Mental Stress Detection Using Machine Learning

This project is submitted to the Department of Computer Science and Engineering, Dhaka International University, in partial fulfillment to the requirements of Bachelor of Science (B. Sc.) in Computer Science and Engineering (CSE).

Submitted By

Name	Reg. No	Roll No
Abu Syed Akash	CS-D-70-21-117328	13
K. M. Jubayer Abdulla	CS-D-70-21-117355	17
Riyaj	CS-D-70-21-117460	29
Susmay Das	CS-D-70-21-117514	37
Abdullah Al Barakat	CS-D-70-21-117547	44

CSE-425, Project Work

Batch: 70th (1st Shift), Session: 2021-22

Supervised By

Md. Mostafijur Rahman
Lecturer

Department of Computer Science and Engineering
Faculty of Science and Engineering
Dhaka International University
Dhaka, Bangladesh

February-2026

Dhaka International University

Department of Computer Science and Engineering

Declaration

We hereby declare that; this project "**Mental Stress Detection using Machine Learning**" has been carried out by us and it has been submitted for the award of the B.Sc. degree. We also certify that this project was prepared by us for the purpose of fulfillment of the requirements for the Bachelor of Science (B.Sc.) in Computer Science and Engineering

Authors

....
Abu Syed Akash
B.Sc. in CSE, Roll No: 13
Reg. No: 117328, Shift: 1st
Batch: 70th, Session: 2021-22
Dhaka International University

....
K. M. Jubayer Abdulla
B.Sc. in CSE, Roll No: 17
Reg. No: 117355, Shift: 1st
Batch: 70th, Session: 2021-22
Dhaka International University

....
Riyaj
B.Sc. in CSE, Roll No: 29
Reg. No: 117460, Shift: 1st
Batch: 70th, Session: 2021-22
Dhaka International University

....
Susmay Das
B.Sc. in CSE, Roll No: 37
Reg. No: 117514, Shift: 1st
Batch: 70th, Session: 2021-22
Dhaka International University

....
Abdullah Al Barakat
B.Sc. in CSE, Roll No: 44
Reg. No: 117547, Shift: 1st
Batch: 70th, Session: 2021-22
Dhaka International University

Supervisor's Signature

Date: ....

....
Md. Mostafijur Rahman
Lecturer
Dept. of Computer Science and Engineering
Dhaka International University

Dhaka International University

Department of Computer Science and Engineering

Supervisor's Statement

This is to certify that the project paper entitled as **"Mental Stress Detection using Machine Learning"** submitted by Abu Syed Akash, Roll No:- 13; K. M. Jubayer Abdulla, Roll No:- 17; Riyaj, Roll No:- 29; Susmay Das, Roll No:- 37; Abdullah Al Barakat, Roll No:-44; has been carried out under my supervision. This project has been prepared in partial fulfillment of the requirement for the Degree of B.Sc. in Computer Science & Engineering, Department of Computer Science & Engineering, Dhaka International University, Dhaka, Bangladesh.

Supervisor's Signature

Date: ……………………….

……..........…...……………………

Md. Mostafijur Rahman
Lecturer
Dept. of Computer Science & Engineering
Dhaka International University

Approval

The project report as **“Mental Stress Detection using Machine Learning”** submitted by Abu Syed Akash, K. M. Jubayer Abdulla, Riyaj, Susmay Das, Abdullah Al Barakat to the Department of Computer Science and Engineering, Dhaka International University, has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents.

Board of Honorable Examiners

1. Chairman

....
Prof.Md. Abdul Based
Professor and Chairman (Acting),
Dept. of Computer Science and Engineering,
Dhaka International University

2. Internal Member

....
Prof. Mst. Jahanara Akhtar
Professor,
Dept. of Computer Science and Engineering,
Dhaka International University

3. Supervisor and Member

....
Md. Mostafijur Rahman
Lecturer,
Dept. of Computer Science and Engineering,
Dhaka International University

Acknowledgements

We would like to pay our gratitude to the Almighty Allah who created us with not only the ability to design and program this system but also the power of practice.

We would also like to express our sincere thanks to our respected supervisor **Md. Mostafijur Rahman, Lecturer,** Department of CSE, Dhaka International University for his continuous encouragement, motivation and professional guidance during the work of this project.

We would like to thank all the faculty members for their valuable time spent on requirements analysis and evaluation of the project work.

We would like to express our sincere and cordial gratitude to the people those who have supported us directly, provided mental encouragement. We are also thankful to our family and friends who have contributed directly or indirectly the development word and its associated activities.

We warmly thank **Prof. Md. Abdul Based,** Associate Professor and Chairman (Acting), Department of Computer Science and Engineering and **Mst. Jahanara Akhtar**, Professor, Department of Computer Science and Engineering, Dhaka International University for their valuable advice and moral support. Their extensive discussion around work and interesting exploration in operations has been very helpful for this study.

Finally, we would like to dedicate this project to our teachers for their love, encouragement and professional guidance throughout the project.

Dedicated to

Our Parents

&

Teachers

Abstract

The problem of mental stress becomes increasingly popular nowadays, and it impacts individuals of various ages and occupations and is associated with a worse wellbeing and productivity. Although the Perceived Stress Scale (PSS-14) is very popular in researching stress, it fails to directly relate the levels of stress when demographic, academic, lifestyle, and psychosocial factors are combined. This paper fills this gap by training supervised machine learning models to predict stress as Low, Moderate, or High. A web-based survey generated 353 responses and used 17 non-PSS features and PSS-14 items. Reproducible pipeline A cleaning step was made, suitable encoding, scaling where necessary, and a stratified 75/25 train-test split were used. The confusion matrices, class-wise precision/recall/F1, and Macro-F1 were evaluated to indicate the performance of models compared under a similar environment that included: Logistic Regression, a balanced RBF-SVM, Histogram-based Gradient Boosting and hard/soft voting ensembles. The best overall results were obtained on the held-out (88 samples) test set with the soft voting ensemble (accuracy 0.8523, Macro-F1 0.7766), and reached more stable decisions at the boundaries of different classes than with single models. An analysis of feature impact indicated that the sleep habits, workload patterns, and wellbeing indicators are the influential factors. On the whole, the findings demonstrate that a mixture of complementary learners can enhance the stress categorization of structured survey data, as well as justify the application of such models to the early, proactive screening in academic and workplaces.

Table of Contents

List of Figures

List of Tables

Chapter 1 Introduction

1.1 Overview

Stress is the most pervasive mental health problem in our fast-paced world today. Academic pressure and workplace competition along with financial insecurity from student loans and a lack of job security have raised the level of stress for many. While some stress is a natural part of life, chronic or poorly managed stress can compromise both mental and physical health. It is associated with anxiety, depression, heart disease, poor sleep and diminished concentration and decision-making. Due to these risks, early detection of stress has become a critical issue in health-care, work environment and educational institutions. Conventional measures of stress are primarily questionnaires and interviews. The Perceived Stress Scale (PSS-14) that assesses the degree to which individuals consider their lives unpredictable, uncontrollable and overwhelming is one of the most commonly used instruments. It is easy to perform, inexpensive, and appropriate for mass screening. But it quantifies stress at a given point in time and cannot monitor future stress or uncover hidden patterns pertinent to behavior and lifestyle. To solve such limitations, this study introduces a machine learning-based approach with the PSS-14 dataset in combination with 17 demographic lifestyle and behavioral factors, including sleep duration, workload, social interaction, caffeine use, and academic satisfaction. The aim is to predict subjects into Low, Moderate and High stress and tackling class imbalance when fairly assessing the model. This feature-based method is different from the sensor-oriented ones and uses self-reported structured data only, which makes it feasible, interpretable, and relevant for early stress detection and prevention.

1.2 Problem Statement

Everyone knows disruptive stress takes a toll on the health of our minds and bodies, but identifying it early is difficult. Many stress evaluations use questionnaire results that categorize people according to prescribed threshold levels. Even though this approach is easy to do, and provides practical results, it does not go below the surface to find associations between daily lifestyle behaviors and stress. The most important problem identified in the dataset during this study was class imbalance. The majority of the participants were labelled as Moderate stress, with considerably fewer samples for Low and High Stress categories. This skewness leads to model chooses the majority class as winner prediction. A model can therefore predict "Moderate" stress for most of the population and still achieve a high overall accuracy, despite poor performance in identifying those with High stress. Mentally, this is dangerous because people in a High stress rate are more prone to feeling of anxiety and depression and experiencing burnout. The question, however, remains which factors are responsible for the majority of stress. Age, gender, marital status, sleep time , and workload as well as social support influence mental health. Yet, the overall 2 predictive value demands systematic evaluation by data driven methods. Accordingly, in the current study we used machine learning techniques and PSS-14 along with lifestyle features to better determine the stress levels. The aim is, not only to enhance prediction performance under class imbalance, but also discovering and interpreting the primary contributors of factors that lead to stress variation

1.3 Reasons for Choosing Stress Level Prediction

For this study, stress was selected because it is increasingly common globally and has serious effects on both mental and physical health. It has been estimated that about 280 million individuals suffer from depression and approximately 301 million people have an anxiety disorder worldwide (World Health Organization,2023). Chronic stress is considered one of the major contributing factors to these conditions. Work related and academic stress are particularly prevalent in modern society. According to the American Institute of Stress (2022), approximately 83% of U.S. workers suffer from work-related stress, costing the economy nearly $300 billion annually due to absenteeism, turnover, and reduced productivity. Furthermore, the American College Health Association (2022) reported that over 60% of college students experience over whelming anxiety.

Long-term stress does not only affect mental health but is also strongly associated with severe physical health problems, such as heart disease and a weakened immune system (Cohen et al., 2007). Due to the widespread and serious impact of stress, early prediction of stress levels using machine learning and data driven approaches can help provide timely support and enable proactive healthcare interventions.

1.4 Research Objectives

This study will help to know more about stress and predict the stress level since stress is connected to numerous severe health and social issues. These are anxiety and depression, cardiovascular diseases, sleeping problems, low productivity, drug abuse, and even more, suicide. Other negative impacts of stress on the cognitive functions include memory and attention, as well as decision making ability.

Early stress detection is one of the primary objectives this study aims to achieve. Early detection of stress can result in early intervention which can prevent physical and mental health losses in the long term. Contrary to neurological diseases like Alzheimer disease, stress is capable of afflicting individuals regardless of their age and social status. It is dynamic and mostly depends on everyday lifestyle, school pressure, work conditions and individual habits. Since stress varies with time and is determined by a variety of factors, then it is an appropriate predictive model based on structured questionnaire data. In this study, data is gathered through survey-based data, PSS14 scale, to construct machine learning models to predict stress. Questionnaire-based stress prediction is less expensive and simpler to execute as compared to sensor-based monitoring systems.

Consequently, the suggested framework will provide a viable, economical, and scalable solution to the early stress detection in both academic and workplace settings.

1.5 Purpose Of Study

This research is primarily aimed at designing and testing machine learning algorithms capable of stress level classification based on structured psychosocial and behavioral data taken out of the PSS-14 questionnaire.

This research paper will find out the main determinants of stress. It compares the various classification algorithms in order to determine the best model. As the dataset includes imbalance in classes, proper methods are used to address the problem and provide fair outcomes. The models are tested on balanced and reliable performance indicators. Moreover, the analysis of the feature importance helps to make the results easy to understand, as well as to display the factors that contribute to the stress prediction the most.

The other significant objective of the research is to demonstrate that stress could be forecasted reasonably with no need of the costly physiological equipment like wearable heart rate monitors or brain sensors. Rather, bare data of questionnaires can be meaningful.

As an illustration, this system can be applied in a university where it is possible to recognize students who are under a lot of stress when examinations are about to begin and provide counseling services early. On the same note, organizations would use this model to check on the welfare of employees and develop stress management initiatives. This enables the suggested method to be practical, cheap and applicable in a real world academic and workplace environment.

1.6 Aims of the Stress Level Prediction Model

The objective of the present research is to create a valid and efficient machine learning model to forecast and categorize the stress level through the information obtained via the PSS-14 questionnaire. The study aims at the utilization of the structured psychosocial and behavioral data to classify stress into Low, Moderate, and High levels.

To this end, the dataset is thoroughly preprocessed, i.e. cleaned, the data is categorical and therefore this is encoded and the features are organized in a way that they can be analyzed using machine learning. Class imbalance is also to be considered by the study to make sure that every category of stress has its fair share in model training because it enhances the overall credibility of the results.

The other significant objective is to use and compare three classification models: Hist Gradient, Balanced SVM and Logistic Regression. To measure their performance, balanced and suitable measures are used like Accuracy, Macro F1-score, Balanced Accuracy and Per-class Recall to guarantee equal performance measurement across the different stress levels.

Moreover, the research will strive to determine the most influential factors that influence the level of stress using the feature of permutation importance. Lastly, the study aims at identifying the most stable and generalizable model that would perform consistently. In general, this research will seek to develop a highly pragmatic, interpretable and effective framework of stress levels prediction based on the data gathered via questionnaires.

1.7 Application of Machine Learning for Stress Classification

In this paper, the author utilized machine learning in an effort to classify stress. Machine learning offers a viable means of finding trends in survey data that cannot be explained by mere assumptions. An example of a stress prediction is the present study in which the Low, Moderate, and High stress groups are separated by the models in relation to the chosen set of demographic, lifestyle, and psychosocial variables. A number of machine learning classifiers were tested in this analysis. The simple baseline was a Logistic Regression, whereas a balanced Support Vector machine (SVM) was implemented due to its ability to acquire the non-linear decision boundaries in the classes by simply maximizing the margin where the decision borders the classes. Because SVM is sensitive to the scale of features, the input variables are normalized in a way that does not give more influence to any particular feature during the training process. Besides individual models, an ensemble approach was also experimented. Gradient Boosting using histograms was added as it is capable of non-linear predictors of tabular data, and a soft voting based classifier was applied to combine the merits of various learners. It fits well in the psychosocial questionnaire data, in which demographic and behavior factors tend to interact instead of acting apart. Following class imbalance treatment on the training set, the voting ensemble gave the most consistent overall results in evaluation criteria, that is, accuracy, and macro-level scores that monitor the performance in all stress categories. The results indicate that an intelligently constructed machine learning pipeline can make a successful prediction of levels of stress based on data gathered by structured questionnaires, even when the distribution of classes is not equal. In general, machine learning was chosen compared to standard statistical methods as it is able to capture non-linear interactions between many predictors simultaneously that provide stress-level categorization with higher reliability. In mental health prediction, transfer learning and machine learning are effectively applied to identify and predict possible mental health risk factors.

1.8 Significance of the Study

The significance of this study lies in the fact that it offers a feasible and cost-effective method of estimating the levels of stress based on the questionnaire-based information, and machine learning tools. This study has implications on the lives of students, organizations, and machine learning.

This research can also be used in early detection of stress by students particularly when there are exams or when they are under academic pressure. Remington of the educational institutions can anticipate the level of stress thus offer prompt help to the students in the form of counseling, mental health workshops and stress management courses. Early diagnosis may enhance academic performance, emotional status, and life quality.

In the case of organizations, the proposed model can be utilized to keep track of the level of stress among employees at work. When the stress is high, it may decrease productivity, raise the absenteeism and impact job satisfaction. Early detection of stress can provide preventive strategies like workload modification, employee wellness, and supportive policies by organizations. This has the potential to make the workplace a healthier and productive place.

This study has proven, in terms of contribution to the research in machine learning, that structured psychosocial data may be effectively applied to predict stress without necessarily involving physiological sensors that are very expensive to acquire. It also makes comparisons and contrasts various classification models and points out the effectiveness of ensemble techniques such as Random Forest. The study is valuable because it is based on an interpretable and scalable framework of predicting mental health with the use of machine learning.

Chapter 2 Literature Review

2.1 Overview

It has been proposed that machine learning will be able to predict stress with accuracy when the model is supplied with variables that can capture both daily activities and pressures experienced in life. In mental-health AI research, more robust findings tend to be presented when the feature-set is meaningful and the modelling-pipeline is rigorous e.g. in split of data, treating class-imbalance and reporting metrics, and explainability of final model. Physiological-based work (i.e. ECG and skin-conductance) demonstrates that even traditional classifiers remain competitive; ensemble approaches like the Random Forest can frequently be effective in detecting stress-related responses in either a two-class or multi-class task. Simultaneously, results of survey-based studies demonstrate that self-report variables could be very informative as well, although careful feature selection is necessary, along with the evaluation of model performance in a manner that could be considered as the generalization of results in real-life scenarios. The repeated statements in the same body of literature indicate that there are various predictors that are strongly related to stress and psychological well-being. All academic and working requirements like attendance of exams or project, excessive study/workload, and lack of faith in ability to cope with tasks would tend to raise the level of stress by adding to mental load and restricting recovery. The primary role also belongs to lifestyle habits: a lack of sleep, lack of exercise, low exposure to relaxation or mindfulness are often linked to worse mental health, whereas healthier habits can make one less vulnerable. Lifestyle habits (caffeine consumption) and smoking status seem to be also relevant, but their influence may depend on the level of influence and the predisposition of a person. Social and family conditions also determine the exposure to stress: stressful events such as separation, divorce, and widowhood are commonly associated with an increased amount of stress, and supportive relationships and stronger connections in society are likely to reduce the influence of stress. The state of economic disadvantage is frequently referred to as a situational factor that adds chronic stress and lowers the resources to cope with. Further, chronic diseases, including hypertension, diabetes, and asthma, tend to be related to emotional distress and potentially interact with daily operations, so they need to be included in the predictive models as covariates. On the whole, the arguments in support of the subjectivity of stress as an assessment of overloading, uncertainty, and lack of control have greater support than those in support of stress as a mere number of events in the environment. The use of such tools as PSS-14 is commonplace because of this reason: the measures of perceived stress are given a consistent method to determine it, and such measures can be used as a useful outcome variable in supervised learning models utilizing demographic, lifestyle, academic/work, social and health-related characteristics.

2.2 Mental Stress And Its Consequences

Mental stress has emerged as one of the greatest public health and productivity issues due to its impact on emotional health, mental functioning, and functioning in everyday activities in the lives of people of various ages and in various types of occupation. With prolonged stressors, the typical effects are impaired focus, less efficiency, symptoms of burnout, as well as wider decreases in the quality of life, in both academic and workplace settings. These effects drive the necessity of effective screening methods that will be able to detect stress at an early stage and facilitate preventive measures.

2.3 Perceived Stress Scale (PSS-14) As a Standard Measure

One of the most commonly used methods of measuring stress is perceived stress, which focuses on the unpredictable and uncontrollable in addition to overloaded nature of the lives of people. The most firmly established tool on this is the Perceived Stress Scale (PSS) by **Cohen**, **Kamarck**, and **Mermelstein** (1983) which is a world-wide scale of perceived stress which can be used in general populations [1] . A version PSS-14 is still popular and is officially published and documented as an internationally used instrument, and shorter versions are also offered (PSS-10 and PSS-4). Nevertheless, syntheses of the evidence report that psychometric performance is also variant with version and context, and evidence shows that PSS is a perceive stress measure and not an explanation of why stress is different in individuals [2]. Due to this, recent validation studies and adaptations tend to study factor structure, reliability, and cultural fit- justification of the opinion that an interpretation of PSS scores is better suited within a context instead of being used as a wholesome standalone explanation [3] .PSS-14 provides a good foundation in labeling, although, in situations where the objective is to acquire more information about stress category and the influential factors not inherent to the scale as such, a combination of PSS items with demographic, academic, lifestyle, and psychosocial features is required.

2.4 Machine Learning for Stress Prediction

Machine learning has been used as a typical approach to stress prediction due to its ability to learn patterns on multi-dimensional inputs (demographics, lifestyle habits, academic behaviors, psychosocial variables) and project them onto the stress outcomes (continuous or categorical). Recent literature has also explicitly utilized online survey data with supervised algorithms to forecast perceived stress results in groups of students [4]. An increasing number of studies are dedicated to multi-class categorization (e.g., low/moderate/high), due to the fact that the categories fit more into the screening and intervention processes as compared to numeric scores. As an illustrative example, the recent publication by Springer related to the classification of student stress emphasizes applying hybrid or integrated ML methods that involve the combination of several factors related to stress (academic, social, environmental) [5].

2.5 Machine Learning Techniques For Mental Health And Stress Prediction

Machine learning techniques such as logistic regression, SVM, gradient boosting, and ensemble voting can predict mental health and stress levels from survey and behavioral data, enabling earlier and more accurate screening.

2.5.1 Systematic Reviews on Machine Learning for Stress Detection:

Mariachiara Di Cosmo et al. [6] summative surveys of stressful detection literature verify the current importance and expansion of machine learning and deep learning in evaluating mental state. Literature review of EEG, ECG, and wearable sensor data displays that ML methods are efficient in detecting stress patterns, although other issues like variable databases and generalization adversely impact. The critical role of strong validation strategies and interpretive models, these reviews highlight the validity of the methodological decisions in this thesis.

2.5.2 Machine Learning in Broader Mental Health Prediction:

Yafei Guo et al. [7] recent reviews on artificial intelligence applications in mental health demonstrate that machine learning can support early stress detection and screening of psychological conditions using diverse feature sets. These studies underline the importance of feature selection, balanced evaluation metrics, and careful model validation to ensure reliable outcomes. The findings provide methodological support for the supervised classification framework adopted in this research.

2.5.3 Stress Classification Using Wearable Sensor Data:

Emma Todd et al. [8] it was shown through research that used ECG, skin temperature, and features of skin conductance that traditional machine learning algorithms like the Random Forest and KNN algorithm are able to classify with high classification accuracy on both binary and multi class. These results underscore the ability of ensemble approaches in resolving complex patterns of stress related to stress to support the appropriateness of random forest in structured stress prediction models.

2.5.4 Machine Learning for Stress Detection Using Survey Data:

Ravinder Ahuja et al. [9] the most recent studies concerning the use of artificial intelligence in mental health highlight that machine learning has the potential to assist in the early diagnosis and screening of mental disorders based on the wide range of features. These papers highlight the significance of the feature selection process, the measures of evaluation that should be balanced, and adequate model validation to guarantee the sound results. The results support the methodological approach of the supervised classification model assumed in this study. This chapter is a review of the literature that has been written on 1) Lifestyle and demographic predictors of stress and mental health 2) Machine learning methods of mental health prediction 3) Feature importance and model interpretability 4) Gaps in current research that support the present study.

2.6 Academic And Occupational Stressors

Academic and occupational stressors include workload pressure, deadlines, performance expectations, time constraints, and role conflicts that can negatively affect wellbeing and productivity.

2.6.1 Future assessments/studies/exam dates

One of the adverse effects of deadline pressure and performance anxiety is well-being since these factors raise cognitive load and evaluation anxiety. The authors (2025) in a study on the relationship between performance anxiety and employee well-being emphasize deadline failures as another significant route between anxiety and low well-being. This framing can also be applied in the student context: exams and deadlines of projects can also create anticipatory stress that can then cause sleep disturbances, less attention to self-care, and elevated perceptions of stress. To model it, the upcoming exams/projects/deadlines may be considered a situational exposure variable and may mediate or strengthen other correlations [10].

2.6.2 Academic/work environment and self-efficacy satisfaction

The perception of being satisfied with the environment is what determines whether demands are seen to be manageable and the presence of support and fairness. In an experiment of the staff of two organizations, the authors state that the mental health (measured using the GHQ-12) is highly correlated with job satisfaction and satisfaction with the cultural environment and workplace behavior; income itself did not play a significant role in their study. In the context of the university populations, Wilcox and Nordstokke study first-year students and demonstrate that life satisfaction, academic self-efficacy, and mental health measures are interconnected, which means that perceived ability in academic performance and perceived adjustment are applicable to the psychological welfare. These papers justify the need to incorporate both the environmental satisfaction and self-efficacy style variables since they may mediate the effects of academic stressors to perceived stress and mental health effects [11].

2.6.3 Perceived work load/ study load

Workload is a direct stressor as it increases demands on time, attention, and performance and usually reduces recovery time. A qualitative review of workload and mental health underlines the fact that high workload in the occupational sectors is associated with an increasing number of stress, anxiety, burnout, and job dissatisfaction. Patterns in academic workload are the same. The authors Azizova et al. investigate the psychological health of students through the established screening theories (such as the Beck Depression Inventory and the SAN wellbeing/activity/mood scale) and have reported that the students are moderate in their anxious state, which is characterized by frequently experienced states of depression, and their poor psychological health, varying as years go by. Combined, these results support our perceived workload/study load variable as a key predictor of stress/mental health.[12], [13]

2.7 Everyday And Lifestyle Practices

Workload, deadlines, pressure to perform, and job demands are common sources of stress that strongly affect how stressed people feel and how well they function day to day.

2.7.1 Average sleep per day

One of the behavioral predictors of mental health that is consistently supported is sleep. In a massive systematic review and meta-analysis involving more than one million participants, Zhang and others discover that short sleep duration is linked to the higher risk of developing mental disorders in both cohort and cross-sectional studies. Subgroup analyses show that they have significantly strong connections with anxiety and depression, and the duration of long sleep is not as consistent as a risk factor of these outcomes. These results can be used to support the idea that poor sleep can be considered an independent predictor (not a symptom) of poor mental health. Mechanism wise, short sleep is able to disrupt emotion regulation, physiological stress reactions, and cognitive ability, which may render normal daily tasks more threatening [14].

2.7.2 Mindfulness, relaxation and entertainment

Relaxation orientated behavior have the capability to function as protective variables by reducing physiological arousal and disrupting rumination. Zollars and colleagues conducted an intervention study with pharmacy students, which assessed mindfulness meditation with the help of the app in the course of four weeks and found an increase in mindfulness and mental well-being and a reduction in perceived stress. Notably, the authors state that the effect continued to be observed even in the cases of the influence of other health-promoting lifestyle behaviors, indicating an independent role of mindfulness practice. This justifies the operationalization of our variable such as frequency of relaxation/meditation/entertaining days per week, whereby more engagement is anticipated to be associated with less perceived stress [15]

2.7.3 Physical activity per week

Physical exercise is a behavior that can be modified, and has a psychological and biological route to improved mental health. General literature on exercise and mental health provides significant evidence of the fact that regular exercise decreases the symptoms of anxiety, depression, and stress. The review combines several processes, such as the alteration of neurotransmission and stress-related physiology, the anti-inflammatory action, and psychological processes, such as the distractive shift in negative thinking and the enhancement of self-efficacy. Such converging processes have the implication that exercise is not merely implicated in improved mood; it can be possible to hypothesize that exercise may lower stress reactivity and bolster resilience. The direction, which is expected in predictive modeling, is that increased physical activity per week lowers perceived stress and health symptom levels of mental health [16].

2.7.4 Caffeine/ tea/ energy drinks/ day

Caffeine possesses dose-related psychological effects thus it is a subtle predictor and not a straightforward risk factor. The review by Lara is a summary of human evidence which can indicate the association between moderate caffeine consumption and improved mood and/or reduced symptoms of depression in certain groups. Nevertheless, the review continues by stating that in few cases high doses may result in anxiety and in other rare cases, psychotic or manic symptoms; the sensitivity to this drug may be more in individuals with panic disorder or social anxiety. Due to this non-linear trend, it is of interest to measure caffeine as an approximate dose (e.g. cups/servings per day) and our analysis can be enhanced with threshold/high-intake category testing [17] .

2.7.5 Smoking habit

There is a strong association between smoking and mental health and there is some evidence which suggests a causal path of smoking towards deteriorated mental health. Plurphanswat et al. discuss the issue of reverse causality by using the US Behavioral Risk Factor Surveillance System data (2000-2010), and employing an instrumental variables design as an instrument to assess smoking behavior, i.e. state cigarette excise taxes. They discover that smoking adds to the days with low mental health, and big impacts on those who report greater mental health issues. In our model, the smoking status could thus be regarded as one of the behavioral risk factors, and it could also co-occur with other risk factors (sleep disruption, high caffeine use), which is a nice consideration to make in the interpretation of features [18] .

2.8 Family and relationship context

Widowhood represents a distinct form of relationship loss that may involve grief, loneliness, and changes in economic security. A cross-sectional study focused on widowed adults (as shared in your provided source) reports substantial levels of depressive symptoms among both widows and widowers, with indications that widows may experience higher levels of depression in some settings. While cross-sectional findings cannot determine whether depression preceded bereavement, they highlight widowhood as a practical marker of elevated psychological risk that may operate through reduced social support and increased role strain. For our study, coding widowhood separately (rather than grouping all 'not married' categories together) can capture these unique stress processes. [19] Relationship disruption is a strong social stressor because it can simultaneously affect emotional support, finances, daily routines, and identity. In a national UK birth cohort analysis, Cherlin, Kiernan, and Chase-Lansdale report that divorce and separation are associated with long-term adverse mental health outcomes, even after taking earlier vulnerability into account. This suggests that marital disruption can act as a lasting stress pathway rather than a short-term emotional shock only. Mechanistically, divorce can increase stress by reducing practical support, escalating co-parenting conflict, and creating chronic uncertainty around housing and income. For a predictive model, marital status and recent marital transition (e.g., divorced/separated) can therefore be treated as social-structural predictors that influence both stress exposure and coping resources. [20] Evidence suggests that the mental-health advantage of marriage is not only about being exposed to fewer stressors; it also relates to how stress is psychologically processed. Kessler and Essex propose that married people may experience a smaller emotional impact from stressful role strains because they can draw on both social resources (e.g., encouragement, tangible help) and intrapsychic resources (e.g., stronger perceived control or meaning). This implies that marital status may reduce stress reactivity even when stress exposure is similar. In modeling terms, marital status can be conceptualized as a protective factor that partially buffers the link between stressors (workload, illness) and outcomes (depression, anxiety, high perceived stress) [21] .

2.9 Socioeconomic And Social Connectedness

Money-related pressures and how connected someone feels to friends, family, and their community can strongly affect stress levels and overall wellbeing**.**

2.9.1 Number of close friends/social contact

Social connection acts as a buffer against stress by providing emotional validation, practical help, and a sense of belonging. An evidence brief on social connection reports that individuals with fewer than five close relationships experienced greater increases in loneliness during the early COVID-19 period. While the context was unusual, the finding aligns with broader theory: limited close ties can increase isolation and intensify the psychological impact of daily stressors. For our model, the number of close friends (or frequency of meaningful social contact) can be treated as a protect.

2.9.2 Monthly family income and poverty-related disadvantage

Socioeconomic conditions influence mental health by shaping exposure to chronic stress and limiting access to protective resources. A review in the Bulletin of the World Health Organization synthesizes evidence linking poverty indicators to common mental disorders and reports that low education shows one of the most consistent associations. This literature suggests that income and education should be treated as background risk variables that may confound or moderate the effects of lifestyle behaviors and social support. In student or early-career samples, family income can also shape stress indirectly through tuition pressure, housing insecurity, and reduced opportunities for rest or recreation [22] .

2.10 Demographic Determinants

Money-related pressures and how connected someone feels to friends, family, and their community can strongly affect stress levels and overall wellbeing.

2.10.1 Age

Age is repeatedly linked to differences in stress exposure and coping capacity. In a study comparing stress across age groups, Gupta reports that middle-aged adults tend to show lower perceived stress than both younger and older participants. A practical interpretation is that midlife often brings more stable routines and stronger coping repertoires (e.g., better emotion regulation, clearer priorities, and greater problem-solving experience), which can reduce day-to-day stress appraisal. In contrast, younger adults may face transition pressures (academic demands, early-career uncertainty), while older adults may face health-related strain and reduced social roles, both of which can increase perceived stress. For our model, age therefore functions as a baseline stratifier of risk and a potential moderator--meaning the same stressor (such as sleep loss or workload) may have different psychological impacts at different life stages.

2.10.2 Gender

Gender differences in mental health outcomes are consistently documented, but they are best understood as differences in how distress is expressed rather than a simple hierarchy of vulnerability. Rosenfield and colleagues argue that women more often show internalizing patterns (such as anxiety and depression), whereas men more often show externalizing patterns (such as substance-related problems and antisocial behaviors). Importantly, they propose that these patterns can represent comparable underlying suffering, even though the symptoms look different. [23]

2.10.3 Self-perceived mental health and measurement

Happiness ratings are often used as a brief proxy for quality of life, but their interpretation can be more complex when mental disorders are present. Bergsma, Veenhoven, ten Have, and de Graaf examine whether people with mental disorders can meaningfully self-rate happiness and whether these ratings remain valid. Their analysis suggests that self-rated happiness is still informative but may be more strongly shaped by current mood in those with mental disorders; for example, unhappiness can be linked with disproportionately lower life satisfaction among people with a disorder compared with unhappy people without a disorder. This supports using self-rated happiness as a concise indicator of well-being while interpreting it as a mood-sensitive measure rather than a purely objective report.[24]

2.10.4 Chronic physical disease and mental disease comorbidity

Hypertension may also be the source of psychological stress directly (with health worry) and indirectly (with medication regimens and involvement in health care). Kretchy and colleagues in a cross-sectional study of hypertensive patients in a Ghanaian hospital evaluate the symptoms of anxiety, depression, and stress and examine how these relate to the adherence to anti-hypertensive medications. The study highlights that negative emotional conditions are inherent to chronic hypertension and they may have an impact on behavior of treatment processes. This justifies the inclusion of hypertension as a predictor, as well as where feasible the fact of whether emotional distress is associated with health behaviors (e.g., sleep, smoking) in our dataset. [25] Chronic diseases may increase stress due to symptom burden, change to lifestyle and uncertainty of further ill health. According to a Science Direct chapter on diabetes and mental health, the relationship is reciprocal, as mental health disorders (in particular, depressive syndromes) and diabetes are often frequently observed together, and are susceptible to either other via behavioral mechanisms (sleep, diet, activity), biological mechanisms (alterations), and treatment mechanisms. Practically, diabetes can raise the felt stress because it can make day-to-day self-management more stressful and more concerned about complications. This contributes to including diabetes status as a feature of a chronic-condition in our predictor set. [26] Cross-national evidence on asthma is also associated with the mental health burden. A survey that employs the information of the World Mental Health Survey reveals that there is a set of prevalent mental disorders that exist with higher frequency among the adults with asthma. The authors also stress clinical relevance: health providers ought to be concerned with co-occurring issues since mental health issues can make it difficult to manage symptoms and quality of life. In our model, asthma can thus be added as a chronic-condition indicator that has the potential of increasing stress and distress symptoms [27] .

2.11 Gaps In Existing Literature

The present study has addressed these gaps relative to other studies by: Predicting stress with a complex interaction between demographics, lifestyle behaviors, academic demands, health and psychosocial issues (PSS14as the outcome measure) Also the study combined multiple demographic, lifestyle, and psychosocial variables within a unified machine learning model (Stratified train test split). Ensemble tree based algorithms in machine learning especially have been demonstrated to provide attractive performance in prediction of

- Lack of interpretability
- Overfitting due to class imbalance
- Data leakage during preprocessing
- Limited evaluation metrics

The research being conducted here fills these gaps in that it: Exploiting a stratified train-test split (75% training, 25% testing) to maintain the conditions of Low, Moderate and High stress classes in the conditioning. Using SMOTE oversampling on training data, to ensure that minority-class learning is improved as training data, without causing the test performance to be unrealistic or owing to artificial samples. Reporting several evaluation metrics, as opposed to focusing purely on accuracy, such as confusion matrices, class-wise precision/recall/F1, and Macro-F1 to indicate performance on a set of stress categories at imbalance. Running permutation-based feature importance analysis to discover what variables in lifestyle, academic and wellbeing most contribute to predicting stress, even when they do not display built-in feature importance in their models. Comparing a number of classifiers within the same pipeline, such as the baseline run of Logistic Regression, the balancing SVM, Histogram Gradient Boosting and giving a vote to an ensemble as the final model based on balanced output (Accuracy and Macro-F1) on the unknown test set. Also, the study adds to stress analytics as it addresses a non-clinical group and applies demographic factors, daily, academic, lifestyle, and psychosocial factors to categorize stresses. Such an orientation promotes a pragmatic screening approach, in which the objective does not include diagnosis of disease but only identifying those who might need attention at an earlier stage. By doing that, the work promotes the preventative support prior to stress becoming more considerable or initiating its impact on academic performance, wellbeing, or physical health.

2.12 Conclusion of Literature Review

Despite numerous past researchers having conducted their work on stress prediction and mental health assessment, the issues of interpretation, generalization and fairness of the models persist. There are those studies that are only interested in the scores of stress and those that use a lot of clinical or sensor-based data. Such methods can restrain the usefulness of stress prediction in real life academic or working environment. The proposed research will build on the prior findings by integrating PSS-14 scores with demographic and behavioral structured variables. The research tries to combine these various categories of data to eliminate methodological weaknesses that have been discovered in the previous research and offer a more impartial and realistic framework of prediction.

Chapter 3 Dataset Description

3.1 Overview

The current study is a cross-sectional survey with the use of PSS-14 and demographic, lifestyle, and psychosocial factors. The primary purpose of the dataset is to determine the factors that are related to low, medium, and high stress levels as well as to create machine learning models that will assist in predicting stress levels. Categories of stress were identified relying on the overall PSS- 14 score and categorized into three groups: Low, Moderate and High. The survey also gathered the data regarding sleep, workload, income, physical activity, marital status, social contacts, substance use, and general life satisfaction along with PSS-14 items. The data set was tailored in such a way that it took a systematic, quantifiable form of emotional and behavioral state associated with stress.

3.2 Dataset Details

Categorization of Input Features Used for Stress Level Prediction (17 Variables)

Category	**Variable**
Demographic Variables	Age
	Gender
	Marital Status
	Monthly Family Income
Academic/Occupational Factors	Perceived workload/study load
	Upcoming exams/projects/deadlines
	Satisfaction with academic/work environment
	Average study/work hours per day
Lifestyle Variables	Average sleep per day
	Physical activity per week
	Caffeine/tea/energy drinks consumption per day
	Smoking habit
	Average daily screen time
Health and Social Factors	Any chronic health condition (diabetes, hypertension, asthma, etc.)
	Number of close friends/social contacts
	Frequency of engagement in entertainment, meditation, or relaxation
	Self-rated happiness level

Table 3.1: Variables Used in the Study

The table summarizes the variables used to predict stress level. The features are grouped into five categories: demographic, academic/occupational, lifestyle, and health/social factors, along with the target variable. The target variable represents stress level with three classes (Low, Moderate, and High). Overall, these variables provide a comprehensive representation of individual background, workload, daily habits, and well-being for stress prediction.

3.3 Dataset Acquisition Process

The nonlinear structured questionnaire was generated in Google Forms and was used to collect data. Voluntarily answered the questionnaire. There were two big sections of the questionnaire:

- The PSS14 questionnaire to measure perceived stress
- Additional structured questions capturing demographic, academic, health, and lifestyle information.

Participation was anonymous to ensure dishonest responses and reduce social desirability bias. The responses were automatically filed and stored in spreadsheet format and later exported as a CSV file for analysis. Ethical considerations were maintained throughout the data collection process. No personally identifiable information was collected.

3.4 Dataset Processing Method

When the data regarding the survey were gathered, there was a series of preprocessing operations, which transformed the data into uniform and machine learning ready data. Initially, the data were checked on missing values, invalid entry and formatting divergence. Then simple cleaning was used such that the dataset was put in order and could be used to further analyze it. The subsequent step consisted of adding the responses of the PSS-14 questionnaire in order to get one cumulative perceived stress score (PSS-Total) of each participant. According to the overall score, the participants were divided into three categories of stress: Low, Moderate and High. Such translation made this conversion a multi-class classification problem. The individual columns of PSS questions were deleted after the construction of the stress labelling. This was to prevent the leak of information, since the stress label was made on the basis of the same items and holding onto them would result in the models learning the target per se instead of learning it with the help of lifestyle and psychosocial predictor variables. As there are both numerical and categorical variables in the dataset, the suitable encoding of the categorical features was carried out to enable them to be processed by machine learning models appropriately. Standard Scaler was also used in scaling the numerical features. This procedure is significant to algorithms that are scale-sensitive (such as SVM and Logistic Regression), and the procedure prevents numeric variables to stay on a similar scale overall modeling pipeline. Moreover, the stress categories were not distributed equally and therefore, it needed to handle the class imbalance. Splitting was thus conducted on the training part of the dataset wherein SMOTE was used. Lastly, a stratified split was done to partition the dataset into training and testing sets to ensure that the proportion of classes were close in both subsets. Generally, these measures made the data clean, consistent and ready to use in developing sound classification models. The analysis of data will involve Exploratory Data Analysis (EDA) (3.4.1). To prepare the dataset in advance before subjecting it to machine learning algorithms, Exploratory Data Analysis was conducted. This step was aimed to monitor the distribution of stress levels, study the distribution of every feature, detect abnormal patterns, as well as extreme values that may influence the learning of the model. The dataset was made easier to interpret using a number of simple visualization and summary techniques. Bar charts were used to check the stress-level distribution and it was apparent that the number of samples in the Moderate stress group exceeded those of the Low and High one. The significance of this imbalance is in the fact that it may affect the model behavior, it may result in the wrongful performance too when the evaluation is based on the accuracy only. As a result, the subsequent model analysis focused on the confusion matrices and the class-sensitive metrics like Macro-F1 so that all of the stress categories were taken into account equally.

3.4.1 Exploratory Data Analysis (EDA)

Exploratory Data Analysis (EDA) was conducted to have a better insight into the dataset prior to the development of the machine learning models. The primary objective was to investigate the stress levels distribution, count the distribution of each feature, and understand the connection between the various variables and the types of stress. The step also assisted in determining whether there were any outliers or anomalous values that may influence the performance of the models. The analysis was made more understandable and readable by applying various visualization techniques. The distribution of stress levels was observed with the help of bar charts. Confusion matrices were used to interpret the behaviour of model prediction and permutation importance plots were applied to determine the most relevant features. In the first analysis, it was noted that the sample size of the Moderate stress category was by far higher than that of the Low and High categories. This is an imbalance in the classes, which might have an adverse effect on the model performance. Thus, this imbalance needed to be managed with great care at the stage of training in order to have predetermined predictions that are fair and reliable and at all stress levels.

3.4.2 Data Cleaning

The Pandas library was used to load the dataset and clean it in an empirical manner that will facilitate further parts of preprocessing and modeling. To avoid the hidden formatting problems that may lead to errors introduced in the course of processing, first unnecessary spaces in the column names were eliminated. The PSS-14 total score (PSS-Total) was then calculated as the sum of the 14 items in the questionnaire that was administered to participants. The individual PSS columns fell after the total score had been used to generate the category of stress (Low/Moderate/High). This move minimized redundancy and served to alleviate leakage, as the label target is made of the same questionnaires. Normalization (Feature Scaling) is considered a vital component of statistical tests. Normalization (Feature Scaling) is perceived to be a major part of statistical tests. Standard-Scaler of the scikit-learn was used to scale the features. The method is used to change numerical variables in such a way that the mean value of the numerical variables is 0 and a standard deviation is 1. Particularly in models which rely on distances or linear margins (as with SVM) and linear baselines (as with Logistic Regression), scaling is also necessary because variations in feature ranges might otherwise dominate the learning process. Lacking scaling, variables that have bigger numeric values might increase their influence on the learning process, which would decrease the model stability. A scaler was only fitted on the training data to achieve fair assessment and to avoid leakage. The scaler was then applied to both the training and the testing sets using the same fitted scaler. This would guarantee that the test data does not play a role into preprocessing decision and makes the evaluation realistic.

3.4.3 Normalization

The standard Scaler technique of the Scikit-learn library was used in the normalization. Standard Scaler normalizes the features by rescaling the data such that the mean of the data is 0 and the standard deviation is 1. This can be used to equalize all features. Normalization is especially critical to models such as Support Vector Machine (SVM) since SVM is afflicted with the fact that it is sensitive to the magnitude of input features. Failure to scale the features appropriately can have the risk of the large values to take over the learning process and influence the model performance. In order to prevent leakage of data, the scaler was only fitted on the training data. Once fitted, both the training and the testing set were transformed in the same way. This makes sure that the model does not pick any information in the test data in the course of training. Generally, normalization enhances the stability and the performance of the model and that all the features play a fair role in the learning process.

3.4.4 Feature Selection

The selection of the features was done in a simple manner depending on the objective of the research. The items of the PSS-14 were applied to obtain the final label of stress by the total score. Once the target variable was developed, the individual PSS item columns were deleted out of the dataset. In the last input (X) the target variable (Stress-Level) was eliminated and the rest 17 demographic, lifestyle, academic and psychosocial variables as predictors were used. The output variable (Y) was referred to as Stress-Level, which portrays 3 category of stress. This made sure that the models were taught the patterns of the stress based on the external factors that influenced it and not the items in the PSS questionnaire that were used to name the data directly. The effect of processor to the dataset will be evaluated. All the preprocessing stages fulfilled a particular purpose of making the dataset better in developing a model. To start with, the scoring of stresses process transformed the continuous PSS total into three distinct categories of stresses which made the task to be treated as a multi-class classification problem. Second, stratified splitting was used to ensure that the ratio of the Low

Stress cases, Moderate stress cases and High stress cases adjusted equally in the training and testing sample. This enhanced the objective of evaluation since the model was conditioned on a distribution that was comparable to the one it was conditioned on. Third, since the data was skewed, only the training set was subjected to SMOTE so as to synthetically increase the representations of the minority-classes. This enhanced the possibility of the model to acquire the trends of Low and High stress groups without altering the initial test distribution. Lastly, standardization served to assure that numeric predictors were more equally useful in the training process, which can be beneficial when a particular model is weighted by scale, as well as to generally increase the pipeline stability. In a nutshell, preprocessing facilitated better quality of data, minimized imbalance bias, and denied variable model performance.

3.4.5 Processing Effect On Dataset

A few preprocessing measures were taken to prepare the data in order to be used to construct the machine learning models, and each preprocessing measure influenced the data significantly. The total stress score (PSS-Total) was firstly transformed into three levels, i.e., Low Stress (0), Moderate Stress (1), and High Stress (2). This redesign made the problem a multi-class classification problem, with the model making predictions of one of the three levels of stresses. The dataset was then split into two parts: training and testing sets in a ratio of 75: 25. During the split, stratified sampling was employed to make sure that every stress category was, on the training and testing datasets, proportionately taken. This assists in judging the model in a much better way. There was an imbalance in the classes in the dataset, and thus SMOTE (Synthetic Minority Oversampling Technique) was used on the training data. Some of the stress categories had less numbers than other ones before SMOTE was applied. Upon the use of SMOTE, the classes became balanced through the creation of artificial samples on the minority groups. This aids in enhancing the power of the model to correctly classify all the classes and the influence on the majority class gets minimized.

Lastly, features standardization was done using Standard-Scaler. This move got features with the mean or 0 as well as standard deviation of 1, which is more critical with models such as Support Vector machine (SVM) which is sensitive to scales of features.

In general, the preprocessing phases enhanced the quality of the dataset, equalized the amount of classes, model stability, and the overall performance of the model.

3.4.6 Train Test Split

A stratified split was used to select the training and testing set. In the research, training/testing ratios of 75 / 25 were applied and fixed random state was used to ensure that the split is reproducible. The stratification had been added to make sure that each of the stress classes (Low, Moderate, High) would be represented in both sets in more or less equal measure as in the entire data set. The method helps to make more trustworthy assessment, particularly in case the data is skewed.

3.5 Dataset Characteristics

The data used in the study involves 17 features predictors which symbolize demographic variables and stress influencing variables, including lifestyle choices, academic workload, social setting, and indicators of wellbeing. The study utilized online survey in which the respondents were taken through questionnaires in which they were required to input the background details and the PSS-14 questionnaire. Stress Level is the target variable and has several categories (Low, Moderate, and High) and thus, the problem can be perceived as a multi-class classification. Since the dataset is of mixed feature type and is imbalanced in terms of classes (with do moderate being the most common), the preprocessing was used to prepare it in a proper way. These involved cleaning, coding of categorical variables and standardization of numeric characteristics. To minimize the imbalance effect in the training, SMOTE oversampling was implemented on the training data only and the test set was not changed in order to conduct realistic evaluation. To compare the models, a number of classifiers were tested within the same pipeline and testing strategy, namely, the Logistic Regression as the baseline, Balanced SVM, and Histogram-based Gradient Boosting. Besides, predictions were aggregated using a Soft Voting ensemble to enhance the overall robustness. Because the accuracy can be inflated by class imbalance, it was evaluated based on class mindful metrics like confusion matrices and Macro-F1 that would allow measuring performance in relation to Low, Moderate and Higher stress categories of the evaluation more evenly.

Chapter 4 Proposed Methodology

4.1 Overview

This chapter describes the entire approach that was taken to develop and test stress level classification system. The aim of the suggested technique is to establish the stress category of a student based on the personal, academic, and lifestyle-related characteristics. The methodology will be a step-by-step pipeline because any of the steps may be reiterated in the same manner and the outcomes will be universal. The first step of the workflow involves the reading of the dataset using Python, and a few initial checks in the formatting are conducted to prevent some typical issues like column names discrepancy. Once the data is loaded, the responses to the Perceived Stress Scale (PSS) are taken and then a total stress score is calculated on each respondent. This overall score is then transformed into a three stress level category label Low, Moderate and High. After generating the label, the PSS question columns are eliminated in the feature set to avoid information leaking as it is directly based on the values in the label. The dataset undergoes cleaning followed by stratification of training and testing. The stratification is essential due to the unequal nature of the dataset similar to the example of one given class being far more presented, as opposed to the rest. In a bid to solve this problem, it only implements oversampling (SMOTE) to the training data. Balancing should be applied prior to splitting which might result in over optimistic results and the pipeline is cautious about it. Lastly, there is a comparison of several classifiers that have been trained through one structure of evaluation. These models are a baseline linear model (Logistic Regression), a non-linear classifier (balanced SVM), a boosting based model (Histogram Gradient Boosting), and an ensemble, which is used to aggregate three models via voting. The last stage aims at validation, permutation importance of features analysis, and selection of the most reliable model based on all types of stress.

4.2 Dataset Description (PSS14 Dataset)

In this research, the data was recorded in a CSV format and loaded in the Pandas library. It includes the student level of information connected to daily habits, academic work load, and wellness indicators, as well as answers to the 14 questions of the PSS questionnaire. It has the middle level of samples, and the careful preprocessing and assessment is of special significance since small datasets may be subject to overfitting, which occurs otherwise. Considering machine learning, the dataset can initially consist of two different types of information (1) the predictors that can possibly affect the stress (sleep and workload), and (2) the answers to the questions of the PSS questionnaire that can directly measure the stress. Considering that the direction of the model is to estimate the level of stress based on lifestyle/academic factors, responses of PSS are regarded as the means to produce the label instead of predicting it. When the PSS total is determined and transformed into a category, the dataset transforms to a conventional supervised learning dataset. The rest of the columns are then transformed to the input feature matrix at this point and the resultant stress level is the desired target label. This is necessary to make sure that the model acquires meaningful relationships and not purely the questionnaire answers.

4.2.1 Target Variable

The target variable that this study will study is the Stress-Level which will be derived out of the cumulative score of PSS. The initial step by the notebook is to find all the PSS question columns in the data and compute a new numeric score called PSS-Total, which involves adding the answers on all the 14 questions on each row. This aggregate score is that of the perceived value of the stress of the student. In order to transform this continuous stress score into a classification task, a rule-based mapping function is used as part of the methodology. The boundaries employed in the mapping are such that all students can fit perfectly in one of the three classes. The students whose PSS total scores are lower are classified as Low stress; in the middle of the range are categorized as Moderate stress, and at the upper level of range are categorized as High stress. The conversion is feasible since it produces discrete groups, which is simpler to classify models as well as simpler to understand in practice (as example, distinguish the high-risk students). The other significant design option is that the question columns of the PSS will be deleted once the label is designed. By having the PSS answers as the input features, it would just learn the label using the identical values that made it. That would artificially swell performance and it would not be a true prediction system. As such, the need to drop those columns necessitates the need to preserve the validity of research.

4.2.2 Input Features (17 Features)

The study uses 17 input features representing demographic, academic/work-related, lifestyle, health, and social well-being factors. These variables were selected because they are plausibly associated with stress and can be collected through a structured questionnaire without administering the PSS scale each time. The features include age, gender, marital status, sleep pattern, screen time, physical activity, entertainment, smoking, drinking habits, health condition, study routine, workload, number of projects, satisfaction level, social connection, income indicator, and happiness.

Serial No.	**Input Feature**
01	Age
02	Gender
03	Marital Status
04	Monthly Family Income
05	Physical Activity per Week
06	Average Sleep per Day
07	Average Study/Work Hours per Day
08	Perceived Workload/Study Load
09	Self-Rated Happiness Level
10	Satisfaction with Academic/Work Environment
11	Number of Close Friends/Social Contacts
12	Upcoming Exams/Projects/Deadlines
13	Smoking Habit
14	Caffeine/Tea/Energy Drinks per Day
15	Any Chronic Health Condition
16	Average Daily Screen Time
17	Days per Week Engaged in Entertainment/Relaxation

Table 4.1: Stress Classification

4.3 Feature Encoding Scheme (17 Features + Target)

All 17 input features were categorized as binary, ordinal, nominal, or numeric. Binary features were encoded as 0/1, ordinal features using ordered integer codes, nominal features via one-hot encoding, and numeric features retained as continuous values and scaled during preprocessing.

Feature	Category	Assigned Value
Age	15–39 years	2
	40–58 years	1
	59 years and above	2
Gender	Male	1
	Female	2
Marital Status	Single	2
	Married	1
	Divorced	3
	Widow/Widower	3
Average Sleep per Day	Less than 5 hours	3
	5–6 hours	2
	6–7 hours	1
	More than 8 hours	0
Average Daily Screen Time	Less than 2 hours	1
	2–4 hours	1.5
	4–6 hours	2
	More than 6 hours	3
Physical Activity per Week	None	3
	1–2 hours	2
	3–5 hours	1
	More than 5 hours	0
Entertainment/Relaxation (Days/Week)	Never	3
	1–2 days	2.5
	3–4 days	2
	5–6 days	1
	Daily	0
Smoking Habit	Yes	2
	Occasionally	1
	No	0
Caffeine/Tea/Energy Drinks	None	0
	1 cup	0

	2 cups	0
	3+ cups	1
Chronic Health Condition	Yes	3
	No	0
Average Study/Work Hours	0–2 hours	0
	3–5 hours	1
	6–8 hours	2
	More than 8 hours	3
Perceived Workload/Study Load	Very low	1
	Low	1.5
	Moderate	2
	High	2.5
	Very high	3
Upcoming Exams/Projects/Deadlines	Yes	2
	No	0
Satisfaction with Academic/Work Environment	Very unsatisfied	3
	Unsatisfied	2.5
	Neutral	2
	Satisfied	1
	Very satisfied	0
Number of Close Friends/Social Contacts	None	3
	1–2	2
	3–5	1
	6+	0
Monthly Family Income (Taka)	Less than 15,000	3
	15,001–30,000	2.5
	30,001–50,000	2
	50,001–80,000	1
	More than 80,000	0
Self-Rated Happiness Level		
	Very low	3
	Low	2.5
	Neutral	2
	High	1
	Very high	0

Table 4.2: Encoding Scheme Variables

4.4 Data Preprocessing

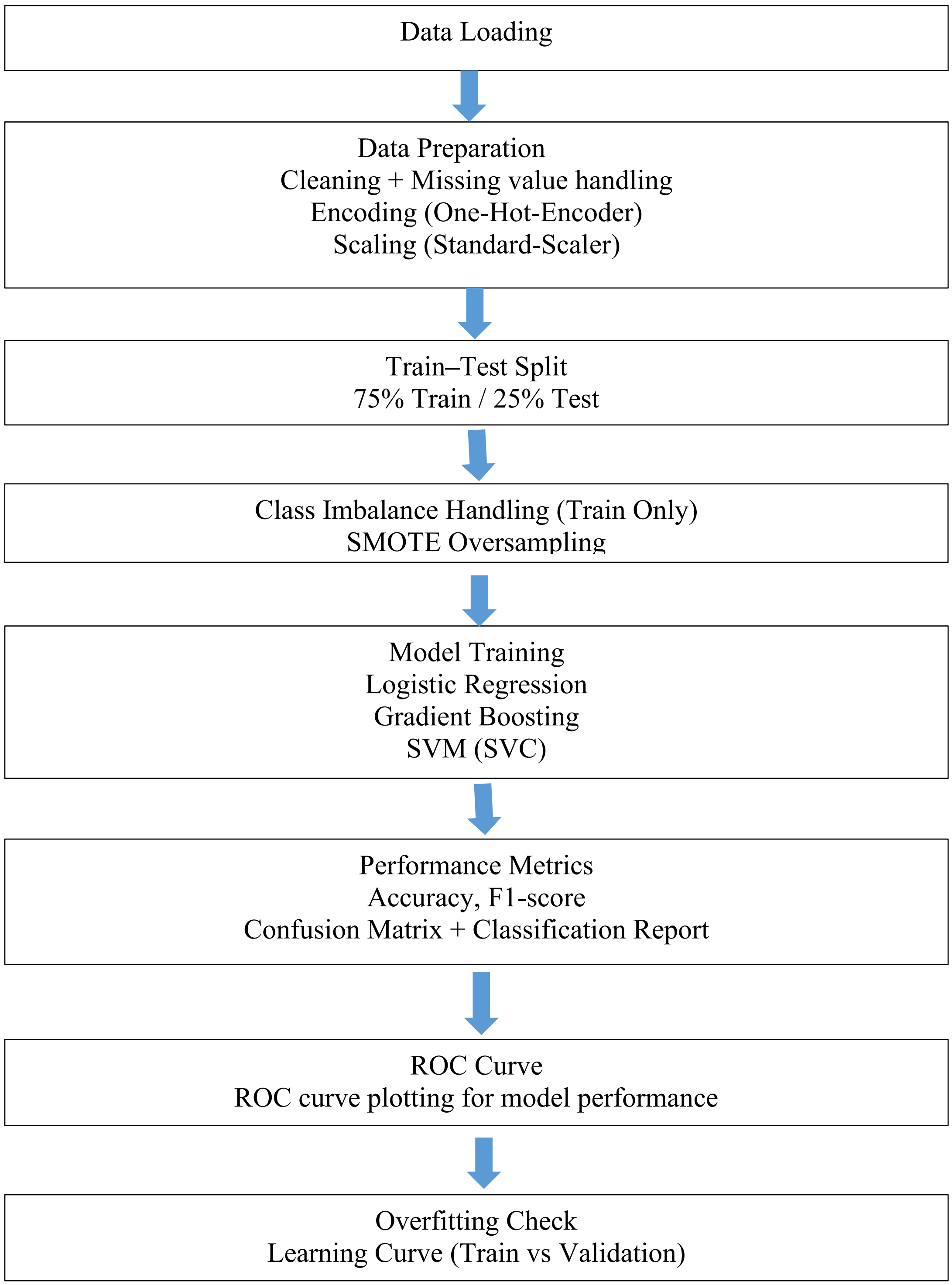
The preprocessing stage serves as an important component of the suggested methodology since the data is mixed in nature with a significant advantage on the target classes. Unless preprocessed, model training can give false results, particularly when the classifier learns to be biased against the majority class. The preprocessing steps are introduced in a systematic way to whereby the pipeline is reasonable, repeatable and in line with best practices.

4.4.1 Initial Class Distribution

The preprocessing step that is done first is to look at the original distribution of the classes of stress. Counts of values to print the count of classes are calculated by the notebook and used through value count. This step shows that the sample size of Moderate stress is significantly more considerable than the sample size of Low and High stress. The significance of this imbalance is that most models are expected to optimize total accuracy. When the data is unbalanced, a model will seem to work depending only on giving most of the time the majority of the results. That would give a high score in accuracy but it would not capture the minority classes appropriately. Low and High stress should be identified in this study and therefore imbalance must not be neglected. Class distribution analysis also gives a point of reference. Comparing the distribution of the distribution prior to and after SMOTE (training only), the researcher can affirm that whether the classes are indeed being oversampled in the way the researcher intends.

Proposed Stress Classification Methodology Framework

This model gives an overview of the stress classification end-to-end workflow, including data preparation and model training, as well as evaluation based on ROC and learning (overfitting) curves.



4.4.2 Handling Class Imbalance

In order to minimize the model bias due to the class imbalance, this paper uses SMOTE (Synthetic Minority Over-sampling Technique) on the training data. SMOTE uses a combination of existing samples to form new synthetic samples of the minority classes by interpolating existing samples in the feature space, as opposed to merely making copies of existing records. This aids the classifier to acquire a more smoother, and general set of decision boundaries particularly of the minority stress categories that get insufficient exposure in training.

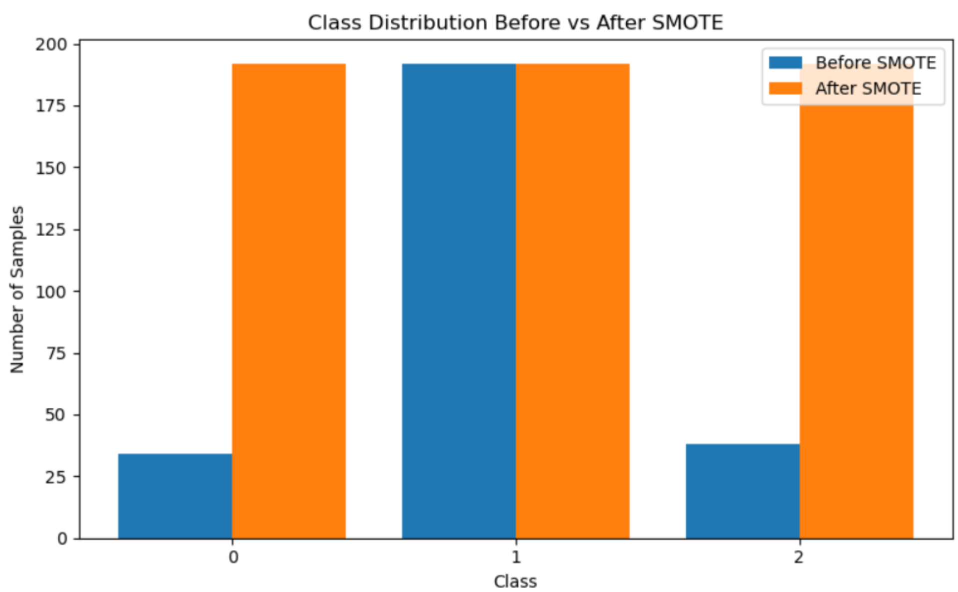


Figure 4.1: Smote Class Distribution

Figure 4.1 (Class Distribution Before vs After SMOTE) has been provided after this paragraph to show the impact of oversampling. The dataset is evidently skewed before SMOTE (blue bars): the number of samples in the middle class (Class 1) is significantly higher, whereas Class 0 and Class 2 contain a significantly smaller number of samples. The number of counts of the classes after the application of SMOTE (orange bars) is more or less equal to the number of counts per all three classes, which means that the training set is balanced. In this application, SMOTE was used with k-neighbors=3 and fixed random seed to make them reproducible.

Besides SMOTE, the SVM model applies class weight = balanced, which puts more weight on the misclassification of the minority classes during training. The combination of the two techniques enhances the strategy through imbalance at two levels: the balancing of data-level (SMOTE) and the weighting of the algorithm-level (balanced loss). SMOTE is only used on the training data and not on the test set since the test set should be a model of the actual distribution in the world to ensure fair and realistic analysis.

4.4.3 Feature Encoding and Scaling

As the dataset has both categorical and numerical variables, there was the need to feature encode. Categorical variables were transformed into numbers like gender and marital status in order to ensure that machine learning algorithms can act on them. The process of encoding was followed by scaling with Standard-Scaler. It is a way of making the data standardized, such that the mean and the standard deviation are 0 and 1 respectively. Scaling can particularly be significant to models such as Support Vector Machine (SVM) which are sensitive to the size of input features. Only the training data was fitted with the scaler and applied to the training and testing set to prevent leakage of data.

4.5 Classification Models

In this study, the comparison of various models is done as opposed to using a single classifier. The rational is that there are varying models which are strong. Other models can be effective when relation is linear whereas others are effective when the relationships are non-linear and complex. The study can defend the final selection on facts and not guess by means of comparing of a number of models with the same preprocessing and evaluation strategy.

4.5.1 Logistic Regression

The Logistic Regression was applied as the baselines classifier in predicting the stress level of the students since it is computationally efficient, stable throughout the training and gives interpretable probability outputs. The stress-level target variable was developed as a result of adding the total PSS score and the PSS question columns were eliminated one at a time to prevent label leakage. The other student-related variables were utilized as the input features. Given that there are three classes (Low, Moderate, High) in the task, a multiclass Logistic Regression set-up was used, and the maximum number of iterations was set to a higher value to make sure that it converged.

Evaluation Metrics	**Score**
Accuracy	0.82
Macro F1-Score	0.70
Weighted F1-Score	0.81

Table 4.3: Logistic Regression Evaluation Metrics

Table 4.3 (Logistic Regression Performance Metrics) singles out the test results. The model had an accuracy of 0.82, which has a Macro F1-score of 0.70 and a Weighted F1-score of 0.81. Due to the unequal distribution of data, Macro F1 is the point of focus in this research as equality of all classes is prioritized, thus, demonstrating the performance of minorities and classes in a more just manner than the sole accuracy of the model. In general, the Logistic Regression can be considered as a robust and interpretable predictive control that can be used for comparison with more developed models, as well as as a component in ensemble methods (particularly soft voting).

4.5.2 Balanced Support Vector machine (SVM)

A Support Vector Machine (SVM) has been taken as a primary classifier in this paper since it is capable of modelling non-linear decision boundaries, which is relevant in stress prediction where the correlation between lifestyle and academic variables and stress level is not necessarily linear. RBF (Radial Basis Function) kernel was chosen to give the flexibility of separation of the three stress classes (Low, Moderate, High). In order to deal with the imbalance in classes, class_weight = "balanced" was used in training the model, which causes the model to be more effective in penalizing error in misclassifiying a minority group, and more objective to the majority (Moderate) group. Since SVM is also sensitive to feature scale, all the input variables were standardized with Standard-Scaler which was only fitted to the training data and applied both to test and training data in order to prevent data leakage. The trained SVM was tested on the unknown test set in terms of accuracy, confusion matrix, and class-wise precision, recall, and F1-score.

Evaluation Metrics	**Score**
Accuracy	0.82
Macro F1-Score	0.75
Weighted F1-Score	0.83

Table 4.4: SVM Evaluation Metrics

The results are indicated in Table 4.4 (Balanced SVM Evaluation Output). The model had a precision of 0.82. The majority class (Class 1: F1 = 0.88) has the best performance, although Class 0 (F1 = 0.75) and Class 2 (F1 = 0.62) are also well classified and more difficult, respectively. This study focuses on Macro F1 (0.75) because the data is uneven and therefore all classes are given equal weight and Micro F1 is more fair than the measure of accuracy

4.5.3 Gradient Boosting using histograms:

As a powerful non-linear predictor, Histogram-based Gradient Boosting (HGB) was applied in this work to predict the degree of stress in students. In contrast to linear models, HGB constructs a sequence of decision trees, each of which is concerned with rectifying the errors of earlier trees. The "histogram-based" method accelerates the training process by grouping continuous attributes into discrete values to facilitate the model to work with large datasets effectively as well as to achieve complex and non-linear relationships between lifestyle/academic variables and stress. The HGB model was trained on the generated feature and tested on an unseen test set with accuracy, confusion matrix and class-wise precision, recall and F1-score (PSS questions were not used to prevent label leakage). Since the data is skewed, Macro F1 was deemed a significant measure since it assigns equal weight to each category of stress.

Evaluation Metrics	Score
Accuracy	0.83
Macro F1-Score	0.75
Weighted F1-Score	0.83

Table 4.5: Gradient Boosting Evaluation Metrics

The results of performance are reported in Table 4.5 (HGB Evaluation Output). The model had an accuracy of 0.83 and it performed slightly better, as compared to the baseline models. The classification report shows the best results with the majority (Class 1) having a F1-score of 0.88. In minority classes, the model obtained an F1-score of 0.70 in Class 0 and 0.67 in Class 2, which is less lopsided in terms of results. The total Macro F1-score is 0.75 indicating that HGB is both offering great multi-class discrimination and is more fair on the minority stress classes.

4.5.4 Ensemble Model (Hard Voting)

Enhanced robustness is also achieved in this study through the application of the ensemble approach to get better than single classifiers. The core concept of an ensemble is that varying models are taught different patterns on the same data: linear models (e.g. Logistic Regression) can be taught the overall trends, whereas non-linear models (e.g. SVM and Histogram Gradient Boosting) can be taught the interaction between the components. The combination of them makes the final prediction more stable and independent of the weakness of any single algorithm. Hard Voting strategy took place. In hard voting, a base model makes predictions considering each model as one of the classes (Low, Moderate, or High) and the ensemble chooses the most voted class. E.g., with Logistic regression predicting Moderate, SVM predicting Moderate and HGB predicting High, the ensemble vote will be Moderate (2/3 votes). This approach is simple to decipher and understandably describes the way the ultimate decision is made. The three models that are used in this work are: Logistic Regression (baseline linear classifier), Balanced SVM (non-linear classifier with class_weight=balanced to eliminate bias in the majority class), and Histogram-based Gradient Boosting (learns non-linear interactions of features effectively). Hard voting is also effective since it is not based on the estimates of probability which in some classifiers can be poorly-calibrated. The hard voting ensemble in general offers an easy and comprehensible form of combining the strengths of the models and minimizing misclassification between stress classes.

4.5.5 Ensemble Model (Soft Voting)

In order to predict the stress level better, the present research employs an ensemble model rather than an individual classifier. The rationale is that various algorithms identify various patterns in the same dataset: logistic regression gets linear trends, SVM (RBF) gets non-linear and complex boundaries, and HGB (Histogram-based Gradient Boosting) gets non-linear interaction of features. Incorporation of these complementary strengths tends to make the prediction more stable and reliable. This ensemble was done using a Soft Voting approach. In contrast to hard voting (utilizing the final class labels only), soft voting is an average of the class probabilities given by each base model. Each of the three stress classes (Low, Moderate and High) is a probability on a per-student record. Averaging these probabilities will then be determined, and the most probable prediction will be the final prediction. It comes in particularly handy in multi-class assignments since it does not simply use the predicted label but also the model confidence.

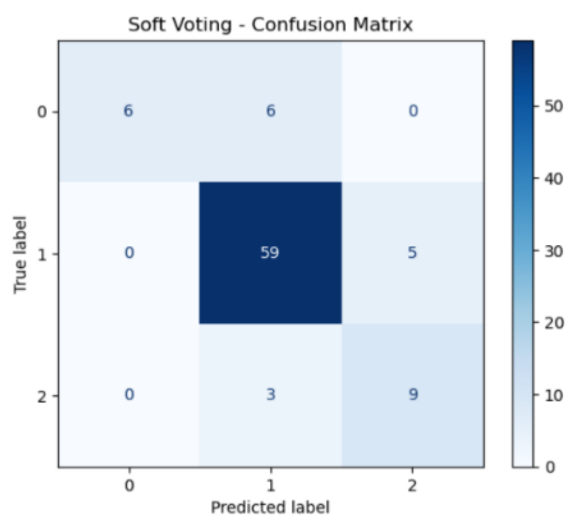


Figure 4.2: Soft Voting Confusion Matrix

Figure 4.2(Soft Voting Confusion Matrix) is also provided here to demonstrate the performance of the soft voting ensemble on the test set. The predictions are mostly on the diagonal (correct classifications) especially the majority Moderate class (59 correct). The misclassifications between the minority classes and the moderate one are also lower in the figure, which implies a more balanced conduct than when using one model. Altogether, the reason why the soft voting ensemble was chosen as the last model was that it delivered more overall constant performance across the three stress classes, eliminating the maximum errors through evidence given by a number of classifiers.

4.6 Feature Importance Analysis

In order to determine the factors that influenced prediction of stress, permutation importance was employed. This technique is useful in explaining the contribution of every feature to the performance of the model. The idea behind permutation importance is in a very simple way, you randomly shuffle the values of a single feature and then measure the amount of degeneration of the model as a result of this shuffling. When the performance decreases greatly following the shuffling of a feature, then the feature is significant in prediction. When the change is minimal or no change at all, the feature is not that influential. This approach enabled us to determine which demographic, lifestyle, and psychosocial factors had a significant influence on the level of stress. This makes this model more open, and simple to interpret and helps us comprehend better the key factors related to stress.

4.6.1 Per Class Importance

Importance feature of permutation of the HGB, Logistic Regression (LR) and the SVM model were calculated by assessing the loss of overall Macro F1-score of the model with each feature randomly permuted. Larger decreases represent the features that the model uses in a more accurate separation of the Low-Moderate-High stress classes.

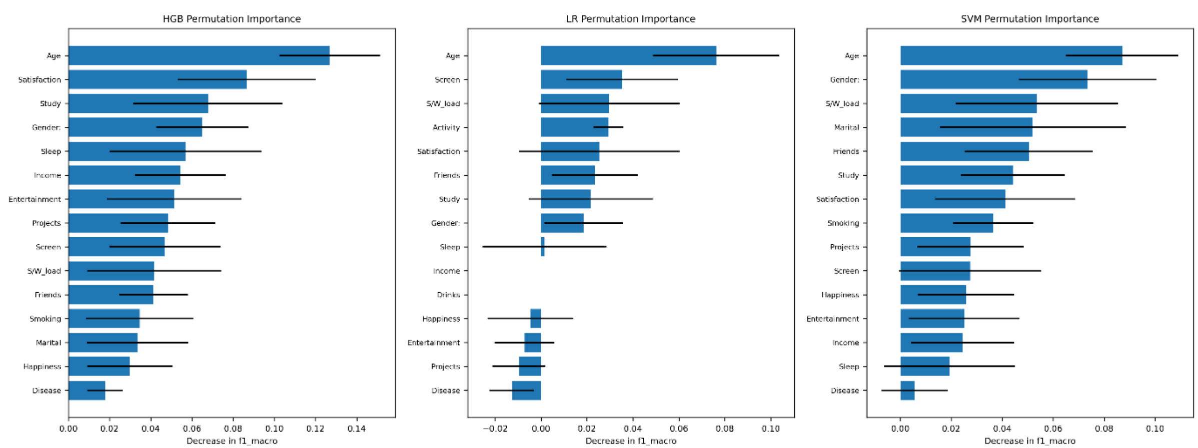


Figure 4.3: Permutation Importance

These results are shown in Figure 4.3 (Permutation Importance of HGB, LR and SVM). There is a definite and constant trend in all three models: Age is the most predictive factor. The Macro F1-score in all plots decreases the most when Age is permuted which indicates that age data is important in preserving the separation of the classes. In the case of HGB model, the significance is shared by a large number of predictors. Satisfaction, Study, Gender, and Sleep are the most powerful variables after Age with routine and workload related variables being Income, Entertainment, Projects, and Screen time, Study/Work load, and Friends. This dispersion implies that HGB enjoys the combination of various demographic and behavioral variables and can include the non-linear relationships. The importance is more centralized in the case of Logistic Regression. Following Age, Screen time, Study/Work load and Activity have the largest effects with Satisfaction, Friends, Study and Gender having minor effects. Some of the remaining variables are of slight significance or slightly negative, and this means that they do not provide a consistent measure of value to this linear classifier. In the case of SVM, the Age is the most dominant indicator with other Gender, Marital status, Friends, and workload indicators (Study/Work load, Study) ranking high as the SVM creates non-linear boundaries with stress classes being overlapping.

4.7 Optimization Strategy

The notebook employs a pragmatic method of optimization that aims at bettering the model configuration, managing the imbalance of classes, and good generalization. Rather than executing a grid-based search that is exhaustive, it will evaluate a small number of carefully selected parameter configurations and examine learning curves and behavior in test-sets to make decisions. The approaches of class-balancing are viewed as a subset of optimization since they have a direct impact on fairness in terms of stress categories. Such methods as SMOTE and equal weights in classes are not merely preprocessing methods, but they assist the models in identifying minority classes (Low and High stress) with greater precision.

In the last model, the soft voting ensemble was chosen not only according to the accuracy. The accuracy itself can be a misleading factor in the situation when the data set is imbalanced and the middle classes prevail. Multiple evaluation reports (confusion matrix to see which classes are being confused- High vs. Moderate in particular), the classification report (the precision, the recall and the F1 value of each class) and above all Macro-F1 (which provides all classes with equal weight) were used to decide on which model to use. The high Macro-F1 implies that the model works in Low, Moderate, and High stresses not only on most of the classes.

The reason why soft voting ensemble was selected is due to the fact that the results of the majority label voting is not used but the combination of the probability outputs of multiple classifiers. This causes it to be more conservative at the edges of classes and not bound to the error of one model. All in all, it was the most appropriate final classifier to predict stress levels in this study, as it was the most balanced in classes and was more likely to be generalized.

4.8 Final Model Selection

The final one was selected among three single classifiers including Logistic Regression (LR), Support Vector Machine (SVM), and Histogram Gradient Boosting (HGB) and two voting ensembles (hard voting and soft voting) employing the same unseen test set (n = 88) which was imbalanced (Low = 12, Moderate = 64, High = 12). As the majority of the samples were the Moderate category, the total accuracy may be good despite the fact that either of the Low or High may have been poorly represented by the model; that is why Macro F1-score and the confusion matrices were chosen to evaluate how fairly each method represented all the three classes. The findings indicate that LR had a macro F1 of approximately 0.70 with an accuracy of 0.8182, but it tended to classify Low-stress as Moderate, which indicates that it has the tendency of favoring the majority classes and it has issues with non-linear classification boundaries. SVM gave the same accuracy (0.8182), but a better macro F1 (0.75), so it was more balanced in terms of the features with proper scaling, though, it still misclassified a few Moderate and High samples near the boundary. HGB was not only the most accurate single model (0.8295) but also had a macro F1 of approximately 0.75, presumably due to its ability to represent more complex patterns and feature interactions; nonetheless, it was also prove to error on Low and High cases, which was largely due to those classes having fewer cases and having significant overlap with Moderate. In order to achieve a more reliable set of predictions, then an ensemble treatment was adopted since LR, SVM, and HGB cannot commit identical errors, and mixing them allows one to mitigate the fault of one model on another. Hard voting did better with accuracy at 0.8409, by making the majority label but it considers all the votes equally and it does not take into account the amount of certainty of each model.

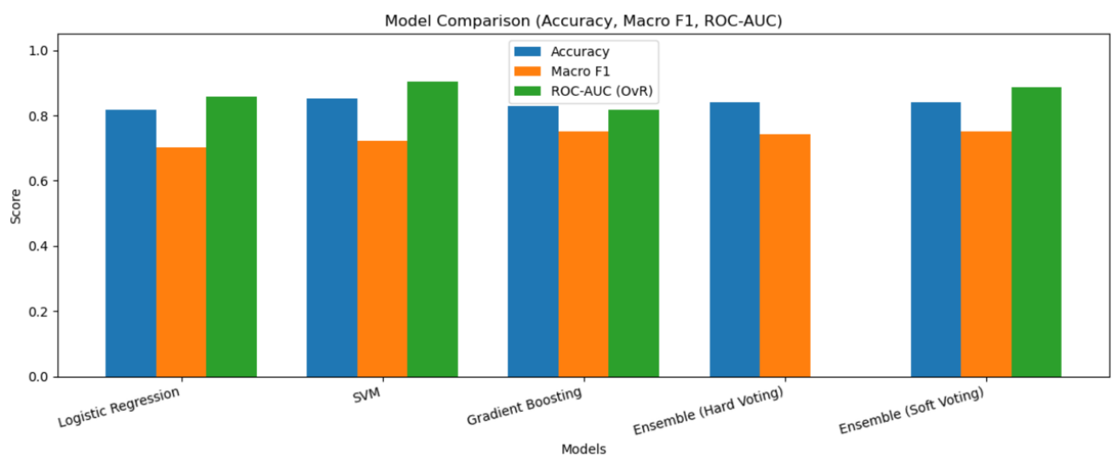


Figure 4.4: Model Comparison

Soft voting, however, is an average of the class probabilities of all base models, and selects the class with the highest average probability which would be more appropriate to this task since categories of stress may overlap, and many test samples are borderline. This probability-based decision resulted in the highest overall performance (accuracy = 0.8523 and Macro F1 = 0.7766) of the soft voting ensemble, which was better than all individual models and hard voting. This option would also be supported by the confusion-matrix patterns. All the methods predict Moderate best as it has the most samples, whereas Low and High are more difficult to predict due to insufficient data and overlap of the boundaries, and soft voting produces the most balanced and consistent result. Thus, in light of its greater accuracy, improved class-balanced Macro F1, and enhanced resilience to imbalance, the final model to be adopted in this thesis was the soft voting ensemble.

4.9 Methodology Summary

We followed an end-to-end machine learning pipeline to classify students into Low, Moderate, or High stress, while keeping the process reproducible and free from leakage—especially important because the classes were imbalanced.

First, we loaded the dataset and cleaned basic formatting (like column names). Stress labels were then created from the 14 PSS questionnaire items by summing them into a PSS total score and converting that score into the three stress categories using a fixed rule. To prevent information leakage, the PSS question columns were removed afterward so the model couldn't "cheat" by learning directly from the same items used to make the label.

Next, we built the final modeling dataset using 17 non-PSS features (demographics, habits, academic workload, social factors, and wellbeing indicators). Data was split into training and test sets using stratified sampling to preserve the stress-level distribution in both sets. Because the data was imbalanced, SMOTE was applied only on the training set to balance the classes, while the test set was kept unchanged for a fair evaluation.

Preprocessing (imputation, encoding, and scaling when needed—especially for SVM) was handled through a structured pipeline. We trained several classifiers (Logistic Regression, Balanced SVM, Histogram Gradient Boosting) and also tested an ensemble approach for more stable performance. Models were evaluated on the unseen test set using accuracy, confusion matrix, and class-wise precision/recall/F1. Since accuracy can be misleading for imbalanced multi-class problems, Macro-F1 was used as the key metric. Learning curves were also checked to watch for overfitting or underfitting.

Based on overall performance and error patterns, we selected a soft-voting ensemble as the final model because it combines predicted probabilities from multiple models, leading to more reliable decisions than a single model or hard voting.

4.10 Stress Level Classification Algorithm

Algorithm 1: Stress Level Classification Procedure (Proposed Method)
Input: CSV dataset D containing 17 student features and PSS responses (PSS_Q1 … PSS_Q14)
Output: Predicted stress class $\hat{y} \in \{0, 1, 2\}$ for each student (0 = Low, 1 = Moderate, 2 = High)

1. **Load dataset**
   1.1 Import the dataset *D* into Python using a data-frame structure.
   1.2 Standardize column names to avoid column-matching errors.
2. **Compute PSS total score**
   2.1 Identify PSS columns: {PSS_Q1, PSS_Q2, …, PSS_Q14}.

2.2 For each student record, compute the total perceived stress score:
PSS-Total = PSS_Q1 + PSS_Q2 + … + PSS_Q14.

3. **Create target label (Stress-Level)**
   3.1 If PSS-Total ≤ 18, assign Stress-Level = 0 (Low).
   3.2 Else if 19 ≤ PSS-Total ≤ 37, assign Stress-Level = 1 (Moderate).
   3.3 Else assign Stress-Level = 2 (High).
4. **Prevent data leakage**
   4.1 Drop all PSS question columns (PSS_Q1 … PSS_Q14) from the data-frame after label creation.
   4.2 Keep only the 17 non-PSS variables as input predictors.
5. **Prepare model inputs**
   5.1 Define feature matrix X using the 17 selected features.
   5.2 Define target vector y using Stress-Level.
6. **Train–test split (stratified)**
   6.1 Split the dataset into training (75%) and testing (25%) sets.
   6.2 Use stratification so the class ratio of Low, Moderate, and High remains similar in both sets.
   6.3 Fix a random seed to keep the split reproducible.
7. **Handle class imbalance (training only)**
   7.1 Check the class distribution in y-train.
   7.2 Apply SMOTE on the training set only to balance the minority classes.
   7.3 Do not modify the test set, so evaluation reflects real unseen data.
8. **Preprocessing (scaling/encoding as required)**
   8.1 Fit preprocessing transformations using training data only (to avoid leakage).
   8.2 Apply the same transformations to both training and test sets.
   8.3 For models sensitive to feature scale (SVM), apply standardization (Standard-Scaler).
   8.4 For the ensemble pipeline, apply structured preprocessing (imputation + scaling, and encoding if categorical fields exist).
9. **Train base classifiers**
   9.1 Train Logistic Regression on the balanced training data.
   9.2 Train Balanced SVM (using class-weight = "balanced") on the balanced training data.
   9.3 Train Histogram Gradient Boosting on the balanced training data.
10. **Train final classifier (Soft Voting Ensemble)**
    10.1 Combine the trained base classifiers using soft voting.

10.2 Compute class probabilities from each base model and average them.
10.3 Assign the final class as the one with the highest averaged probability.

11. **Testing and evaluation**
    11.1 Predict stress class labels for the unseen test set: ŷ-test = Model(X-test).
    11.2 Evaluate performance using: Accuracy, Confusion Matrix, Classification Report (Precision/Recall/F1), and Macro-F1.
    11.3 Select the final model based on balanced performance across all classes and reduced misclassification of minority classes.

Chapter 5 Experimental Result Analysis

5.1 Introduction

The chapter tells about the results of the experiment conducted on the proposed stress-level classification pipeline. The aim of the experiments was to estimate the quality of classifying students into Low (0), Moderate (1), and High (2) stress categories with the help of the chosen 17 features by various machine learning models. Due to the fact that the dataset is not balanced, and the dominant class is the Moderate one, the performance is not evaluated based on the accuracy only. In addition to accuracy, the experiments are based on confusion matrices, class-wise F1-score, precision/recall, and Macro-F1, ROC-AUC (one-v-rest), and learning curves to learn the behavior and reliability of the models. The reported findings are founded on the held out test set of 88 samples, in which the distribution of classes in testing was 12 (Low), 64 (Moderate), and 12 (High). Such a split matters as it is an indication of the actual imbalance trend in the data set and this serves to validate whether minority classes can still be identified by the models in the same way. 5.2 Impact of Behavior and Class Distribution of Data. The samples of Moderate-stress are naturally high in number as compared to Low and High. This unbalance has two impacts on the model training. First, a model may achieve an apparently good accuracy, that predicts Moderate stress often, since that category is the highest occurring class. Second, the minority classes (Low and High) have higher chances of being wrongly classified: into Moderate, particularly in cases of overlapping features between them. To minimize this impact, the training set was post-train test split into SMOTE balanced sets and sets. This is a choice that left the test set intact, which is required in the honest assessment. The test distribution was still skewed following balancing of the training data hence the use of Macro-F1 and class-wise recall was necessary to find out whether the models had really learned to recognize Low and High stress and not merely the majority one.

5.2 Dataset Behaviour And Class Distribution Impact

The initial data was skewed in terms of classes with the number of samples in the Moderate stress group being greater than the number of samples in the Low and High stress groups. This bias may impact the learning process since machine learning models are more likely to be effective on the majority class. In order to overcome this problem and retain the realistic conditions of evaluation, the training dataset only was balanced with the methods of oversampling. The test data was stored in their natural distribution form. This will make sure that in the course of training, the models are taught with balanced data, but assessed with data that is representing the real-world conditions. The study was also able to determine the extent to which the models generalize to the unknown data by maintaining the test set the same way. This arrangement offers a better and more reliable explanation of the real-life performance of models in the category of stress levels.

5.3 Performance Analysis

Performance analysis evaluates how well the models classify stress levels using metrics like the confusion matrix, precision, recall, F1-score, and Macro-F1 to ensure balanced results across all classes.

5.3.1 Logistic Regression Test Performance

The performance of the logistic regression is shown below (5.3.1). This was yielded with an overall test accuracy of 0.8182 by Logistic Regression. Although the accuracy seems to be good, in the confusion matrix, it can be seen that the model was not very capable to differentiate between Low stress and Moderate stress: Confusion Matrix (Weighted Meaning of Logistic Regression) [[5, 7, 0], [0, 59, 5], [0, 4, 8]] This implies that among 12 Low-stress cases, it was only in the 5 cases that it was correct and 7 cases that were wrongly predicted as Moderate. Moderate was relatively successfully dealt with (59 correct out of 64) and even High was also successfully dealt with (8 correct out of 12), but some of the High samples were still erroneously given to Moderate. This is shown by the report of the class-wise. The diagnosis was excellent in the Moderate type of class, and recall was low in the Low type of class. Consequently the Macro-F1 of Logistic Regression was 0.70 which is not as high as it would suggest given the accuracy of models. Results of the ROC Analysis (one-vs-rest) indicate that the Logistic regression attained: AUC (Class 0) = 0.886 AUC (Class 1) = 0.836 AUC (Class 2) = 0.853 Micro-average AUC = 0.931 The behavior of the learning curve of Logistic Regression is more balanced then that of other models: as training size increases the training score drops, but validation improves and levels of. This is an indication that the model is not overfitting intensively but rather its weakness is predominantly due to the linearity of its decision boundary.

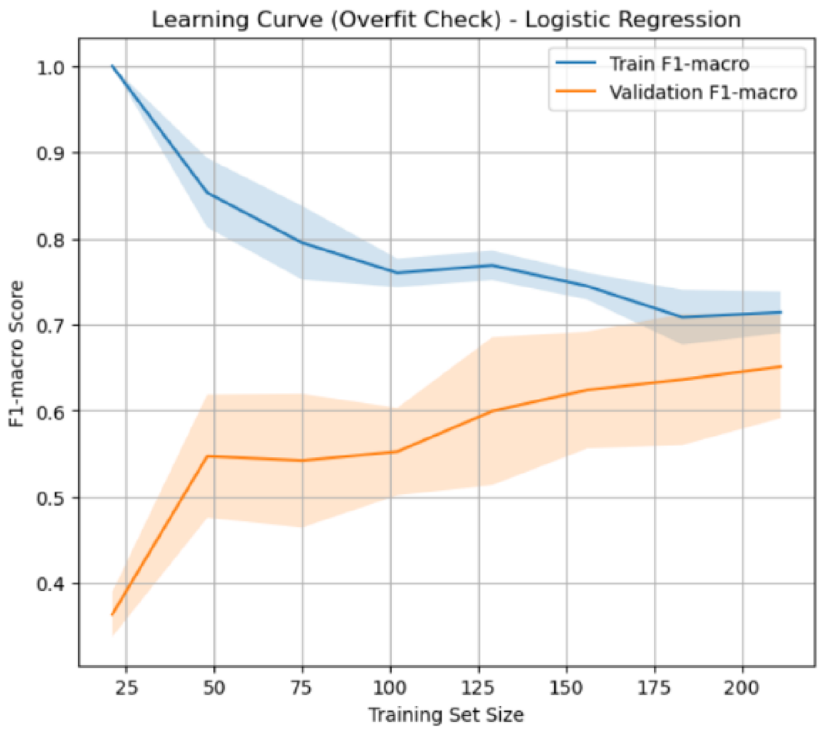


Figure 5.1: Learning Curve (Overfit Check) — Logistic Regression

Figure 5.1 learning curve of Logistic Regression indicates that the training Macro-F1 of the Logistic Regression model declines with the increase in samples whereas the validation Macro-F1 increases and takes a more steady form with increase in the training size. The disconnect between the performance manifested during training and the validation performance diminishes with time, an indicator that Logistic Regression is not particularly overfitting and most likely generalizes well. The validation performance is yet clueless than the training performance thus showing that more data or richer features can still enhance its performance of separating the three classes of stress in a better manner.

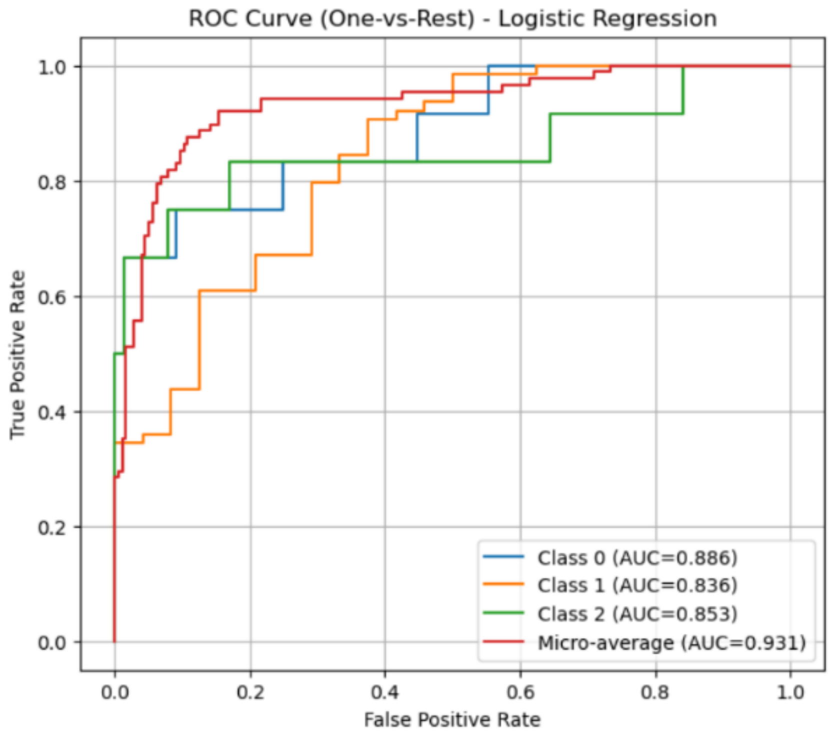


Figure 5.2: ROC Curve (One-vs-Rest) Regression Logistic.

Figure 5.2 ROC curves show that there is a good overall ranking ability by the Logistic Regression to separate the three classes. The micro-average AUC is large that indicates that the model is able to discriminate across thresholds between classes. The other classes such as Moderate stress have a stable behavior since it has a larger sample size whereas Low and High stress are more sensitive since they have fewer test samples. In general, the ROC analysis proves that the predictions of the hard use of Logistic Regression are a good baseline, although they continue to confuse the Low stress with the Moderate in a number of cases.

Logistic Regression Performance

Class-wise Performance:

Stress Level	Precision	Recall	F1-Score	Support
Low (0)	1.00	0.42	0.59	12
Moderate (1)	0.84	0.92	0.88	64
High (2)	0.62	0.67	0.64	12

Overall Performance:

Metric	Score
Accuracy	0.82
Macro Avg Precision	0.82
Macro Avg Recall	0.67
Macro Avg F1-Score	0.70
Weighted Avg Precision	0.83
Weighted Avg Recall	0.82
Weighted Avg F1-Score	0.81

Table 5.1: Logistic Regression Performance

5.3.2 Balanced SVM Performance

Balanced SVM also obtained a test accuracy of 0.8182, although the pattern of error was different, and in certain classes, more balanced: Balanced SVM (Confusion Matrix) [[9, 3, 0], [2, 54, 8], [1, 2, 9]] The SVM also significantly enhanced the recognition of Low stress compared to the Logistic Regression (9 correct of 12). Nonetheless, it brought some additional confusion within the Moderate group: 8 Moderate samples were valued to be placed into the High classification, and 2 were valued to be placed into the Low classification. It brought some positive improvement in the detection of high stress as the sample size with the correct results was raised (9 out of 12), yet some of the samples with high stress were classified as Low or Moderate. The Meso-F1 of SVM was 0.75 meaning that it has a better balance between the three classes than the Logistic Regression. This is in line with the fact that SVM is able to provide non-linear boundaries and the class-weight=balanced allows discouragement of bias favors the majority class. The SVM values of the ROC-AUC were high: AUC (Class 0) = 0.917 AUC (Class 1) = 0.865 AUC (Class 2) = 0.863 Micro-average AUC = 0.938 There is an evident gap between the score of training and validation in the learning curve. The level of training performance is also high, with validation increasing at a slower rate and being lower. This shows that SVM is capturing strong patterns during training but it is also showing overfitting particularly in cases where the size of the training set is small.

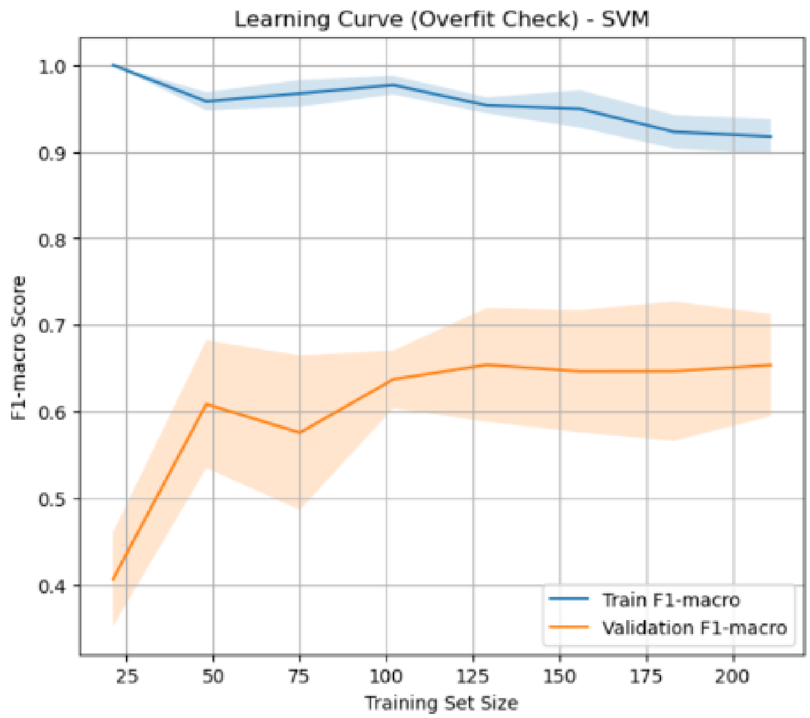


Figure 5.3: Learning Curve (Overfit Check) SVM.

There is a steady high training Macro-F1 in the SVM learning curve, whereas the validation Macro-F1 is increasing steadily, although still significantly lower. This recurrent gap indicates that SVM conforms well to the training data and some overfitting is observed to a certain extent especially when the size of the training is minimal. The more training data increases the better the validation performance is and the more stable it becomes and this implies that the larger the data set the better the model will perform.

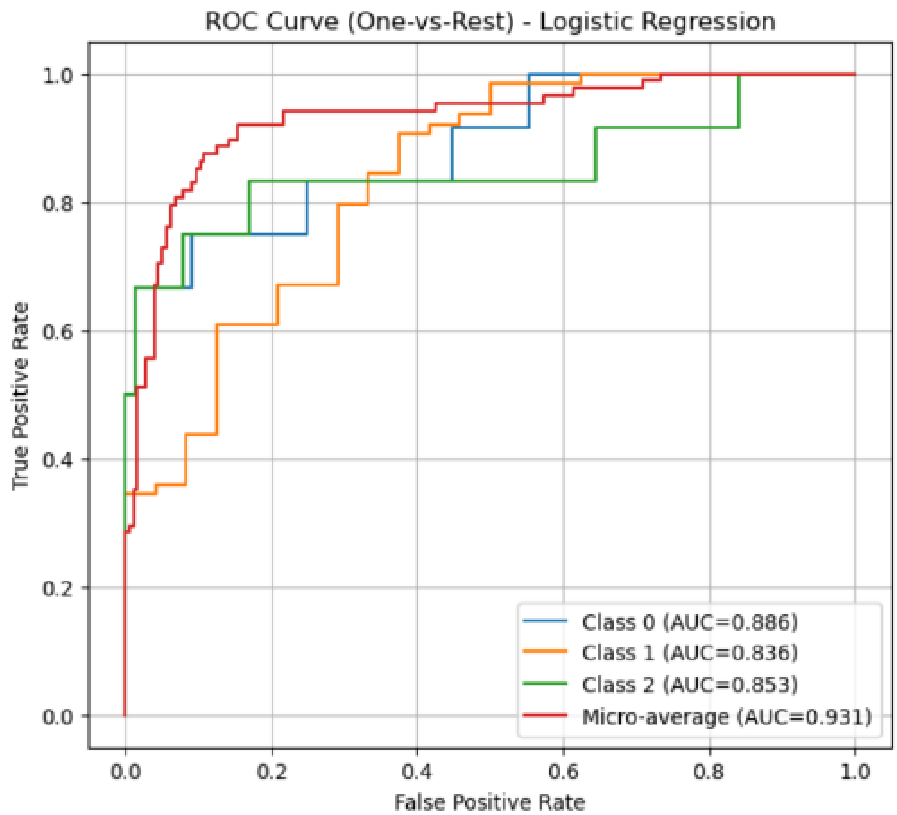


Figure 5.4: One-vs-Rest Curve ROC SVM.

These curves demonstrate that the SVM ROC curves exhibit high performance in terms of separating the classes, and the micro-average AUC and AUC values are high with all three classes. This shows that SVM is effective in discriminating Low, Moderate and High levels of stress at various levels. The findings justify the application of SVM as a powerful non-linear classifier in the study particularly due to its effectiveness in ranking minority-class over other simple linear models.

Balanced SVM Test Performance

Class-wise Performance:

Stress Level	Precision	Recall	F1-Score	Support
Low (0)	0.75	0.75	0.75	12
Moderate (1)	0.92	0.84	0.88	64
High (2)	0.53	0.75	0.62	12

Overall Performance:

Metric	Score
Accuracy	0.82
Macro Avg Precision	0.73
Macro Avg Recall	0.78
Macro Avg F1-Score	0.75
Weighted Avg Precision	0.84
Weighted Avg Recall	0.82
Weighted Avg F1-Score	0.83

Table 5.2: Balanced SVM Performance

5.3.3 Histogram Gradient Boosting Performance

Histogram Gradient Boosting gave the best accuracy of the three single models, having test accuracy of 0.8295: Confusion Matrix (HGB) [[7, 5, 0], [1, 57, 6], [0, 3, 9]] This model forecasted Moderate stress with a high level of accuracy (57 correct out of 64), and it responded to High stress (9 correct out of 12). Low stress was again the main weakness with 5 of 12 being predicted as Moderate. Low stress recall did not go as high as the other classes in the classification report. HGB Macro-average performance was: Macro F1 [?] 0.75 (from the report) The HGB results of the ROC-AUC were less than the SVM and Logistic Regression: AUC (Class 0) = 0.840 AUC (Class 1) = 0.798 AUC (Class 2) = 0.815 Micro-average AUC = 0.907 The HGB learning curve indicates the greatest tendency to overfitting of the tested models. The F1 training increases quickly and almost to near-perfect training, whereas validation increases much slower and also is considerably less. Such a gap represents that the model is overfitting training patterns, which is typical of boosting techniques in cases when the dataset is not large and overlapping boundaries between classes.

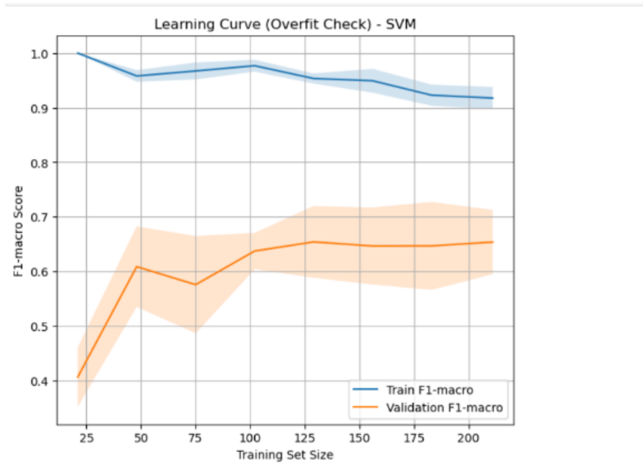


Figure 5.5: Learning Curve (Overfit Check) -Hist Gradient Boosting Classifier.

The learning curve of Histogram Gradient Boosting demonstrates that the training Macro-F1 increases speedily and approaches to the value of almost perfect, whereas validation Macro-F1 increases at a lower rate and is much lower. The fact that the training and validation curves are far apart, suggests that there is more overfitting propensity. This implies that the model is memorizing the specific patterns of the training set that are not completely generalized to unknown data, which might be due to the small size of the dataset and inter-section between classes of stress.

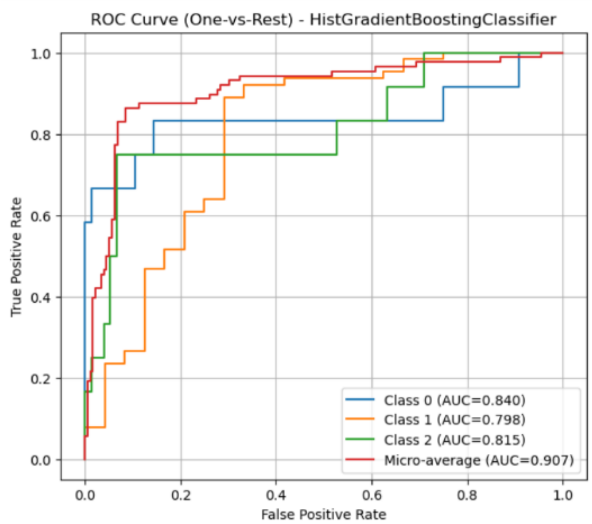


Figure 5.6: ROC Curve (One-vs-Rest) -- Hist gradient Boosting Classifier.

Histogram Gradient Boosting has moderate levels of class discrimination represented in the ROC curves with the smallest values or less values of class-wise AUC than SVM and Logistic Regression. As far as the model performs well in terms of test accuracy, results on AUC show that its probability ranking is not as consistent across classes. This is significant since the stability of probability is significant under the instance where the model will subsequently be utilized in risk scoring or threshold-based alerts. The micro-average AUC supports that in general discrimination is not the strongest of all the evaluated classifiers.

Hist gradient Boosting Test Performance

Class-wise Performance:

Stress Level	Precision	Recall	F1-Score	Support
Low (0)	0.88	0.58	0.70	12
Moderate (1)	0.88	0.89	0.88	64
High (2)	0.60	0.75	0.67	12

Overall Performance:

Metric	Score
Accuracy	0.83
Macro Avg Precision	0.78
Macro Avg Recall	0.74
Macro Avg F1-Score	0.75
Weighted Avg Precision	0.84
Weighted Avg Recall	0.83
Weighted Avg F1-Score	0.83

Table 5.3: Hist gradient Boosting Performance

5.3.4 Ensemble (Hard and Soft) Performance

As a way to enhance stability and organizational independence on the basis of a single classifier a voting ensemble was explored, with either hard or soft voting. The combination of the base models took place in the ensemble and led to the subsequent results. Hard Voting Accuracy = 0.8409 Confusion Matrix [[7, 5, 0], [0, 58, 6], [0, 3, 9]] Hard voting had better accuracy than the individual models, and a smaller number of misclassifications but was still also subject to the same phenomenon that Low samples were shifting to Moderate. Finally Selected Model Soft Voting. Accuracy = 0.8523 Macro-F1 = 0.7766 Confusion Matrix [[7, 5, 0], [0, 59, 5], [0, 3, 9]] Soft voting performed the highest total results in the experiment. The overall progress can be seen in the accuracy and Macro-F1, implying that the model did not only gain even more balanced general performance, but it also performs better in terms of balance within classes. The output indicates that the probability averaging gave a small chance to soft voting that provide more stable results on the borders of classes than hard voting.

5.4 Comparison Of Models

Models The comparative review demonstrates that the choice of models relies on the overall accuracy as well as class fairness. The Logistic Regression is stable, has high Moderate results but Low-stress recall is low, and this lowers macro-level performance. Balanced SVM achieves greater ROC performance and Low stress detection, but is found to move some Moderate samples into High stress. The highest single-model accuracy is based on Hist Gradient Boosting, though the learning curve implies the stronger overfitting and the AUC values are relatively lower, which indicates the lack of the stability of the probability ranking. This combination of models does have its profit as the results of the ensemble prove. Hard voting was more accurate than single models, whereas soft voting had the optimal result. The test accuracy and Macro-F1 of soft voting were the highest implying that it not only predicts better things, but it also has better balance in minority and majority classes. This is why soft voting was chosen as the last classifier of the proposed methodology.

Model	**Accuracy**	**Macro-F1**	**Micro-Avg AUC**
Logistic Regression	0.82	0.70	0.931
Balanced SVM	0.82	0.75	0.938
HGB	0.83	0.75	0.907
Ensemble (Hard)	0.84	------	-----
Ensemble (Soft)	0.85	0.77	-----

Table 5.4: Comparison of Models

5.5 Feature Importance Analysis

The values of features that are important were analyzed in a permutation-based method (also consistent with the project design, as neither the SVM nor voting ensembles gives explicit feature importance in a homogeneous manner). This analysis aims at analyzing the influence of the input variables on the model predictions it is hoped to comprehend which input variables are the most significant in forecasting performance by measuring the change in performance when the model is randomized by randomly shuffling the feature values. When the stress is predicted, the most significant elements are usually associated with the aspects related to everyday life and academic issues, i.e., sleep-related behavior, workload/study patterns, satisfaction, and happiness indicators, as such variables are conceptually related to stress outcomes. The features look more significant in separating Low and Moderate, and others appear more important when distinguishing between High stress and between the High and Moderate stress levels. This is in line with the confusion matrices, with the majority of the errors around the boundaries of the classes called as Moderate.

5.6 Impact Of Class Imbalance

The skew in data had a significant influence over the results of models. Without balancing, the models will leant towards the Moderate predictions. All models, including SVM, which also utilized class weighting, showed improvement in minority-class detection after SMOTE had been applied on the training data. But the test set had been perversely balanced, i.e. the difficulty was still there: the minority classes are expected to be more difficult to identify reliably, as there are fewer real representatives of them, and they often confound with Moderate ones. This is why Macro-F1 ought to have been regarded as a key measure. The Macro-F1 (0.7766) had the highest value, in the case of the soft voting ensemble, which means the influence of balancing and model combination allowed reducing the impact of the majority-class dominance.

5.7 Robustness And Reliability

Learning curves and ROC curves were used to assess robustness and not individual test scores. Learning curves demonstrated that the Logistic Regression is more consistent and less prone to overfitting whereas the SVM and the HGB actually have larger distances between the training and validation performance. The ROC-AUC analysis also indicated the validity of SVM and Logistic Regression whose micro-average AUC values were in the range of over 0.93. HGB demonstrated lower values of AUC, implying that its ranking probabilities were more inconsistent between classes. The voting ensemble of recidivism was soft, and this increased reliability because it did not rely on a single model. When a single base classifier suffers an uncertain prediction, it is able to be fixed up by the ensemble via probability averaging in intention to stabilize decisions at the boundaries of classes.

5.8 Practical Implications

The findings have shown that the classification of stress levels based on lifestyle and academic characteristics is feasible, particularly when the model is trained with imbalance management and measured with balanced measures. Having the accuracy of more than 85% and improved Macro-F1, the soft voting ensemble can be deemed as appropriate in situations of the screening, where the objective is to identify students who might need some attention. Nevertheless, the system is supposed to be utilized as a decision-support tool as opposed to professional evaluation. In practice, it may be useful to detect patterns early, but decision making must have human examination and ethical measures.

5.9 Limitations Observed During Experiments

A number of limitations were noted when conducting experiments:

1.The test set has small minority classes. There are 12 samples of Low and High stress, which restricts the degree of stability per-class metrics can have.
2.Moderate stress, is overlapping with the Low and High. Majority of the misclassifications lie in and around Moderate which is an ambiguity of the real world and not necessarily just a weakness of the model.
3.Overfitting of complicated models. HGB performs highly with low validation in the learning curves, which implies the risk of overfitting.
4.Detail is restricted by feature coding. Most of the variables are coded into survey numbers. This makes granularity smaller and finer behavioral differences can be concealed.
5.Single-dataset limitation. Findings are made on a single dataset and hence results might vary with students of varying situations, colleges and ages.

Although these restrictions exist, the experiments confirm that the proposed pipeline, especially the soft voting ensemble, has the best balanced and reliable performance on this dataset.

Chapter 6 Conclusion And Future Scopes

6.1 Conclusion

This study also sought to come up with a realistic machine learning framework that would be used to classify the stress of students into three categories namely: Low, Moderate, and High. The overall stress score was calculated and put into the target class labels by using the PSS-14 questionnaire it was then followed by the prediction based on 17 demographic, academic, lifestyle, and psychosocial variables. Remaining in the feature set were PSS item responses, which helped in avoiding label leakage and also made the models be assessed in a realistic manner. In order to maintain the organizational process, the research followed a full pipeline in the way outlined in the preceding chapters. These steps involved checking and cleaning the data, creating the labels of stress, encoding categorical variables, scaling (where relevant) and stratified division of the data into the training and test sets. Since the imbalance in the dataset is inherently present, SMOTE was used to the training data only to ensure that minority classes are better learnt but leave the test set distribution as is. Multiple classifiers have been trained and compared such as Logistic Regression, a balanced SVM, Histogram Gradient Boosting and ensembles based on voting. Owing to the fact that the majority of the samples fall in the Moderate group, it was not assessed based on the accuracy. As an alternative, confusion matrices, class-specific precision, recall, and F1-scores have been taken into account in addition to Macro-F1 to comprehend that each of the models dealt with all stress categories. The total performance of the soft voting ensemble was the most consistent on the held-out test set of 88 samples (12 High, 64 Moderate and 12 Low). It scored 0.8523 and 0.7766 on the accuracy and Macro- F1 score respectively, demonstrating the improved balance across the classes than to the individual models. As anticipated, the Moderate stress category was the only one identified best with Low and High being the most difficult because there is no high sample but close to the Moderate line. The discussion concerning the feature impact also gives the idea that stress is not manipulated by only one factor. Daily habits variables and academic pressure variables, including sleep-related behavior, workload/study pattern, and measure of wellbeing demonstrated significant predictive values. This helps the notion that stress depending on personal perception and daily situation. Altogether, even well-organized survey-based data can be successfully used in machine learning to categorize stress levels in a valid and replicable way.

Despite the potential introduction of bias in the form of self-reported data, the pipeline, as proposed and the results suggest, will be of a valuable basis in a subsequent study and interventions to apply in practice to carry out mental health screening in students.

6.2 Future Work

Despite the excellent outcomes of the proposed approach, there are some aspects through which this piece of work can be enhanced going forward. To start with, it is necessary to increase the scale of the dataset and diversify it with more participants. The external validity could be enhanced by gathering data of various institutions, disciplines, and socio-cultural backgrounds. More to the point, adding Low and High stress samples, would assist the model in learning the minority patterns more faithfully and mitigating the figures of confusing the two, with the one of a minority with the one of a Moderate stress. Second, additional imbalance-handling strategies can be experimented. In this study, SMOTE was used on the training set alone, a move that enhanced the learning of the minorities at the expense of the test evaluation without compromising the realism. The variants of SMOTE could be compared with future work, and hybrid sampling (excessive and restrictive sampling) or cost-sensitive training could be used to determine which approach gives the most reasonable generalization and the ability to control overfitting. Third, more robust classes and powerful ensemble designs can be investigated in future research. The soft voting ensemble did not do badly in this research, although methods like stacking or blending can lead to higher performance. It might also be explored using deep learning models that are created to work with tabular data, particularly when a bigger dataset is possible. Fourth, the existing data is cross-sectional and does not qualify to provide a classification of stress once. Further work in this area could involve the collection of longitudinal data in future such that stress is modeled as a processes that changes with time. This would enable time sensitive analysis and possibly early warning systems, through monitoring of stress-related behavior. Last but not least, the model can be incorporated into a practical web or mobile-based screening device that will be used in student support environments. Nevertheless, any practical implementation must put privacy, transparency and ethical protection first, and the system must not be applied as a diagnosing tool, but still can serve as a helping screening tool to the practitioner.

References

[1] S. Cohen, T. Kamarck, and R. Mermelstein, “A Global Measure of Perceived Stress,” *Journal of Health and Social Behavior*, vol. 24, no. 4, p. 385, Dec. 1983, doi: 10.2307/2136404.

[2] E.-H. Lee, “Review of the Psychometric Evidence of the Perceived Stress Scale,” *Asian Nursing Research*, vol. 6, no. 4, pp. 121–127, Dec. 2012, doi: 10.1016/j.anr.2012.08.004.

[3] C. Shi, Y. Guo, H. Ma, and M. Zhang, “Psychometric validation of the 14-item perceived stress scale in Chinese medical residents,” *Curr Psychol*, vol. 38, no. 6, pp. 1428–1434, Dec. 2019, doi: 10.1007/s12144-019-00285-y.

[4] 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.

[5] V. Doma, A. Abd Almisreb, E. Yaman, S. Amanzholova, and N. Ismail, “Classification of Student Stress Levels Using a Hybrid Machine Learning Model,” in *AI-Driven: Social Media Analytics and Cybersecurity*, vol. 1180, W. M. S. Yafooz and Y. Al-Gumaei, Eds., in Studies in Computational Intelligence, vol. 1180. , Cham: Springer Nature Switzerland, 2025, pp. 77–104. doi: 10.1007/978-3-031-80334-5_5.

[6] M. Di Cosmo, S. Campanella, M. Bernardini, A. Mancini, and L. Palma, “Mental health risk prediction in autoimmune patients in primary care: A class-sensitive machine learning approach,” *Smart Health*, vol. 39, p. 100634, Mar. 2026, doi: 10.1016/j.smhl.2026.100634.

[7] Y. Guo *et al.*, “Exploring nonlinear and interaction effects of macro- and micro-environmental factors on age-friendliness of streets through explainable machine learning,” *Building and Environment*, vol. 285, p. 113659, Nov. 2025, doi: 10.1016/j.buildenv.2025.113659.

[8] E. Todd *et al.*, “Lifestyle factors and other predictors of common mental disorders in diagnostic machine learning studies: A systematic review,” *Computers in Biology and Medicine*, vol. 185, p. 109521, Feb. 2025, doi: 10.1016/j.compbiomed.2024.109521.

[9] R. Ahuja and A. Banga, “Mental Stress Detection in University Students using Machine Learning Algorithms,” *Procedia Computer Science*, vol. 152, pp. 349–353, 2019, doi: 10.1016/j.procs.2019.05.007.

[10] PhD. Scholar, Department of Management Sciences, Superior University, Lahore, Punjab, Pakistan and R. Jabbar, “Performance Anxiety, Deadlines, and Employee Well-Being: A Moderated Mediation Model,” *JLSS*, vol. 7, no. 3, pp. 154–176, Sep. 2025, doi: 10.52279/jlss.07.03.154176.

[11] G. Wilcox and D. Nordstokke, “Predictors of University Student Satisfaction with Life, Academic Self-Efficacy, and Achievement in the First Year,” *CJHE*, vol. 49, no. 1, pp. 104–124, Apr. 2019, doi: 10.47678/cjhe.v49i1.188230.

[12] M. S. Syahrir and F. Falah, “Occupational Burnout and Workload: Mental Health Challenges in the Workplace and Solutions,” *J Health Lit Qual Res*, vol. 1, no. 2, pp. 69–83, Sep. 2021, doi: 10.61194/jhlqr.v1i2.528.

[13] A. F. Azizova, L. O. Baimoldina, Z. N. Zhangaliyeva, and N. Y. Kabirova, “THE IMPACT OF ACADEMIC WORKLOAD ON STUDENTS’ PSYCHOLOGICAL HEALTH: A DIAGNOSTIC ASPECT,” *BULLETIN Series Psychology*, vol. 82, no. 1, Mar. 2025, doi: 10.51889/2959-5967.2025.82.1.014.

[14] J. Zhang, M. He, X. Wang, H. Jiang, J. Huang, and S. Liang, “Association of sleep duration and risk of mental disorder: a systematic review and meta-analysis,” *Sleep Breath*, vol. 28, no. 1, pp. 261–280, Mar. 2024, doi: 10.1007/s11325-023-02905-1.

[15] I. Zollars, T. I. Poirier, and J. Pailden, “Effects of mindfulness meditation on mindfulness, mental well-being, and perceived stress,” *Currents in Pharmacy Teaching and Learning*, vol. 11, no. 10, pp. 1022–1028, Oct. 2019, doi: 10.1016/j.cptl.2019.06.005.

[16] K. Mikkelsen, L. Stojanovska, M. Polenakovic, M. Bosevski, and V. Apostolopoulos, “Exercise and mental health,” *Maturitas*, vol. 106, pp. 48–56, Dec. 2017, doi: 10.1016/j.maturitas.2017.09.003.

[17] D. R. Lara, “Caffeine, Mental Health, and Psychiatric Disorders,” *JAD*, vol. 20, no. s1, pp. S239–S248, Apr. 2010, doi: 10.3233/JAD-2010-1378.

[18] N. Plurphanswat, R. Kaestner, and B. Rodu, “The Effect of Smoking on Mental Health,” *am j health behav*, vol. 41, no. 4, pp. 471–483, Jul. 2017, doi: 10.5993/AJHB.41.4.12.

[19] V. Kavya, B. Suganya, M. Alagarsamy, A. Kalaivanan, and P. Chandy, “A Cross-Sectional Study to Identify the Depression among Widowed in a Rural Community,” *Indian Journal of Continuing Nursing Education*, vol. 23, no. 2, pp. 201–204, Jul. 2022, doi: 10.4103/ijcn.ijcn_106_21.

[20] M. Richards, R. Hardy, and M. Wadsworth, “The effects of divorce and separation on mental health in a national UK birth cohort,” *Psychol. Med.*, vol. 27, no. 5, pp. 1121–1128, Sep. 1997, doi: 10.1017/S003329179700559X.

[21] R. C. Kessler and M. Essex, “Marital Status and Depression: The Importance of Coping Resources,” *Social Forces*, vol. 61, no. 2, p. 484, Dec. 1982, doi: 10.2307/2578238.

[22] V. Patel and A. Kleinman, “Poverty and common mental disorders in developing countries,” *Bulletin of the World Health Organization*, 2003.

[23] S. Rosenfield and D. Mouzon, “Gender and Mental Health,” in *Handbook of the Sociology of Mental Health*, C. S. Aneshensel, J. C. Phelan, and A. Bierman, Eds., in Handbooks of Sociology and Social Research. , Dordrecht: Springer Netherlands, 2013, pp. 277–296. doi: 10.1007/978-94-007-4276-5_14.

[24] A. Bergsma, R. Veenhoven, M. Ten Have, and R. De Graaf, “Do They Know How Happy They Are? On the Value of Self-Rated Happiness of People With a Mental Disorder,” *J Happiness Stud*, vol. 12, no. 5, pp. 793–806, Oct. 2011, doi: 10.1007/s10902-010-9227-5.

[25] I. A. Kretchy, F. T. Owusu-Daaku, and S. A. Danquah, “Mental health in hypertension: assessing symptoms of anxiety, depression and stress on anti-hypertensive medication adherence,” *Int J Ment Health Syst*, vol. 8, no. 1, p. 25, Dec. 2014, doi: 10.1186/1752-4458-8-25.

[26] D. J. Robinson, M. Luthra, and M. Vallis, “Diabetes and Mental Health,” *Canadian Journal of Diabetes*, vol. 37, pp. S87–S92, Apr. 2013, doi: 10.1016/j.jcjd.2013.01.026.

[27] K. M. Scott *et al.*, “Mental disorders among adults with asthma: results from the World Mental Health Survey,” *General Hospital Psychiatry*, vol. 29, no. 2, pp. 123–133, Mar. 2007, doi: 10.1016/j.genhosppsych.2006.12.006.

Appendix A

Survey / Questionnaire (Google Form + PSS items)

A1. Survey Title
A survey to assess mental stress in the context of Bangladesh

A2. Participant Notice / Privacy Statement
All information provided in this survey will be used solely for research purposes, and your personal privacy will always be protected. Please respond thoughtfully based on your own experiences.

A3. Form Settings
* Indicates required question

A4. Section 1: Language Preference
Do you want to fill up the form in Bangla or English?
• Bangla
• English

A5. Section 2: Basic Information
We are collecting some basic information for our research paper.
"Please respond thoughtfully based on your experience"
Email *
• (Respondent email address)
Age *
• (Age)
Gender *
• Male
• Female
Marital Status *
• Single
• Married
• Divorced
• Widow/Widower
Average sleep per day? *
• Less than 5 hours
• 5–6 hours
• 6–7 hours
• More than 8 hours

Average daily screen time (Facebook, Instagram, TikTok, YouTube, other social media)? *

- Less than 2 hours
- 2–4 hours
- 4–6 hours
- More than 6 hours

Physical activity per week *

- None
- 1–2 hours
- 3–5 hours
- More than 5 hours

How many days per week do you engage in entertainment, meditation, or relaxation? *

- Never
- 1–2 days
- 3–4 days
- 5–6 days
- Daily

Smoking habit *

- Yes
- No
- Occasionally

Caffeine/tea/energy drinks per day *

- None
- 1 cup
- 2 cups
- 3+ cups

Any chronic health condition (diabetes, hypertension, asthma, others)? *

- Yes
- No

Average study/work hours per day? *

- 0–2 hours
- 3–5 hours
- 6–8 hours
- More than 8 hours

Perceived workload/study load *

- Very low
- Low
- Moderate
- High

• Very high

Upcoming exams/projects/deadlines? *

• Yes
• No

Satisfaction with academic/work environment *

• Very unsatisfied
• Unsatisfied
• Neutral
• Satisfied
• Very satisfied

Number of close friends/social contacts *

• None
• 1–2
• 3–5
• 6+

Monthly Family Income (in Taka) *

• Less than 15,000
• 15,001 – 30,000
• 30,001 – 50,000
• 50,001 – 80,000
• More than 80,000

Self-rated happiness level *

• Very low
• Low
• Neutral
• High
• Very high

A6. Section 3: Perceived Stress Questions (14 Items)

Instruction: In the last month, how often have you experienced the following? Response options for all items: Never / Almost Never / Sometimes / Fairly Often / Very Often.

1. In the last month, how often have you been upset because of something that happened unexpectedly? *

• Never
• Almost Never
• Sometimes
• Fairly Often
• Very Often

2. In the last month, how often have you felt that you were unable to control the important things in your life? *

• Never
• Almost Never
• Sometimes

• Fairly Often
• Very Often
3. In the last month, how often have you felt nervous and “stressed”? *
• Never
• Almost Never
• Sometimes
• Fairly Often
• Very Often
4. In the last month, how often have you dealt successfully with day to day problems and annoyances? *
• Never
• Almost Never
• Sometimes
• Fairly Often
• Very Often
5. In the last month, how often have you felt that you were effectively coping with important changes that were occurring in your life? *
• Never
• Almost Never
• Sometimes
• Fairly Often
• Very Often
6. In the last month, how often have you felt confident about your ability to handle your personal problems? *
• Never
• Almost Never
• Sometimes
• Fairly Often
• Very Often
7. In the last month, how often have you felt that things were going your way? *
• Never
• Almost Never
• Sometimes
• Fairly Often
• Very Often
8. In the last month, how often have you found that you could not cope with all the things that you had to do? *
• Never
• Almost Never
• Sometimes

• Fairly Often
• Very Often
9. In the last month, how often have you been able to control irritations in your life? *
• Never
• Almost Never
• Sometimes
• Fairly Often
• Very Often
10. In the last month, how often have you felt that you were on top of things? *
• Never
• Almost Never
• Sometimes
• Fairly Often
• Very Often
11. In the last month, how often have you been angered because of things that happened that were outside of your control? *
• Never
• Almost Never
• Sometimes
• Fairly Often
• Very Often
12. In the last month, how often have you found yourself thinking about things that you have to accomplish? *
• Never
• Almost Never
• Sometimes
• Fairly Often
• Very Often
13. In the last month, how often have you been able to control the way you spend your time? *
• Never
• Almost Never
• Sometimes
• Fairly Often
• Very Often
14. In the last month, how often have you felt difficulties were piling up so high that you could not overcome them? *
• Never
• Almost Never
• Sometimes

• Fairly Often
• Very Often

A7. Optional Comment

If you have any comments or feedback, please write them here.

APPENDIX B

Dataset Codebook (17 Features + Label/Encoding)

B1. Dataset Overview

The dataset was collected through a web-based survey in Bangladesh. It contains responses to a structured questionnaire covering demographic, lifestyle, academic/work, health, and social well-being factors. The Perceived Stress Scale (PSS-14) items were used to compute a total stress score and then generate the final categorical stress label. After label creation, PSS item columns are removed to avoid label leakage, and only the 17 non-PSS variables are used as input predictors.

B2. Target Variable (Label)

Target name: Stress_Level (derived from PSS-Total).
Classes: Low, Moderate, High.
Numeric encoding used in experiments: Low = 0, Moderate = 1, High = 2.

B3. Feature List and Codebook (17 Input Features)

Feature	Description / Options	Type	Encoding
Age	Age group of respondent	Ordinal	15–39 → 0; 40–58 → 1; 59+ → 2
Gender	Male / Female	Binary	Male → 0; Female → 1
Marital_Status	Single / Married / Divorced / Widow/Widower	Nominal	One-hot encoding (4 columns)
Sleep_Per_Day	Less than 5 hours / 5–6 hours / 6–7 hours / More than 8 hours	Ordinal	<5 → 0; 5–6 → 1; 6–7 → 2; >8 → 3
Screen_Time	Less than 2 hours / 2–4 hours / 4–6 hours / More than 6 hours	Ordinal	<2 → 0; 2–4 → 1; 4–6 → 2; >6 → 3
Physical_Activity	None / 1–2 hours / 3–5 hours / More than 5 hours (per week)	Ordinal	None → 0; 1–2 → 1; 3–5 → 2; >5 → 3
Relaxation_Days	Never / 1–2 days / 3–4 days / 5–6 days / Daily	Ordinal	Never → 0; 1–2 → 1; 3–4 → 2; 5–6 → 3; Daily → 4
Smoking_Habit	Yes / No / Occasionally	Nominal	One-hot encoding (3 columns)
Caffeine_Intake	None / 1 cup / 2 cups / 3+ cups (per day)	Ordinal	None → 0; 1 → 1; 2 → 2; 3+ → 3
Chronic_Condition	Any chronic health condition (diabetes, hypertension, asthma, others): Yes / No	Binary	No → 0; Yes → 1

Study_Work_Hours	0–2 / 3–5 / 6–8 / More than 8 hours (per day)	Ordinal	0–2 → 0; 3–5 → 1; 6–8 → 2; >8 → 3
Perceived_Workload	Very low / Low / Moderate / High / Very high	Ordinal	Very low → 0; Low → 1; Moderate → 2; High → 3; Very high → 4
Upcoming_Deadlines	Upcoming exams/projects/deadlines: Yes / No	Binary	No → 0; Yes → 1
Env_Satisfaction	Very unsatisfied / Unsatisfied / Neutral / Satisfied / Very satisfied	Ordinal	Very unsatisfied → 0; Unsatisfied → 1; Neutral → 2; Satisfied → 3; Very satisfied → 4
Social_Contacts	None / 1–2 / 3–5 / 6+	Ordinal	None → 0; 1–2 → 1; 3–5 → 2; 6+ → 3
Monthly_Income	<15,000 / 15,001–30,000 / 30,001–50,000 / 50,001–80,000 / >80,000 (Taka)	Ordinal	<15000 → 0; 15001–30000 → 1; 30001–50000 → 2; 50001–80000 → 3; >80000 → 4
Happiness_Level	Very low / Low / Neutral / High / Very high	Ordinal	Very low → 0; Low → 1; Neutral → 2; High → 3; Very high → 4

B4. General Encoding Rules Used

• Binary variables are encoded as 0/1.
• Ordinal variables are encoded using ordered integer codes (lower value indicates lower intensity/amount).
• Nominal variables are encoded using one-hot encoding to avoid imposing an artificial order.
• If any continuous numeric fields exist, they are kept numeric and scaled during preprocessing (e.g., StandardScaler).

B5. Notes on Data Cleaning (Summary)

• Records with missing/invalid values are handled during the cleaning step before model training.
• PSS_Q1–PSS_Q14 columns are used only for label creation and then removed from the predictor set to prevent leakage.
• Encoding and scaling are fit on the training set and applied to the test set to maintain a fair evaluation setting.

APPENDIX C

Data Preprocessing Details (Cleaning, Scaling, SMOTE)

C1. Overview

This appendix documents the preprocessing pipeline used to prepare the survey dataset for machine learning. The goal was to ensure a reproducible workflow and prevent information leakage, while addressing class imbalance.

C2. Data Loading and Initial Cleaning

• The dataset was stored in CSV format and loaded using Python (Pandas).
• Basic formatting checks were performed (e.g., consistent column names, removal of obvious formatting issues).

• Records were reviewed for missing/invalid values before modeling.

C3. Stress Score Computation (PSS-Total)

• Identify PSS columns: {PSS_Q1, PSS_Q2, …, PSS_Q14}.
• For each respondent, compute total perceived stress score:
PSS-Total = PSS_Q1 + PSS_Q2 + … + PSS_Q14.

C4. Target Label Creation (Stress-Level)

The continuous PSS-Total score was converted into a 3-class label using a fixed rule:
• If PSS-Total $\leq 18 \rightarrow$ Stress-Level = 0 (Low)
• If $19 \leq$ PSS-Total $\leq 37 \rightarrow$ Stress-Level = 1 (Moderate)
• If PSS-Total $\geq 38 \rightarrow$ Stress-Level = 2 (High)
This mapping converts the problem into a multi-class classification task.

C5. Preventing Leakage (Dropping PSS Items)

After generating the target label, the individual PSS question columns (PSS_Q1–PSS_Q14) were removed from the feature set. This ensures the model learns stress patterns from the 17 external factors (demographic, lifestyle, academic/work, social, and wellbeing) rather than directly from the same items used to create the label.

C6. Encoding of Categorical Features

• Binary variables were encoded as 0/1.
• Ordinal variables were encoded using ordered integer codes.
• Nominal variables (e.g., Marital Status, Smoking Habit) were one-hot encoded so no artificial order is imposed.

C7. Train–Test Split (Stratified)

The data was split into training and testing sets using a 75:25 ratio. Stratified sampling was applied so that Low, Moderate, and High stress categories were proportionally represented in both sets. This supports a fair evaluation under the naturally imbalanced distribution.

C8. Scaling / Standardization

Standardization was applied to ensure predictors are comparable in scale, particularly important for scale-sensitive models (e.g., SVM, Logistic Regression). Scaling parameters were fit on the training set and then applied to the test set to avoid data leakage.

C9. Handling Class Imbalance with SMOTE (Training Only)

Because the dataset was imbalanced (Moderate stress was the majority class), SMOTE (Synthetic Minority Over-sampling Technique) was applied only to the training data.

- SMOTE generated synthetic samples for minority classes to balance the training distribution.
- The test set was not modified to preserve a realistic evaluation setting.
- In this study, SMOTE was used with k-neighbors = 3 and a fixed random seed for reproducibility

.

C10. Final Prepared Dataset

After preprocessing, the final modeling dataset consisted of:

- X: 17 non-PSS predictors (encoded and scaled as required)
- y: Stress-Level label (0 = Low, 1 = Moderate, 2 = High)

The same preprocessing order was kept for all models to enable fair comparison.

APPENDIX D

Model Hyperparameters (LR, SVM, HGB, Voting)

D1. Summary

This appendix lists the model configurations (hyperparameters) used for the four models reported in the thesis: Logistic Regression (LR), Balanced Support Vector Machine (SVM), Histogram-based Gradient Boosting (HGB), and Voting Ensembles (Hard Voting and Soft Voting). The thesis notes the use of class weighting for the SVM (class_weight = “balanced”) and the selection of soft voting as the final ensemble. Where an exact value is not explicitly stated in the book, the default setting of the scikit-learn implementation was used.

D2. Logistic Regression (LR)

Hyperparameter	Value / Setting
Model	Logistic Regression (baseline classifier)
Class weighting	Not specified (default)
Other hyperparameters	Default scikit-learn settings (not explicitly specified in the thesis)

D3. Balanced Support Vector Machine (SVM)

Hyperparameter	Value / Setting
Model	Support Vector Machine (Balanced SVM)
class_weight	“balanced” (explicitly stated)
Kernel	Not specified in the thesis (default scikit-learn SVC)
Other hyperparameters (C, gamma, etc.)	Default scikit-learn settings (not explicitly specified in the thesis)

D4. Histogram-based Gradient Boosting (HGB)

Hyperparameter	Value / Setting
Model	Histogram-based Gradient Boosting (HGB)
Learning rate / depth / iterations	Not explicitly specified in the thesis (default scikit-learn settings)

D5. Voting Ensembles (Hard Voting and Soft Voting)

Hyperparameter	Value / Setting
Hard Voting	Majority vote of predicted class labels from {LR, SVM, HGB} (described in the thesis)

Soft Voting (final model)	Average of predicted class probabilities from {LR, SVM, HGB}; class with highest mean probability is selected (described in the thesis)
Base estimators	Logistic Regression + Balanced SVM + Histogram Gradient Boosting
Ensemble weights	Not specified (assumed equal weights)

APPENDIX E
Additional Results (Confusion Matrices and Reports)

E1. Confusion Matrices (Test Set)
Rows = Actual class, Columns = Predicted class.

Logistic Regression (LR)

	Low (0)	Moderate (1)	High (2)
Low (0)	5	7	0
Moderate (1)	0	59	5
High (2)	0	4	8

Balanced SVM

	Low (0)	Moderate (1)	High (2)
Low (0)	9	3	0
Moderate (1)	2	54	8
High (2)	1	2	9

Histogram Gradient Boosting (HGB)

	Low (0)	Moderate (1)	High (2)
Low (0)	7	5	0
Moderate (1)	1	57	6
High (2)	0	3	9

Hard Voting Ensemble

	Low (0)	Moderate (1)	High (2)
Low (0)	7	5	0
Moderate (1)	0	58	6
High (2)	0	3	9

Soft Voting Ensemble (Selected Model)

	Low (0)	Moderate (1)	High (2)
Low (0)	7	5	0
Moderate (1)	0	59	5
High (2)	0	3	9

E2. Summary of Reported Test Results (from the thesis)

Model	Accuracy	Macro F1	Micro-average AUC	Notes / Class-wise AUC
Logistic Regression (LR)	0.8182	0.70	0.931	AUC0=0.886, AUC1=0.836, AUC2=0.853
Balanced SVM	0.8182	0.75	0.938	AUC0=0.917, AUC1=0.865, AUC2=0.863
HGB	0.8295	≈0.75 (reported)	0.907	AUC0=0.840, AUC1=0.798, AUC2=0.815
Hard Voting	0.8409	—	—	Confusion matrix reported
Soft Voting (final)	0.8523	0.7766	—	Selected final model

E3. Notes

• The Moderate class is the majority class in the dataset; therefore Macro-F1 is emphasized as a balanced metric.

• The soft voting ensemble was selected as the final model due to improved overall accuracy and better class balance (Macro-F1).

APPENDIX F

Feature Importance Output (Permutation Importance: Full Table)

F1. Method (as described in the thesis)

• Permutation importance was used to explain how each feature contributes to model performance.

• For each feature, values are randomly shuffled and the decrease in model performance is measured.

• In this thesis, importance was computed as the loss in overall Macro F1-score when each feature is permuted.

• Permutation importance was computed for three single models: Logistic Regression (LR), Balanced SVM, and Histogram Gradient Boosting (HGB).

F2. Interpretation Guide

• Higher importance score (larger Macro-F1 decrease) ⇒ the model relies more on that feature.

• Near-zero importance ⇒ the feature contributes little to the prediction.

• The thesis reports a consistent pattern across models, with Age as the most predictive factor in the permutation plots.

F3. Figure Reference

Figure 4.3 (Permutation Importance of HGB, LR and SVM) is referenced in the thesis as the visual summary of these importance outputs.

APPENDIX G

Source Code / Notebook (Link + Key Snippets)

G1. Repository Link

GitHub Repository: https://github.com/dasusmay/Mental_Stress_Detection.git

G2. Repository Structure (Main Files)

The repository contains the following main files:

• Mental_Stress.ipynb — Jupyter Notebook containing the full analysis pipeline
• Data_CSV.csv — Dataset used in the notebook (survey responses)
• Mental_Stress_Book.docx — Project book/report document
• README.md — Project overview and usage notes
• LICENSE (MIT) — License information

G3. Key Code Snippets (from the Notebook)

The following snippets summarize the core steps used in the project workflow.

G3.1 Load data and inspect columns

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

df = pd.read_csv("Data_022.csv")
print(df.columns.tolist())
```

G3.2 Compute PSS total score

```
pss_columns = df.filter(like='PSS_')
df['PSS_Total'] = pss_columns.sum(axis=1)
df[['PSS_Total']].head()
```

G3.3 Create 3-class stress label

```
def categorize_stress(score):
    if score <= 18:
        return 0  # Low Stress
    elif score <= 37:
        return 1  # Moderate Stress
    else:
        return 2  # High Stress
```

```
df['Stress_Level'] = df['PSS_Total'].apply(categorize_stress)
df[['PSS_Total', 'Stress_Level']].head()
```

G3.4 Remove PSS columns to prevent leakage

```
df.columns = df.columns.str.strip()  # delete extra space
df = df.drop(columns=[col for col in df.columns if col.startswith('PSS_')])
```

G3.5 Train-test split (stratified)

```
X = df.drop(columns=['Stress_Level'])
y = df['Stress_Level']

from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(
    X, y,
    test_size=0.25,
    random_state=40,
    stratify=y
)
```

G3.6 SMOTE oversampling (train only)

```
from imblearn.over_sampling import SMOTE
smote = SMOTE(k_neighbors=3, random_state=42)
X_train_smote, y_train_smote = smote.fit_resample(X_train, y_train)
```

G3.7 Standard scaling (for SVM/LR pipelines)

```
from sklearn.preprocessing import StandardScaler
scaler = StandardScaler()
X_train_scaled = scaler.fit_transform(X_train)
X_test_scaled = scaler.transform(X_test)
```

G3.8 Model objects (as shown in notebook outputs)

```
from sklearn.linear_model import LogisticRegression
lr = LogisticRegression(max_iter=2000, random_state=40)

from sklearn.svm import SVC
svm = SVC(class_weight='balanced', probability=True, random_state=42)

from sklearn.ensemble import HistGradientBoostingClassifier
hgb = HistGradientBoostingClassifier(max_depth=6, max_iter=300, random_state=42)
```

G4. Notes

• The notebook computes the stress label from PSS-14 responses and then removes PSS columns before model training to avoid leakage.
• A stratified train-test split and SMOTE are used to address the imbalanced class distribution.
• Models reported include Logistic Regression, Balanced SVM, Histogram Gradient Boosting, and voting ensembles (described in the thesis).