
			Severe	0.73	0.62	0.96
			Extremely Severe	0.97	0.83	0.99

(A)

	RF				
	N	M	M	S	E
N	126	7	0	0	0
M	20	53	5	0	0
M	1	14	103	2	0
S	0	0	21	36	3
E	0	0	1	6	62

(B)

	DT				
	N	M	M	S	E
N	122	11	0	0	0
M	21	50	7	0	0
M	2	16	87	13	2
S	0	0	18	36	6
E	0	0	2	13	54

(C)

	SVM				
	N	M	M	S	E
N	124	9	0	0	0
M	0	76	2	0	0
M	0	3	116	1	0
S	0	0	8	50	2
E	0	0	0	4	64

(D)

	NB				
	N	M	M	S	E
N	111	22	0	0	0
M	0	64	14	0	0
M	0	7	107	6	0
S	0	0	19	41	0
E	0	0	0	13	56

(E)

	KNN				
	N	M	M	S	E
N	126	7	0	0	0
M	23	47	8	0	0
M	2	13	102	3	0
S	0	0	21	37	2
E	0	0	1	11	57

Figure 4.10: Confusion matrices of ML models for the Hybrid dataset for Depression (1:3)

4.8.2 *Depression (1: 7)*

For synthetic dataset for Depression ratio 1:7, RF correctly classified 290 out of 297 normal cases, 76 out of 126 mild cases, 184 out of 219 moderate cases, 86 out of 119 severe cases, and 143 out of 159 extremely severe cases. DT correctly classified 273 out of 297 normal cases, 70 out of 126 mild cases, 152 out of 219 moderate cases, 83 out of

119 severe cases, and 134 out of 159 extremely severe cases. SVM correctly classified 219 out of 297 normal cases, 119 out of 126 mild cases, 214 out of 219 moderate cases, 109 out of 60 severe cases, and 152 out of 159 extremely severe cases. NB correctly classified 268 out of 297 normal cases, 99 out of 126 mild cases, 206 out of 219 moderate cases, 101 out of 159 severe cases, and 143 out of 159 extremely severe cases. KNN correctly classified 289 out of 297 normal cases, 57 out of 126 mild cases, 165 out of 219 moderate cases, 75 out of 119 severe cases, and 144 out of 159 extremely severe cases. The best performing model came out to be SVM with an F1 score of 96 %. The results of all these five models for depression and their confusion matrices are given in the 7 and Figure 4.11 respectively.

Table 4.17: Performance evaluation of the five ML models on the Hybrid data for Depression (1:7)

Model	Accuracy	F1 Score	Classes	Precision	Recall	Specificity
RF	0.85	0.84	Normal	0.88	0.98	0.94
			Mild	0.68	0.60	0.95
			Moderate	0.83	0.84	0.95
			Severe	0.81	0.72	0.98
			Extremely Severe	0.95	0.90	0.99
DT	0.77	0.77	Normal	0.89	0.92	0.91
			Mild	0.52	0.56	0.92
			Moderate	0.76	0.68	0.72
			Severe	0.64	0.70	0.67
			Extremely Severe	0.88	0.86	0.98
SVM	**0.96**	**0.96**	**Normal**	**0.98**	**0.98**	**0.99**
			Mild	**0.92**	**0.94**	**0.99**
			Moderate	**0.96**	**0.98**	**0.99**
			Severe	**0.94**	**0.92**	**0.99**
			Extremely Severe	**0.97**	**0.96**	**0.97**
NB	0.89	0.89	Normal	1.00	0.90	1.00
			Mild	0.71	0.79	0.95
			Moderate	0.83	0.94	0.88
			Severe	0.85	0.85	0.98
			Extremely Severe	0.99	0.90	1.00
KNN	0.79	0.79	Normal	0.84	0.97	0.91
			Mild	0.51	0.45	0.93
			Moderate	0.77	0.75	0.97

			Severe	0.78	0.63	0.97
			Extremely Severe	0.93	0.91	0.99

(A)

	RF				
	N	**M**	**M**	**S**	**E**
N	290	7	0	0	0
M	39	76	11	0	0
M	1	29	184	5	0
S	0	0	25	86	8
E	0	0	1	15	143

(B)

	DT				
	N	**M**	**M**	**S**	**E**
N	273	273	23	1	0
M	32	32	70	24	0
M	3	3	37	152	25
S	0	0	1	18	83
E	0	0	0	2	23

(C)

	SVM				
	N	**M**	**M**	**S**	**E**
N	291	6	0	0	0
M	5	119	2	0	0
M	0	5	214	0	0
S	0	0	6	109	4
E	0	0	0	7	152

(D)

	NB				
	N	**M**	**M**	**S**	**E**
N	268	29	0	0	0
M	1	99	26	0	0
M	0	11	206	2	0
S	0	0	17	101	1
E	0	0	0	16	143

(E)

	KNN				
	N	**M**	**M**	**S**	**E**
N	289	8	0	0	0
M	54	57	15	0	0
M	1	47	165	6	0
S	0	0	33	75	11
E	0	0	0	15	144

Figure 4.11: Confusion matrices of ML models for the Hybrid dataset for Depression (1:7)

4.8.3 *Depression (1:11)*

For synthetic dataset for Depression ratio 1:11, RF correctly classified 461 out of 478 normal cases, 132 out of 193 mild cases, 265 out of 324 moderate cases, 101 out of 163 severe cases, and 197 out of 222 extremely severe cases. DT correctly classified 450 out of 478 normal cases, 125 out of 193 mild cases, 220 out of 324 moderate cases, 105 out of 163 severe cases, and 194 out of 222 extremely severe cases. SVM correctly classified 471 out of 478 normal cases, 190 out of 193 mild cases, 321 out of 324 moderate cases, 143 out of 163 severe cases, and 216 out of 222 extremely severe cases. NB correctly classified 434 out of 478 normal cases, 163 out of 193 mild cases, 304 out of 324 moderate cases, 116 out of 163 severe cases, and 195 out of 222 extremely severe cases. KNN correctly classified 470 out of 478 normal cases, 102 out of 193 mild cases, 257 out of 324 moderate cases, 99 out of 163 severe cases, and 197 out of 222 extremely severe cases. The best performing model came out to be SVM with an F1 score of 97 %. The results of all these five models for depression and their confusion matrices are given in the Table 4.19 and Figure 4.12 respectively.

Table 4.19: Performance evaluation of the five ML models on the Hybrid data for Depression (1:11)

Model	Accuracy	F1 Score	Classes	Precision	Recall	Specificity
RF	0.84	0.84	Normal	0.92	0.96	0.96
			Mild	0.69	0.68	0.95
			Moderate	0.879	0.82	0.93
			Severe	0.70	0.62	0.96
			Extremely Severe	0.94	0.89	0.99
DT	0.79	0.79	Normal	0.91	0.94	0.95
			Mild	0.62	0.65	0.94
			Moderate	0.77	0.68	0.94
			Severe	0.57	0.64	0.93
			Extremely Severe	0.90	0.87	0.98
SVM	**0.97**	**0.97**	**Normal**	**1.00**	**0.99**	**1.00**
			Mild	**0.95**	**0.99**	**0.98**
			Moderate	**0.95**	**0.99**	**0.98**
			Severe	**0.95**	**0.88**	**0.99**

			Extremely Severe	**0.97**	**0.97**	**0.99**
NB	0.88	0.88	Normal	1.00	0.91	1.00
			Mild	0.73	0.84	0.95
			Moderate	0.80	0.94	0.93
			Severe	0.81	0.71	0.98
			Extremely Severe	0.99	0.88	1.00
KNN	0.82	0.81	Normal	0.87	0.98	0.92
			Mild	0.63	0.53	0.95
			Moderate	0.79	0.79	0.93
			Severe	0.72	0.61	0.97
			Extremely Severe	0.93	0.89	0.99

(A)

	RF				
	N	**M**	**M**	**S**	**E**
N	461	17	0	0	0
M	40	132	21	0	0
M	0	41	265	18	0
S	0	0	49	101	13
E	0	0	0	25	197

(B)

	DT				
	N	**M**	**M**	**S**	**E**
N	450	28	0	0	0
M	40	125	28	0	0
M	3	48	220	52	1
S	0	1	36	105	21
E	0	0	0	28	194

(C)

	SVM				
	N	**M**	**M**	**S**	**E**
N	471	7	0	0	0
M	0	190	3	0	0
M	0	2	321	1	0
S	0	0	14	143	6
E	0	0	0	6	216

(D)

	NB				
	N	**M**	**M**	**S**	**E**
N	434	42	2	0	0
M	0	163	30	0	0
M	0	19	304	1	0
S	0	0	45	116	2
E	0	0	0	27	195

(E)

	KNN				
	N	**M**	**M**	**S**	**E**
N	470	8	0	0	0
M	71	102	20	0	0
M	1	52	257	14	0
S	0	0	49	99	15
E	0	0	0	25	197

Figure 4.12: Confusion matrices of ML models for the Hybrid dataset for Depression (1:11)

Below are the graphs showing performance evaluation (Accuracy, Precision, Recall and Specificity) of Depression for each of the model individually.

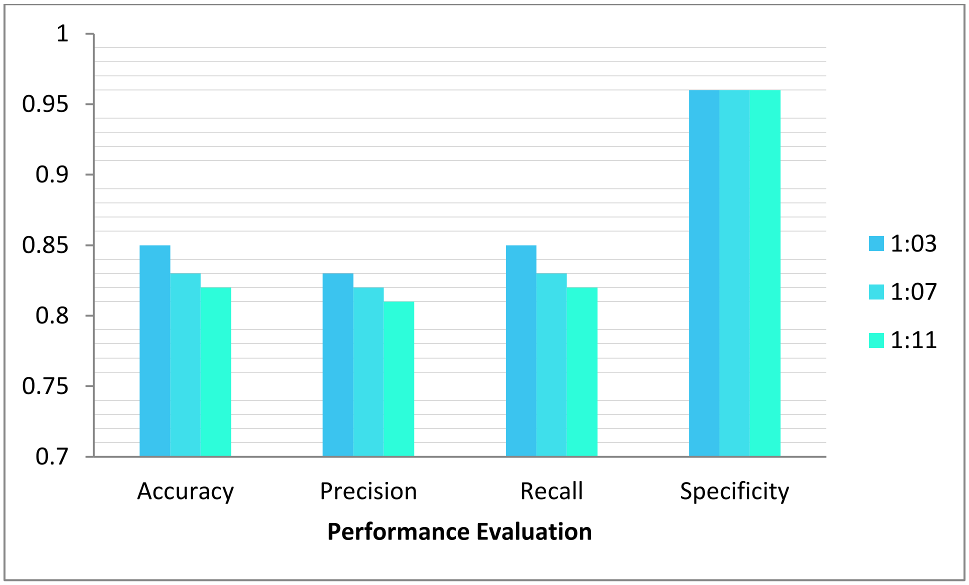


Figure 4.13: Performance Evaluation of RF for Depression across Different Ratios

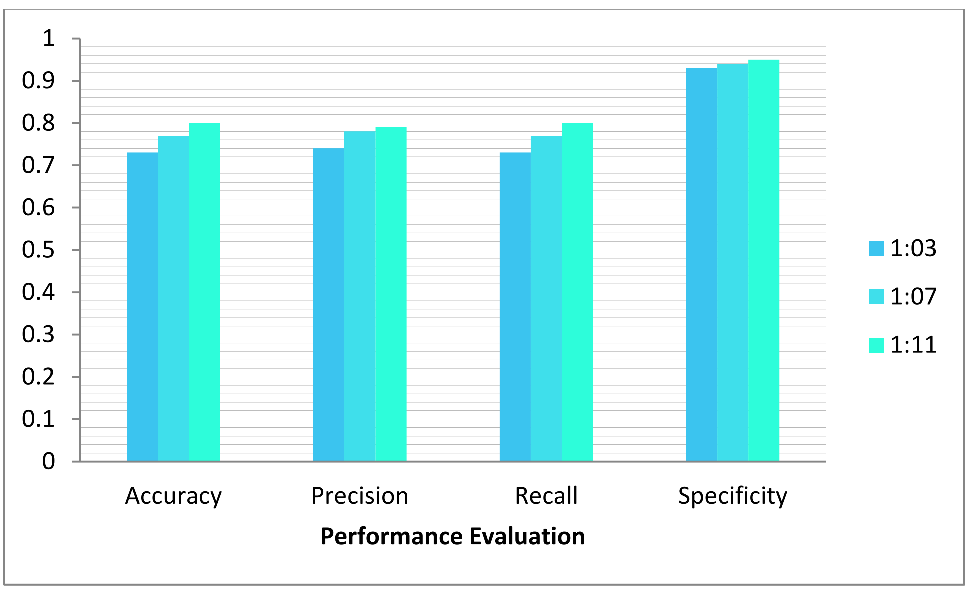


Figure 4.14: Performance Evaluation of DT for Depression across Different Ratios

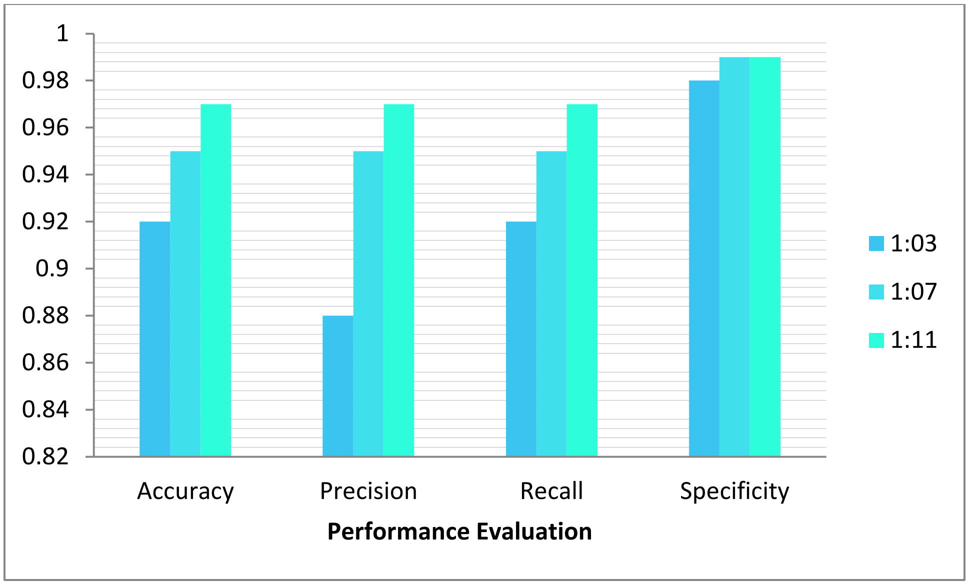


Figure 4.15: Performance Evaluation of SVM for Depression across Different Ratios

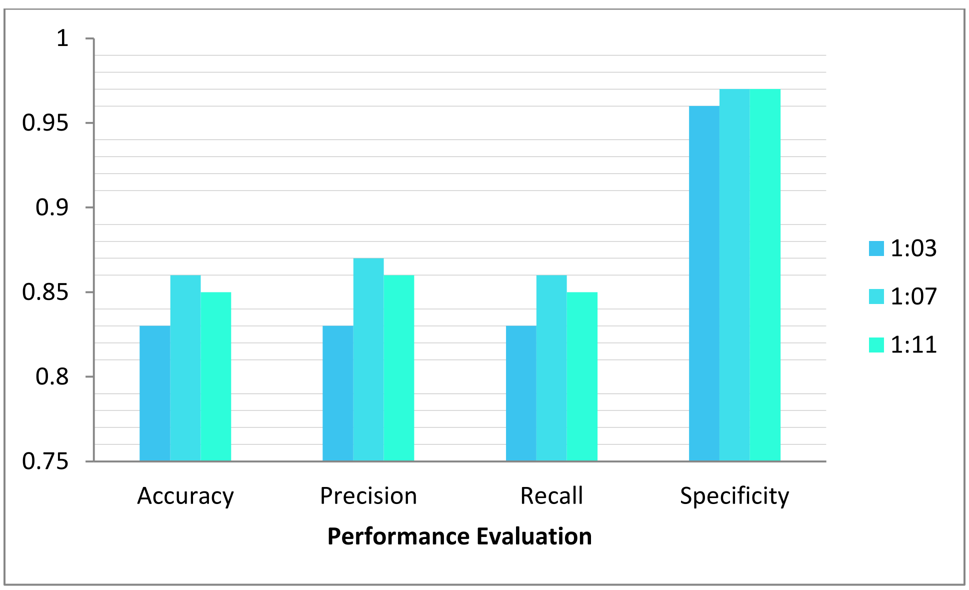


Figure 4.16: Performance Evaluation of NB for Depression across Different Ratios

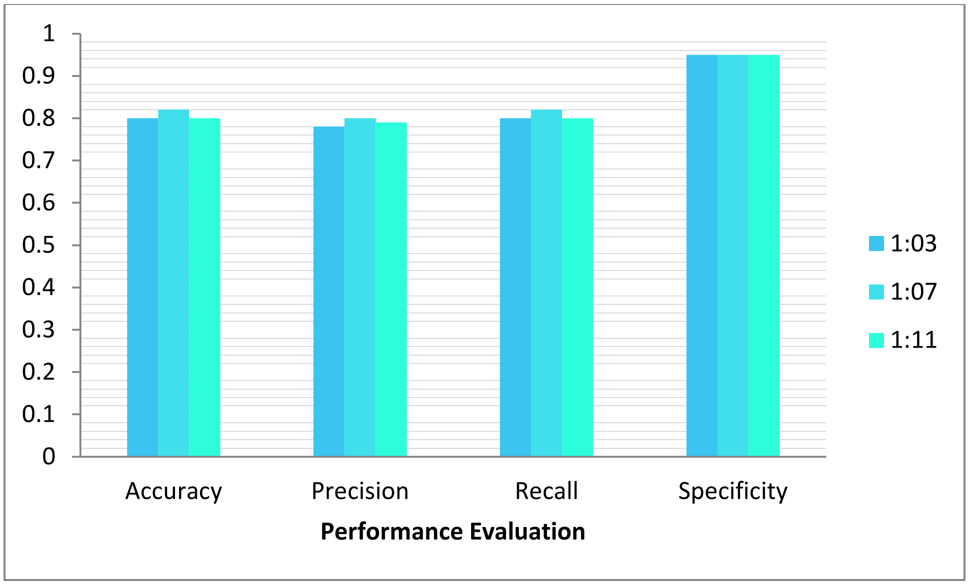


Figure 4.17: Performance Evaluation of KNN for Depression across Different Ratios

4.8.4 *Anxiety (1:3)*

For synthetic dataset for Anxiety ratio 1:3, RF correctly classified 125 out of 128 normal cases, 4 out of 22 mild cases, 71 out of 85 moderate cases, 27 out of 54 severe cases, and 162 out of 171 extremely severe cases. DT correctly classified 117 out of 128 normal cases, 8 out of 22 mild cases, 48 out of 85 moderate cases, 21 out of 54 severe cases, and 143 out of 171 extremely severe cases. SVM correctly classified 127 out of 128 normal cases, 0 out of 22 mild cases, 85 out of 85 moderate cases, 49 out of 54 severe cases, and 160 out of 171 extremely severe cases. NB correctly classified 117 out of 128 normal cases, 1 out of 22 mild cases, 79 out of 85 moderate cases, 35 out of 54 severe cases, and 152 out of 171 extremely severe cases. KNN correctly classified 128 out of 67 normal cases, 1 out of 12 mild cases, 62 out of 41 moderate cases, 19 out of 20 severe cases, and 158 out of 90 extremely severe cases. The best performing model came out to be SVM with an F1 score of 89%. The results of all these five models for Depression and their confusion matrices are given in the Table 4.19 and Figure 4.18 respectively.

Table 4.19: Performance evaluation of the five ML models on the Hybrid data for Anxiety (1:3)

Model	Accuracy	F1 Score	Classes	Precision	Recall	Specificity
RF	0.85	0.83	Normal	0.87	0.98	0.94
			Mild	0.50	0.18	0.99
			Moderate	0.81	0.84	0.95
			Severe	0.64	0.50	0.96
			Extremely Severe	0.91	0.95	0.94
DT	0.73	0.74	Normal	0.88	0.91	0.95
			Mild	0.27	0.36	0.95
			Moderate	0.61	0.56	0.92
			Severe	0.36	0.39	0.91
			Extremely Severe	0.90	0.84	0.94
SVM	**0.92**	**0.89**	**Normal**	**0.92**	**0.99**	**0.97**
			Mild	**0.00**	**0.00**	**1.00**
			Moderate	**0.88**	**1.00**	**0.97**
			Severe	**0.82**	**0.91**	**0.97**
			Extremely Severe	**0.98**	**0.94**	**0.99**
NB	0.83	0.83	Normal	0.97	0.91	0.99
			Mild	0.12	0.05	0.98
			Moderate	0.69	0.93	0.90
			Severe	0.59	0.65	0.94
			Extremely Severe	0.97	0.89	0.98
KNN	0.80	0.78	Normal	0.80	1.00	0.90
			Mild	0.10	0.05	0.98
			Moderate	0.67	0.73	0.92
			Severe	0.59	0.35	0.97
			Extremely Severe	0.96	0.92	0.98

(A)

	RF				
	N	M	M	S	E
N	125	2	1	0	0
M	14	4	4	0	0
M	5	2	71	6	1
S	0	0	12	27	15
E	0	0	0	9	162

(B)

	DT				
	N	M	M	S	E
N	117	5	5	1	0
M	6	8	8	0	0
M	9	17	48	10	1
S	1	0	17	21	15
E	0	0	1	27	143

(C)

	SVM				
	N	M	M	S	E
N	127	1	0	0	0
M	11	0	11	0	0
M	0	0	85	0	0
S	0	0	1	49	4
E	0	0	0	11	160

(D)

	NB				
	N	M	M	S	E
N	117	6	5	0	0
M	4	1	17	0	0
M	0	1	79	5	0
S	0	0	14	35	5
E	0	0	0	19	152

(E)

	KNN				
	N	M	M	S	E
N	128	0	0	0	0
M	20	1	1	0	0
M	13	9	69	1	0
S	0	0	29	19	6
E	0	0	1	12	158

Figure 4.18: Confusion matrices of ML models for the Hybrid dataset for Anxiety (1:3)

4.8.5 *Anxiety (1:7)*

For synthetic dataset for Anxiety ratio 1:7, RF correctly classified 245 out of 251 normal cases, 22 out of 66 mild cases, 138 out of 170 moderate cases, 42 out of 96 severe cases, and 319 out of 337 extremely severe cases. DT correctly classified 236 out of 251 normal cases, 30 out of 66 mild cases, 112 out of 170 moderate cases, 43 out of 96 severe cases, and 290 out of 337 extremely severe cases. SVM correctly classified 249 out of 251 normal cases, 55 out of 66 mild cases, 157 out of 170 moderate cases, 85 out of 96 severe cases, and 331 out of 337 extremely severe cases. NB correctly classified 227 out of 251 normal cases, 28 out of 66 mild cases, 155 out of 170 moderate cases, 64 out of 96 severe cases, and 317 out of 337 extremely severe cases. KNN correctly classified 246 out of 251 normal cases, 11 out of 66 mild cases, 130 out of 170 moderate cases, 47 out of 96 severe cases, and 321 out of 337 extremely severe cases. The best performing model came out to be SVM with an F1 score of 95%. The results of all these five models for Depression and their confusion matrices are given in the Table 4.15 and Figure 4.19 respectively.

Table 4.15: Performance evaluation of the five ML models on the Hybrid data for Anxiety (1:7)

Model	Accuracy	F1 Score	Classes	Precision	Recall	Specificity
RF	0.85	0.83	Normal	0.89	0.98	0.95
			Mild	0.45	0.33	0.97
			Moderate	0.77	0.81	0.94
			Severe	0.61	0.44	0.97
			Extremely Severe	0.92	0.95	0.95
DT	0.77	0.78	Normal	0.91	0.94	0.97
			Mild	0.41	0.45	0.95
			Moderate	0.70	0.66	0.94
			Severe	0.41	0.45	0.92
			Extremely Severe	0.90	0.86	0.94
SVM	**0.95**	**0.95**	**Normal**	**0.98**	**0.99**	**0.99**
			Mild	**0.87**	**0.83**	**0.99**
			Moderate	**0.94**	**0.92**	**0.99**
			Severe	**0.87**	**0.89**	**0.98**
			Extremely Severe	**0.98**	**0.98**	**0.99**
NB	0.86	0.86	Normal	0.96	0.90	0.99
			Mild	0.53	0.42	0.97
			Moderate	0.71	0.91	0.92
			Severe	0.69	0.67	0.96
			Extremely Severe	0.99	0.94	0.99
KNN	0.82	0.81	Normal	0.81	0.98	0.92
			Mild	0.28	0.17	0.97
			Moderate	0.75	0.76	0.94

			Severe	0.67	0.49	0.97
			Extremely Severe	0.96	0.95	0.98

(A)

	RF				
	N	**M**	**M**	**S**	**E**
N	245	6	0	0	0
M	31	22	13	0	0
M	0	21	138	10	1
S	0	0	28	42	26
E	0	0	1	17	319

(B)

	DT				
	N	**M**	**M**	**S**	**E**
N	236	13	2	0	0
M	19	30	17	0	0
M	3	30	112	19	6
S	0	1	25	43	27
E	0	0	4	43	290

(C)

	SVM				
	N	**M**	**M**	**S**	**E**
N	249	2	0	0	0
M	6	55	5	0	0
M	0	6	157	7	0
S	0	0	5	85	6
E	0	0	0	6	331

(D)

	NB				
	N	**M**	**M**	**S**	**E**
N	227	20	4	0	0
M	9	28	29	0	0
M	0	5	155	10	0
S	0	0	29	64	3
E	0	0	1	19	317

(E)

	KNN				
	N	**M**	**M**	**S**	**E**
N	246	5	0	0	0
M	47	11	8	0	0
M	9	23	130	8	0
S	0	0	35	47	14
E	0	0	1	15	321

Figure 4.19: Confusion matrices of ML models for the Hybrid dataset for Anxiety

4.8.6 *Anxiety (1:11)*

For synthetic dataset for Anxiety ratio 1:11, RF correctly classified 348 out of 355 normal cases, 127 out of 192 mild cases, 94 out of 164 moderate cases, 78 out of 157 severe cases, and 484 out of 512 extremely severe cases. DT correctly classified 338 out of 355 normal cases, 123 out of 192 mild cases, 94 out of 164 moderate cases, 80 out of 157 severe cases, and 466 out of 512 extremely severe cases. SVM correctly classified 354 out of 361 normal cases, 188 out of 176 mild cases, 150 out of 181 moderate cases, 142 out of 161 severe cases, and 508 out of 501 extremely severe cases. NB correctly classified 320 out of 361 normal cases, 161 out of 176 mild cases, 91 out of 181 moderate cases, 119 out of 161 severe cases, and 485 out of 501 extremely severe cases. KNN correctly classified 346 out of 361 normal cases, 102 out of 176 mild cases, 90 out of 181 moderate cases, 77 out of 161 severe cases, and 486 out of 501 extremely severe cases. The best performing model came out to be SVM with an F1 score of 96%. The results of all these five models for Depression and their confusion matrices are given in the Table 4.21 and Figure 4.20 respectively.

Table 4.21: Performance evaluation of the five ML models on the Hybrid data for Anxiety (1:11)

Model	Accuracy	F1 Score	Classes	Precision	Recall	Specificity
RF	0.82	0.81	Normal	0.91	0.98	0.97
			Mild	0.67	0.66	0.95
			Moderate	0.58	0.57	0.94
			Severe	0.63	0.50	0.96
			Extremely Severe	0.93	0.95	0.94
DT	0.80	0.80	Normal	0.90	0.95	0.96
			Mild	0.66	0.64	0.95
			Moderate	0.59	0.57	0.95
			Severe	0.53	0.51	0.94
			Extremely Severe	0.92	0.91	0.95
SVM	**0.96**	**0.96**	**Normal**	**1.00**	**1.00**	**1.00**
			Mild	**0.95**	**0.98**	**0.99**
			Moderate	**0.94**	**0.91**	**0.99**
			Severe	**0.93**	**0.90**	**0.99**
			Extremely Severe	**0.98**	**0.99**	**0.99**
NB	0.83	0.84	Normal	0.66	0.84	0.93
			Mild	0.63	0.55	0.96
			Moderate	0.63	0.55	0.96

			Severe	0.70	0.76	0.96
			Extremely Severe	0.99	0.95	0.97
KNN	0.78	0.77	Normal	0.84	0.97	0.93
			Mild	0.58	0.53	0.94
			Moderate	0.53	0.55	0.93
			Severe	0.68	0.49	0.97
			Extremely Severe	0.96	0.95	0.98

(A)

	RF				
	N	**M**	**M**	**S**	**E**
N	348	7	0	0	0
M	35	127	30	0	0
M	0	52	94	18	0
S	0	3	37	78	39
E	0	0	1	27	484

(B)

	DT				
	N	**M**	**M**	**S**	**E**
N	338	17	0	0	0
M	37	123	28	4	0
M	1	39	94	27	3
S	0	7	33	80	37
E	0	0	5	41	466

(C)

	SVM				
	N	**M**	**M**	**S**	**E**
N	354	1	0	0	0
M	0	188	4	0	0
M	0	8	150	6	0
S	0	0	5	142	10
E	0	0	0	4	508

(D)

	NB				
	N	**M**	**M**	**S**	**E**
N	320	35	0	0	0
M	10	161	21	0	0
M	0	49	91	24	0
S	0	0	33	119	5
E	0	0	0	27	485

(E)

	KNN				
	N	**M**	**M**	**S**	**E**
N	346	9	0	0	0
M	68	102	22	0	0
M	0	64	90	10	0
S	0	1	58	77	21
E	0	0	0	26	486

Figure 4.20: Confusion matrices of ML models for the Hybrid dataset for Anxiety(1:11)

Below are the graphs showing performance evaluation (Accuracy, Precision, Recall and Specificity) of Anxiety for each of the model individually.

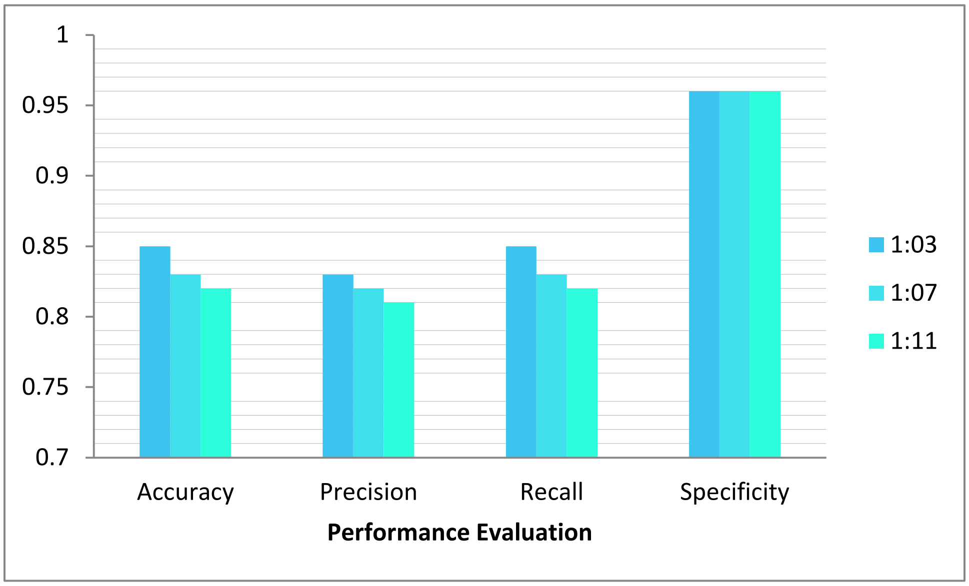


Figure 4.21: Performance Evaluation of RF for Anxiety across Different Ratios

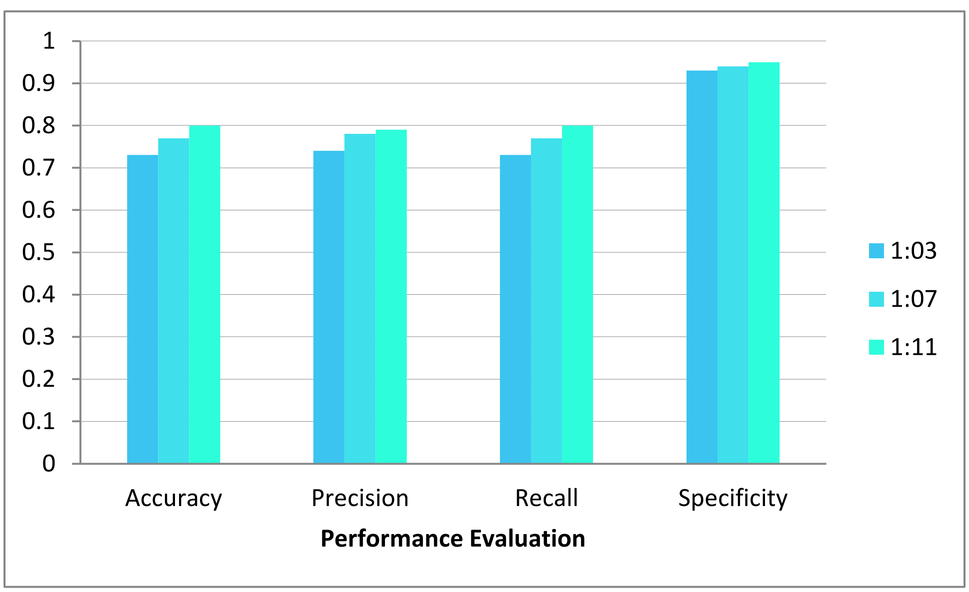


Figure 4.22: Performance Evaluation of DT for Anxiety across Different Ratios

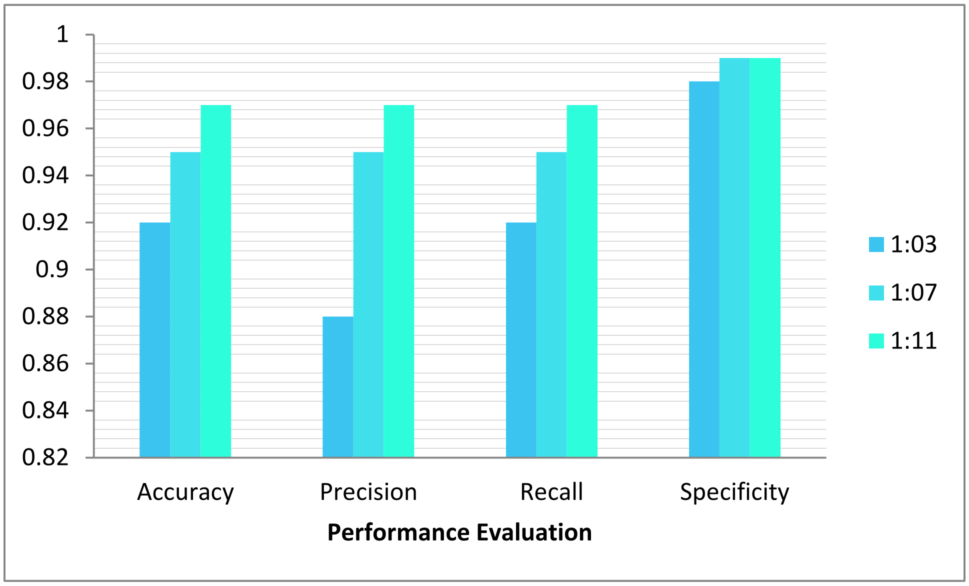


Figure 4.23: Performance Evaluation of SVM for Anxiety across Different Ratios

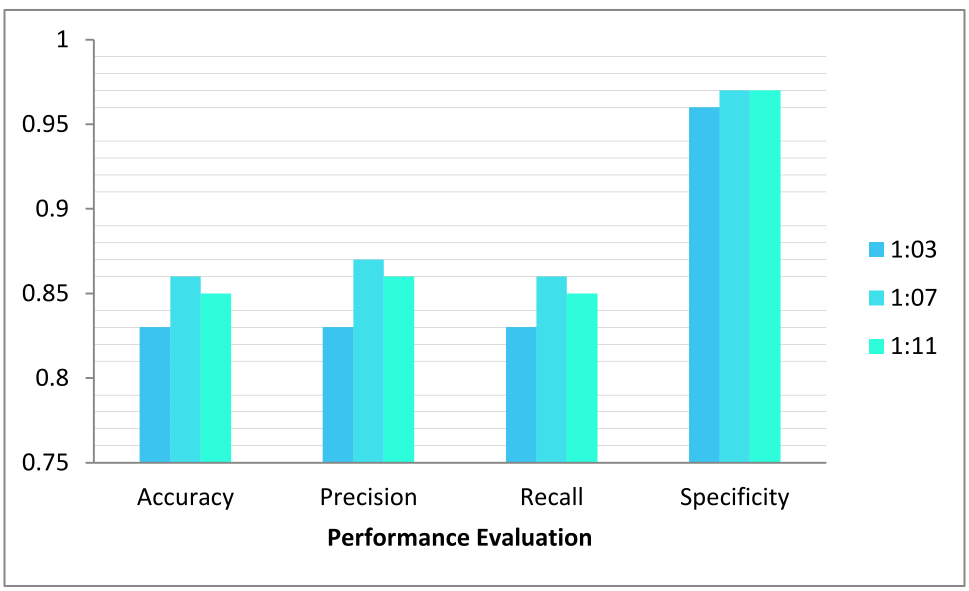


Figure 4.24:Performance Evaluation of NB for Anxiety across Different Ratios

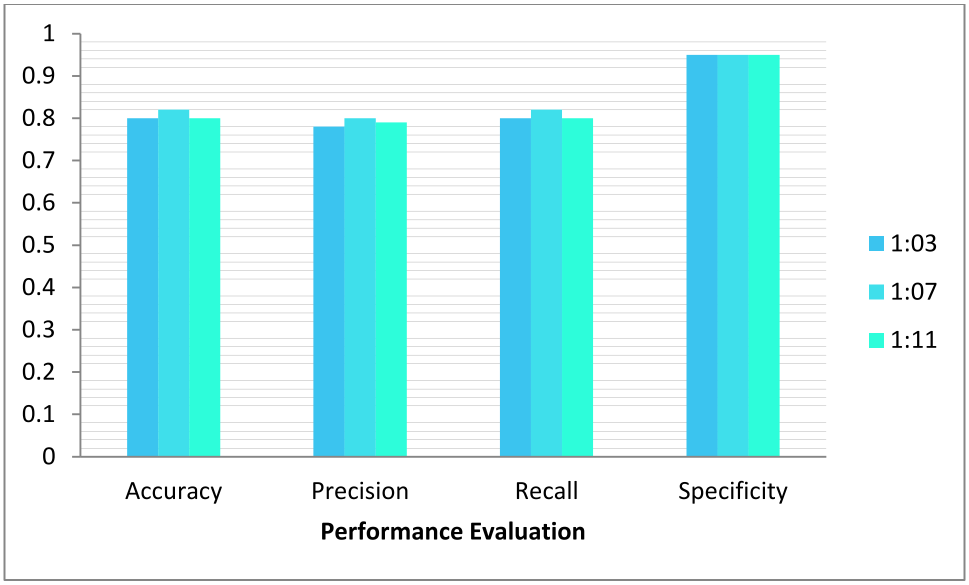


Figure 4.25: Performance Evaluation of KNN for Anxiety across Different Ratios

4.8.7 *Stress (1:3)*

For synthetic dataset for stress ratio 1:3, RF correctly classified 211 out of 217 normal cases, 38 out of 65 mild cases, 40 out of 69 moderate cases, 54 out of 73 severe cases, and 28 out of 36 extremely severe cases. DT correctly classified 198 out of 217 normal cases, 33 out of 65 mild cases, 41 out of 69 moderate cases, 50 out of 73 severe cases, and 27 out of 36 extremely severe cases. SVM correctly classified 214 out of 217 normal cases, 57 out of 65 mild cases, 57 out of 69 moderate cases, 67 out of 73 severe cases, and 30 out of 36 extremely severe cases. NB correctly classified 199 out of 217 normal cases, 53 out of 65 mild cases, 49 out of 69 moderate cases, 61 out of 73 severe cases, and 29 out of 36 extremely severe cases. KNN correctly classified 209 out of 217 normal cases, 30 out of 65 mild cases, 36 out of 69 moderate cases, 53 out of 73 severe cases, and 30 out of 36 extremely severe cases. The best performing model came out to be SVM with F1 score of 92 %. The results of all these five models for stress and their confusion matrices are given in the Table 4.22 and Figure 4.26 respectively.

Table 4.22: Performance evaluation of the five ML models on the Hybrid data for Stress (1:3)

Model	Accuracy	F1 Score	Classes	Precision	Recall	Specificity
RF	0.81	0.80	Normal	0.93	0.97	0.93
			Mild	0.64	0.58	0.95
			Moderate	0.62	0.58	0.94
			Severe	0.70	0.74	0.94
			Extremely Severe	0.88	0.78	0.99
DT	0.76	0.76	Normal	0.93	0.91	0.94
			Mild	0.49	0.51	0.91
			Moderate	0.55	0.59	0.91
			Severe	0.70	0.68	0.95
			Extremely Severe	0.79	0.75	0.98
SVM	**0.92**	**0.92**	**Normal**	**0.99**	**0.99**	**0.99**
			Mild	**0.86**	**0.88**	**0.98**
			Moderate	**0.86**	**0.83**	**0.98**
			Severe	**0.85**	**0.92**	**0.97**
			Extremely Severe	**0.94**	**0.83**	**1.00**
NB	0.85	0.86	Normal	1.00	0.92	1.00
			Mild	0.62	0.82	0.92
			Moderate	0.71	0.71	0.95
			Severe	0.84	0.84	0.97
			Extremely Severe	0.88	0.81	0.99
KNN	0.78	0.77	Normal	0.87	0.96	0.87
			Mild	0.53	0.46	0.93
			Moderate	0.65	0.52	0.95
			Severe	0.75	0.73	0.95
			Extremely Severe	0.83	0.83	0.99

(A)

	RF				
	N	M	M	S	E
N	211	6	0	0	0
M	16	38	11	0	0
M	0	14	40	15	0
S	0	1	14	54	4
E	0	0	0	8	28

(B)

	DT				
	N	M	M	S	E
N	198	19	0	0	0
M	13	33	19	0	0
M	2	14	41	12	0
S	0	1	15	50	7
E	0	0	0	9	27

(C)

	SVM				
	N	M	M	S	E
N	214	3	0	0	0
M	3	57	5	0	0
M	0	6	57	6	0
S	0	0	4	67	2
E	0	0	0	6	30

(D)

	NB				
	N	M	M	S	E
N	199	18	0	0	0
M	0	53	12	0	0
M	0	15	49	5	0
S	0	0	8	61	4
E	0	0	0	7	29

(E)

	KNN				
	N	M	M	S	E
N	209	8	0	0	0
M	30	30	5	0	0
M	2	19	36	12	0
S	0	0	14	53	6
E	0	0	0	6	30

Figure 4.26: Confusion matrices of ML models for the Hybrid dataset for Stress (1:3)

4.8.8 Stress (1:7)

For synthetic dataset for stress ratio 1:7, RF correctly classified 454 out of 467 normal cases, 76 out of 129 mild cases, 104 out of 135 moderate cases, 108 out of 125 severe cases, and 63 out of 71 extremely severe cases. DT correctly classified 427 out of

467 normal cases, 80 out of 129 mild cases, 84 out of 135 moderate cases, 90 out of 125 severe cases, and 58 out of 71 extremely severe cases. SVM correctly classified 462 out of 467 normal cases, 111 out of 129 mild cases, 124 out of 135 moderate cases, 122 out of 71 severe cases, and 66 out of 71 extremely severe cases. NB correctly classified 437 out of 467 normal cases, 105 out of 129 mild cases, 109 out of 135 moderate cases, 112 out of 125 severe cases, and 62 out of 287 extremely severe cases. KNN correctly classified 457 out of 467 normal cases, 80 out of 129 mild cases, 99 out of 135 moderate cases, 147 out of 125 severe cases, and 5 out of 71 extremely severe cases. The best performing model came out to be SVM with an F1 score of 96%. The results of all these five models for stress and their confusion matrices are given in the Table 4.23 and Figure 4.27 respectively.

Table 4.23: Performance evaluation of the five ML models on the Hybrid data for Stress (1:7)

Model	Accuracy	F1 Score	Classes	Precision	Recall	Specificity
RF	0.88	0.87	Normal	0.94	0.97	0.93
			Mild	0.67	0.62	0.95
			Moderate	0.78	0.77	0.96
			Severe	0.89	0.86	0.98
			Extremely Severe	0.94	0.89	1.00
DT	0.80	0.81	Normal	0.95	0.92	0.95
			Mild	0.55	0.65	0.92
			Moderate	0.65	0.62	0.94
			Severe	0.73	0.72	0.96
			Extremely Severe	0.82	0.82	0.98
SVM	**0.96**	**0.96**	**Normal**	**0.99**	**0.99**	**0.99**
			Mild	**0.90**	**0.90**	**0.98**
			Moderate	**0.93**	**0.92**	**0.99**
			Severe	**0.94**	**0.98**	**0.99**
			Extremely Severe	**1.00**	**0.93**	**1.00**
NB	0.90	0.90	Normal	1.00	0.94	1.00
			Mild	0.68	0.85	0.94
			Moderate	0.79	0.81	0.96
			Severe	0.90	0.90	0.98
			Extremely Severe	0.95	0.87	1.00
KNN	0.88	0.88	Normal	0.93	0.98	0.93
			Mild	0.67	0.65	0.95

			Moderate	0.80	0.73	0.97
			Severe	0.90	0.86	0.98
			Extremely Severe	0.93	0.92	0.99

(A)

RF	N	M	M	S	E
N	454	12	0	0	0
M	37	76	16	0	0
M	0	25	104	6	0
S	0	0	13	108	4
E	0	0	0	8	63

(B)

DT	N	M	M	S	E
N	427	34	4	1	0
M	22	80	21	0	0
M	1	31	84	19	0
S	1	0	21	90	13
E	0	0	0	13	58

(C)

SVM	N	M	M	S	E
N	462	4	0	0	0
M	6	111	6	0	0
M	0	8	124	3	0
S	0	0	3	122	0
E	0	0	0	5	66

(D)

NB	N	M	M	S	E
N	437	28	1	0	0
M	0	105	18	0	0
M	0	22	109	4	0
S	0	0	10	112	3
E	0	0	0	9	62

(E)

KNN	N	M	M	S	E
N	457	9	0	0	0
M	32	80	11	0	0
M	0	30	99	6	0
S	0	0	48	147	11
E	0	0	13	107	5

Figure 4.27: Confusion matrices of ML models for the Hybrid dataset for Stress (1:7)

4.8.9 *Stress (1:11)*

For synthetic dataset for stress ratio 1:11, RF correctly classified 632 out of 661 normal cases, 129 out of 190 mild cases, 166 out of 235 moderate cases, 154 out of 198 severe cases, and 76 out of 96 extremely severe cases. DT correctly classified 621 out of 661 normal cases, 119 out of 190 mild cases, 146 out of 235 moderate cases, 141 out of 198 severe cases, and 80 out of 96 extremely severe cases. SVM correctly classified 652 out of 661 normal cases, 186 out of 190 mild cases, 228 out of 235 moderate cases, 181 out of 198 severe cases, and 87 out of 96 extremely severe cases. NB correctly classified 604 out of 661 normal cases, 159 out of 190 mild cases, 209 out of 235 moderate cases, 175 out of 198 severe cases, and 74 out of 96 extremely severe cases. KNN correctly classified 637 out of 661 normal cases, 113 out of 190 mild cases, 163 out of 235 moderate cases, 163 out of 198 severe cases, and 85 out of 96 extremely severe cases. The best performing model came out to be SVM with F1 score of 97 %. The results of all these five models for stress and their confusion matrices are given in the Table 4.24 and Figure 4.28 respectively.

Table 4.24: Performance evaluation of the five ML models on the Hybrid data for Stress (1:11)

Model	Accuracy	F1 Score	Classes	Precision	Recall	Specificity
RF	0.84	0.84	Normal	0.94	0.96	0.95
			Mild	0.65	0.68	0.94
			Moderate	0.72	0.71	0.94
			Severe	0.76	0.78	0.96
			Extremely Severe	0.94	0.79	1.00
DT	0.80	0.80	Normal	0.93	0.94	0.93
			Mild	0.58	0.63	0.93
			Moderate	0.68	0.62	0.94
			Severe	0.74	0.71	0.96
			Extremely Severe	0.80	0.83	0.98
SVM	**0.97**	**0.97**	**Normal**	**1.00**	**0.99**	**1.00**
			Mild	**0.93**	**0.98**	**0.99**
			Moderate	**0.94**	**0.97**	**0.99**
			Severe	**0.94**	**0.91**	**0.99**
			Extremely Severe	**0.95**	**0.91**	**1.00**

NB	0.88	0.89	Normal	1.00	0.91	1.00
			Mild	0.68	0.84	0.94
			Moderate	0.79	0.89	0.95
			Severe	0.86	0.88	0.98
			Extremely Severe	1.00	0.77	1.00
KNN	0.84	0.84	Normal	0.92	0.96	0.92
			Mild	0.59	0.59	0.94
			Moderate	0.78	0.69	0.96
			Severe	0.84	0.82	0.97
			Extremely Severe	0.89	0.89	0.99

(A)

	RF				
	N	M	M	S	E
N	632	29	0	0	0
M	36	129	25	0	0
M	1	40	116	28	0
S	0	0	39	154	5
E	0	0	0	20	76

(B)

	DT				
	N	M	M	S	E
N	621	37	3	0	0
M	44	119	27	0	0
M	4	49	146	35	1
S	0	1	37	141	19
E	0	0	1	15	80

(C)

	SVM				
	N	M	M	S	E
N	652	9	0	0	0
M	1	186	3	0	0
M	0	4	228	3	0
S	0	0	12	181	5
E	0	0	0	9	87

(D)

	NB				
	N	M	M	S	E
N	604	56	1	0	0
M	0	159	31	0	0
M	0	20	209	6	0
S	0	0	23	175	0
E	0	0	0	22	74

(E)

	KNN				
	N	M	M	S	E
N	637	24	0	0	0
M	55	113	22	0	0
M	0	53	163	19	0
S	0	0	24	163	11
E	0	0	0	11	85

Figure 4.28: Confusion matrices of ML models for the Hybrid dataset for Stress (1:11)

Below are the graphs showing performance evaluation (Accuracy, Precision, Recall and Specificity) of Stress for each of the model individually.

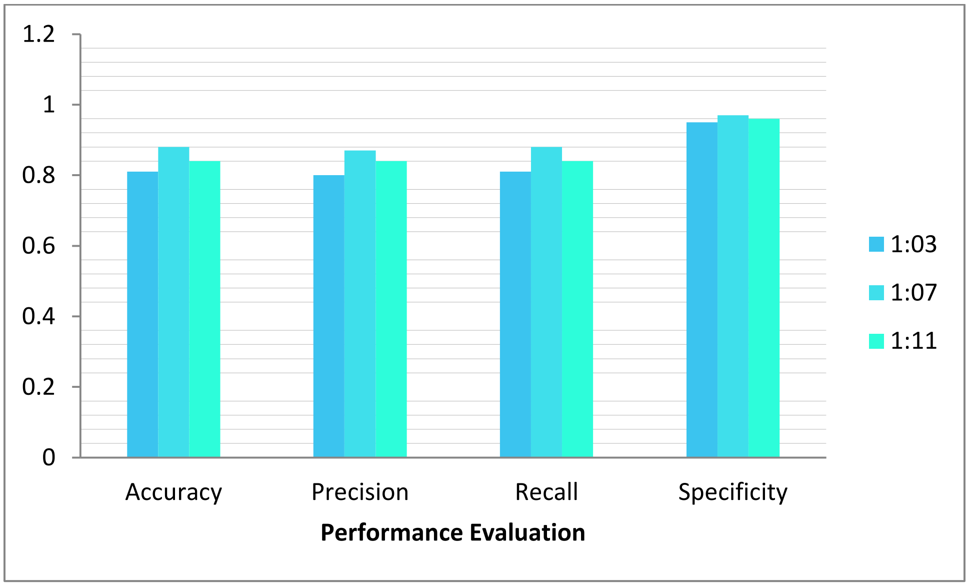


Figure 4.29: Performance Evaluation of RF for Stress across Different Ratio

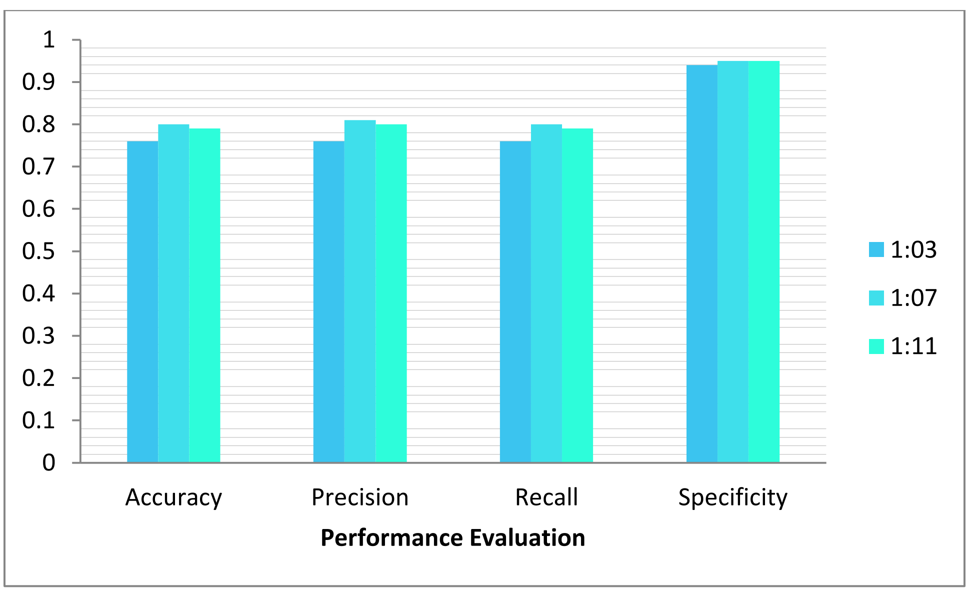


Figure 4.30: Performance Evaluation of DT for Stress across Different Ratios

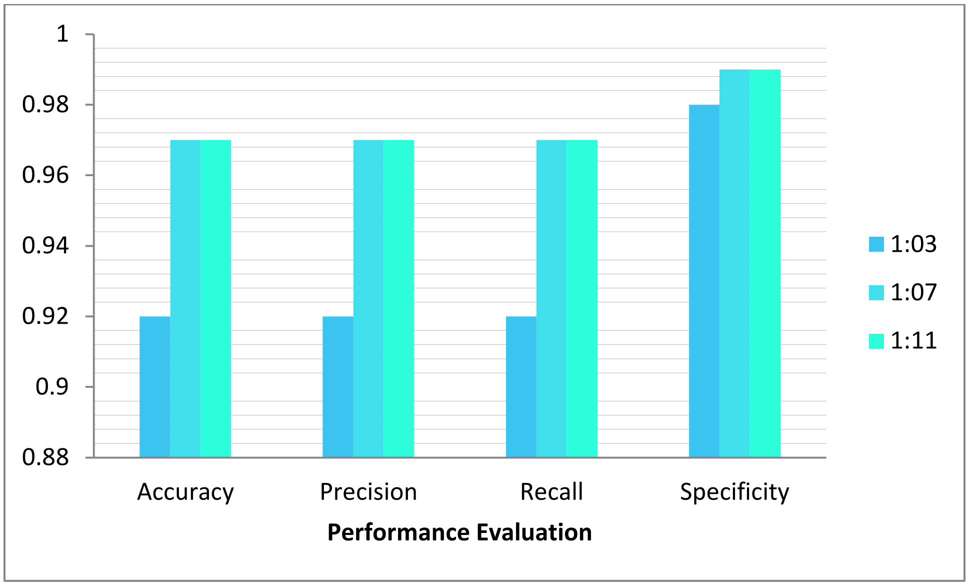


Figure 4. 31: Performance Evaluation of SVM for Stress across Different Ratios

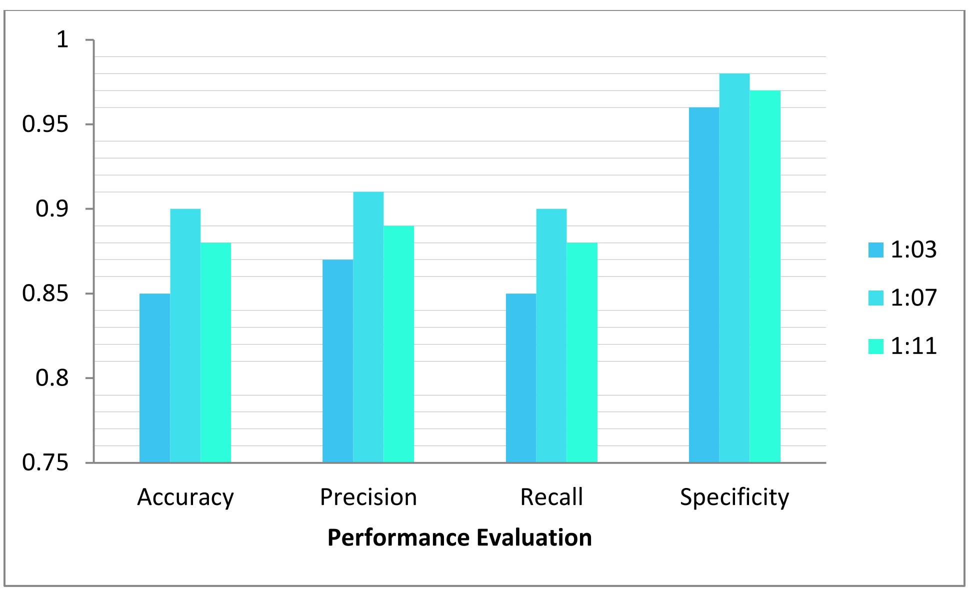


Figure 4.32: Performance Evaluation of NB for Stress across Different Ratios

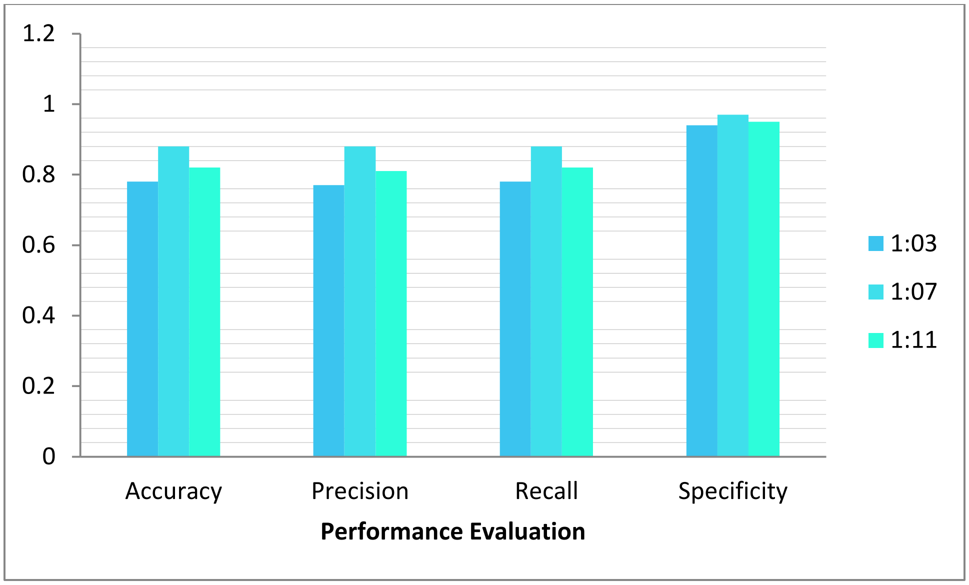


Figure 4.33: Performance Evaluation of KNN for Stress across Different Ratios

SVM demonstrate superior performance compared to other models when evaluated on different mental health scales and data ratios. The SVM demonstrate remarkable F1 score of 96%, 96%, and 97% for depression when the ratios were 1:3, 1:7, and 1:12, respectively. Regarding anxiety, the F1 score achieved are 92%, 95%, and 98% for the same ratios. This indicates a consistent pattern of enhanced performance as the datasets increase in size. Finally, in terms of stress, the SVM model achieved accuracy F1 score of 92%, 96%, and 98% respectively. The results emphasize the SVM's resilience and its superior capacity to manage diverse data distributions in various mental health assessments.there was also miss-classification of class mild for anxiety scale.Because there were less number of instances for mild class which lead to less recall rate. Below are the graph and table showing the F1 score comparison of different ML models across different ratios for all three scales (Depression, Anxiety and Stress).

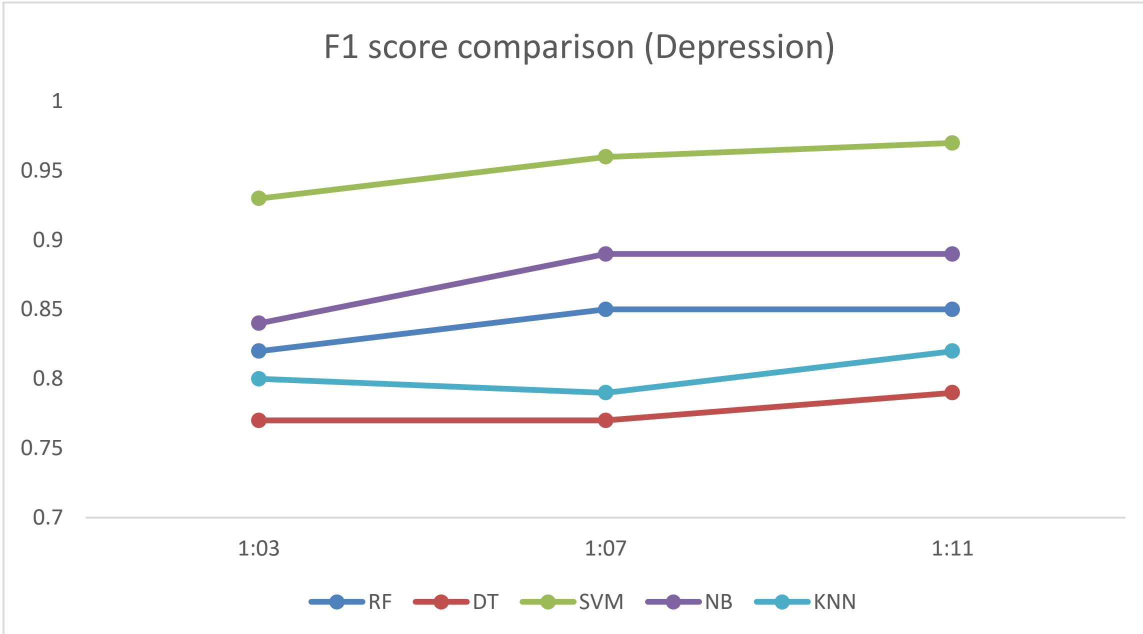


Figure 4.34: F1 Score Comparison of ML Models for Depression across Different Ratios

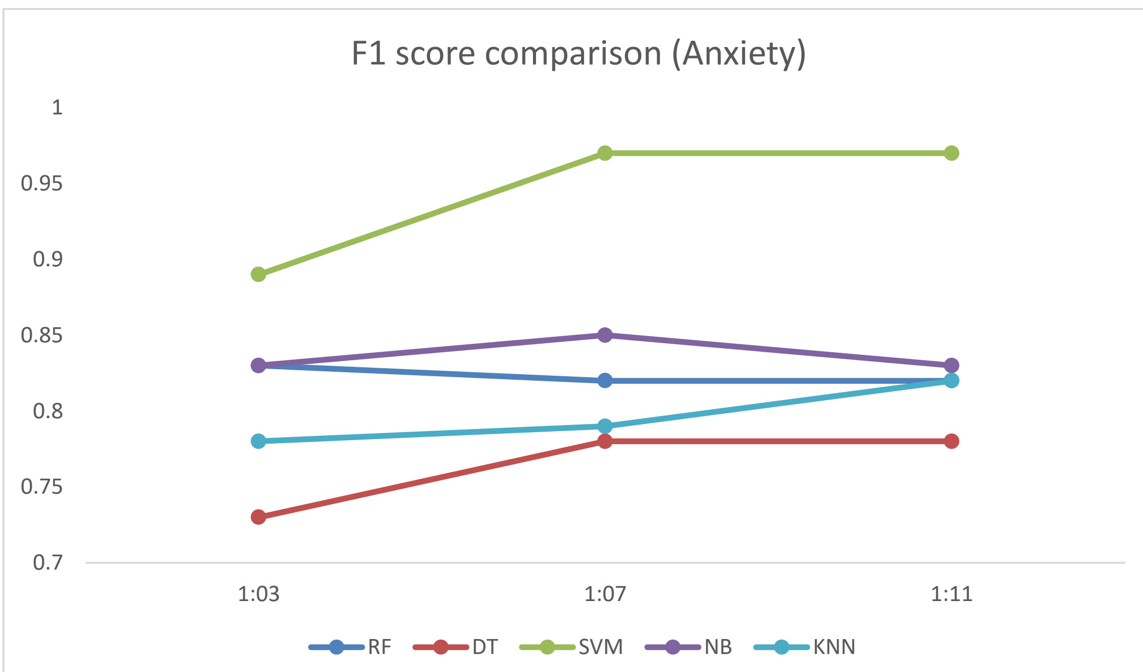


Figure 4.35: F1 Score Comparison of ML Models for Anxiety across Different Ratios

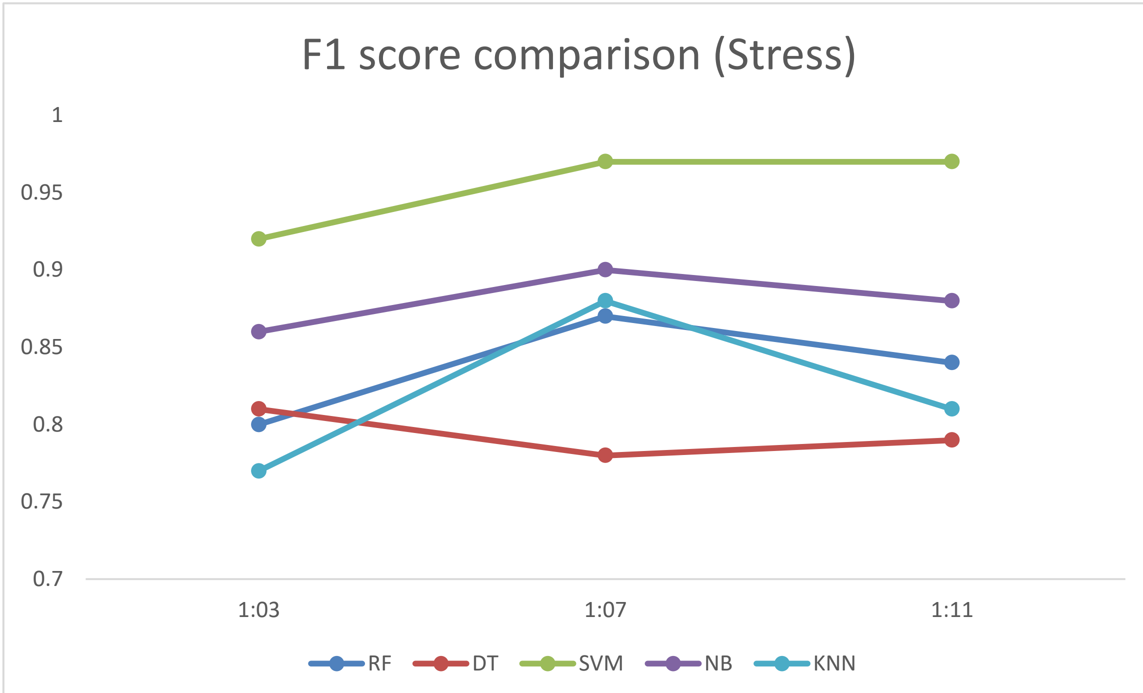


Figure 4.36: F1 Score Comparison of ML Models for Stress across Different Ratios

Table 4.25: F1 Scores of the Five ML Models on the Hybrid Data for Depression, Anxiety and Stress

Model	**Ratio**								
	1:3			**1:7**			**1:11**		
	D	A	S	D	A	S	D	A	S
RF	82	83	80	84	83	87	84	81	84
DT	76	73	76	77	78	81	79	80	80
SVM	**96**	**92**	**92**	**96**	**95**	**96**	**97**	**96**	**97**
NB	83	83	86	89	86	90	88	84	89
KNN	80	80	77	79	81	88	81	77	84

5. CONCLUSIONS AND FUTURE RECOMMENDATIONS

5.1 Key Findings and Strengths

The key results of this research reveal that ML prediction models can successfully screen students for depression, anxiety and stress with an F1 score of 0.97, 0.96 and 0.97. Among the five algorithms tested, SVM is the most successful, showing higher F1 score, accuracy, precision and specificity in predicting depression, anxiety and stress. Feature selection process demonstrates that each feature is critical to the ML prediction of depression, anxiety and stress. These findings are consistent with previous research on students' mental health and the use of ML to anticipate mental health difficulties, showing the efficacy of ML in mental health informatics. This work is the first to develop synthetic hybrid data exclusively for predicting mental health issues, representing a major advancement in the area. This new breakthrough is likely to spark future research and development on the integration of ML into mental healthcare, an area where present research is limited. By addressing issues such as social stigma, restricted access to services, and the financial expenses associated with mental health consultations, this study's contributions seek to minimize the burden of mental health impairments and promote overall well-being.

5.2 Limitations

This research has some significant limitations that must be carefully considered. One major problem is the use of a small and class imbalanced dataset to generate synthetic data, which may have an impact on the prediction models' reliability and applicability. Future study should prioritize getting a larger and more varied original dataset in order to increase the quality of synthetic data and eliminate bias. To address these concerns, it is critical to convert the created technique into an end-user application.

5.3 Future Recommendations

The proposed predictive models can be used in hospitals, clinics, universities, and even at home using the free self-rated version of DASS21. Within hospitals, these models might be used in psychiatric departments, where patients often have mental health difficulties. However, optimizing these models to overcome the highlighted constraints is critical to their effective application.

Mental health education should be incorporated into university curriculum to improve access to mental health resources such counseling services, safe disclosure spaces, and user-friendly online support networks. Universities can significantly reduce stigma associated with mental health issues and promote overall student well-being by incorporating mental health education into academic programs, which aligns with previous proposals for preventative measures in higher education settings. Additionally, digital monitoring tools and solutions employed in this research are advised. These systems allow continuous monitoring and personalized treatments in mental health.

For further investigation, alternate approaches for generating synthetic data, especially for multi-classification such as those experienced in this research. Several deep learning techniques such as Generative Adversarial Networks (GANs) can be used to generate synthetic data that properly represents the complicated class distributions and relationships seen in mental health datasets [50]. Given the nature of multi-class predictions for mental health disorders, novel ways to producing synthetic data may include strategies that retain the nuanced class distributions and relationships seen in the original dataset. This method assures that synthetic data closely resembles real-world circumstances, which improves the robustness and generalizability of prediction models in mental health informatics.

6. REFERENCE

[1] M. Wada *et al.*, “University students’ perspectives on mental illness stigma,” *Ment. Health Prev.*, vol. 14, p. 200159, Jun. 2019, doi: 10.1016/j.mph.2019.200159.

[2] D. Wheatley, “2. STRESS, ANXIETY AND DEPRESSION,” *Stress Med.*, vol. 13, no. 3, pp. 173–177, Jul. 1997, doi: 10.1002/(SICI)1099-1700(199707)13:3<173::AID-SMI739>3.0.CO;2-6.

[3] J. Lally, A. Ó Conghaile, S. Quigley, E. Bainbridge, and C. McDonald, “Stigma of mental illness and help-seeking intention in university students,” *The Psychiatrist*, vol. 37, no. 8, pp. 253–260, Aug. 2013, doi: 10.1192/pb.bp.112.041483.

[4]. Mossie, D. Kindu, and A. Negash, “Prevalence and Severity of Depression and Its Association with Substance Use in Jimma Town, Southwest Ethiopia,” *Depression Research and Treatment*, vol. 1–7, 2016, doi: 10.1155/2016/3460462.

[5] A. Javed, M. N. S. Khan, A. Nasar, and A. Rasheed, “Mental healthcare in Pakistan,” *Taiwan. J. Psychiatry*, 2020.

[6] Kieling C, Baker-Henningham H, Belfer M, Conti G, Ertem I, Omigbodun O, Rohde LA, Srinath S, Ulkuer N, Rahman A., “Child and adolescent mental health worldwide: evidence for action,” *The Lancet*, vol. 378, no. 9801, pp. 1515–1525, 2011, doi: https://doi.org/10.1016/S0140-6736(11)60827-1.

[7] z, “Fall 2017 Reference Group Executive Summary,” 2017. [Online]. Available: https://www.acha.org/documents/ncha/NCHA-II_FALL_2017_REFERENCE_GROUP_EXECUTIVE_SUMMARY.pdf

[8] M. O. Husain et al., ““Stigma toward mental and physical illness: attitudes of healthcare professionals, healthcare students and the general public in Pakistan,” *BJPsych Open*, vol. 6, no.6, p.e81, 2020.

[9] R. Parekh, “What is Mental Illness?,” What is Mental Illness? [Online]. Available: https://www.psychiatry.org/patients-families/what-is-mental-illness#:~:text=Mental%20illnesses%20are%20health%20conditions

[10] F.Torres, “What is Depression?” [Online]. Available: https://www.psychiatry.org/patients-families/depression/what-is-depression

[11] Cleveland Clinic, “Anxiety Disorders.” [Online]. Available:

https://my.clevelandclinic.org/health/diseases/9536-anxiety-disorders

[12] "Stress." [Online]. Available: https://www.who.int/news-room/questions-and-answers/item/stress#:~:text=What%20is%20Stress%3F

[13] "Mental Health Screening Tools." Accessed: Nov. 26, 2023. [Online]. Available: https://www.annabellepsychology.com/screening-tools

[14] B. Löwe *et al.*, "Validation and Standardization of the Generalized Anxiety Disorder Screener (GAD-7) in the General Population," *Med. Care*, vol. 46, no. 3, pp. 266–274, Mar. 2008, doi: 10.1097/MLR.0b013e318160d093.

[15] R. L. Spitzer, K. Kroenke, J. B. W. Williams, and B. Löwe, "A Brief Measure for Assessing Generalized Anxiety Disorder The GAD-7".

[16] Easton SD, Safadi NS, Wang Y, Hasson RG 3rd, "The Kessler psychological distress scale: translation and validation of an Arabic version," *Health Qual Life Outcomes*, Oct. 2017, doi: 10.1186/s12955-017-0783-9.

[17] D. P. Goldberg et al., "The validity of two versions of the GHQ in the WHO study of mental illness in general health care," *Psychol Med*, vol. 27, no.1, pp. 191–197, Jan. 1997, doi: 10.1017/S0033291796004242.

[18] M. Sterling, "General Health Questionnaire – 28 (GHQ-28)," *J. Physiother.*, vol. 57, no. 4, p. 259, 2011, doi: 10.1016/S1836-9553(11)70060-1.

[19] Lovibond, S. H., & Lovibond, P. F., "Manual for the Depression Anxiety Stress Scales," *Sydney: Psychology Foundation*, 1995.

[20] "What is Artificial Intelligence, Machine Learning, and Deep Learning?" Accessed: Nov. 26, 2023. [Online]. Available: https://rapidminer.com/blog/artificial-intelligence-machine-learning-deep-learning/

[21] M. Srividya, S. Mohanavalli, and N. Bhalaji, "Behavioral Modeling for Mental Health using Machine Learning Algorithms," *J. Med. Syst.*, vol. 42, no. 5, p. 88, May 2018, doi: 10.1007/s10916-018-0934-5.

[22] Babcock University *et al.*, "Supervised Machine Learning Algorithms: Classification and Comparison," *Int. J. Comput. Trends Technol.*, vol. 48, no. 3, pp. 128–138, Jun. 2017, doi: 10.14445/22312803/IJCTT-V48P126.

[23] T. Hofmann, "Unsupervised Learning by Probabilistic Latent Semantic Analysis".

[24] S. Graham *et al.*, "Artificial Intelligence for Mental Health and Mental Illnesses: an

Overview," *Curr. Psychiatry Rep.*, vol. 21, no. 11, p. 116, Nov. 2019, doi: 10.1007/s11920-019-1094-0.

[25] Nayan, M. I. H., Uddin, M. S. G., Hossain, M. I., Alam, M. M., Zinnia, M. A., Haq, I., ... & Methun, M. I. H, "Comparison of the performance of machine learning-based algorithms for predicting depression and anxiety among University Students in Bangladesh: A result of the first wave of the COVID-19 pandemic," *Asian Journal of Social Health and Behavior*, vol. 5(2), pp. 75–84, 2022.

[26] S. I. Khan, A. Islam, A. Hossen, T. I. Zahangir, and A. S. Md. Latiful Hoque, "Supporting the Treatment of Mental Diseases using Data Mining," in *2018 International Conference on Innovations in Science, Engineering and Technology (ICISET)*, Chittagong, Bangladesh: IEEE, Oct. 2018, pp. 339–344. doi: 10.1109/ICISET.2018.8745591.

[27] Samarkand State University, A. Voris Yunusovich, F. Ahmedov, K. Norboyev, and F. Zakirov, "Analysis of Experimental Research Results Focused on Improving Student Psychological Health," *Int. J. Mod. Educ. Comput. Sci.*, vol. 14, no. 2, pp. 14–30, Apr. 2022, doi: 10.5815/ijmecs.2022.02.02.

[28] I. J. Ratul, M. M. Nishat, F. Faisal, S. Sultana, A. Ahmed, and M. A. Al Mamun, "Analyzing Perceived Psychological and Social Stress of University Students: A Machine Learning Approach," *Heliyon*, vol. 9, no. 6, p. e17307, Jun. 2023, doi: 10.1016/j.heliyon.2023.e17307.

[29] J. Chung and J. Teo, "Mental Health Prediction Using Machine Learning: Taxonomy, Applications, and Challenges," *Appl. Comput. Intell. Soft Comput.*, vol. 2022, pp. 1–19, Jan. 2022, doi: 10.1155/2022/9970363.

[30] S. Saleem, Z. Mahmood, and M. Naz, "Mental Health Problems in University Students: A Prevalence Study".

[31] Dr. E. Sandhu, "COMPUTE DEPRESSION AND ANXIETY AMONG STUDENTS IN PAKISTAN, USING MACHINE LEARNING," *Lahore Garrison Univ. Res. J. Comput. Sci. Inf. Technol.*, vol. 6, no. 04, Dec. 2022, doi: 10.54692/lgurjcsit.2022.0604402.

[32] G. VanRossum and F. L. Drake, "The python language reference," *Python Softw. Found. Amst.*, vol. 561, Netherlands 2010.

[33] Viorel Ciorniciuc, “DASS 21 - Depression Anxiety Stress Scale Test,” 2015, [Online]. Available: https://www.thecalculator.co/health/DASS21-Depression-Anxiety-Stress-Scale-Test-938.html

[34] Altamash Institute of Dental Medicine, Karachi and A. Salahuddin, “Depression, Stress and Anxiety in Post-Graduate Dentists,” *Isra Med. J.*, vol. 14, no. 4, pp. 152–156, Jul. 2023, doi: 10.55282/imj.oa1339.

[35] D. G. Bonett and T. A. Wright, “Cronbach’s alpha reliability: Interval estimation, hypothesis testing, and sample size planning,” *J Organ Behav*, vol. 36, no. 1, pp. 3–15, 2015.

[36] M. Tavakol and R. Dennick, “Making sense of Cronbach’s alpha,” *Int J Med Educ*, vol. 2, p. 53, 2011.

[37] B. Damasceno and B. Damasceno, “‘Hypothesis Testing,’ Research on Cognition Disorders: Theoretical and Methodological Issues,” pp. 157–166, 2020.

[38] N. Nnamoko, F. Arshad, D. England, J. Vora, and J. Norman, “Evaluation of filter and wrapper methods for feature selection in supervised machine learning,” *Age Omaha*, vol. 21, no. 81, pp. 32–33, 2014.

[39] A. Bommert, X. Sun, B. Bischl, J. Rahnenführer, and M. Lang, “Benchmark for filter methods for feature selection in high-dimensional classification data,” *Comput Stat Data Anal*, vol. 143, p. 106839, 2020, doi: https://doi.org/10.1016/j.csda.2019.106839.

[40] N. Pandis, “‘The chi-square test,’ American journal of orthodontics and dentofacial orthopedics,” vol. 150, no. 5, pp. 898–899, 2016.

[41] S.-K. Hong, H. Kim, S. Lee, and Y.-S. Moon, “Secure multiparty computation of chi-square test statistics and contingency coefficients,”,” *IEEE*, pp. 53–56, 2017.

[42] B. F. Darst, K. C. Malecki, and C. D. Engelman, ““Using recursive feature elimination in random forest to account for correlated variables in high dimensional data,” *BMC Genet*, vol. 19, pp. 1–6, 2018.

[43] A. Gonzales, G. Guruswamy, and S. R. Smith, “Synthetic data in health care: a narrative review,” *PLOS Digit. Health*, vol. 21, no. 1, p. e0000082, 2023.

[44] F. Pedregosa et al, “‘Scikit-learn: Machine learning in Python,’ the Journal of machine Learning research,” vol. 12, pp. 2825–2830, 2011.

[45] A. Cutler, D. R. Cutler, and J. R. Stevens, "'Random forests,' Ensemble machine learning: Methods and applications," pp. 157–175, 2012.

[46] S. Chen, G. I. Webb, L. Liu, and X. Ma, "A novel selective naïve Bayes algorithm," *Knowl.-Based Syst.*, vol. 192, p. 105361, Mar. 2020, doi: 10.1016/j.knosys.2019.105361.

[47] L. Zhang, L. Jiang, C. Li, and G. Kong, "Two feature weighting approaches for naive Bayes text classifiers," *Knowl.-Based Syst.*, vol. 100, pp. 137–144, May 2016, doi: 10.1016/j.knosys.2016.02.017.

[48] G. Guo, H. Wang, D. Bell, Y. Bi, and K. Greer, "KNN Model-Based Approach in Classification," in *On The Move to Meaningful Internet Systems 2003: CoopIS, DOA, and ODBASE*, vol. 2888, R. Meersman, Z. Tari, and D. C. Schmidt, Eds., in Lecture Notes in Computer Science, vol. 2888. , Berlin, Heidelberg: Springer Berlin Heidelberg, 2003, pp. 986–996. doi: 10.1007/978-3-540-39964-3_62.

[49] V. E. Staartjes and J. M. Kernbach, "Significance of external validation in clinical machine learning: let loose too early?," *Spine J.*, vol. 20, no. 7, pp. 1159–1160, Jul. 2020, doi: 10.1016/j.spinee.2020.02.016.

[50] Z. Pan, W. Yu, X. Yi, A. Khan, F. Yuan, and Y. Zheng, "Recent progress on generative adversarial networks (GANs): A survey," *IEEE Access*, vol. 7, pp. 36322–36333, 2019.

APPENDIX

A. Code for Testing Goodness of fit using Multinomial Probability Distribution

```
# Load libraries
library(MASS)  # for chisq.test
# Read data
data=read.csv("DATA.csv",header=TRUE)
# Check for missing values (not recommended for multinomial)
if (any(is.na(data$DASS01))) {
  stop("Data contains missing values. Multinomial not suitable!")
}
# Get observed counts
observed <- table(data$DASS01)
# Define total number of trials (observations)
n <- sum(observed)
# Define the vector of unequal probabilities (replace with your values)
p <- c(0.41, 0.30, 0.21, 0.6)  # Probabilities for categories 0, 1, 2, 3
# Check if probability vector length matches category count
if (length(p) != nrow(table(data$rDASS01))) {
  stop("Probability vector length must match number of categories!")
```

}

 # Calculate expected counts using the probabilities

expected <- n * p

Create table for observed and expected counts

counts_table <- cbind(Category = names(observed), Observed = observed, Expected = expected)

Print the table

cat("Table of Observed and Expected Counts:\n")

print(counts_table)

 # Perform Chi-squared goodness-of-fit test

chisq.result <- chisq.test(observed, p = expected/n)

Print results

cat("Chi-squared test for multinomial distribution with unequal probabilities:\n")

print(chisq.result)

Interpret results

if (chisq.result$p.value > 0.05) {

  cat("p-value =", chisq.result$p.value,

      "\nWe fail to reject the null hypothesis of multinomial fit.\n")

} else {

  cat("p-value =", chisq.result$p.value,

      "\nWe reject the null hypothesis of multinomial fit.\n")

}

B. Code for Generating Synthetic data with Correlation

```
# Start with an identity matrix and then fill off-diagonal with the target correlation

cor_matrix <- diag(num_series) * (1 - target_correlation) + target_correlation

# Ensure the matrix is positive definite

# This can be adjusted to ensure positive definiteness

cor_matrix <- nearPD(cor_matrix, corr = TRUE)$mat

# Generate multivariate normal distribution with the specified correlation matrix

normals <- mvrnorm(n, mu = rep(0, num_series), Sigma = cor_matrix)

# Transform the normal variables to uniform using the CDF

uniforms <- pnorm(normals)

# Define the probabilities for each series

probs_list <- list(

  c(0.20, 0.40, 0.20, 0.20),

  c(0.38, 0.31, 0.18, 0.13),

  c(0.31, 0.26, 0.22, 0.21),

  c(0.28, 0.32, 0.19, 0.21),

  c(0.44, 0.23, 0.17, 0.16),

  c(0.21, 0.35, 0.23, 0.21),

  c(0.30, 0.29, 0.23, 0.18)

)
```

```
# Function to map uniform variables to multinomial
uniform_to_multinomial <- function(u, probs) {
  return(findInterval(u, cumsum(probs), rightmost.closed = TRUE))
}
# Generate the series of multinomial random numbers
series_list <- lapply(1:num_series, function(i) {
  sapply(uniforms[, i], uniform_to_multinomial, probs = probs_list[[i]])
})
# Combine the series into a data frame
result <- do.call(cbind, series_list)
colnames(result) <- paste0("series", 1:num_series)
# Save to a CSV file
write.csv(result, file = "Ran_Mul.csv", row.names = FALSE)
# Verify the average correlation
cor_matrix_result <- cor(result)
average_correlation <- mean(cor_matrix_result[upper.tri(cor_matrix_result)])
print(average_correlation)
```

C. Feature Selection

 # Uploading File

```
from google.colab import files
```

```
import pandas as pd

# Uploading the CSV file from computer

uploaded = files.upload()

filename = 'Stress.csv'

df = pd.read_csv(filename)

print(df)

# Separate the features and target variables

X = df.drop('Target', axis=1)

y = df['Target']
```

 # Random Forest

```
import numpy as np

import matplotlib.pyplot as plt

from sklearn.datasets import make_classification

from sklearn.model_selection import cross_val_score, StratifiedKFold

from sklearn.feature_selection import RFE

from sklearn.tree import DecisionTreeClassifier

from sklearn.ensemble import RandomForestClassifier

from sklearn.pipeline import Pipeline

# Function to evaluate a model using stratified 5-fold cross-validation

def evaluate_model(model, X, y):

    cv = StratifiedKFold(n_splits=5, shuffle=True, random_state=42)
```

```
    scores = cross_val_score(model, X, y, scoring='accuracy', cv=cv, n_jobs=-1)
    return np.mean(scores)
# Initialize a RandomForestClassifier as the estimator
estimator = RandomForestClassifier(n_estimators=100, random_state=42)
# Initialize a list to store the mean accuracies for each number of selected features
mean_accuracies = []
# Initialize a list to store the names of the optimal number of selected features
optimal_feature_counts = []
# Loop through different numbers of selected features
for i in range(1, 8):
    rfe = RFE(estimator=estimator, n_features_to_select=i)
    model = RandomForestClassifier()
    pipeline = Pipeline(steps=[('s', rfe), ('m', model)])
    mean_accuracy = evaluate_model(pipeline, X, y)

    mean_accuracies.append(mean_accuracy)
    optimal_feature_counts.append(i)
    print(f"Number of Selected Features: {i}, Mean Accuracy: {mean_accuracy:.3f}")
# Find the index of the maximum mean accuracy
optimal_idx = np.argmax(mean_accuracies)
optimal_features = optimal_feature_counts[optimal_idx]
```

print(f"Optimal Number of Selected Features: {optimal_features}")

Plot feature importances using the RandomForestClassifier

estimator.fit(X, y)

importances = estimator.feature_importances_

print(importances)

plt.figure(figsize=(10, 6))

plt.title("Feature Importances")

plt.bar(range(len(importances)), importances, tick_label=np.arange(1, len(importances) + 1))

plt.xlabel("Feature Number")

plt.ylabel("Importance")

plt.show()

D. Machine Learning Models

D.1 Random Forest

from google.colab import files

import numpy as np

import pandas as pd

from sklearn.ensemble import RandomForestClassifier

from sklearn.metrics import accuracy_score, precision_score, recall_score, f1_score, confusion_matrix, make_scorer

from sklearn.model_selection import cross_val_predict, StratifiedKFold

```
import seaborn as sns

import matplotlib.pyplot as plt

# Uploading the CSV file from computer

uploaded = files.upload()

filename = 'Stress.csv'

data = pd.read_csv(filename)

print(data)

# Separate features (questions) and target variable (evaluation)

X = data.drop('Target', axis=1)

y = data['Target']

# Initialize the Random Forest model

model = RandomForestClassifier(n_estimators=100, random_state=42)

# Use Stratified K-Fold cross-validation

skf = StratifiedKFold(n_splits=5, shuffle=True, random_state=42)

# Cross-validation predictions

y_pred = cross_val_predict(model, X, y, cv=skf)

# Calculate metrics

accuracy = accuracy_score(y, y_pred)

precision = precision_score(y, y_pred, average='weighted')

f1 = f1_score(y, y_pred, average='weighted')

recall = recall_score(y, y_pred, average='weighted')
```

conf_matrix = confusion_matrix(y, y_pred)

 # Calculate specificity for each class

def calculate_specificity(conf_matrix, class_index):

true_negatives = np.sum(conf_matrix) - np.sum(conf_matrix[class_index, :]) - np.sum(conf_matrix[:, class_index]) + conf_matrix[class_index, class_index]

false_positives = np.sum(conf_matrix[:, class_index]) - conf_matrix[class_index, class_index]

specificity = true_negatives / (true_negatives + false_positives)

return specificity

specificity_per_class = [calculate_specificity(conf_matrix, i) for i in range(conf_matrix.shape[0])]

overall_specificity = np.mean(specificity_per_class)

Calculate precision, recall, specificity, and F1-score for each class

precision_per_class = precision_score(y, y_pred, average=None)

recall_per_class = recall_score(y, y_pred, average=None)

f1_per_class = f1_score(y, y_pred, average=None)

 # Print the results

print(f"Random Forest Results:")

print(f"Accuracy: {accuracy:.2f}")

print(f"Precision (Weighted): {precision:.2f}")

print(f"F1-Score (Weighted): {f1:.2f}")

print(f"Recall (Weighted): {recall:.2f}")

```
print(f"Overall Specificity: {overall_specificity:.2f}")
print("")
# Print class-specific metrics
class_labels = ['Normal', 'Mild', 'Moderate', 'Severe', 'Extremely Severe']
for i, label in enumerate(class_labels):
    print(f"Class: {label}")
    print(f"Precision: {precision_per_class[i]:.2f}")
    print(f"Recall: {recall_per_class[i]:.2f}")
    print(f"Specificity: {specificity_per_class[i]:.2f}")
    print(f"F1-Score: {f1_per_class[i]:.2f}")
    print("")
# Visualize the confusion matrix in a table form
conf_matrix_df = pd.DataFrame(conf_matrix, index=class_labels, columns=class_labels)
plt.figure(figsize=(10, 7))
sns.set(font_scale=1.2)  # Adjust the font scale for better readability
sns.heatmap(conf_matrix_df, annot=True, fmt='d', cmap='Blues', annot_kws={"size": 14})  # Adjusted annotation size
plt.ylabel('Actual', fontsize=12)
plt.xlabel('Predicted', fontsize=12)
plt.title('Confusion Matrix', fontsize=15)
plt.show()
```

D.2 Decision Tree

 # Import necessary libraries

```
from google.colab import files

import numpy as np

import pandas as pd

from sklearn.model_selection import cross_val_score, cross_val_predict, KFold

from sklearn.metrics import accuracy_score, precision_score, recall_score, f1_score,
confusion_matrix

from sklearn.tree import DecisionTreeClassifier

import seaborn as sns

import matplotlib.pyplot as plt

# Uploading the CSV file from computer

uploaded = files.upload()

filename = 'Stress.csv'

data = pd.read_csv(filename)

print(data)

# Separate features (questions) and target variable (evaluation)

X = data.drop('Target', axis=1)

y = data['Target']

# Use Stratified K-Fold cross-validation

skf = StratifiedKFold(n_splits=5, shuffle=True, random_state=42)
```

```
# Initialize the Decision Tree model
model = DecisionTreeClassifier()
# Perform cross-validation
y_pred = cross_val_predict(model, X, y, cv=skf)
# Calculate metrics
accuracy = accuracy_score(y, y_pred)
precision = precision_score(y, y_pred, average='weighted')
f1 = f1_score(y, y_pred, average='weighted')
recall = recall_score(y, y_pred, average='weighted')
conf_matrix = confusion_matrix(y, y_pred)
# Calculate specificity for each class
def calculate_specificity(conf_matrix, class_index):
    true_negatives = np.sum(conf_matrix) - np.sum(conf_matrix[class_index, :]) - np.sum(conf_matrix[:, class_index]) + conf_matrix[class_index, class_index]
    false_positives = np.sum(conf_matrix[:, class_index]) - conf_matrix[class_index, class_index]
    specificity = true_negatives / (true_negatives + false_positives)
    return specificity
specificity_per_class = [calculate_specificity(conf_matrix, i) for i in range(conf_matrix.shape[0])]
overall_specificity = np.mean(specificity_per_class)
```

 # Calculate precision, recall, specificity, and F1-score for each class

```
precision_per_class = precision_score(y, y_pred, average=None)

recall_per_class = recall_score(y, y_pred, average=None)

f1_per_class = f1_score(y, y_pred, average=None)
```

 # Print the results

```
print(f"Decision Tree Results:")

print(f"Accuracy: {accuracy:.2f}")

print(f"Precision (Weighted): {precision:.2f}")

print(f"F1-Score (Weighted): {f1:.2f}")

print(f"Recall (Weighted): {recall:.2f}")

print(f"Overall Specificity: {overall_specificity:.2f}")

print("")

# Print class-specific metrics

class_labels = ['Normal', 'Mild', 'Moderate', 'Severe', 'Extremely Severe']

for i, label in enumerate(class_labels):

    print(f"Class: {label}")

    print(f"Precision: {precision_per_class[i]:.2f}")

    print(f"Recall: {recall_per_class[i]:.2f}")

    print(f"Specificity: {specificity_per_class[i]:.2f}")

    print(f"F1-Score: {f1_per_class[i]:.2f}")

    print("")
```

 # Visualize the confusion matrix in a table form

conf_matrix_df = pd.DataFrame(conf_matrix, index=class_labels, columns=class_labels)

plt.figure(figsize=(10, 7))

sns.set(font_scale=1.2) # Adjust the font scale for better readability

sns.heatmap(conf_matrix_df, annot=True, fmt='d', cmap='Blues', annot_kws={"size": 14}) # Adjusted annotation size

plt.ylabel('Actual', fontsize=12)

plt.xlabel('Predicted', fontsize=12)

plt.title('Confusion Matrix', fontsize=15)

plt.show()

D.3 Support Vector Machines

 # Import necessary libraries

from google.colab import files

import numpy as np

import pandas as pd

from sklearn.model_selection import cross_val_score, cross_val_predict, StratifiedKFold

from sklearn.metrics import accuracy_score, precision_score, recall_score, f1_score, confusion_matrix

from sklearn.svm import SVC

import seaborn as sns

import matplotlib.pyplot as plt

```
# Uploading the CSV file from computer

uploaded = files.upload()

filename = 'Stress.csv'

data = pd.read_csv(filename)

print(data)

# Separate features (questions) and target variable (evaluation)

X = data.drop('Target', axis=1)

y = data['Target']

# Use Stratified K-Fold cross-validation

skf = StratifiedKFold(n_splits=5, shuffle=True, random_state=42)

# Initialize the SVM model

model = SVC()

# Perform cross-validation

y_pred = cross_val_predict(model, X, y, cv=skf)

# Calculate metrics

accuracy = accuracy_score(y, y_pred)

precision = precision_score(y, y_pred, average='weighted')

f1 = f1_score(y, y_pred, average='weighted')

recall = recall_score(y, y_pred, average='weighted')

conf_matrix = confusion_matrix(y, y_pred)

# Calculate specificity for each class
```

```
def calculate_specificity(conf_matrix, class_index):
    true_negatives = np.sum(conf_matrix) - np.sum(conf_matrix[class_index, :]) - np.sum(conf_matrix[:, class_index]) + conf_matrix[class_index, class_index]
    false_positives = np.sum(conf_matrix[:, class_index]) - conf_matrix[class_index, class_index]
    specificity = true_negatives / (true_negatives + false_positives)
    return specificity
specificity_per_class = [calculate_specificity(conf_matrix, i) for i in range(conf_matrix.shape[0])]
overall_specificity = np.mean(specificity_per_class)
# Calculate precision, recall, specificity, and F1-score for each class
precision_per_class = precision_score(y, y_pred, average=None)
recall_per_class = recall_score(y, y_pred, average=None)
f1_per_class = f1_score(y, y_pred, average=None)
# Print the results
print(f"Support Vector Machiness Results:")
print(f"Accuracy: {accuracy:.2f}")
print(f"Precision (Weighted): {precision:.2f}")
print(f"F1-Score (Weighted): {f1:.2f}")
print(f"Recall (Weighted): {recall:.2f}")
print(f"Overall Specificity: {overall_specificity:.2f}")
print("")
```

Print class-specific metrics

```
class_labels = ['Normal', 'Mild', 'Moderate', 'Severe', 'Extremely Severe']

for i, label in enumerate(class_labels):

    print(f"Class: {label}")

    print(f"Precision: {precision_per_class[i]:.2f}")

    print(f"Recall: {recall_per_class[i]:.2f}")

    print(f"Specificity: {specificity_per_class[i]:.2f}")

    print(f"F1-Score: {f1_per_class[i]:.2f}")

    print("")
```

Visualize the confusion matrix in a table form

```
conf_matrix_df = pd.DataFrame(conf_matrix, index=class_labels, columns=class_labels)

plt.figure(figsize=(10, 7))

sns.set(font_scale=1.2)  # Adjust the font scale for better readability

sns.heatmap(conf_matrix_df, annot=True, fmt='d', cmap='Blues', annot_kws={"size": 14})  # Adjusted annotation size

plt.ylabel('Actual', fontsize=12)

plt.xlabel('Predicted', fontsize=12)

plt.title('Confusion Matrix', fontsize=15)

plt.show()
```

D.4 Naive Bayes

Import necessary libraries

```
from google.colab import files
```

```
import numpy as np

import pandas as pd

from sklearn.model_selection import cross_val_score, cross_val_predict, StratifiedKFold

from sklearn.metrics import accuracy_score, precision_score, recall_score, f1_score, confusion_matrix

from sklearn.naive_bayes import GaussianNB

import seaborn as sns

import matplotlib.pyplot as plt

# Uploading the CSV file from computer

uploaded = files.upload()

filename = 'Stress.csv'

data = pd.read_csv(filename)

print(data)

# Separate features (questions) and target variable (evaluation)

X = data.drop('Target', axis=1)

y = data['Target']

# Use Stratified K-Fold cross-validation

skf = StratifiedKFold(n_splits=5, shuffle=True, random_state=42)

# Initialize the Naive Bayes model

model = GaussianNB()
```

```
# Perform cross-validation

y_pred = cross_val_predict(model, X, y, cv=skf)

# Calculate metrics

accuracy = accuracy_score(y, y_pred)

precision = precision_score(y, y_pred, average='weighted')

f1 = f1_score(y, y_pred, average='weighted')

recall = recall_score(y, y_pred, average='weighted')

conf_matrix = confusion_matrix(y, y_pred)

# Calculate specificity for each class

def calculate_specificity(conf_matrix, class_index):

    true_negatives = np.sum(conf_matrix) - np.sum(conf_matrix[class_index, :]) - np.sum(conf_matrix[:, class_index]) + conf_matrix[class_index, class_index]

    false_positives = np.sum(conf_matrix[:, class_index]) - conf_matrix[class_index, class_index]

    specificity = true_negatives / (true_negatives + false_positives)

    return specificity

specificity_per_class = [calculate_specificity(conf_matrix, i) for i in range(conf_matrix.shape[0])]

overall_specificity = np.mean(specificity_per_class)

# Calculate precision, recall, specificity, and F1-score for each class

precision_per_class = precision_score(y, y_pred, average=None)

recall_per_class = recall_score(y, y_pred, average=None)
```

```
f1_per_class = f1_score(y, y_pred, average=None)
```

 # Print the results

```
print(f"Naive Bayes Results:")
print(f"Accuracy: {accuracy:.2f}")
print(f"Precision (Weighted): {precision:.2f}")
print(f"F1-Score (Weighted): {f1:.2f}")
print(f"Recall (Weighted): {recall:.2f}")
print(f"Overall Specificity: {overall_specificity:.2f}")
print("")
```

 # Print class-specific metrics

```
class_labels = ['Normal', 'Mild', 'Moderate', 'Severe', 'Extremely Severe']
for i, label in enumerate(class_labels):
    print(f"Class: {label}")
    print(f"Precision: {precision_per_class[i]:.2f}")
    print(f"Recall: {recall_per_class[i]:.2f}")
    print(f"Specificity: {specificity_per_class[i]:.2f}")
    print(f"F1-Score: {f1_per_class[i]:.2f}")
    print("")
```

 # Visualize the confusion matrix in a table form

```
conf_matrix_df = pd.DataFrame(conf_matrix, index=class_labels, columns=class_labels)
plt.figure(figsize=(10, 7))
```

sns.set(font_scale=1.2) # Adjust the font scale for better readability

sns.heatmap(conf_matrix_df, annot=True, fmt='d', cmap='Blues', annot_kws={"size": 14}) # Adjusted annotation size

plt.ylabel('Actual', fontsize=12)

plt.xlabel('Predicted', fontsize=12)

plt.title('Confusion Matrix', fontsize=15)

plt.show()

D.5 k-Nearest Neighbors
 # Import necessary libraries

from google.colab import files

import numpy as np

import pandas as pd

from sklearn.model_selection import cross_val_score, cross_val_predict, StratifiedKFold

from sklearn.metrics import accuracy_score, precision_score, recall_score, f1_score, confusion_matrix

from sklearn.neighbors import KNeighborsClassifier

import seaborn as sns

import matplotlib.pyplot as plt

Uploading the CSV file from computer

uploaded = files.upload()

filename = 'Stress.csv'

data = pd.read_csv(filename)

```
print(data)

# Separate features (questions) and target variable (evaluation)

X = data.drop('Target', axis=1)

y = data['Target']

# Use Stratified K-Fold cross-validation

skf = StratifiedKFold(n_splits=5, shuffle=True, random_state=42)

# Initialize the KNN model

model = KNeighborsClassifier()

# Perform cross-validation

y_pred = cross_val_predict(model, X, y, cv=skf)

# Calculate metrics

accuracy = accuracy_score(y, y_pred)

precision = precision_score(y, y_pred, average='weighted')

f1 = f1_score(y, y_pred, average='weighted')

recall = recall_score(y, y_pred, average='weighted')

conf_matrix = confusion_matrix(y, y_pred)

# Calculate specificity for each class

def calculate_specificity(conf_matrix, class_index):

    true_negatives = np.sum(conf_matrix) - np.sum(conf_matrix[class_index, :]) - np.sum(conf_matrix[:, class_index]) + conf_matrix[class_index, class_index]

    false_positives = np.sum(conf_matrix[:, class_index]) - conf_matrix[class_index, class_index]
```

```
    specificity = true_negatives / (true_negatives + false_positives)
    return specificity
specificity_per_class = [calculate_specificity(conf_matrix, i) for i in range(conf_matrix.shape[0])]
overall_specificity = np.mean(specificity_per_class)
# Calculate precision, recall, specificity, and F1-score for each class
precision_per_class = precision_score(y, y_pred, average=None)
recall_per_class = recall_score(y, y_pred, average=None)
f1_per_class = f1_score(y, y_pred, average=None)
# Print the results
print(f"k-Nearest Neighbors Results:")
print(f"Accuracy: {accuracy:.2f}")
print(f"Precision (Weighted): {precision:.2f}")
print(f"F1-Score (Weighted): {f1:.2f}")
print(f"Recall (Weighted): {recall:.2f}")
print(f"Overall Specificity: {overall_specificity:.2f}")
print("")
# Print class-specific metrics
class_labels = ['Normal', 'Mild', 'Moderate', 'Severe', 'Extremely Severe']
for i, label in enumerate(class_labels):
    print(f"Class: {label}")
```

```
    print(f"Precision: {precision_per_class[i]:.2f}")
    print(f"Recall: {recall_per_class[i]:.2f}")
    print(f"Specificity: {specificity_per_class[i]:.2f}")
    print(f"F1-Score: {f1_per_class[i]:.2f}")
    print("")
```

 # Visualize the confusion matrix in a table form

```
conf_matrix_df = pd.DataFrame(conf_matrix, index=class_labels, columns=class_labels)
plt.figure(figsize=(10, 7))
sns.set(font_scale=1.2)  # Adjust the font scale for better readability
sns.heatmap(conf_matrix_df, annot=True, fmt='d', cmap='Blues', annot_kws={"size": 14})  # Adjusted annotation size
plt.ylabel('Actual', fontsize=12)
plt.xlabel('Predicted', fontsize=12)
plt.title('Confusion Matrix', fontsize=15)
plt.show()
```