Independent University **Bangladesh (IUB)**

IUB Academic Repository

IUB Library Thesis

2025-01

Towards Development of an Empathetic AI Companion: Using Multi-Modal Passive Sensing Data to Predict and Understand Student Mental Stress

Halder, Nabarun

IUB

https://ar.iub.edu.bd/handle/11348/1262

Downloaded from IUB Academic Repository



Towards Development of an Empathetic AI Companion: Using Multi-Modal Passive Sensing Data to Predict and Understand Student Mental Stress

By

Nabarun Halder

Student ID: **2312720**

Summer, 2025

Supervisor:

Ashraful Islam, PhD

Assistant Professor

Department of Computer Science & Engineering

Co-Supervisor:

M. Ashraful Amin, PhD

Professor

Department of Computer Science & Engineering

September 1, 2025

Dissertation submitted in partial fulfillment for the degree of Master of Science in Computer Science

Department of Computer Science & Engineering

Independent University, Bangladesh

Attestation

I declare that this thesis, titled **"Towards Development of an Empathetic AI Companion: Using Multi-Modal Passive Sensing Data to Predict and Understand Student Mental Stress"**, is the result of my own original work and research. I completed this work as part of the requirements for the degree of Master of Science in Computer Science and Engineering at Independent University, Bangladesh (IUB).

I also confirm that this work, or any part of it, has not been submitted before for any other degree or qualification at any other university or institution. All of the sources and references that I used in writing this thesis have been properly acknowledged.

I understand that academic dishonesty is a serious offense. I accept that if I have violated the university's rules, I will be subject to disciplinary action.

Signature	Date

Nabarun Halder

Name

Acknowledgement

I want to show my heartfelt gratitude to my supervisor. Without his guidance and feedback, I would not have been able to complete this study. He is the person who motivated me and kept motivating me to finish this work. On the days I felt I was falling behind and could not believe in myself, he would ask about my progress again and again.

Not just for this thesis, he was always there from the beginning of my master's journey. He even motivated me to do the master's in the first place. He is like a mentor to me; his presence in my life is invaluable. Thanking him is never enough. He brought me to the light I always hoped for.

A very special thanks goes to all the participants, who patiently used our passive sensing app and allowed us to use their data. Without their contribution, this project would have failed at the beginning.

Most importantly, I am grateful to my family. Without their support, I would not have been able to come this far. My study is about stress, but at the times when I felt stressed, they were always there for me.

I would also like to acknowledge the use of AI to refine the tone and grammar of this writing, and I carefully reviewed whether the meaning remains the same.

Letter of Transmittal

Dr. Ashraful Islam
Assistant Professor
Department of Computer Science and Engineering
Independent University, Bangladesh

Dear Dr. Ashraful Islam,

I am pleased to submit my thesis titled **"Towards Development of an Empathetic AI Companion: Using Multi-Modal Passive Sensing Data to Predict and Understand Student Mental Stress"**, prepared in partial fulfillment of the requirements for the degree of *Master of Science in Computer Science* at Independent University, Bangladesh.

This thesis is the result of my own research, which I conducted under your supervision. My work involved developing a smartphone application to understand student well-being. The app uses passive sensing to collect data on daily behaviors and an AI companion to provide personalized support. The study shows a strong link between this passively collected data and a student's self-reported stress levels.

I would like to express my sincere gratitude for your invaluable guidance, feedback, and encouragement throughout the duration of this project.

I respectfully submit this thesis for your review and hope it fulfills the academic standards and expectations of the department.

Sincerely,

Nabarun Halder
ID: 2312720
Department of Computer Science and Engineering
Independent University, Bangladesh

Evaluation Committee

Supervision Panel

.....	.....
Supervisor (Dr. Ashraful Islam)	Co-supervisor (Prof. M. Ashraful Amin)

Examiners

.....	.....
Dr. Saadia Binte Alam (Internal)	Dr. AKM Mahbubur Rahman (Internal)

.....
External Examiner (Dr. Md. Mosaddek Khan, University of Dhaka)

Office Use

.....	.....
Program Coordinator	Head of the Department

.....
Director, Graduate Studies, Research and Industry Relations

.....
Library

Abstract

Smartphones offer a unique opportunity to understand human well-being in daily life. This thesis presents a comprehensive study conducted with university students in Bangladesh to investigate the link between passively sensed daily behaviors and mental health. We developed a multi-modal Android application that collects data in two ways: passively (continuously gathering data from accelerometer, gyroscope, light, and GPS sensors, alongside phone usage statistics) and actively (prompting users with the weekly DASS-21 mental health survey and a daily stress diary). A total of 88 undergraduate students (47 male, 41 female) from Independent University, Bangladesh (IUB), participated in the data collection study. The participants were from various academic years and schools, enrolled in CSE100 and CIS101 courses.

Using the collected passive data, we engineered features representing physical activity, mobility, and digital behavior. We then developed machine learning models to predict self-reported stress. For daily stress prediction, our models achieved an accuracy of 74% and F1 macro score of 0.67 in a 3-class classification using a Random Forest model. For a 5-class classification, the best model achieved 59% accuracy and 0.49 F1 macro score. For weekly stress prediction, an LSTM model achieved 71% accuracy and 0.57 F1 macro score in the 3-class setup, and 46% accuracy with 0.41 F1 macro score for the 5-class version. These results demonstrate a meaningful connection between passively measured behaviors and students' subjective stress levels.

Based on these results, we designed and built an AI Wellbeing Companion (WellBeing Mate). This chatbot uses passive data and stress predictions to provide personalized, empathetic conversations. To evaluate this feature, we conducted a separate usability study with 27 students. The AI Companion received an excellent System Usability Scale (SUS) score of 87.31, with participants highlighting its ability to remember past conversations and its supportive tone. This work presents a successful end-to-end system, ranging from real-world, longitudinal data collection to a well-accepted, context-aware AI Assistant, offering a new model for digital phenotyping and mental health support.

Contents

Attestation i

Acknowledgement ii

Letter of Transmittal iv

Evaluation Committee i

1 Introduction 1

1.1 Mental Stress . . . 1
1.2 Navigating Stress and Its Associated Challenges . . . 1
1.3 Student Stress . . . 2
1.4 Passive Sensing . . . 3
1.5 Smartphone-Based Passive Sensing: How Technology Gathers Data Without You Doing Anything . . . 3
1.6 Mobile Apps for Stress Management . . . 4
1.7 Objectives and Research Questions . . . 5
1.7.1 Objectives . . . 5
1.7.2 Research Questions and Rationale . . . 6
1.8 Contributions of the Thesis . . . 7
1.9 Thesis Project Management . . . 8
1.9.1 WBS . . . 8
1.9.2 Thesis Workflow: A Seven-Phase Approach from Concept to Completion . . . 8

2 Literature Review 11

2.1 Literature Review of Papers . . . 11
2.1.1 Introduction . . . 11
2.1.2 Methods . . . 11
2.1.3 Results . . . 13
2.1.4 Discussion . . . 16
2.1.5 Research Gaps . . . 17

3 Research Methodology **18**
3.0.1 Study Design and Participants . . . 18
3.0.2 Study Procedure . . . 19
3.0.3 Data Collection Methods . . . 19
3.0.4 Data Processing and Feature Preparation . . . 29
3.0.5 Modeling Approach . . . 31
3.0.6 Model Evaluation and Validation . . . 32
3.1 The AI Well-being Companion . . . 34
3.1.1 Purpose of the Companion . . . 34
3.1.2 How the Companion Works: A Smart and Personal Approach . . . 34
3.1.3 An Example Conversation Flow . . . 36
3.1.4 Why This Approach is Important . . . 37

4 Results and Analysis **38**
4.1 Key Patterns in Student Data . . . 38
4.2 Results: Predicting Daily Student Stress . . . 40
4.2.1 Experiment 1: Predicting Five Levels of Stress . . . 41
4.2.2 Experiment 2: Predicting Three Levels of Stress . . . 41
4.2.3 Summary of Daily Prediction Findings . . . 42
4.3 Results: Predicting Weekly Stress Severity . . . 42
4.3.1 Training and Evaluation Method . . . 42
4.3.2 Experiment 1: Predicting Five Levels of Weekly Stress . . . 47
4.3.3 Experiment 2: Predicting Three Levels of Weekly Stress . . . 47
4.3.4 Summary of Weekly Prediction Findings . . . 48
4.4 Usability Evaluation of the AI Companion (WellBeing Mate) . . . 48
4.4.1 Participants and Method . . . 48
4.4.2 System Usability Scale (SUS) Score . . . 50
4.4.3 Key Qualitative Feedback from Participants . . . 50

5 Discussion **53**
5.1 Summary of Key Findings . . . 53
5.2 Principal Implications . . . 53
5.2.1 Smartphones are Viable Tools for Mental Health Monitoring . . . 54
5.2.2 Personalization is Key for AI Companions . . . 54
5.2.3 A Model for Future Digital Health Apps . . . 54
5.3 Broader Impacts and Ethical Considerations . . . 54
5.3.1 Social Effect . . . 54
5.3.2 Environmental Impact and Sustainability . . . 55
5.3.3 Addressing Ethical Issues . . . 55

5.4 Limitations and Challenges 55
5.4.1 Data from a Specific Population 55
5.4.2 Technical Challenges of Real-World Sensing 56
5.4.3 Privacy and User Trust 56
5.5 Threats to Validity 56
5.5.1 Self-Selection Bias 56
5.5.2 The Hawthorne Effect (Changing Behavior) 56
5.5.3 Self-Report Bias 56

6 Conclusion and Future Work **58**
6.1 Conclusion 58
6.2 Future Work 59
6.2.1 Moving from Prediction to Intervention 59
6.2.2 Enhancing the AI Companion 59
6.2.3 Expanding Data Sources and Populations 59
6.2.4 Long-Term Impact Studies 59

Appendix **60**
Publications From This Research & Relevant Work 60

Bibliography **72**

List of Figures

1.1 Thesis Work Breakdown Structure (WBS) 9

2.1 PRISMA flow diagram of the literature search and selection process. . . . 12

3.1 Screenshot of the 'WellBeing' App Main Dashboard, showing cards for different data collection modules. 23
3.2 Screenshot of the 'WellBeing' App Usage Stats Screen, showing screen time, unlock count, and top apps. 25
3.3 Screenshot of the 'WellBeing' App DASS-21 survey interface, showing one question per page. 26
3.4 Screenshot of the 'WellBeing' App DASS-21 history chart, showing a line graph for Depression, Anxiety, and Stress scores over time. 27
3.5 Screenshot of the 'WellBeing' App Daily Diary (EMA) interface, showing one of the steps like the stressor checklist. 28
3.6 A screenshot of the AI Companion's chat interface. It shows a friendly greeting from WellBeing Mate and a simple chat window. 35

4.1 Average Student Stress Over Spring 2025 Semester 38
4.2 Sleep and Study Duration on Weekdays vs. Weekends 39
4.3 Proportion of Primary Stressors: Exam vs. Non-Exam Weeks 40
4.4 Confusion Matrices for the **Logistic Regression** Model. 43
4.5 Confusion Matrices for the **Random Forest** Model. 44
4.6 Confusion Matrices for the **Support Vector Machine (SVM)** Model. . 45
4.7 Confusion Matrices for the **MLP** Model. 46
4.8 Confusion Matrices for the **Advanced LSTM Model** on the Weekly Classification Tasks. The model's performance is visibly stronger on the simpler 3-class task (b). 49
4.9 Usability Score of Our AI Companion. 51
4.10 Participant voting on 5-point Likert scales for individual SUS questions. . 52

List of Tables

2.1 Comparison of key studies on passive smartphone sensing for stress and mental health identification in university students. 14

3.1 The Depression, Anxiety, and Stress Scale-21 (DASS-21) Items and Subscales. The response scale for each item is 0 (*Did not apply at all*), 1 (*Applied to some degree*), 2 (*Applied a considerable degree*), or 3 (*Applied very much*). The final score for each subscale is the sum of its item scores multiplied by two. 20
3.2 Structure of the Daily Diary (Ecological Momentary Assessment, EMA) . 21

4.1 Performance of Models for 5-Class Daily Stress Prediction 41
4.2 Performance of Models for 3-Class Daily Stress Prediction 42
4.3 Performance of the LSTM Model for 5-Class Weekly Stress Prediction . . 47
4.4 Performance of the LSTM Model for 3-Class Weekly Stress Prediction . . 47
4.5 The System Usability Scale (SUS) Questionnaire Items. 50

Chapter 1

Introduction

1.1 Mental Stress

Mental stress is conceptualized as a multifaceted physiological and psychological response that emerges when environmental demands exceed an individual's adaptive capacity [1]. While brief exposure to mild stress can be normal and may even enhance performance [2], persistent or chronic stress is associated with adverse outcomes such as depression, burnout, and various diseases [1, 3, 4, 5]. Early adulthood, a period marked by rapid physiological and psychological transitions, is particularly vulnerable to these effects.

Today, stress has emerged as a major public health concern, with stress-related healthcare costs exceeding $300 million annually [6]. Surveys further indicate that more than half of Americans report higher stress levels compared to previous years [7]. In essence, stress can be described as the body's generalized, non-specific response to any factor that overwhelms or threatens to overwhelm its regulatory mechanisms aimed at maintaining homeostasis.

More broadly, stress encompasses negative experiences related to demand, threat, or harm. Such experiences trigger biological and behavioral responses intended to either modify the situation or facilitate adaptation to the stressor [8]. Beyond impairing academic performance, stress is closely linked to a range of detrimental emotional states and health conditions, including depression, obesity, and cardiovascular disease [9, 10, 11].

1.2 Navigating Stress and Its Associated Challenges

Stress is a universal experience, particularly among students [12, 13] and workers in developed capitalist societies [14, 15, 16, 17], where constant pressure for rationality and efficiency is the norm. According to the World Health Organization (WHO), over 90% of suicides are linked to mental disorders [18]. Both the WHO and the World Bank

recognize mental health as not only a global health crisis but also a significant economic challenge [16]. In Japan, worker mental health has received growing attention, prompting government initiatives such as the nationwide self-reporting stress check program [19]. In Bangladesh [13], many students face high levels of stress due to academic pressure, financial worries, and social expectations. These challenges can affect their mental health, leading to anxiety, sadness, and trouble concentrating. Research shows that students often struggle to balance studies with personal life, and many do not have enough support to manage stress well.

Mental disorders are diverse, but anxiety is a core symptom that often emerges in the early stages of illness [20]. Importantly, anxiety has been identified as a risk factor for later development of psychosis [21, 22, 23], underscoring the need for early detection of subclinical anxiety. Traditional approaches rely on self-assessment, yet frequent monitoring is difficult to sustain due to the time and effort required.

To address these limitations, several physiological methods for stress detection have been proposed. Research suggests that indicators such as blood pressure, salivary amylase and cortisol, as well as heart rate measures like RRI, can reliably estimate anxiety levels [24, 25]. However, while physiological indices provide robust measurements, daily application remains impractical for routine use.

Individuals experiencing stress, anxiety, or mild depression may develop chronic mental health symptoms that affect mobility, life satisfaction, and social interactions [26, 27]. If these symptoms are not detected early, they can worsen, leading to greater impacts and an increased need for medication or hospitalization [28, 29, 30]. This highlights mild mental health symptoms as an important target for digital phenotyping, which aims to enable early detection and timely intervention.

1.3 Student Stress

Early adulthood represents a developmental period marked by pronounced physiological and psychological transitions. This stage is further compounded by novel challenges such as relocation to unfamiliar environments, stringent academic or occupational deadlines, heightened social demands, and the need to balance competing responsibilities across education, employment, and personal life. Collectively, these pressures render university students particularly susceptible to stress [31, 32, 33]. Empirical evidence consistently demonstrates that this population is at elevated risk of stress-induced psychological distress, manifesting in diminished academic performance, strained interpersonal relationships, depression, burnout, self-injurious behaviors, and, in severe cases, suicide [34, 35].

Academics are a major source of stress for students. In one study, class attendance and homework were identified as leading stressors among college students [36]. Research also

indicates that students experience stress related to performance pressure and concerns about postgraduate planning [37]. Beyond academics, other stressors exist in the college environment. Interpersonal challenges, such as forming friendships, fitting in socially, and managing roommate relationships, have been reported as significant sources of stress [38]. Additionally, environmental factors, like financial difficulties and changes in eating or sleeping patterns, further contribute to student stress levels [38, 39].

1.4 Passive Sensing

Passive sensing is a method of data collection that uses sensors to gather information about an environment or a user without requiring their active participation [40]. Unlike active sensing, which involves the user directly interacting with a device (like tapping a button or speaking a command), passive sensing operates in the background, often unnoticed. This technology leverages a variety of sensors already present in common devices, such as smartphones, smartwatches, and smart home devices. These sensors continuously collect data streams, including: Location data (GPS, Wi-Fi, Bluetooth) to track movement and location. Acoustic data (microphones) to detect sounds like speech, laughter, or the presence of other people. Motion data (accelerometers, gyroscopes) to identify activities such as walking, running, or sitting. Environmental data (light sensors, thermometers) to monitor ambient conditions. Physiological data (heart rate monitors, electrodermal activity sensors) to measure biological responses related to stress or emotion [41]. By analyzing patterns and correlations within this data, systems can infer a great deal about a person's behavior, routine, and state of mind. For example, a sudden drop in physical activity combined with a change in social interactions detected through a phone's sensors might indicate a shift in a user's mood or well-being [42].

1.5 Smartphone-Based Passive Sensing: How Technology Gathers Data Without You Doing Anything

Passive sensing technologies gather data by detecting various phenomena, such as environmental light or user movement. These methods have been widely used in research on physical activity, sleep, and cardiovascular health, employing devices such as polysomnography systems, implantable cardiovascular monitors, pedometers, commercial wristband trackers, smartwatches, and smartphones [43, 44, 45, 46]. Consumer-grade mobile health technologies that collect data passively have been highlighted in medical literature for their potential to reduce user burden and enhance healthcare delivery [47]. In particular,

smartphones have been increasingly considered as a platform for passive sensing [48].

Smartphones stand out among passive sensing devices because of their widespread use and increasing computational capabilities [48]. By 2018, smartphone ownership in the United States had reached 94% among adults aged 18–29 and 89% among those aged 30–49 [49]. Due to their ubiquity, smartphones provide a less intrusive option for sensing compared to specialized wearable devices. Furthermore, modern smartphones are equipped with multiple sensors that can capture diverse data and combine them into complex, meaningful features. They also record behavioral information, including call logs, text messaging, and app usage patterns [50].

A wide range of self-help mobile apps have been developed to support university students facing mental health challenges [51]. These apps allow distressed students to quickly access affordable or free tools that provide coping strategies without requiring traditional face-to-face therapy sessions. Since smartphones are widely used among young people, they offer a convenient way to deliver digital mental health services [52]. Younger adults, in particular, have shown strong interest in such apps, with demand increasing further during the COVID-19 pandemic [53, 54, 55]. Students are especially drawn to these applications because they are readily available, convenient, confidential, and help reduce the stigma often associated with in-person mental health care [52, 56]. Moreover, academic stress and the challenges of transitioning into higher education encourage many students to use mental health apps as a coping resource [57].

An additional advantage of mental health smartphone apps is their ability to collect large amounts of user data. Modern smartphones are equipped with passive sensors that can monitor aspects of daily life. For example, GPS can track location, accelerometers and gyroscopes can measure movement patterns, and step counters can estimate physical activity. Growing interest has been placed on using such passive sensing data to identify markers that may predict mental health status and outcomes.

1.6 Mobile Apps for Stress Management

Mobile apps for stress management are widely available, but only a small fraction are supported by scientific evidence or rigorous evaluation. Systematic reviews show that while thousands of apps exist, less than 10% have any research backing, and only about 2% have published, peer-reviewed evidence of efficacy or feasibility [58, 59]. The most common evidence-based strategies in these apps include mindfulness, meditation, breathing exercises, and psychoeducation, with mindfulness-based apps like Headspace having the most research support [58, 60, 59]. Meta-analyses and randomized controlled trials indicate that these apps can produce small but significant reductions in self-reported and physiological stress, as well as improvements in well-being and coping self-efficacy, especially when used regularly over several weeks [61, 62, 63, 64, 65]. However, the

overall quality of information, data privacy, and long-term availability of these apps remain moderate to low, and many apps lack transparency in development or evidence-based content [60, 59, 63]. Despite these limitations, mobile stress management apps show promise as scalable tools for supporting mental health, particularly when they incorporate a variety of evidence-based techniques and are integrated into broader wellness strategies [61, 62, 65].

1.7 Objectives and Research Questions

University students often face significant challenges, and their mental well-being, particularly stress, is a growing concern. In this research, we aimed to explore how everyday smartphone technology can be used in a new way to both understand and support the mental health of these students. Our vision was to create a complete system: starting from carefully gathering different types of information from smartphones, then using this information to build a smart and helpful tool for students themselves. To achieve this broad goal, our study was guided by a set of clear objectives and specific research questions.

1.7.1 Objectives

1. **To Develop a Reliable Data Collection Tool:** Our first objective was to design and build a single, reliable mobile app that can collect multiple types of data. This includes passive data from sensors (like movement and phone usage) and active data from user self-reports. The active data included two specific types of surveys: the Depression, Anxiety, and Stress Scale-21 (DASS-21) [66], and a Daily Diary, also known as Ecological Momentary Assessment (EMA). This data collection tool needed to be stable enough to run consistently in the background on a user's phone for an entire semester.

2. **To Explore the Link Between Behavior and Stress:** Our second objective was to analyze the data collected by the app. We wanted to use advanced machine learning techniques to see if we could find a clear connection between a student's daily behaviors (such as how much they moved, how long they slept, or how they used their phone) and their self-reported feelings of stress.

3. **To Create a User-Centered AI Application:** Our third and most important objective was to use our findings to build a helpful feature for the user. We aimed to create an AI-powered chatbot, named WellBeing Mate, that could use the patterns from the passive data to offer personalized and empathetic support, moving beyond just collecting data to create a truly interactive tool for well-being.

1.7.2 Research Questions and Rationale

To achieve these objectives, we asked the following research questions (RQs):

1. **RQ1: Can a single, reliable mobile app be developed to effectively collect both passive sensor data and active self-reported well-being information from university students over a full academic semester?**
   Rationale: Developing a stable and efficient data collection tool is the very first and most critical step for any study using smartphone sensing. If the app does not work well (for example, if it crashes often, uses too much battery, or fails to collect data consistently), then all other parts of the research become impossible. This question addresses the fundamental technical success and dependability of our app, ensuring that we can gather the high-quality, continuous data needed for our later analysis. It tests whether our technical design allows for real-world, long-term data collection without significant problems for the student participants.

2. **RQ2: Can machine learning models, trained on passively collected smartphone behavioral data (such as activity levels, sleep patterns, and phone usage), accurately predict a student's self-reported daily and weekly stress levels?**
   Rationale: This question is at the heart of understanding whether passive smartphone data holds real value for insights into mental health. It goes beyond simply *collecting* data to asking if this data can actually *tell us something important* about a student's internal state. If we can find strong and reliable connections between objective behaviors (like how much a student moves or uses their phone) and their subjective feelings of stress, it opens the door for early detection and personalized support without needing constant self-reporting. Successfully answering this question helps validate the use of passive sensing as a powerful tool for monitoring well-being.

3. **RQ3: How can an AI-powered conversational companion utilize both long-term memory of past interactions and real-time insights from passive smartphone data to offer personalized, empathetic, and context-aware support for student well-being?**
   Rationale: This question focuses on the innovative design of our AI companion, WellBeing Mate. Standard chatbots often give general or unhelpful responses. We wanted to see if we could create an AI that feels more like a supportive friend, one that remembers your past conversations and subtly understands how you might be feeling based on your phone data (even if you do not explicitly say it). The goal is to move beyond simple information delivery to create a truly personalized and engaging interaction. This is crucial for making the AI companion feel helpful and

not just a generic computer program, which in turn encourages students to use it regularly for self-reflection and support.

4. **RQ4: What are university students' perceptions regarding the usability, helpfulness, and overall acceptance of a smartphone-based AI Well-being Companion that is informed by their passive behavioral data?**
   Rationale: Even the most technically advanced system is only valuable if people find it easy to use and genuinely helpful in their daily lives. This question directly addresses the user's experience. It explores whether students feel comfortable interacting with WellBeing Mate, if they understand how it works, and most importantly, if they perceive it as a beneficial tool for managing their stress and improving their well-being. Understanding user perception is critical for ensuring that such digital health tools are actually used and have a real-world positive impact, guiding future improvements and design choices.

1.8 Contributions of the Thesis

This thesis makes several important contributions to the field of mobile health and well-being. The main contributions of this work are presented below:

1. **A Comprehensive, All-in-One Sensing Platform:**
   This research presents the design and development of a single, robust mobile app that can be used for complex, real-world studies. Before this work, a researcher might have needed separate tools to collect sensor data, deliver surveys, and interact with users. Our application combines all of these functions into one easy-to-use platform. It is designed to run 24/7 and collect many different types of data (passive and active) simultaneously. This platform can be used by other researchers in the future to conduct their own studies.

2. **New Evidence Linking Behavior to Stress in a Specific Context:**
   This study provides new and important evidence that passively collected data from a smartphone is connected to the mental well-being of university students in Bangladesh. While other studies have been done in North America or Europe, our work is one of the first to explore these connections in this specific cultural and academic environment. Our findings show that it is possible to use this data to build machine learning models that can predict a student's stress levels with high accuracy.

3. **A Novel Design for a Data-Informed AI Companion:**
   This thesis introduces a new design for AI chatbots in mental health. We present an AI Well-being Companion named *WellBeing Mate* that uses the passive data

collected from the user's phone to make its conversations more personal and empathetic. This is a significant step beyond simple chatbots that only respond to text. WellBeing Mate's ability to understand the context of a user's day (for example, by knowing their predicted stress level) allows it to provide a more supportive and meaningful interaction. The successful usability evaluation of WellBeing Mate proves that users find this data-informed approach to be helpful and engaging.

4. **An End-to-End Research Framework:**
   Finally, this work provides a complete "end-to-end" model for research in this area. We show the entire process from the very beginning to the very end: from the first idea and building the tool, to collecting and analyzing the data in the real world, and finally to using those findings to create a helpful application for the user. This thesis serves as a complete guide for other researchers who want to conduct similar projects in the future.

1.9 Thesis Project Management

This document outlines a comprehensive project management plan for a thesis, designed to guide the research and development process from initial concept to final submission. The plan uses a Work Breakdown Structure (WBS) to organize the project into a logical and manageable workflow.

1.9.1 WBS

This thesis was divided into 7 phases, and the usage of WBS ensured the completion of all phases systematically. Figure 1.1 shows the workflow of the study.

1.9.2 Thesis Workflow: A Seven-Phase Approach from Concept to Completion

Phase 1: Foundation & Literature Review

- Problem Definition & Research Questions
- Comprehensive Literature Review
- Study Design & Ethical Approval

Phase 2: System Development & Implementation

- App Architecture Design

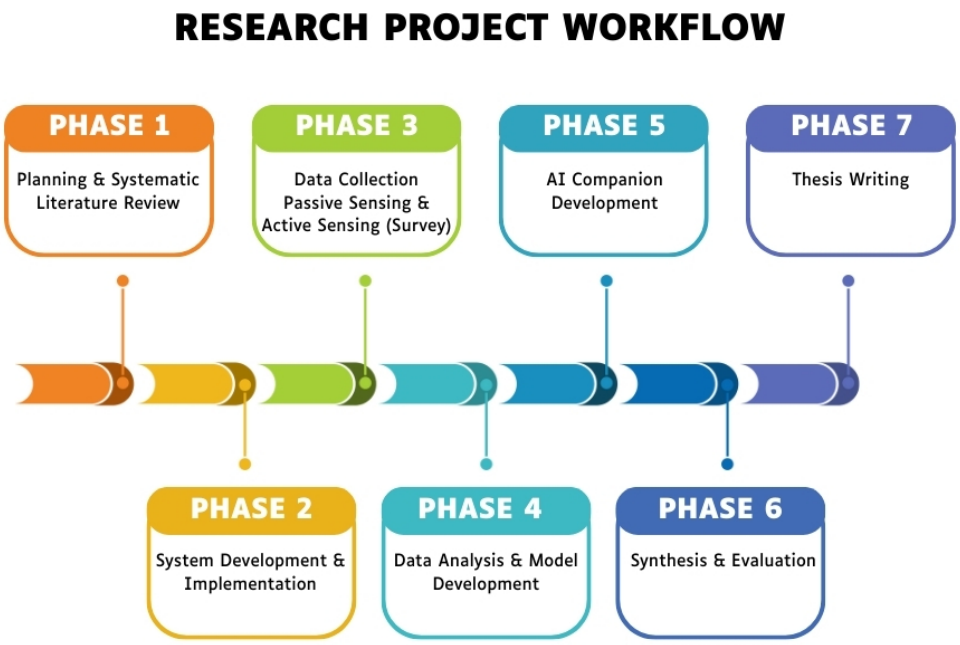


Figure 1.1: Thesis Work Breakdown Structure (WBS)

- Passive Sensing Module Development
- Active Sensing Module Development

Phase 3: Data Collection & Field Study

- Participant Recruitment & Onboarding
- Longitudinal Data Collection
- Data Export & Anonymization

Phase 4: Data Analysis & Model Development

- Data Preprocessing & Cleaning
- Feature Engineering
- Predictive Modeling

Phase 5: AI Companion Development

- AI System Design

- Integration of Gemini API
- Data-Informed Personalization

Phase 6: Synthesis & Evaluation

- AI Companion Usability Study
- Analysis of Evaluation Data
- Discussion of Overall Findings

Phase 7: Thesis Writing

- Chapter Drafting
- Creation of Visuals
- Review, Revision & Submission

Chapter 2

Literature Review

2.1 Literature Review of Papers

2.1.1 Introduction

The identification of stress in university and college students using passive smartphone sensing has become a rapidly expanding field, utilizing the ubiquity of smartphones and their embedded sensors to unobtrusively monitor behavioral and physiological markers of stress. Recent studies demonstrate that features such as GPS-derived mobility, accelerometer-based activity, screen usage, call logs, and app usage patterns can be effectively combined with machine learning models to predict stress and related mental health outcomes in student populations [67, 68, 69, 70, 71, 72, 73, 74, 40, 75]. Systematic reviews and meta-analyses confirm the feasibility and moderate-to-high accuracy of these approaches, especially when continuous data collection is paired with periodic self-report surveys [68, 72, 76]. Methodological advances, including transfer learning, deep learning, and personalized modeling, have further improved prediction accuracy and generalizability across diverse student cohorts and study durations [77, 75, 78]. However, challenges remain regarding data integrity, battery consumption, sample size, and the need for standardized methodologies [67, 70, 72]. This review synthesizes findings from recent empirical studies, systematic reviews, and meta-analyses, focusing on passive sensing modalities, modeling approaches, real-world longitudinal studies in student populations, and key methodological considerations.

2.1.2 Methods

This systematic literature review was conducted and reported in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 guidelines. Figure 2.1 is the PRISMA flow diagram of the literature search and selection process.

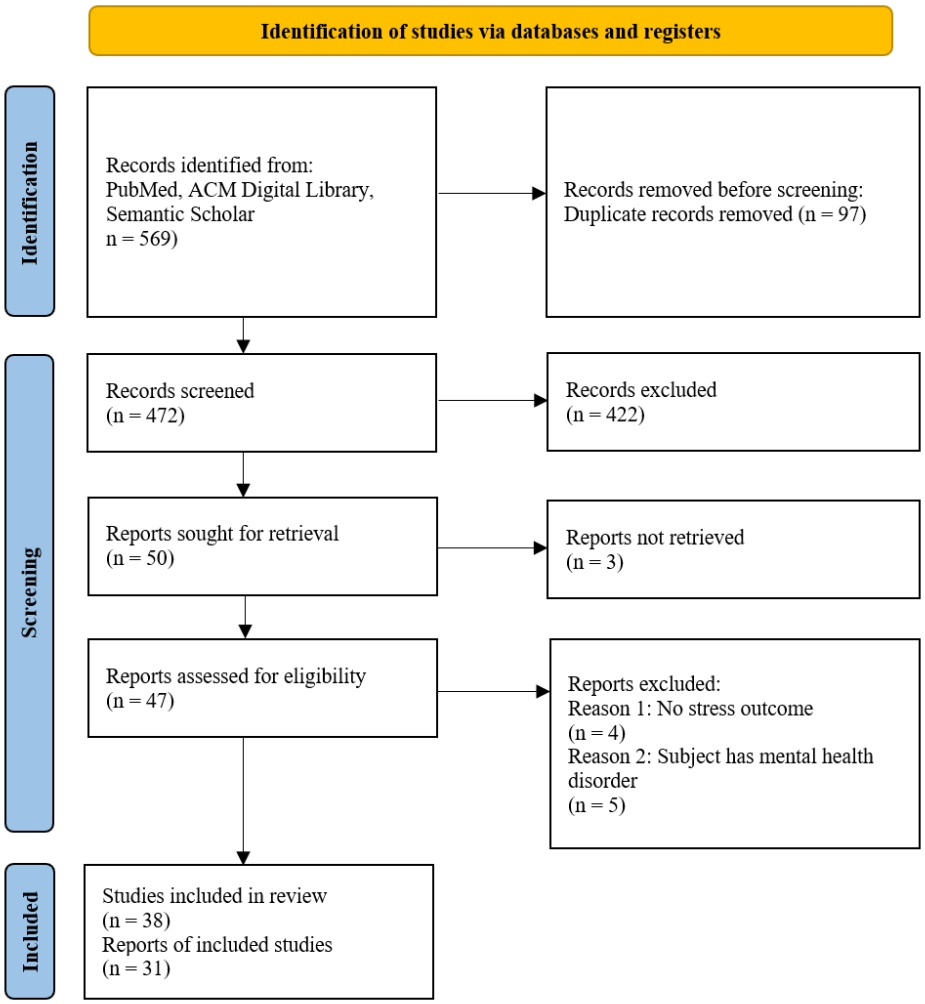


Figure 2.1: PRISMA flow diagram of the literature search and selection process.

Eligibility Criteria

Studies were included in this review if they met the following criteria: (1) the study population consisted of university or college students; (2) the study utilized passive smartphone sensing or digital phenotyping to collect behavioral data; (3) the study's primary outcome was related to the measurement, prediction, or understanding of stress; (4) the study employed machine learning or deep learning models for analysis; and (5) the article was published in English. We excluded conference abstracts and studies where stress was not a primary measured outcome.

Information Sources and Search Strategy

A comprehensive search was conducted across 2 electronic databases: PubMed and ACM Digital Library, and also searched for literature on Semantic Scholar. The search was performed to include all articles published between 2010 and 2025. To ensure a comprehensive search, the following search string was used across all databases:

```
("(digital phenotyping" OR "passive sensing") AND (stress) AND (students)
AND ("machine learning" OR "deep learning"))
```

Data Extraction

From each included study, I extracted key information, including study design, participant demographics, duration of data collection, types of passive data collected, self-report measures used, machine learning models applied, and primary findings related to stress prediction.

2.1.3 Results

Passive Sensing Modalities and Features

Studies consistently utilize a range of smartphone sensors, including GPS, accelerometer, gyroscope, microphone, screen usage, call logs, and Wi-Fi to capture behavioral patterns relevant to stress [67, 69, 70, 71, 73, 40, 75, 79]. Key features include mobility (number of locations visited, distance traveled), physical activity (steps, movement), sleep proxies, app usage (especially social/messenger apps), and communication patterns (calls, texts) [67, 71, 73, 79]. Some studies also integrate wearable sensor data (e.g., skin conductance, heart rate) for enhanced physiological monitoring [70, 80, 81].

Machine Learning and Statistical Modeling Approaches

The majority of studies employ supervised machine learning models, including random forests, support vector machines, neural networks, and deep learning architectures

(LSTM, CNN, hybrid models), to predict stress or anxiety from passive sensing data [67, 82, 71, 83, 75, 78]. Recent work highlights the benefits of personalized and idiographic models, transfer learning, and dynamic clustering to address inter-individual variability and improve prediction accuracy [77, 71, 75, 84, 78]. Reported classification accuracies range from 74% to over 90%, depending on the model, features, and study design [82, 85, 80, 86, 75, 78].

Real-World, Longitudinal, and Student-Focused Studies

Several large-scale, real-world studies have tracked university students over periods ranging from several weeks to multiple semesters, combining continuous passive data collection with periodic self-report surveys [67, 69, 70, 71, 72, 73, 74, 40, 75]. These studies confirm that stress levels fluctuate with academic workload and life events, and that passive sensing features can capture these dynamics with moderate-to-high accuracy [67, 69, 74, 40, 75]. The StudentLife dataset and similar resources have become benchmarks for the field [74, 40, 75].

Methodological Considerations: Feasibility, Data Integrity, and Limitations

Key methodological challenges include ensuring data integrity (handling missing data, device variability), minimizing battery consumption, achieving sufficient sample sizes, and maintaining participant adherence [67, 70, 72, 75]. Studies report that combining passive sensing with periodic surveys improves data quality and model performance [68, 69, 71, 72]. However, heterogeneity in study designs, sensor configurations, and reporting standards remains a barrier to reproducibility and generalizability [67, 72, 75, 87].

Table 2.1: Comparison of key studies on passive smartphone sensing for stress and mental health identification in university students.

Paper	Sensing Modalities	Study Duration & Population	Modeling Approach	Key Results
[67]	GPS, accelerometer, step count, gyroscope, activity monitoring	30 days; 409 Australian university students	Lasso, Ridge, KNN, RF, ANN, SDA	Steps and location variety predicted stress; stressed students showed distinct mobility profiles.

Paper	Sensing Modalities	Study Duration and Population	Modeling Approach	Key Results
[75]	GPS, phone use, audio, activity (StudentLife)	9 weeks; Dartmouth College students	Branched Deep Learning (CALM-Net), LSTM Autoencoder	Cold-start F1 = 87%; personalized modeling boosted performance by 37% over baselines.
[70]	Wearables (EDA, temp), phone usage, geolocation	1 month; 201 students	LASSO, SVM (linear/RBF)	Wearable sensor features most predictive of stress; accuracy up to 78.3%.
[73]	GPS location traces	2 weeks; 228 undergraduates	Mobility feature extraction + MLP	More socially anxious students had lower entropy; model accuracy = 85%.
[71]	App usage logs, sleep proxy features	Up to 60 days; 224 students	LASSO, SVR, RF, SHAP (Explainable AI)	Key predictors: app category usage and sleep; idiographic models more accurate.
[69]	GPS, ambient audio, sleep (StudentLife)	1 academic term; 95 students	PGEE (longitudinal)	Study locations increased stress; travel/audio amplitude at night reduced stress.
[82]	Illuminance, motion, app usage	1 month; 20 participants	RF, XGBoost; categorical feature combination	Co-occurring sensor contexts improved anxiety prediction (F1 = 74.2%).
[88]	StudentLife: sleep, activity, location, conversation	10 weeks; 48 students	Correlation with PHQ-9, PSS scores	Detected "term lifecycle": stress and poor sleep increased over semester.
[89]	EMA + GPS, acceleration, step count	6 weeks; 243 students (exam period)	Logistic Regression	Problematic smartphone use linked to worse mood and social anxiety; more mobility.

Paper	Sensing Modalities	Study Duration and Population	Modeling Approach	Key Results
[90]	iPhone screen time (user screenshots)	12 weeks; 187 students	Multilevel Models	No week-to-week link with stress; minor between-person mood effects.
[42]	GPS, heart rate (camera), calls, light, weather	8 days; 31 depressed students	Idiographic RF, Nomothetic XGBoost	Smartphone data predicted depressed mood hour-to-hour (r = 0.587).
[79]	Accelerometer, call/text logs	2 weeks; 59 students	XGBoost (leave-one-out CV)	r = 0.702 for social anxiety severity; movement oscillation 6s highly predictive.
[91]	Typing-time sensors: accelerometer, gyroscope, touchscreen	Non-longitudinal; 110 participants	Genetic Algorithm + kNN	Stress prediction via typing: 89.61% accuracy; real-time app implemented.
[92]	GPS, accelerometer, activity, call/text logs	5 weeks; 80 anxious students	RF, CatBoost, MERF + imputation (MICE, LOCF, Matrix)	Anxiety subtypes accurately predicted; matrix completion reduced error by 22%.

2.1.4 Discussion

The literature demonstrates that passive smartphone sensing, when combined with robust machine learning models, can accurately identify stress and related mental health states in university students, especially in real-world, longitudinal settings [67, 68, 69, 70, 71, 72, 75]. The strongest evidence comes from studies using multi-modal sensor data and advanced modeling techniques, which achieve high classification accuracy and offer insights into behavioral patterns associated with stress [67, 68, 69, 70, 75, 78]. Systematic reviews confirm the generalizability of these findings across diverse student populations and study durations [68, 72]. However, the field faces challenges related to data quality, participant adherence, and methodological heterogeneity, which can limit reproducibility and real-world deployment [67, 70, 72, 75, 87]. There is a clear need for standardized protocols, larger and more diverse samples, and further research into the integration of passive sensing with just-in-time interventions and privacy-preserving analytics.

2.1.5 Research Gaps

Passive smartphone sensing, especially when combined with machine learning and periodic surveys, offers a promising, scalable approach for identifying stress in university students. The field is supported by robust empirical evidence and systematic reviews, but faces ongoing challenges related to data quality, standardization, and real-world implementation.

Despite significant progress, research gaps remain in the standardization of methodologies, long-term feasibility studies, integration of privacy-preserving analytics, and the development of interventions based on real-time stress detection. Future work should focus on standardizing methodologies, improving data integrity, and developing real-time, privacy-preserving interventions. Key open questions include:

- **How can passive smartphone sensing protocols be standardized for stress detection in university students?**
  - Standardization would improve reproducibility, comparability, and facilitate large-scale deployment.
- **What are the long-term effects and feasibility of continuous passive sensing in real-world student settings?**
  - Longitudinal studies are needed to assess adherence, battery impact, and sustained accuracy over full academic terms.
- **How can privacy-preserving analytics be integrated into real-time stress detection and intervention systems?**
  - Addressing privacy concerns is critical for ethical deployment and user acceptance.

In summary, passive smartphone sensing is a validated and promising approach for stress identification in university students, but further research is needed to address methodological, ethical, and practical challenges for widespread adoption.

Chapter 3

Research Methodology

This chapter details the comprehensive methodology employed in our study to understand student stress using a combination of self-reported surveys and passively collected smartphone data. We describe the study design, participant recruitment, data collection protocols, the advanced smartphone application developed for data collection, the steps involved in preparing and transforming the raw data into meaningful features, the machine learning and deep learning models used, and the metrics for evaluating their performance.

3.0.1 Study Design and Participants

This study followed a group of 88 university students for one full school semester (Spring 2025). Our main goal was to understand student stress by combining two types of information: what students reported in surveys (their feelings) and what their smartphones collected automatically (their behaviors). This kind of study, where we observe people over a period of time without making any changes or interventions, is called a prospective observational study. By using both surveys and phone sensors, we followed modern approaches in digital health research to get a full picture of student well-being [93, 88].

We recruited 88 undergraduate students, aged 18-25, from CSE100 and CIS101 courses. To join the study, students needed to own an Android smartphone and agree to let us collect data using the app we developed. We did not include students with known mental disorders and whose phones were not compatible with our app. This group size is large enough to find important results and is similar to other long-term smartphone studies [88, 94].

3.0.2 Study Procedure

After giving their consent online, each student installed the study app on their smartphone and filled out a short form asking for basic information about themselves. The study lasted for 99 days (January 28, 2025, to May 06, 2025), which is 14 weeks. During this time, the following data were collected:

- **Passive Sensing:** The app collected data from the phone's sensors automatically in the background, all day and night.
- **Daily Diary (EMA):** Every evening around 9:00 PM, the app sent a notification asking students to complete a very short survey about their day.
- **Weekly Survey (DASS-21):** Every Sunday around 9:00 PM, the app prompted them to complete a standard questionnaire about their mood over the past week.

Students could pause the data collection or skip any survey question at any time. We used short surveys and gentle reminders to make participation easy and encourage students to stay in the study for the full semester [95, 96].

3.0.3 Data Collection Methods

We used a combination of self-report surveys and passive smartphone sensing to gather rich and varied data about student stress and daily life.

Self-Report Surveys

We used two different types of surveys to measure stress and related feelings directly from the students.

First, we used the **DASS-21** every week. This is a widely used and scientifically proven questionnaire that measures negative emotions. It helped us get a reliable score of a student's stress level over the past week. Table 3.1 shows all the questions in the DASS-21. While our main focus was the stress score, we also used the depression and anxiety scores to get a complete understanding of a student's mental health [97, 98].

Second, we used **EMA**. This was a very short survey sent each evening to capture how students felt on that specific day. This method helps reduce memory errors, because students report on their day right after it happens. The daily diary asked about their stress level, the main causes of any stress, and a few optional questions about their day. Table 3.2 provides the full details of the daily diary questions. This approach is common in mobile health research for tracking daily experiences [99, 96].

Table 3.1: The Depression, Anxiety, and Stress Scale-21 (DASS-21) Items and Subscales. The response scale for each item is 0 (*Did not apply at all*), 1 (*Applied to some degree*), 2 (*Applied a considerable degree*), or 3 (*Applied very much*). The final score for each subscale is the sum of its item scores multiplied by two.

Item No.	**Item Statement**	**Subscale**
1	I found it hard to wind down	Stress
2	I was aware of dryness of my mouth	Anxiety
3	I couldn't seem to experience any positive feeling at all	Depression
4	I experienced breathing difficulty (e.g., excessively rapid breathing)	Anxiety
5	I found it difficult to work up the initiative to do things	Depression
6	I tended to over-react to situations	Stress
7	I experienced trembling (e.g., in the hands)	Anxiety
8	I felt that I was using a lot of nervous energy	Stress
9	I was worried about situations in which I might panic	Anxiety
10	I felt that I had nothing to look forward to	Depression
11	I found myself getting agitated	Stress
12	I found it difficult to relax	Stress
13	I felt down-hearted and blue	Depression
14	I was intolerant of anything that kept me from getting on with what I was doing	Stress
15	I felt I was close to panic	Anxiety
16	I was unable to become enthusiastic about anything	Depression
17	I felt I wasn't worth much as a person	Depression
18	I felt that I was rather touchy	Stress
19	I was aware of the beating of my heart in the absence of physical exertion	Anxiety
20	I felt scared without any good reason	Anxiety
21	I felt that life was meaningless	Depression

Table 3.2: Structure of the Daily Diary (Ecological Momentary Assessment, EMA)

Component	**Prompt and Items**	**Response Format**
Primary Outcome (Daily Stress)	"Overall, how stressed did you feel **today**?"	5-point Likert scale: 0 = *Not at all* 1 = *A little* 2 = *Moderately* 3 = *Quite a bit* 4 = *Extremely*
Primary Stressor Identification	"What were the main stressor(s) today?" • **Academic:** class quiz/test, midterm/final, assignment/project, group work, time pressure • **Personal:** relationship, family, physical health, mental health • **Social/Financial:** conflict with peers, loneliness, finances, part-time job • **Other (please specify):**	Multi-select checklist plus an optional open-text field.
Optional Context	• Hours studied today • Approx. bedtime last night / wake time today • Two affect items: – "calm ↔ tense" – "Happy ↔ Sad"	Numeric slider (0–12) Time input (HH:MM) Visual analog sliders (0–100)

Passive Smartphone Sensing: The "WellBeing" Mobile Application

To collect continuous, objective data about student behavior, we developed a specialized mobile application called "WellBeing." This app was built for Android phones and combined both passive sensing (automatic data collection) and active sensing (survey prompts). It was designed to run reliably 24/7 in the background, minimizing battery drain while maximizing data collection.

System Architecture The "WellBeing" app has a modern and robust design to ensure reliable data collection.

- **Frontend (User Interface):** The part of the app that users interact with was built using *Flutter*. This allowed for a clean, modern design and ensured a smooth user experience. The user interface (UI) showed the status of data collection, displayed collected data in simple charts, and presented surveys.
- **Backend (Data Collector):** The core work of collecting data was handled by native *Android code*. This made the data collection very reliable and efficient in terms of battery use.
  - A *Foreground Service* was used for continuous, always-on data collection (like motion sensors). This service showed a persistent notification to the user, indicating that the app was running, which is the standard Android method for important background tasks.
  - *WorkManager* was used for tasks that happen at specific times but not constantly, such as sending daily diary notifications or checking app usage hourly. This method is very battery-efficient.
- **Data Storage:** All collected data was stored locally and securely on the user's phone in a structured *Room database*. This ensured privacy before data was exported.
- **Communication:** A "Method Channel" served as a bridge, allowing the Flutter UI to communicate with the native Android backend. This enabled the UI to start/stop services, request data, and show the correct status.

Key Features and Data Collection Modules The "WellBeing" application was organized into several modules, each responsible for collecting a specific type of data. The main dashboard for data collection provided a central overview, allowing users to check data collection status, enable survey reminders, view data history, and export data. An example of the dashboard is shown in Figure 3.1.

The app collected the following types of passive data:

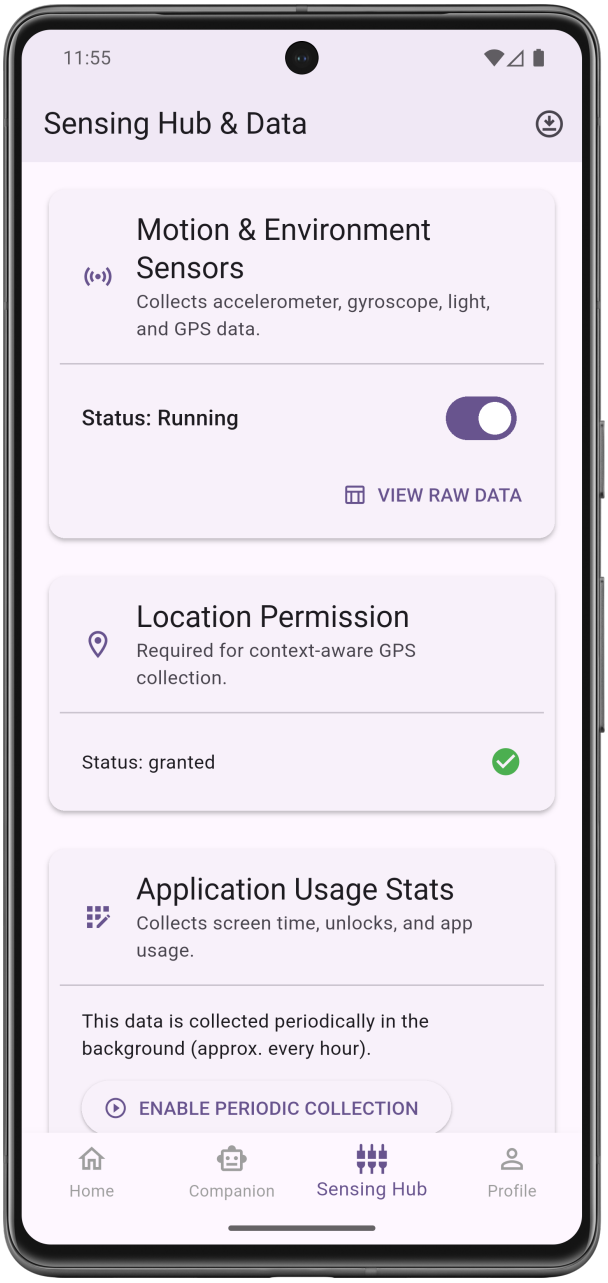


Figure 3.1: Screenshot of the 'WellBeing' App Main Dashboard, showing cards for different data collection modules.

- **Movement and Environment (Sensors Module):** This module continuously ran in the background, collecting data from various sensors.
  - *Data Collected:* Accelerometer (for physical movement like walking or running), gyroscope (for phone rotation and orientation), and light sensor (to detect if the user was in a dark or bright environment).
  - *Frequency:* Data was sampled at 5 times per second (5 Hz), providing a good balance between detail and battery life.
  - *Purpose:* This data helped us understand a student's physical activity levels

and their immediate surroundings.

- **Location (GPS Module):** Understanding mobility is key, but GPS uses significant battery. We implemented a "smart" GPS collection system.
  - *How it Works:* The app used motion sensors to detect if the user was moving. GPS was only activated when significant movement was detected and turned off when the user became still.
  - *Data Collected:* When active, it recorded the user's location (latitude and longitude).
  - *Purpose:* This smart system allowed us to track a student's mobility patterns throughout the day while greatly saving battery power.
- **Phone Usage (Usage Module):** This module provided insights into a student's digital habits. An example of the usage stats screen is shown in Figure 3.2.
  - *How it Works:* A scheduled task periodically summarized phone usage for the day.
  - *Data Collected:* Total screen-on time, the number of times the phone was unlocked, and a list of which apps were used and for how long.
  - *Purpose:* This data helped us understand digital habits, which can be linked to aspects like sleep patterns or study habits.
- **Sleep Proxy:** We estimated sleep time by looking for long periods of phone inactivity at night, combining information from movement sensors and screen-on/off times.
- **Communication:** The app recorded the number and length of phone calls and text messages (metadata only, without recording content, to protect privacy).

The "WellBeing" app also facilitated the active sensing (survey) components:

- **Weekly Mental Well-being Survey (DASS-21 Module):**
  - *What it is:* The standard DASS-21 survey. An example of the survey interface is shown in Figure 3.3, and a history chart in Figure 3.4.
  - *How it Works:* Every Sunday around 9 PM, a notification prompted the user. The survey presented one question at a time for ease of completion.
  - *Data Visualization:* Users could view their past scores on a chart within the app, observing changes in depression, anxiety, and stress over time.
- **Daily Diary (EMA Module):**

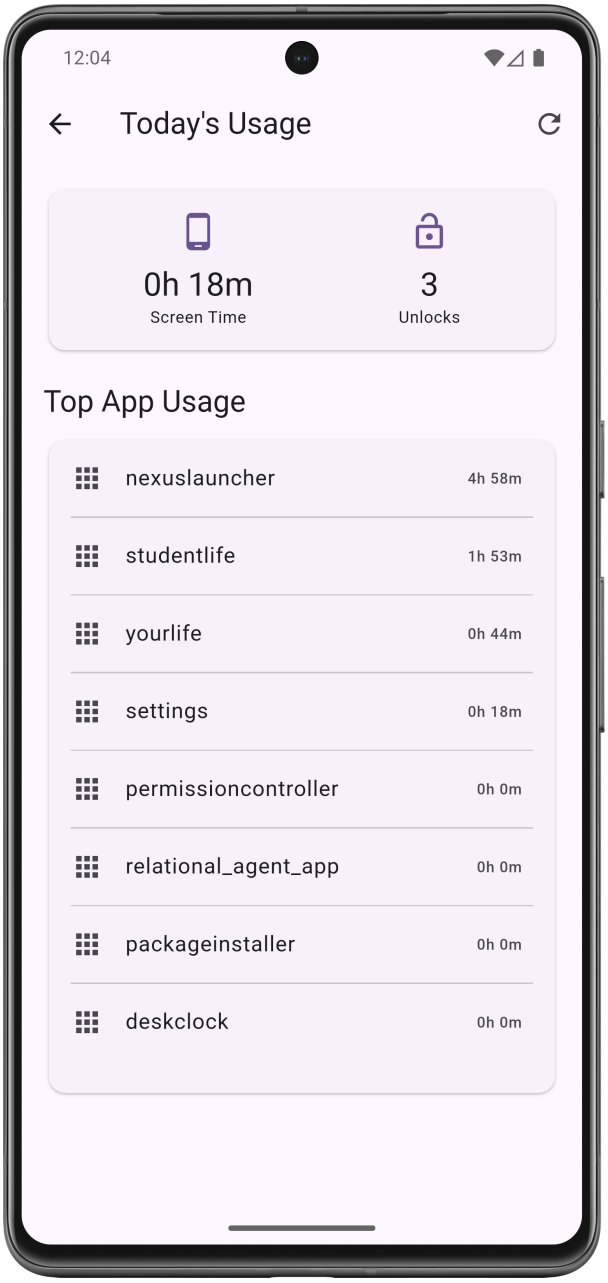


Figure 3.2: Screenshot of the 'WellBeing' App Usage Stats Screen, showing screen time, unlock count, and top apps.

- *What it is:* A short, daily diary for in-the-moment stress reporting. An example of the daily diary interface is shown in Figure 3.5.
- *How it Works:* Every evening around 9 PM, a notification prompted the user. The survey was a multi-step form asking about: overall stress level (5-point scale), main stressors (checklist), and optional context like hours studied, sleep times, and current mood.
- *Purpose:* This provided detailed, day-to-day information linking a person's feelings of stress to specific events or activities.

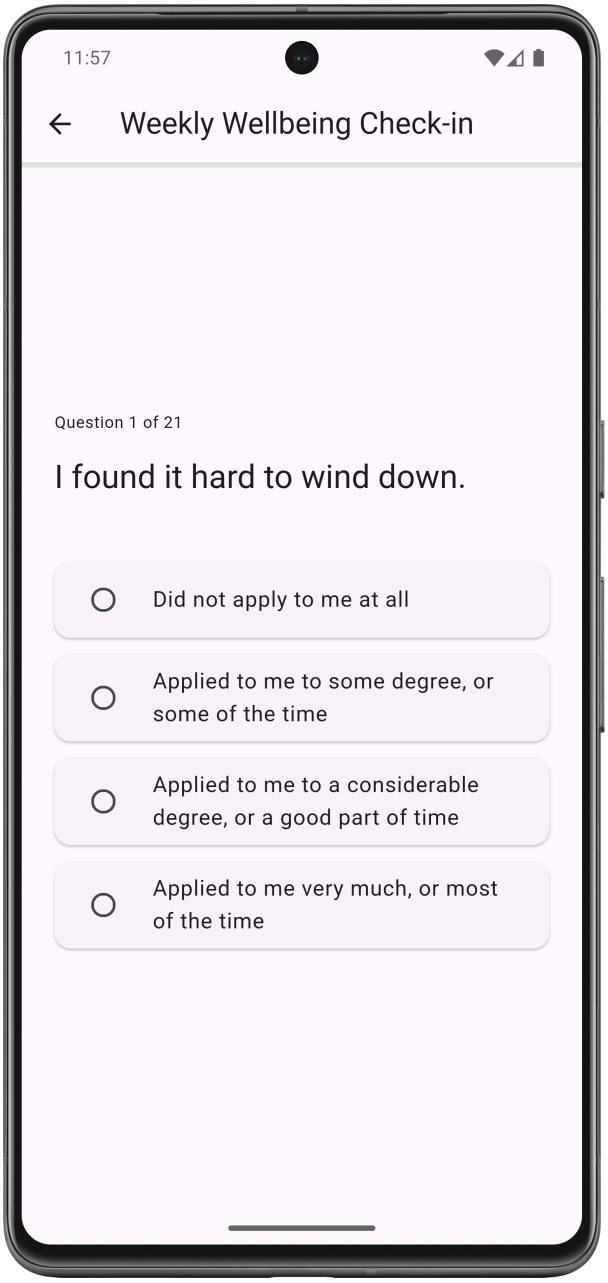


Figure 3.3: Screenshot of the 'WellBeing' App DASS-21 survey interface, showing one question per page.

Data Management and Export All data collected by the "WellBeing" app was stored securely on the user's device. To make this data useful for researchers, the app included a simple "Export All Data" button on the main dashboard. When tapped, the app gathered all information from its database and created four separate CSV (Comma-Separated Values) files. These files are simple text files that can be easily opened in programs like Excel or used for analysis in software like R or Python.

- *passive_sensors.csv*: Contained all timestamped accelerometer, gyroscope, light, and GPS data.

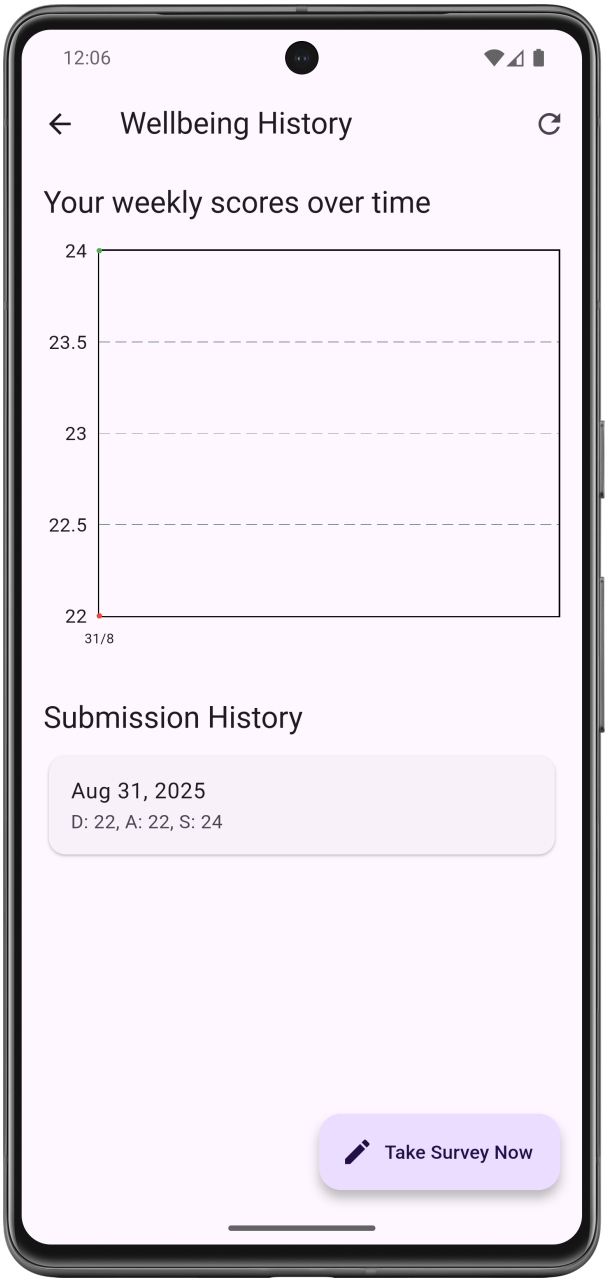


Figure 3.4: Screenshot of the 'WellBeing' App DASS-21 history chart, showing a line graph for Depression, Anxiety, and Stress scores over time.

- *usage_stats.csv*: Contained daily summaries of screen time, phone unlocks, and app usage.
- *dass21_surveys.csv*: Contained weekly DASS-21 scores and raw answers.
- *ema_diaries.csv*: Contained all responses from the daily stress diary.

This feature made it very easy for researchers to collect the data from participants at the end of a study.

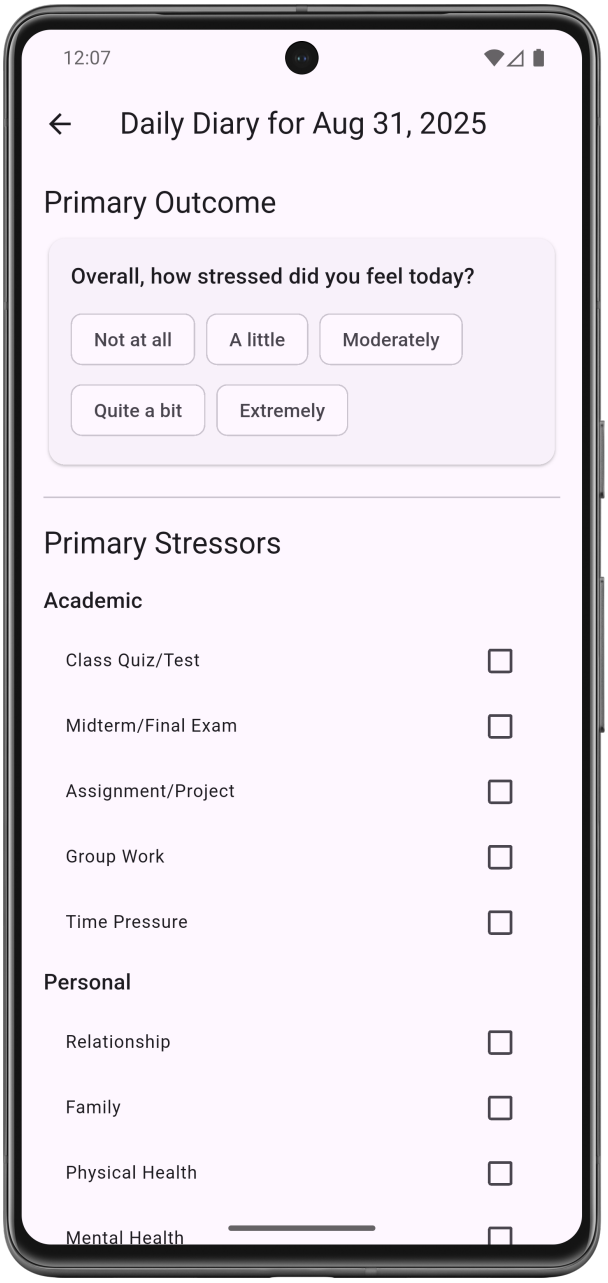


Figure 3.5: Screenshot of the 'WellBeing' App Daily Diary (EMA) interface, showing one of the steps like the stressor checklist.

System Robustness Ensuring continuous background operation is a challenge for sensing apps, as modern phones often close apps to save battery. The "WellBeing" app included several features to be as robust as possible:

- **Foreground Service:** As mentioned, this was the primary tool to prevent the operating system (OS) from stopping sensor collection.
- **Restart on App Swipe:** If a user closed the app from the "recent apps" screen, an *AlarmManager* automatically restarted the service within one second.
- **Restart on Phone Reboot:** A *BootReceiver* listened for when the phone finished

restarting and automatically started the sensor collection service again.

Despite these measures, some phone manufacturers have additional battery-saving software that can still interfere. The most reliable solution for this is to ask users to add the app to an "exception list" in their phone's settings.

3.0.4 Data Processing and Feature Preparation

The raw data collected from the "WellBeing" app and surveys was complex and required several steps to be processed into meaningful numerical summaries, called "features," ready for machine learning models.

Feature Creation and Aggregation

We created two main datasets, each tailored for our specific prediction tasks:

Daily Prediction Dataset For each student, for each day of the study, we calculated a set of summary features from their passive smartphone data. These daily features included:

- Total distance traveled (calculated from GPS data).
- Total minutes of screen time.
- Number of phone unlocks.
- Estimated hours of sleep.
- Information from the student's academic calendar (e.g., a flag indicating an *is_exam_day* or *is_weekend*).
- Aggregates of call and text message counts and durations.
- Summaries of app usage categories (e.g., social media, productivity apps).

The "target" or "label" we wanted to predict for this dataset was the student's self-reported stress level from their daily diary for that specific day.

Weekly Prediction Dataset For each student, for each week of the study, we prepared a single data sample. This sample consisted of the sequence of all the daily features collected during that week (e.g., a sequence of 7 days of daily feature data). The "target" or "label" for this dataset was the stress severity category derived from the student's weekly DASS-21 survey score for that particular week.

Handling Missing Data and Outliers

- **Handling Missing Data:** Sometimes, a phone might be off, or the app might be temporarily stopped, leading to gaps in data. We identified these missing periods and used standard statistical methods, such as mixed-effects models and multiple imputation, to estimate and fill in these gaps [93, 100].
- **Controlling for Outliers:** Some data points can be unusually extreme (e.g., an exceptionally high number of phone unlocks on one day). We used a method to cap these extreme values, preventing them from having too much influence on our models. We also compared each student's daily data to their own personal average behavior, which helped to account for individual differences [88].

Preparing Labels for Classification

For our models that predicted stress categories (classification models), we transformed the survey scores into discrete labels:

For Daily Prediction Labels We used the original 5-point Likert scale from the daily diary (0 = *Not at all*, 1 = *A little*, 2 = *Moderately*, 3 = *Quite a bit*, and 4 = *Extremely*). For some analyses, we also created a simpler 3-level version:

- **Low Stress:** Original levels 0 and 1.
- **Medium Stress:** Original level 2.
- **High Stress:** Original levels 3 and 4.

For Weekly Prediction Labels We converted the DASS-21 stress subscale scores into categories based on standard clinical cut-offs. We used a 5-level scale:

- Normal
- Mild
- Moderate
- Severe
- Extremely Severe

Additionally, we prepared a simpler 3-level scale for weekly stress:

- **Low Stress:** Combining Normal and Mild categories.
- **Elevated Stress:** Moderate category.
- **High Stress:** Combining Severe and Extremely Severe categories.

3.0.5 Modeling Approach

Our main goal was to build models that could predict student stress using the features created from their smartphone data. We used different types of machine learning models for our daily and weekly prediction tasks, as each task required understanding different patterns in the data.

Models for Daily Stress Prediction

For the daily prediction task, where we aimed to predict a student's stress level for a single day using that day's features, we used four well-known machine learning models. These models are effective at finding patterns in individual data points.

- **Logistic Regression (LR):** A simple and fast statistical model that serves as a good baseline. It estimates the probability of an outcome by fitting data to a logistic function [101].
- **Random Forest (RF):** A powerful ensemble model made up of many "decision trees." It combines predictions from multiple trees to make a more accurate and robust final prediction, excellent for finding complex patterns [102].
- **Support Vector Machine (SVM):** This model finds the best boundary (a hyperplane in high-dimensional space) that separates different categories of data, maximizing the margin between them [103].
- **Multi Layer Perceptron (MLP):** A model inspired by the human brain, consisting of multiple layers of interconnected "neurons." MLPs can learn very complex relationships within the data, making them suitable for non-linear patterns [104].

Model for Weekly Stress Prediction

For the weekly prediction task, we needed a model that could understand patterns in sequences of data over time (e.g., how daily behaviors change throughout a week). For this, we used a specialized type of deep learning model:

- **Long Short-Term Memory (LSTM) Network:** An LSTM is a type of recurrent neural network (RNN) specifically designed to learn from data that has a time sequence, like the daily features over a week [105]. It can "remember" important information from earlier parts of the sequence and use it to make predictions later. We used an advanced LSTM with multiple layers to capture intricate temporal dependencies. To address the challenge of imbalanced data (where some stress categories were much less common than others), we used a technique called *Weighted Random Sampling* during model training. This technique assigns higher importance

to the less common stress categories, ensuring the model learns to identify them properly.

3.0.6 Model Evaluation and Validation

To check how well our models performed and to ensure their reliability, we used a fair and strict evaluation process. Our focus was on predicting stress categories, so we primarily used metrics suitable for classification tasks.

Evaluation Protocol: Leave-One-Subject-Out (LOSO)

For all of our models, we used **LOSO** cross-validation. This is a very strict and robust method to test how well a model can generalize to new people. The process was as follows:

1. To test the model's performance on a specific student (e.g., Student A), we first trained the model exclusively on the data from all the other 87 students.
2. After training, the model was tested only on the data of Student A, which it had never seen before during training.
3. This entire process was repeated 88 times, once for each student in the study, ensuring that every student's data was used as the test set exactly once.

This LOSO method is considered the best way to determine if a model can work for a completely new person it has never learned from, which is the most realistic scenario for real-world applications [106, 107].

Evaluation Metrics

We used several evaluation metrics to measure the performance of our classification models. Since our primary goal was to predict categories of stress, we focused on metrics that are suitable for classification tasks. The performance of a classification model can be described in terms of four fundamental outcomes, derived from the confusion matrix:

- **True Positive (TP):** The number of positive instances that are correctly predicted as positive by the model.
- **True Negative (TN):** The number of negative instances that are correctly predicted as negative by the model.
- **False Positive (FP):** The number of negative instances that are incorrectly predicted as positive (also referred to as Type I error).

- **False Negative (FN):** The number of positive instances that are incorrectly predicted as negative (also referred to as Type II error).

These four measures serve as the foundation for deriving higher-level evaluation metrics, including Accuracy, Precision, Recall, and the F1-score [108].

- **Accuracy:** This is a straightforward metric that tells us the percentage of all predictions the model got correct. Accuracy quantifies the overall correctness of the model by measuring the proportion of correctly classified instances relative to the total number of instances:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}. \tag{3.1}$$

- **Precision:** Precision, also referred to as the positive predictive value, measures the fraction of instances predicted as positive that are actually positive:

$$\text{Precision} = \frac{TP}{TP + FP}. \tag{3.2}$$

  High precision reflects a model's ability to minimize false positives, making it especially relevant in applications where false alarms are costly.

- **Recall:** Recall, also known as sensitivity or true positive rate, measures the fraction of actual positive instances that are correctly identified by the model:

$$\text{Recall} = \frac{TP}{TP + FN}. \tag{3.3}$$

  A higher recall implies that the model successfully detects a larger proportion of positive cases, which is crucial in applications where missing a positive instance has severe consequences.

- **F1-score (Macro):** This was our most important metric for evaluating performance across different stress categories. The F1-score measures the balance between two other scores, precision and recall. A "Macro" F1-score calculates the F1-score for each individual stress category and then averages them, giving equal weight to all classes (even rare ones). A high F1 Macro score means the model is good at identifying all the different stress categories, not just the most common ones.

$$F1 - Score = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}. \tag{3.4}$$

  We can obtain the F1-score in terms of TP, FP, and FN:

$$F1 - Score = \frac{2TP}{2TP + FP + FN}. \tag{3.5}$$

- **Classification Report:** This is a detailed table that provides the precision, recall, and F1-score for each individual stress category (class), along with their averages. It helps us understand where the model performs well and where it might struggle.
- **Confusion Matrix:** This is a visual diagram that provides a summary of the model's performance. It shows us exactly how many predictions were correct for each class and what kinds of mistakes the model made (e.g., if it often confused "Medium" stress with "High" stress). This matrix helps in understanding the specific strengths and weaknesses of the model.

3.1 The AI Well-being Companion

A key feature of our application is the AI Well-being Companion, which we have named "WellBeing Mate". While passive sensing is good for collecting objective data, we wanted to create a feature that allows users to actively engage with their own well-being in a personal and interactive way. WellBeing Mate is a chatbot designed to be a supportive and friendly partner in the user's mental health journey.

3.1.1 Purpose of the Companion

The main goal of WellBeing Mate is to provide a safe space for users to talk about their feelings and daily life. It is not a replacement for a doctor or therapist, but it serves as an "always-on" tool for self-reflection and emotional support. It can help users:

- Understand their own stress patterns.
- Feel heard and less alone.
- Receive gentle encouragement and suggestions.
- Connect their daily activities to their feelings.

3.1.2 How the Companion Works: A Smart and Personal Approach

What makes WellBeing Mate special is that it is not just a standard chatbot. It uses a combination of memory, passive data, and a powerful language model (Google's Gemini API) to have more meaningful and personalized conversations with the user. Although the "brain" of the AI is on a server, it feels like it is part of the app. This is how it works:



Figure 3.6: A screenshot of the AI Companion's chat interface. It shows a friendly greeting from WellBeing Mate and a simple chat window.

- **Using a Base Prompt:** We give the AI a personality and a set of rules. We tell it: "You are WellBeing Mate, an empathetic AI companion. Your goal is to be a supportive friend to a university student. Listen, validate feelings, and never give medical advice." This base instruction, or "prompt," ensures that WellBeing Mate always acts in a safe and helpful way.
- **Remembering the User:** To feel connected, a friend needs to remember past conversations. WellBeing Mate has two types of memory:
  - *Short-Term Memory:* WellBeing Mate remembers the last few messages in

the current conversation, just like a person would. This helps the conversation flow naturally.

  - *Long-Term Memory:* WellBeing Mate can extract important facts from conversations, like "the user has an important exam next week" or "the user enjoys playing the guitar." It remembers these facts for future conversations, making the user feel recognized and understood.

- **Using Passive Data as a Cue:** The app uses the passively collected data (like activity level and phone usage) to predict the user's likely stress level for the day. This "predicted stress" information is given to WellBeing Mate as a quiet hint, or a "subtle cue," about the user's mood. WellBeing Mate does not say "My data shows you are stressed!" Instead, it uses this information to choose a more appropriate tone.

3.1.3 An Example Conversation Flow

Let's imagine a user opens the app.

- **The Greeting:** The first thing the user sees is a personalized greeting from WellBeing Mate. To create this greeting, the app sends the user's recent conversation history and their predicted stress level to the Gemini API.

  - If they talked yesterday: WellBeing Mate will remember the topic and ask about it. For example: "Hey, you mentioned you had a big project due. How did it go?" This makes the user feel like they are continuing a conversation with a friend.
  - If it's the first chat in a while: WellBeing Mate will use the stress level as a guide. If stress is predicted to be 'High', the greeting will be very gentle: "Hey there, it seems like today might be a bit heavy. I'm here to listen if you'd like to talk."

- **The Conversation:** When the user types a message, the app sends their message, along with the short-term and long-term memories, to the Gemini API. The base prompt guides the AI to respond in a caring way. For example, if a user says "I am so tired," WellBeing Mate will not just say "Okay." It will use its instructions to be empathetic and ask an open-ended question, like: "I'm sorry to hear that. It sounds like you're feeling worn out. Is there anything specific that's on your mind?"

This entire process happens through an API, which is a way for the app to talk to a server. This design keeps the app itself simple and fast, while allowing us to use a very powerful AI model on the backend. The user just experiences a smooth, continuous conversation with their companion, WellBeing Mate.

3.1.4 Why This Approach is Important

The AI Well-being Companion bridges the gap between passive data collection and active user engagement. Passive data alone is just a collection of numbers. The Companion uses this data to make its conversations more personal and empathetic.

By remembering facts and adapting its tone, WellBeing Mate feels less like a computer program and more like a connected, supportive friend. This helps the user feel more comfortable sharing their thoughts and feelings, turning our application from just a tool for collecting data into a true partner for improving daily well-being.

Chapter 4

Results and Analysis

This chapter explains the findings from our study. We collected data from 88 students during the Spring 2025 semester. First, we present basic patterns observed in the data. Next, we report results from models that aimed to predict student stress each day. Finally, we describe our attempt to predict weekly stress survey scores.

4.1 Key Patterns in Student Data

Before building prediction models, we examined the data for logical patterns to ensure its quality. We identified three key patterns.

Pattern 1: Stress is Higher During Exams

We hypothesized that students would experience higher stress during midterm and final exams. To evaluate this, we plotted the average stress for all students across the semester.

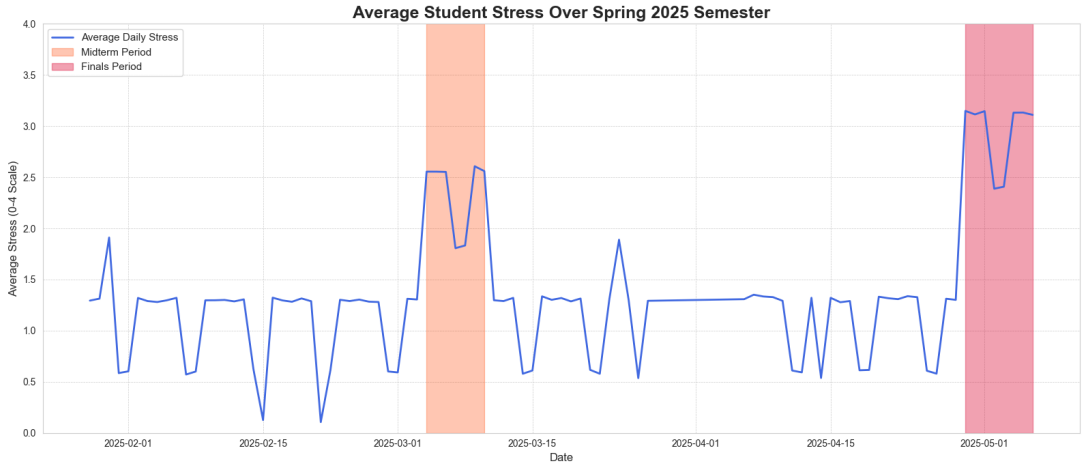


Figure 4.1: Average Student Stress Over Spring 2025 Semester

Figure 4.1 shows sharp increases in stress during midterm and final exam weeks. A statistical t-test confirmed that these increases were significant ($p < .001$), indicating that exams are a major contributor to student stress.

Pattern 2: Students Behave Differently on Weekends

We examined changes in daily habits, specifically sleep and study durations, between weekdays and weekends.

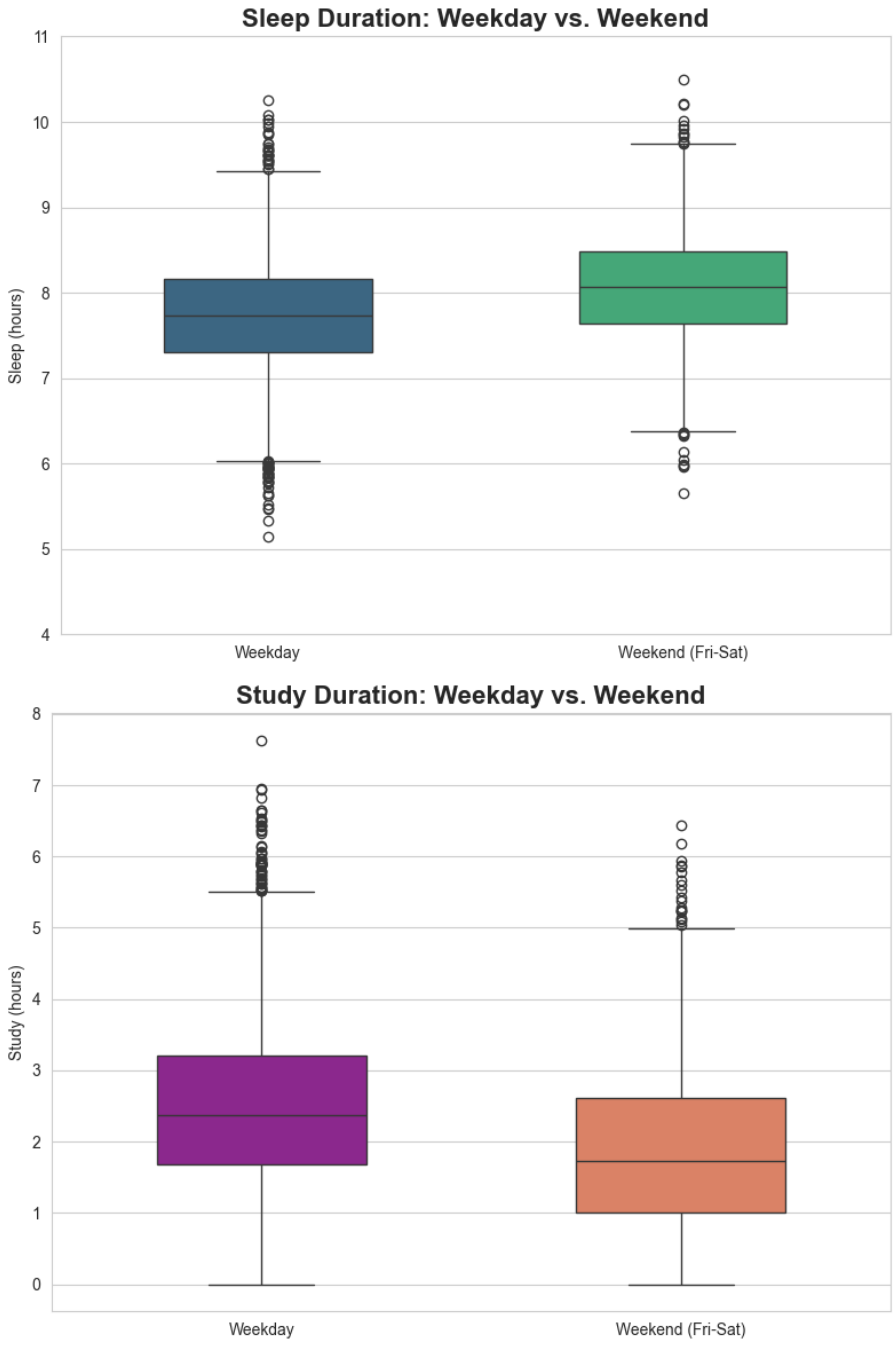


Figure 4.2: Sleep and Study Duration on Weekdays vs. Weekends

As shown in Figure 4.2, students slept more and studied less on weekends. A t-test showed that the increase in sleep on weekends was statistically significant ($p < .001$), suggesting a consistent weekly routine among students.

Pattern 3: School Work is the Main Cause of Stress

We analyzed responses from daily surveys where students reported their primary sources of stress.

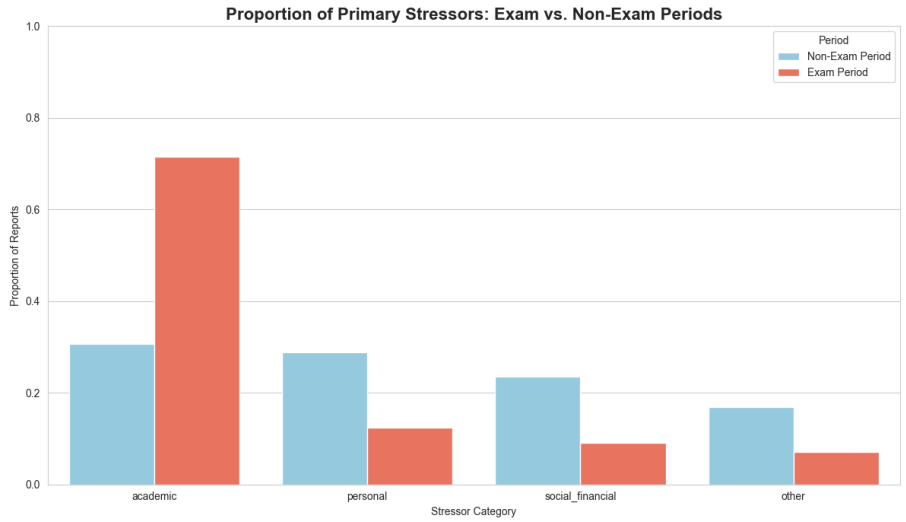


Figure 4.3: Proportion of Primary Stressors: Exam vs. Non-Exam Weeks

Figure 4.3 shows that during normal weeks, stress stemmed from various sources including academics, personal life, and social or financial issues. However, during exam weeks, academic concerns overwhelmingly dominated. This confirms that school-related factors are the primary drivers of stress during peak academic periods.

These findings demonstrate that the collected data accurately reflects the lived experiences of students.

4.2 Results: Predicting Daily Student Stress

This section presents the results of our efforts to predict daily student stress using machine learning models. The models were trained on passive sensor and calendar data collected from 88 university students. Our primary goal was to determine how accurately we could predict a student's self-reported stress level on any given day.

We tested four different types of machine learning models to see which performed best: LR, RF, SVM, MLP. To ensure our tests were fair and realistic, we used a strict method called LOSO cross-validation. This method always tests the model on a student it has never seen before.

Table 4.1: Performance of Models for 5-Class Daily Stress Prediction

Model	Accuracy	F1 Macro Score
LR	0.54	0.51
RF	0.59	0.49
SVM	0.52	0.49
MLP	0.56	0.51

We conducted two main experiments. First, we tried to predict the original five levels of stress reported by students. Second, we simplified the problem by grouping the stress levels into three broader categories: "Low", "Medium", and "High".

4.2.1 Experiment 1: Predicting Five Levels of Stress

In our first experiment, we trained the models to predict the exact stress level reported by students on the 5-point scale: "Not at all", "A little", "Moderately", "Quite a bit", and "Extremely". Table 4.1 summarizes the performance of the four models on this difficult task. The "F1 Macro" score is the most important metric because it shows how well the model performs across all five stress levels, treating each one as equally important.

The results show that this is a challenging task. While all models performed much better than random guessing, the Random Forest model achieved the highest overall accuracy at 59%. However, the F1 Macro scores for all models were around 0.50. This indicates that the models struggled to tell the difference between some of the stress levels. This is visualized in the confusion matrices, such as the one for Logistic Regression shown in Figure 4.4 (a). We can see that the model often made mistakes between adjacent categories, for example, confusing "A little" with "Moderately" stressed. The models also had difficulty correctly identifying the "Extremely" stressed days, which were the least common in the dataset.

4.2.2 Experiment 2: Predicting Three Levels of Stress

Because the models found it difficult to distinguish between five very specific stress levels, we simplified the problem by grouping them into three broader, more practical categories: "Low", "Medium", and "High". We then trained the same four models on this new 3-class problem. The performance of all models improved significantly, as shown in Table 4.2.

The results for the 3-class task were much stronger. The Random Forest model was the top performer with an accuracy of 74%. The F1 Macro scores also improved greatly for all models, showing they were much better at reliably identifying all three categories.

The benefit of simplifying the problem is clearly visible in the confusion matrices shown in Figures 4.4 through 4.7. For every model, the subfigure on the bottom (b),

Table 4.2: Performance of Models for 3-Class Daily Stress Prediction

Model	Accuracy	F1 Macro Score
LR	0.71	0.70
RF	**0.74**	**0.67**
SVM	0.70	0.68
MLP	0.72	0.67

which shows the 3-class results, has a much clearer diagonal pattern than the subfigure on the top (a). This means the models made far more correct predictions.

For instance, in Figure 4.5 (b), the Random Forest model correctly identified 4554 of the 4975 "Low Stress" days and 592 of the 808 "High Stress" days. Most of its errors were in confusing "Low" and "Medium" stress, which is a less severe error. Importantly, the model almost never confused a "Low Stress" day with a "High Stress" day, demonstrating its reliability for identifying students in need of support.

4.2.3 Summary of Daily Prediction Findings

Our experiments showed that it is possible to predict a student's daily stress level from passive data, but the task becomes much more reliable when using broader stress categories. The Random Forest model proved to be the most effective algorithm for this problem. The success of the 3-class prediction task, with an accuracy of 74%, demonstrates a strong potential for using this technology in real-world applications to identify students who may be experiencing elevated or high levels of stress.

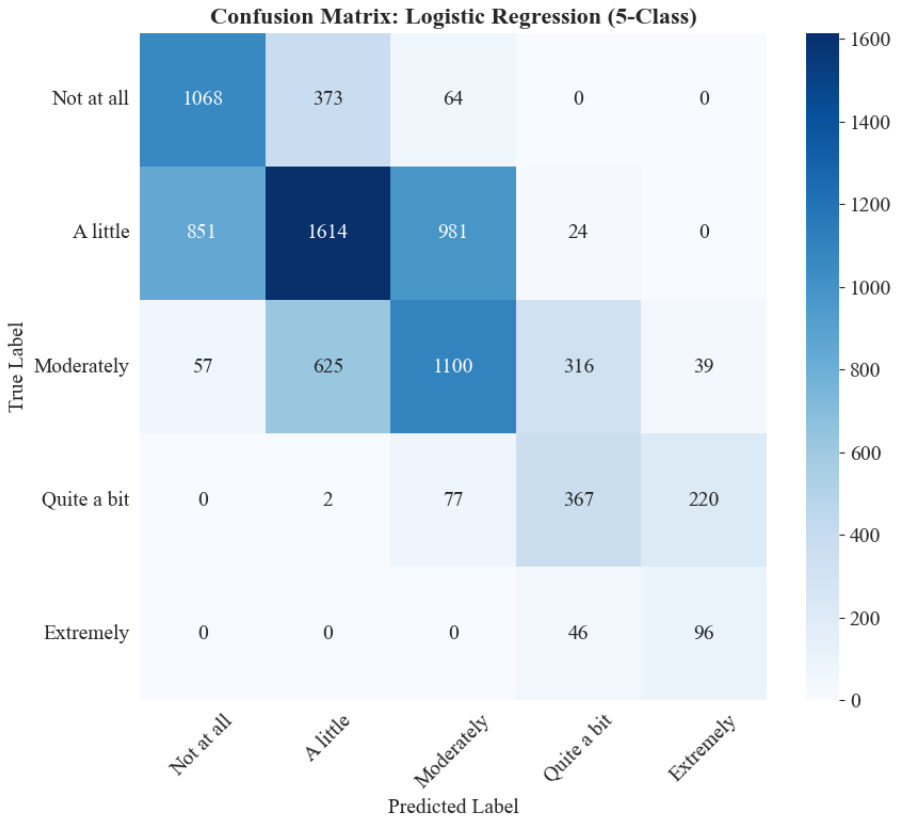
4.3 Results: Predicting Weekly Stress Severity

Following the analysis of daily stress, our second major goal was to determine if we could predict a student's overall stress level for an entire week. For this task, we used the weekly DASS-21 survey score, which we categorized into different severity levels. We used a more advanced model called a Long Short-Term Memory (LSTM) network, which is specifically designed to find patterns in sequences of data, such as a week of daily behaviors.

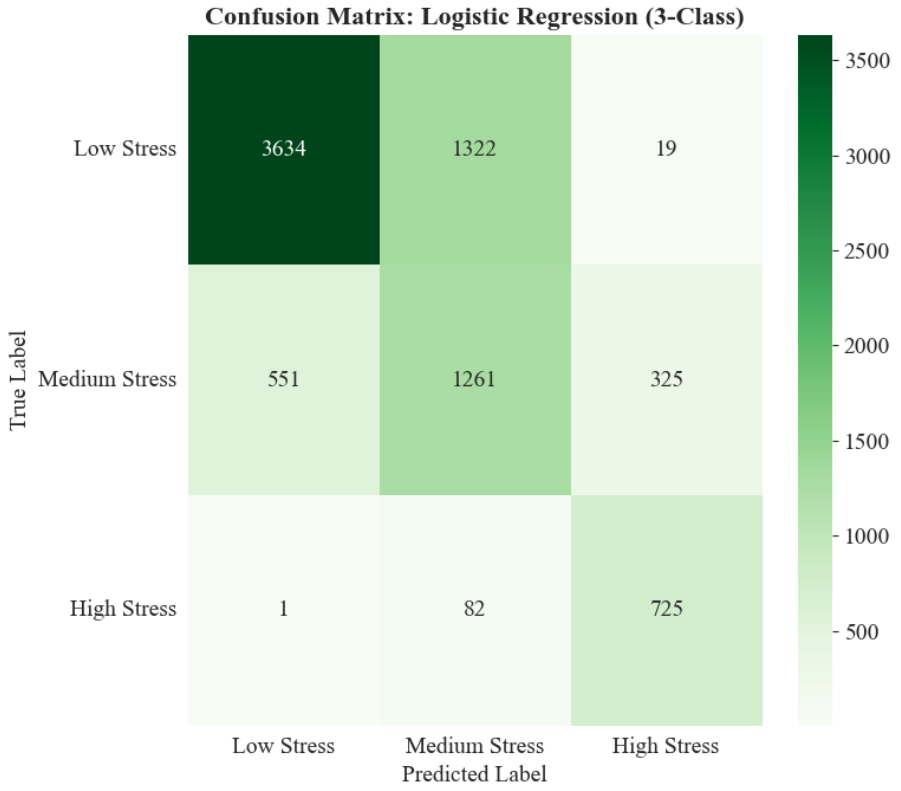
4.3.1 Training and Evaluation Method

The evaluation method was the same LOSO protocol used for the daily models. The LSTM was trained on the weekly data sequences from 87 students and then tested on the one student it had not seen.

A key challenge with this data was class imbalance. There were many weeks where students reported "Normal" stress but very few weeks in the "Severe" or "Extremely Se-

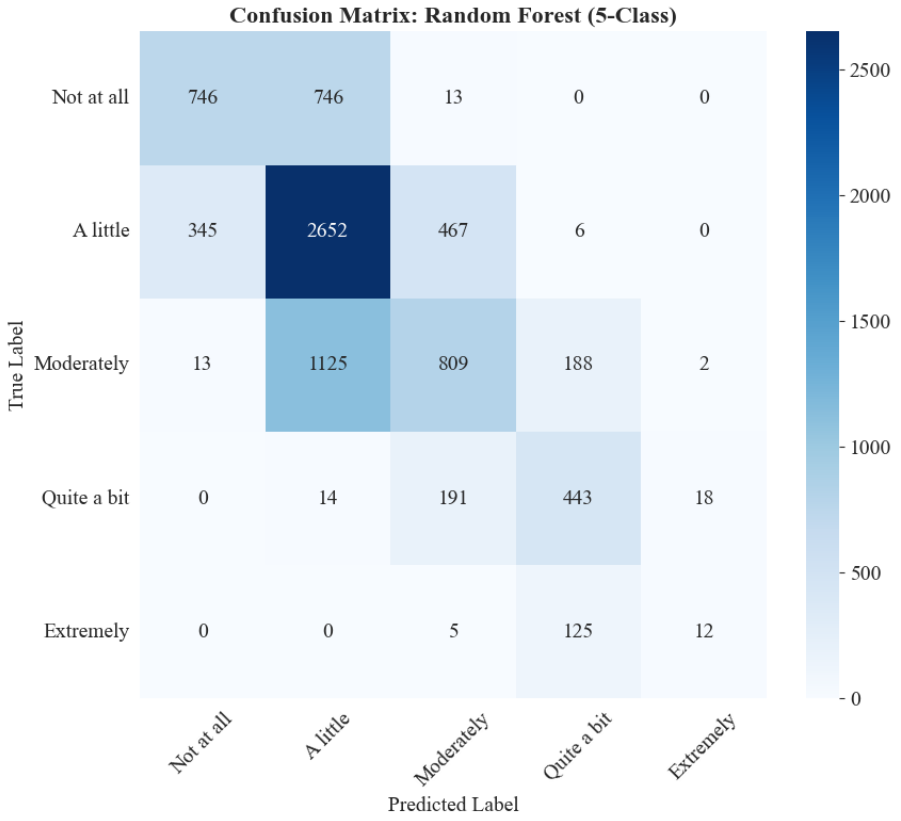


(a) 5-Class Prediction

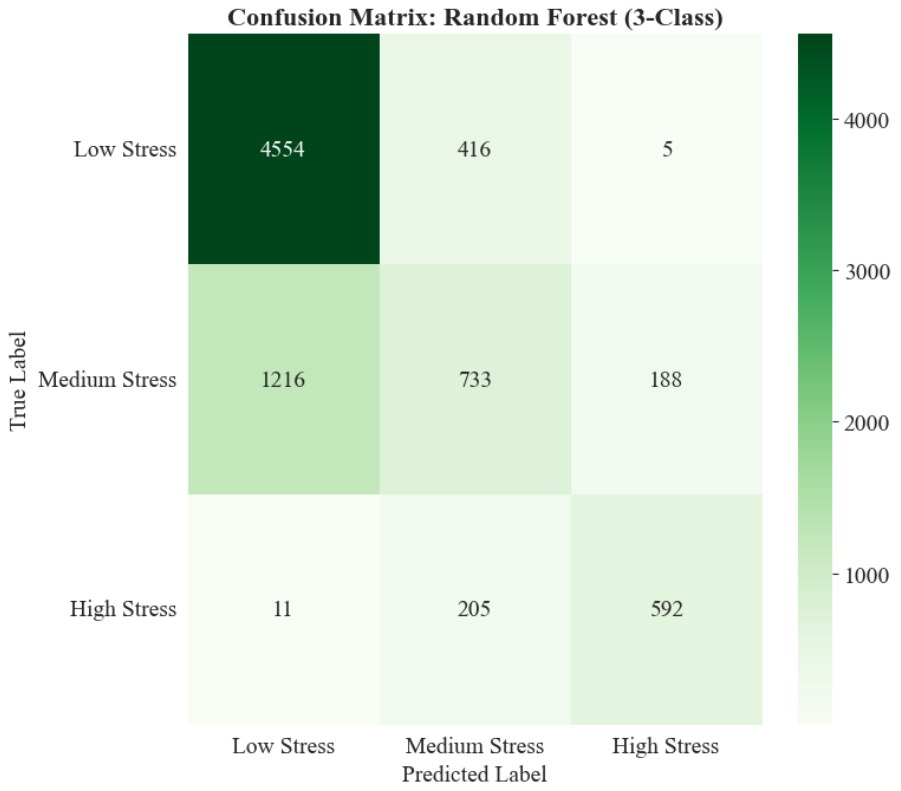


(b) 3-Class Prediction

Figure 4.4: Confusion Matrices for the **Logistic Regression** Model.

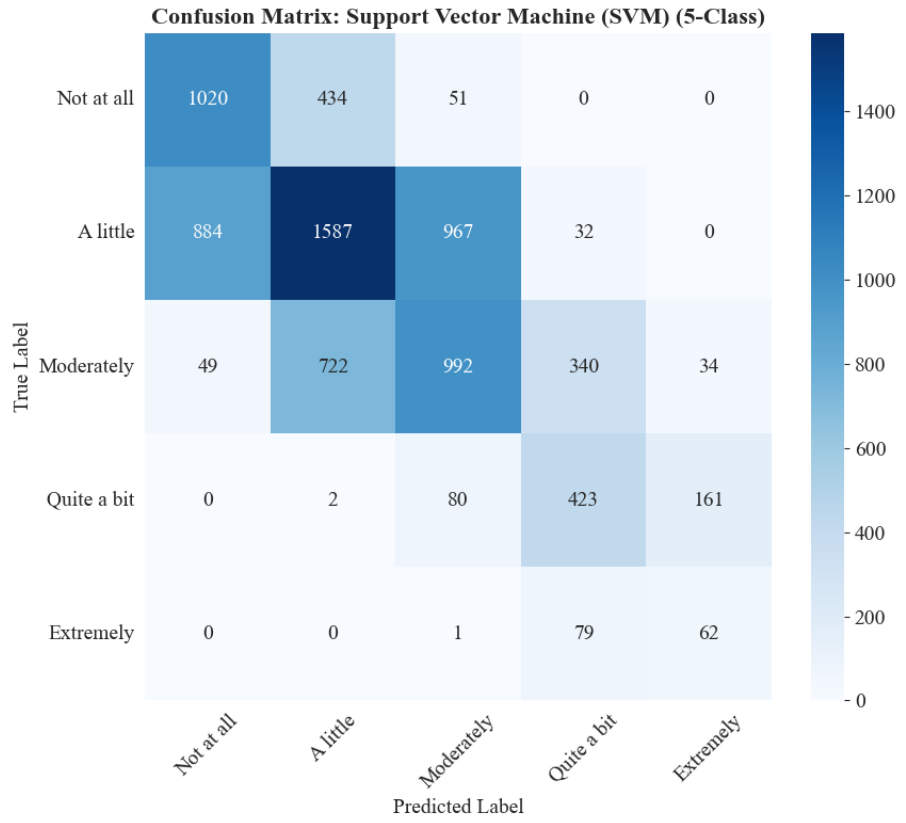


(a) 5-Class Prediction

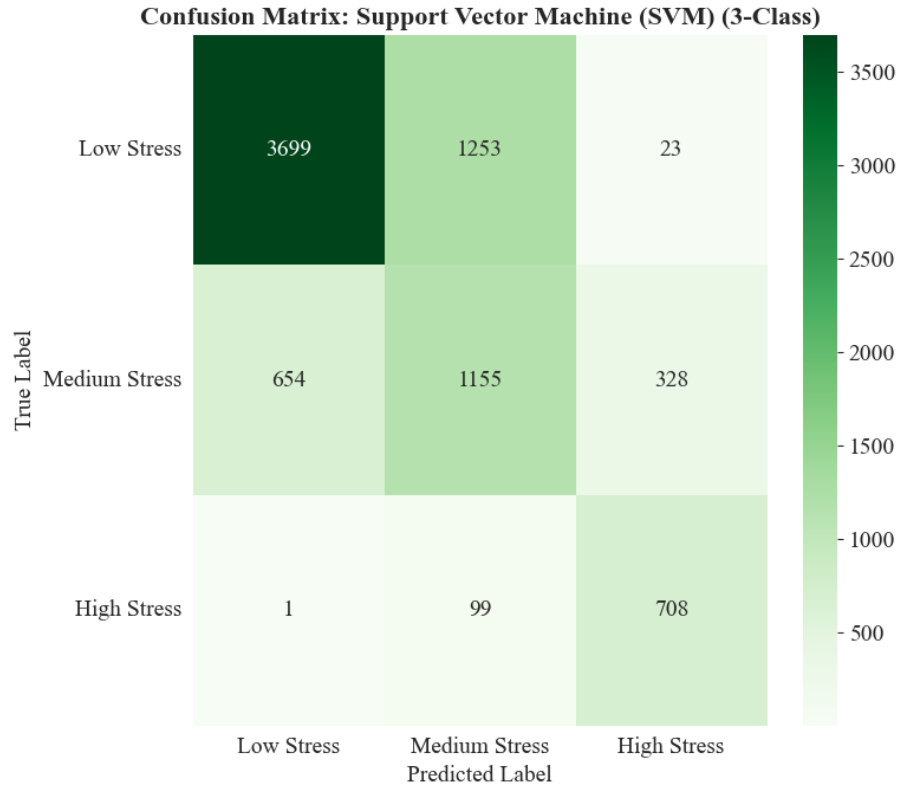


(b) 3-Class Prediction

Figure 4.5: Confusion Matrices for the **Random Forest** Model.

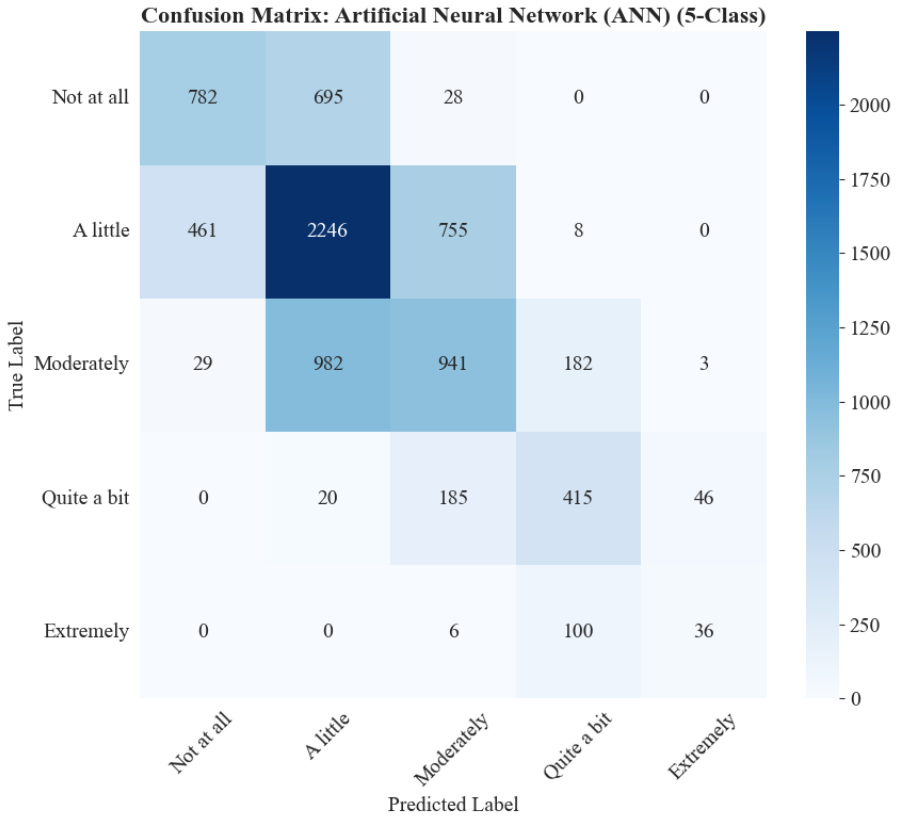


(a) 5-Class Prediction

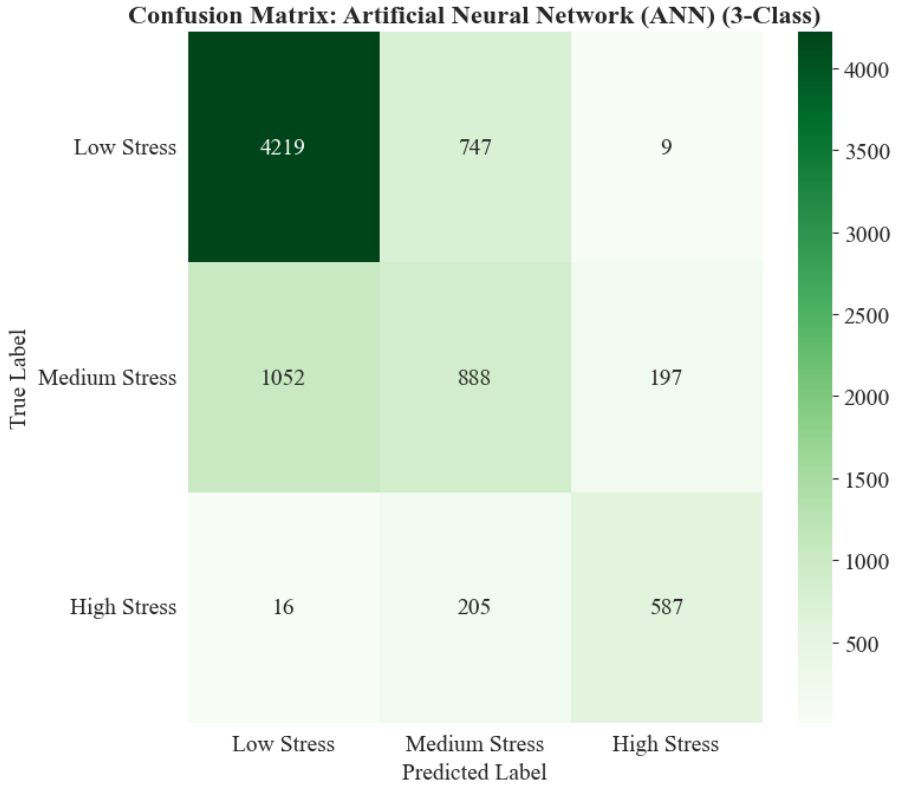


(b) 3-Class Prediction

Figure 4.6: Confusion Matrices for the **Support Vector Machine (SVM)** Model.



(a) 5-Class Prediction



(b) 3-Class Prediction

Figure 4.7: Confusion Matrices for the **MLP** Model.

Table 4.3: Performance of the LSTM Model for 5-Class Weekly Stress Prediction

Metric	Score
Accuracy	0.46
F1 Macro Score	0.41

Table 4.4: Performance of the LSTM Model for 3-Class Weekly Stress Prediction

Metric	Score
Accuracy	0.71
F1 Macro Score	0.57

vere" categories. To address this, we used a technique called Weighted Random Sampling during training. This method ensures that the model pays equal attention to all stress categories, including the rare but clinically important high-stress cases.

We conducted two experiments with the LSTM model. First, we tried to predict the standard five levels of stress severity. Second, we used a simpler three-level scale.

4.3.2 Experiment 1: Predicting Five Levels of Weekly Stress

Our first experiment was to classify each week into one of the five standard DASS-21 stress categories: "Normal", "Mild", "Moderate", "Severe", and "Extremely Severe". Table 4.3 summarizes the model's performance.

The results show that this is a very difficult task. The F1 Macro score of 0.41 indicates that the model struggled to reliably distinguish between the five different levels. The confusion matrix in Figure 4.8 (a) helps explain why. The model was best at identifying the "Normal" category but made many mistakes between the "Mild", "Moderate", and "Severe" categories. For example, out of 261 actual "Moderate" weeks, the model only correctly identified 26. This is likely because the behavioral patterns for these middle categories are very similar and hard to separate.

4.3.3 Experiment 2: Predicting Three Levels of Weekly Stress

Recognizing the difficulty of the 5-class problem, we simplified the task into three more distinct categories: "Low", "Elevated", and "High". On this task, the LSTM model's performance improved substantially. Table 4.4 shows the results.

With an accuracy of 71% and an F1 Macro score of 0.57, the model demonstrated a much better ability to classify weekly stress into these broader categories. This performance is promising and shows that the weekly sequence of behaviors contains a clear signal related to a student's overall stress state.

The improvement is clearly visible in the confusion matrix in Figure 4.8 (b). The model was very effective at identifying the "Low Stress" category, correctly classifying

756 out of 946 weeks. More importantly, it showed a strong ability to detect the "High Stress" category. Of the 57 actual "High Stress" weeks, the model correctly identified 38 (a recall of 67%). This is a very important result, as identifying students in the highest stress category is a primary goal for a mental health application. While the model had some difficulty with the middle "Elevated" category, its success at identifying the two extremes (Low and High) is a significant finding.

4.3.4 Summary of Weekly Prediction Findings

Predicting a student's weekly stress severity from a sequence of daily data is a challenging task. Our advanced LSTM model, which was trained to handle the imbalance in the data, was not able to reliably distinguish between five fine-grained levels of stress.

However, when the task was simplified to classifying stress into three broader levels (Low, Elevated, High), the model's performance improved significantly. It achieved an F1 Macro score of 0.57 and was particularly good at identifying students in the "Low" and "High" stress categories. This demonstrates the potential of using longitudinal passive data to monitor weekly well-being and flag individuals who may be experiencing clinically significant levels of stress.

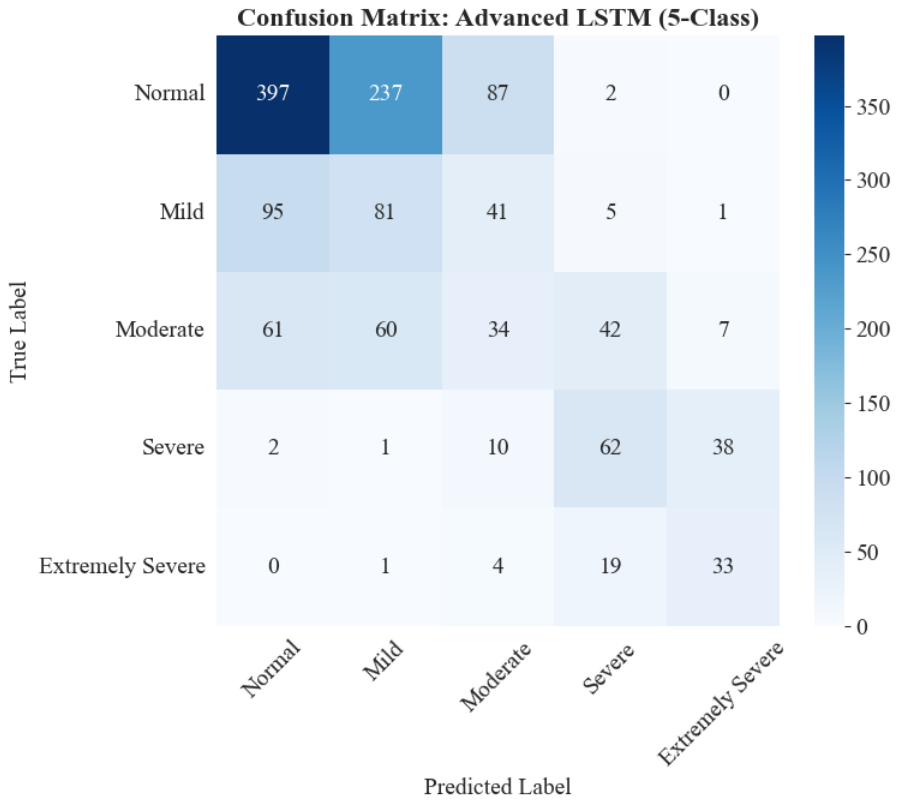
4.4 Usability Evaluation of the AI Companion (Well-Being Mate)

After developing the AI Companion, "WellBeing Mate," we wanted to understand how users felt about interacting with it. We conducted a usability study to get feedback on its design, its conversational ability, and its overall usefulness.

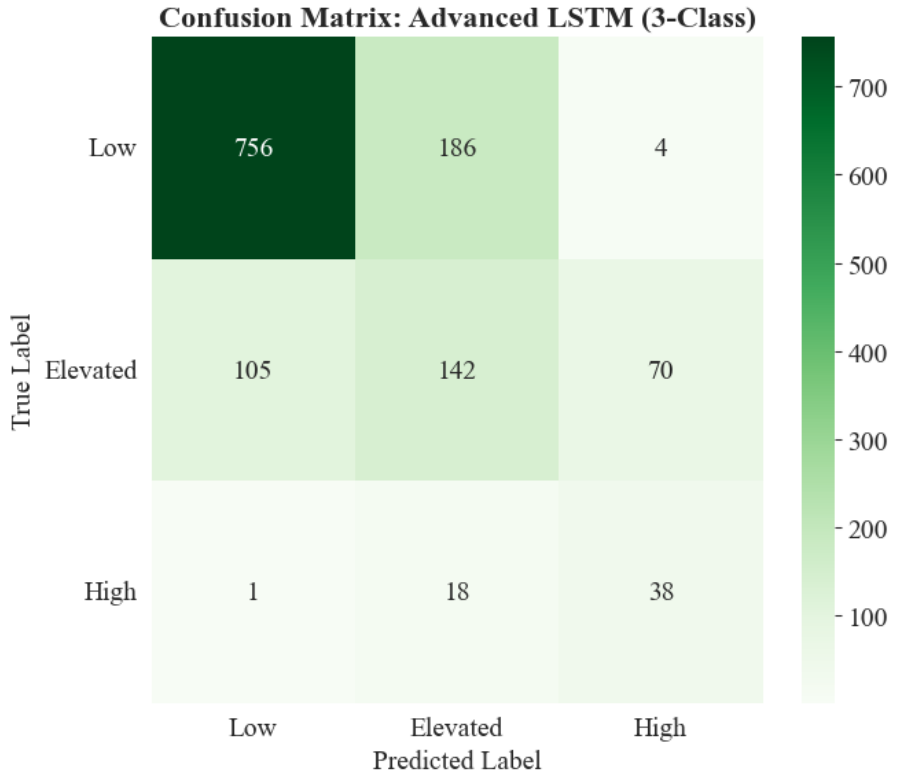
4.4.1 Participants and Method

A total of 27 university students participated in our evaluation. Each participant was asked to interact with WellBeing Mate over a period of several days. After their interaction period, they completed a standard questionnaire to measure the usability of the system. We used standard SUS questionnaires.

The SUS is a simple, ten-question survey that gives a single score representing the overall usability of a product, Table 4.5 [109]. The questions ask about things like ease of use, system complexity, and user confidence. The final score ranges from 0 to 100. A higher score means better usability.



(a) 5-Class Prediction



(b) 3-Class Prediction

Figure 4.8: Confusion Matrices for the **Advanced LSTM Model** on the Weekly Classification Tasks. The model's performance is visibly stronger on the simpler 3-class task (b).

No.	Questionnaire Item
1	I think that I would like to use this system frequently.
2	I found the system unnecessarily complex.
3	I thought the system was easy to use.
4	I think that I would need the support of a technical person to be able to use this system.
5	I found the various functions in this system were well integrated.
6	I thought there was too much inconsistency in this system.
7	I would imagine that most people would learn to use this system very quickly.
8	I found the system very cumbersome to use.
9	I felt very confident using the system.
10	I needed to learn a lot of things before I could get going with this system.

Table 4.5: The System Usability Scale (SUS) Questionnaire Items.

4.4.2 System Usability Scale (SUS) Score

The average SUS score for WellBeing Mate across all 27 participants was 87.31, Figure 4.9, and participant voting on 5-point Likert scales for individual SUS questions is shown in Figure 4.10; figures are generated using [110]. An SUS score of 87.31 is considered excellent. According to established research on SUS scores, any score above 80.3 is in the top 10% of all products tested and is given an "A" grade. This high score tells us that, overall, the participants found WellBeing Mate to be very easy to use, well-integrated, and not unnecessarily complex. It suggests that the design of the chat interface and the flow of the conversation were successful.

4.4.3 Key Qualitative Feedback from Participants

In addition to the numerical score from the SUS, we also collected written feedback from the participants about their experience. This qualitative data helps us understand why the usability score was high and what specific features users liked. We found several common themes in their responses:

- **Feeling of Being Remembered:** Many participants mentioned that they were pleasantly surprised when WellBeing Mate remembered details from previous conversations. One user wrote, "I was having a stressful week with my exam, and the next day, WellBeing Mate asked me how my exam went. It felt like I was talking to a friend who actually listens." This shows that the long-term memory feature was very effective at building a connection.

- **Empathetic and Non-Judgmental Tone:** Participants felt that WellBeing Mate provided a safe space to share their thoughts. A common piece of feedback was that the Companion's tone was "gentle" and "supportive." One participant said, "It's easier to talk about feeling stressed with WellBeing Mate because I know it will

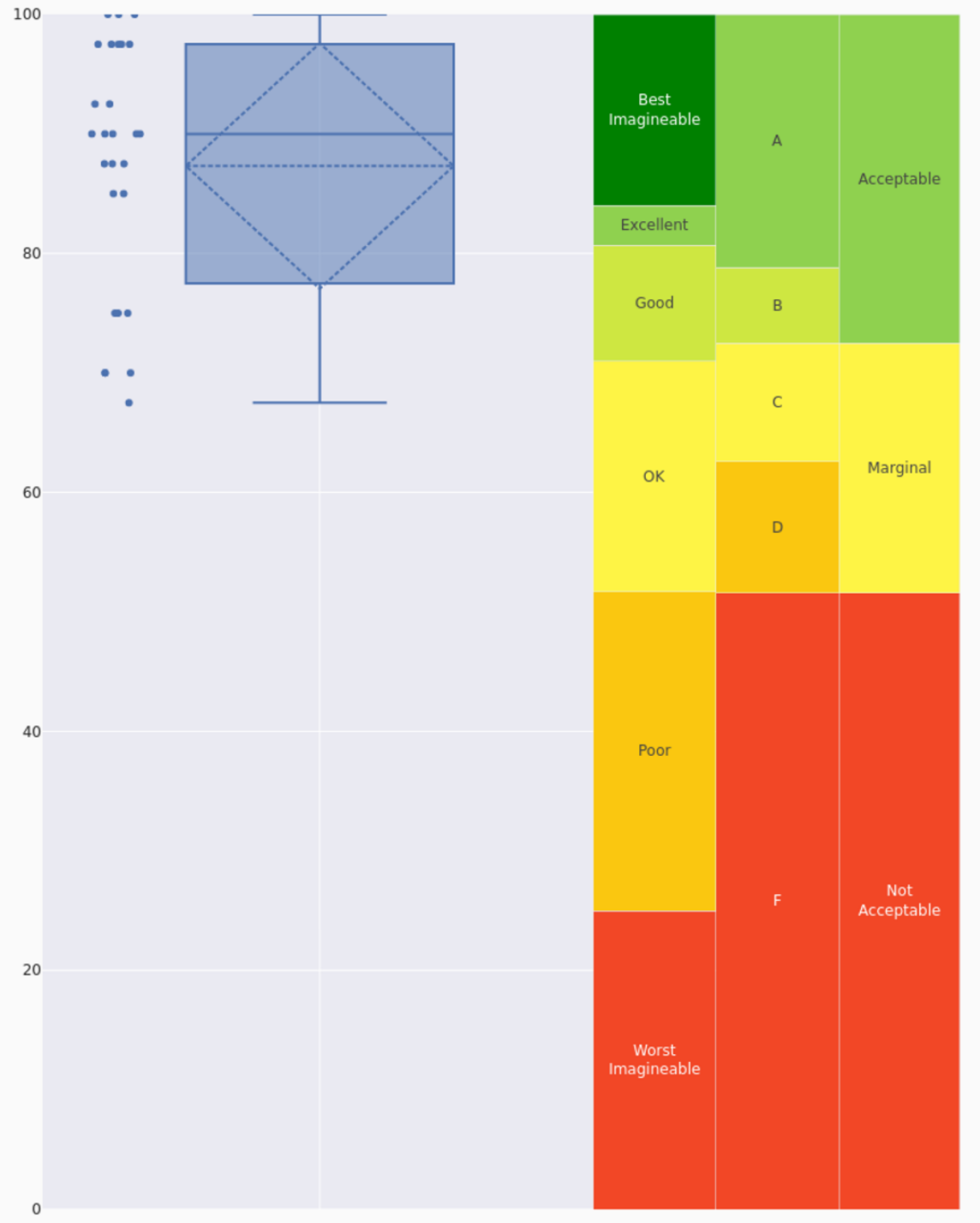


Figure 4.9: Usability Score of Our AI Companion.

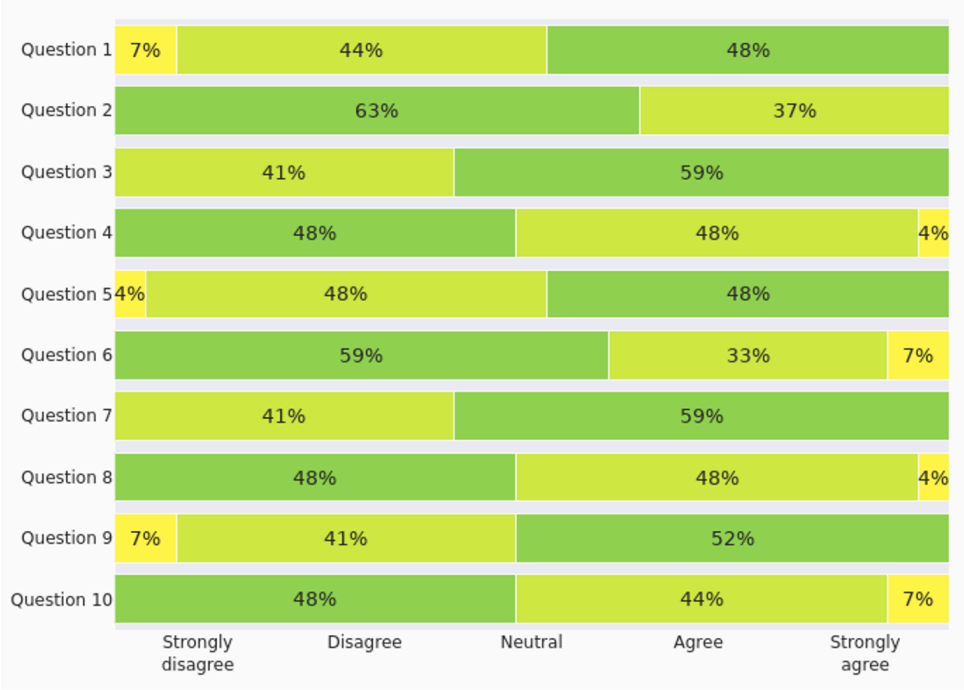


Figure 4.10: Participant voting on 5-point Likert scales for individual SUS questions.

not judge me. It just listens." This suggests that the base prompt and the data-informed conversational style were successful.

- **Personalized Greetings:** The feature where WellBeing Mate provides a personalized greeting based on past conversations or predicted stress was highly valued. Users reported that it made opening the app feel more welcoming and personal. "It was not just the same 'Hello' every day. It felt like it knew how my day was going," a participant commented.

- **Simplicity of the Interface:** Users found the chat interface to be clean and straightforward. They did not feel overwhelmed by buttons or options. The focus was entirely on the conversation, which they appreciated.

The results of our usability study strongly suggest that the AI Companion, WellBeing Mate, is a well-designed and highly usable feature of our application. The excellent SUS score of 87.31, combined with the positive qualitative feedback, indicates that users found the system easy to use and, more importantly, felt a personal connection with it. The data-informed approach, where WellBeing Mate uses memory and passive data to personalize conversations, appears to be a key reason for its success in creating an engaging and supportive user experience.

Chapter 5

Discussion

In this chapter, we discuss the deeper meaning of our findings. We will explore what our results imply, acknowledge the challenges and limitations of our study, and consider potential threats to the validity of our study.

5.1 Summary of Key Findings

Before we discuss the broader meaning of our work, it is helpful to review our most important results. This study produced several key findings:

1. **Daily Stress is Predictable:** We successfully trained machine learning models to predict a student's daily stress. Our best model, a RF, was able to correctly classify a student's stress into three categories (Low, Medium, High) with 74% accuracy. This is a strong and significant result.

2. **Weekly Stress Can Be Classified:** Predicting weekly stress was more difficult, but still possible. Our advanced LSTM model was able to classify a student's weekly stress severity into three levels (Low, Elevated, High). Most importantly, it correctly identified students in the "High" stress category 67% of the time (a recall of 0.67).

3. **The AI Companion was Well-Received:** Our AI-powered chatbot, "WellBeing Mate," received very positive feedback from students. They found it easy to use, helpful, and appreciated its ability to remember past conversations, which made the experience feel personal and supportive.

5.2 Principal Implications

Our study has several important implications for researchers and developers in the field of mental health technology.

5.2.1 Smartphones are Viable Tools for Mental Health Monitoring

Our success in predicting daily stress with 74% accuracy confirms that smartphones are powerful tools for understanding mental well-being. It shows that we do not always need to ask someone "How are you?" to get a sense of their mental state. By looking at their digital and physical "footprints" (their daily behaviors like sleep and study, which our models found to be key predictors), we can get valuable clues. This opens the door to creating systems that can offer help at the right moment, without being intrusive.

5.2.2 Personalization is Key for AI Companions

The high usability score of our AI Companion, WellBeing Mate, was not just because it was a chatbot. The positive feedback from users showed that they valued its ability to **remember** and **personalize** the conversation. Knowing that WellBeing Mate would remember a previous chat about an exam, or that its greeting would be gentler on a stressful day, made the experience feel more human and less robotic. This implies that for future mental health chatbots, simply having a good language model is not enough. The most successful companions will be those that are deeply integrated with the user's personal context.

5.2.3 A Model for Future Digital Health Apps

The architecture of our application, combining a robust passive sensing backend with an engaging, data-informed (stress level and past conversations) AI frontend, can serve as a model for other digital health tools. We have shown that it is possible to create a single app that serves two purposes: it can be a powerful data collection instrument for researchers, and at the same time, a genuinely helpful and supportive tool for the end-user.

5.3 Broader Impacts and Ethical Considerations

Beyond our technical findings, it is important to consider the wider impact of this kind of technology.

5.3.1 Social Effect

A tool like this has the potential for positive social impact. It could make mental health support more accessible, especially for young people who are comfortable with technology and may be hesitant to seek traditional help. The AI Companion, WellBeing

Mate, offers a private, non-judgmental space for students to talk, which could help reduce the stigma around mental health. However, we must also be careful. An over-reliance on digital tools could potentially reduce face-to-face interaction. The goal should be to use this app as a support tool, not a replacement for human connection.

5.3.2 Environmental Impact and Sustainability

The environmental impact of a software application is mainly related to the energy it uses. Our app was designed with sustainability in mind. By using "smart" GPS that only turns on during movement and by using efficient background tasks (WorkManager), we tried to minimize the app's battery consumption. Less battery use means the phone needs to be charged less often, which saves a small amount of electricity. On a larger scale, if millions of people used such an app, this focus on energy efficiency would become very important.

5.3.3 Addressing Ethical Issues

This research involves highly sensitive personal data, so ethics are extremely important. Our main ethical principle was privacy-by-design. All data was collected and stored securely on the user's own device. No data was uploaded to a server without their clear knowledge. Before participating, every student was given a detailed consent form that explained exactly what data was being collected and how it would be used. The AI Companion was also programmed with safety in mind; it was designed to never give medical advice and to gently guide users to seek professional help if they expressed severe distress. Any future development of this technology must continue to prioritize user privacy, consent, and safety above all else.

5.4 Limitations and Challenges

No study is perfect, and it is important to recognize the limitations of our work.

5.4.1 Data from a Specific Population

Our study was conducted with university students from a single university in Bangladesh. The patterns of stress and behavior we found might be very specific to this group. For example, "academic stress" was a major factor for our participants. The results might be very different for working adults, older people, or students in another country. Therefore, we must be careful not to generalize our findings to everyone.

5.4.2 Technical Challenges of Real-World Sensing

Collecting data 24/7 on a smartphone is very difficult. Even though we designed our app to be robust, we faced challenges. Some phone brands have aggressive battery-saving software that can stop the app from running. Also, sensor data can sometimes be noisy or missing (for example, if GPS cannot get a signal indoors). We had to develop methods to clean this data, but it is a constant challenge.

5.4.3 Privacy and User Trust

Our app collects a lot of sensitive, personal information. We stored all data on the user's phone to protect their privacy. However, for the data to be used in a larger study, it would need to be uploaded to a server. This creates a big responsibility to ensure the data is encrypted and handled ethically. Building and maintaining user trust is perhaps the biggest challenge for any project like this.

5.5 Threats to Validity

Threats to validity are factors that could make our conclusions incorrect. We must consider them carefully.

5.5.1 Self-Selection Bias

The students who agreed to participate in our study might be different from those who did not. For example, students who are more interested in technology or more open about their mental health might have been more likely to sign up. This means our group of participants might not represent the entire student population.

5.5.2 The Hawthorne Effect (Changing Behavior)

It is possible that some students changed their behavior because they knew they were being monitored. For example, a student might have used their phone less or walked more because they knew the app was tracking them. This could affect the accuracy of our data.

5.5.3 Self-Report Bias

The "ground truth" for our stress models was the students' own self-reports from the surveys and diaries. People are not always perfect at reporting their feelings. They might forget how they felt, or they might not want to admit they felt very stressed. This can introduce errors into the data that our models were trained on.

By acknowledging these limitations and threats, we can be more confident in what we have learned and also provide a clear path for other researchers to improve upon our work in the future.

Chapter 6

Conclusion and Future Work

This study set out to explore the potential of using smartphones to understand and support the mental well-being of university students. We successfully designed, developed, and evaluated a complete system that moves from passive data collection to a personalized, AI-driven user experience. This final chapter summarizes our key contributions, discusses the broader impacts of this technology, and looks ahead to the exciting possibilities for future work.

6.1 Conclusion

In this research, we have demonstrated that a smartphone can be transformed into a powerful instrument for digital phenotyping. Our key contributions are threefold.

- **Passively collected behavioral data:** We showed that passively collected behavioral data is meaningfully connected to a student's self-reported stress. Our machine learning models were able to predict weekly and daily stress with high accuracy, confirming that digital footprints can reflect real-world well-being.
- **Mobile sensing platform:** We developed a robust and usable mobile sensing platform that successfully collected rich, multi-modal data in a challenging real-world environment. The application's architecture proved stable enough for continuous, long-term data collection.
- **AI Well-being Companion:** Perhaps most significantly, we created an innovative, data-informed AI Well-being Companion. By using passive data to provide context, our chatbot, WellBeing Mate, was able to offer a personalized and empathetic experience. The excellent usability results confirmed that users found WellBeing Mate to be an engaging and supportive tool, validating our human-centered design approach.

In summary, this project has presented a successful end-to-end pipeline: from collecting raw sensor data, to building predictive models, to using those models to power an interactive and supportive application.

6.2 Future Work

This project serves as a strong foundation for many exciting future directions.

6.2.1 Moving from Prediction to Intervention

The next logical step is to use our stress prediction model to create real-time, personalized interventions. For example, if the app detects a high probability of stress, it could proactively send a notification suggesting a calming activity, such as a short walk, a breathing exercise, or listening to music (the smartband can be used as a supportive tool for passively sensing these and predicting real-time stress).

6.2.2 Enhancing the AI Companion

The capabilities of WellBeing Mate can be greatly expanded. Future versions could help users set and track personal well-being goals (e.g., "I want to try to get more sleep this week"). It could also be trained to recognize more complex emotional states and provide an even wider range of supportive conversational strategies.

6.2.3 Expanding Data Sources and Populations

To make our models more accurate and generalizable, we could incorporate additional data sources, such as sleep patterns from a wearable device or calendar information about a student's schedule. Furthermore, deploying this application with different populations, such as high school students, working professionals, or older adults, would help us understand how daily behaviors and stress manifest across different life stages.

6.2.4 Long-Term Impact Studies

The most important future work would be to conduct a long-term study. By having users interact with the app and WellBeing Mate over several months, we could begin to answer the most critical question: *Can a tool like this actually lead to a measurable improvement in a person's long-term mental well-being?*

We believe that the future of digital health lies in creating systems that are not just smart but also empathetic, ethical, and truly personal. Our work represents one step forward on that promising path.

Appendix

Publications From This Research & Relevant Work

1. N. Halder, J. H. Setu, L. Rafid, A. Islam, and M. A. Amin, "Smartwatch-Based Human Stress Diagnosis Utilizing Physiological Signals and LSTM-Driven Machine Intelligence," 2024 Advances in Science and Engineering Technology International Conferences (ASET), pp. 01–08, 2024 [41].

2. N. Masrur, N. Halder, S. Rashid, J. H. Setu, A. Islam, and T. Ahmed, "Performance Analysis of Ensemble and DNN Models for Decoding Mental Stress Utilizing ECG-Based Wearable Data Fusion," 2024 IEEE International Black Sea Conference on Communications and Networking (BlackSeaCom), pp. 276–279, 2024 [111].

3. S. T. Pasha, N. Halder, T. Badrul, J. H. Setu, A. Islam, and M. Z. Alam, "Physiological Signal Data-Driven Workplace Stress Detection Among Healthcare Professionals Using BiLSTM-AM and Ensemble Stacking Models," 2024 Advances in Science and Engineering Technology International Conferences (ASET), pp. 1–10, 2024 [17].

4. M. Alam, J. M. Sadik-Ul Islam Smaron, M. Tashfiq Islam, N. Halder, A. Islam, and M. A. Amin, "User-Centered Design and Evaluation of a Mobile Mental Health App for Stress Management and Relaxation," 2025 5th International Conference on Emerging Smart Technologies and Applications (eSmarTA), pp. 1–8, 2025 [112].

5. N. Halder, A. Alam, J. H. Setu, A. Mahmood, A. Islam, and M. A. Amin, "Analyzing Machine Learning Models for Multiclass Classification of Human Stress Levels: Comparison Between Balanced and Imbalanced Data," 2025 17th International Conference on Knowledge and Smart Technology (KST), pp. 104–109, 2025 [113].

6. N. Halder, A. Alam, J. H. Setu, A. Islam, and M. A. Amin, "Emotion Recognition in Bangla Text: An Ensemble Approach with Data Augmentation using BanglaBERT and MultiBERT," 2025 17th International Conference on Computer and Automation Engineering (ICCAE), pp. 6–11, 2025 [114].

7. N. Halder, T. P. Proma, J. H. Setu, A. Noor, A. Islam, and M. A. Amin, "Using Transformers for Emotion Recognition in Bangla Text: A Comparative Study of MultiBERT and BanglaBERT with Data Augmentation," 2024 International Conference on Machine Learning and Applications (ICMLA), pp. 1657–1662, 2024 [115].

Bibliography

[1] S. Cohen, D. Janicki-Deverts, and G. E. Miller, "Psychological stress and disease," *JAMA*, vol. 298, no. 14, pp. 1685–1687, 2007.

[2] E. Y. Yuen, W. Liu, I. N. Karatsoreos, J. Feng, B. S. McEwen, and Z. Yan, "Acute stress enhances glutamatergic transmission in prefrontal cortex and facilitates working memory," *PNAS*, vol. 106, no. 33, pp. 14075–14079, 2009.

[3] S. Melamed, T. Kushnir, and A. Shirom, "Burnout and risk factors for cardiovascular diseases," *Behav Med.*, vol. 18, no. 2, pp. 53–60, 1992.

[4] A. Weber and A. Jackel-Reinhard, "Burnout syndrome: A disease of modern societies?," *Occup Med.*, vol. 50, no. 7, pp. 512–517, 2000.

[5] R. Bianchi, D. Truchot, E. Laurent, R. Brisson, and I. S. Schonfeld, "Is burnout solely job-related? a critical comment," *Scandanavian J Psychol.*, vol. 55, no. 4, pp. 357–361, 2014.

[6] The American Institute of Stress, "Financial costs of job stress," 2016. Accessed Feb. 1, 2016.

[7] American Psychological Association, "Stress in america 2007," 2007. PsycEXTRA Dataset.

[8] A. Baum, "Stress, intrusive imagery, and chronic distress," *Health Psychol*, vol. 9, no. 6, pp. 653–675, 1990. PMID: 2286178.

[9] S. Moylan, M. Berk, O. M. Dean, Y. Samuni, L. J. Williams, A. O'Neil, and et al., "Oxidative & nitrosative stress in depression: why so much stress?," *Neurosci Biobehav R*, vol. 45, pp. 46–62, 2014.

[10] A. Steptoe and M. Kivimäki, "Stress and cardiovascular disease," *Nat Rev Cardiol*, vol. 9, no. 6, pp. 360–370, 2012.

[11] A. J. Tomiyama, "Stress and obesity," *Ann R Psychol*, vol. 70, no. 1, pp. 703–718, 2019.

[12] J. Hunt and D. Eisenberg, "Mental health problems and help-seeking behavior among college students," *J. Adolesc. Health*, vol. 46, no. 1, pp. 3–10, 2010.

[13] M. Firoz, M. M. Islam, M. Shidujaman, A. Islam, and M. T. Habib, "University student's mental stress detection using machine learning," in *Seventh International Conference on Mechatronics and Intelligent Robotics (ICMIR 2023)* (S. Patnaik and T. Shen, eds.), vol. 12779, p. 127792O, International Society for Optics and Photonics, SPIE, 2023.

[14] S. G. Hofmann, M. G. Newman, A. Ehlers, and W. T. Roth, "Psychophysiological differences between subgroups of social phobia," *J. Abnorm. Psychol.*, vol. 104, no. 1, pp. 224–231, 1995.

[15] S. M. Turner, D. C. Beidel, and K. T. Larkin, "Situational determinants of social anxiety in clinic and nonclinic samples: physiological and cognitive correlates," *J. Consult. Clin. Psychol.*, vol. 54, no. 4, pp. 523–527, 1986.

[16] Editorial, "Mind matters," *Nature*, vol. 532, p. 6, 2016.

[17] S. T. Pasha, N. Halder, T. Badrul, J. H. Setu, A. Islam, and M. Z. Alam, "Physiological signal data-driven workplace stress detection among healthcare professionals using bilstm-am and ensemble stacking models," *2024 Advances in Science and Engineering Technology International Conferences (ASET)*, pp. 1–10, 2024.

[18] World Health Organization, "Preventing suicide: a global imperative," tech. rep., World Health Organization, 2014.

[19] N. Kawakami and A. Tsutsumi, "The stress check program: a new national policy for monitoring and screening psychosocial stress in the workplace in japan," *J. Occupat. Health*, vol. 58, no. 1, pp. 1–6, 2016.

[20] A. Bittner, R. D. Goodwin, H. U. Wittchen, K. Beesdo, M. Höfler, and R. Lieb, "What characteristics of primary anxiety disorders predict subsequent major depressive disorder?," *J. Clin. Psychiatry*, vol. 65, no. 5, pp. 618–626, 2004.

[21] N. M. Docherty, A. St-Hilaire, J. M. Aakre, J. P. Seghers, A. McCleery, and M. Divilbiss, "Anxiety interacts with expressed emotion criticism in the prediction of psychotic symptom exacerbation," *Schizophr. Bull.*, vol. 37, no. 3, pp. 611–618, 2011.

[22] L. Krabbendam, I. Janssen, M. Bak, R. V. Bijl, R. de Graaf, and J. van Os, "Neuroticism and low self-esteem as risk factors for psychosis," *Social Psychiatry Psychiat. Epidemiol.*, vol. 37, pp. 1–6, 2002.

[23] S. S. Olin and S. A. Mednick, "Risk factors of psychosis: identifying vulnerable populations premorbidly," *Schizophr. Bull.*, vol. 22, no. 2, pp. 223–240, 1996.

[24] D. C. Beidel, S. M. Turner, and C. V. Dancu, "Physiological, cognitive and behavioral aspects of social anxiety," *Behav. Res. Ther.*, vol. 23, no. 2, pp. 109–117, 1985.

[25] R. Vineetha, K. M. Pai, M. Vengal, K. Gopalakrishna, and D. Narayanakurup, "Usefulness of salivary alpha amylase as a biomarker of chronic stress and stress related oral mucosal changes - a pilot study," *J. Clin. Exp. Dentist.*, vol. 6, no. 2, pp. e132–e137, 2014.

[26] D. T. Lardier, C. Y. Lee, J. M. Rodas, P. Garcia-Reid, and R. J. Reid, "The effect of perceived college-related stress on depression, life satisfaction, and school satisfaction: the coping strategies of hispanic college students from a hispanic serving institution," *Educ Urban Soc.*, vol. 52, no. 8, pp. 1204–1222, 2020.

[27] A. Rezaei and E. M. Jeddi, "Relationship between wisdom, perceived control of internal states, perceived stress, social intelligence, information processing styles and life satisfaction among college students," *Curr Psychol.*, vol. 39, pp. 927–933, 2018.

[28] J. Futo, "Dealing with mental health issues on campus starts with early recognition and intervention," *Campus Law Enforc J.*, vol. 41, no. 3, p. 22, 2011.

[29] P. K. Jonason, D. Talbot, M. L. Cunningham, and J. Chonody, "Higher-order coping strategies: who uses them and what outcomes are linked to them," *Pers Individ Differ.*, vol. 155, p. 109755, 2020.

[30] A. F. Jorm, "Mental health literacy: empowering the community to take action for better mental health," *Am Psychol.*, vol. 67, no. 3, pp. 231–243, 2012.

[31] L. Lu, "University transition: Major and minor life stressors, personality characteristics and mental health," *Psychol Med.*, vol. 24, pp. 81–87, 1994.

[32] P. Verger, J.-B. Combes, V. Kovess-Masfety, M. Choquet, V. Guagliardo, F. Rouillon, and et al., "Psychological distress in first year university students: Socioeconomic and academic stressors, mastery, and social support in young men and women," *Soc Psychiatry Psychiatr Epidemiol.*, vol. 44, pp. 643–650, 2008.

[33] S. H. Hamaideh, "Stressors and reactions to stressors among university students," *Int J Soc Psychiatry.*, vol. 57, pp. 69–80, 2011.

[34] J. S. Ribeiro, R. Pereira, I. V. Freire, B. G. D. Oliveira, C. A. Casotti, and E. N. Boery, "Stress and quality of life among university students: A systematic literature review," *Health Professions Educ.*, vol. 4, pp. 70–77, 2018.

[35] J. Sharp and S. Theiler, "A review of psychological distress among university students: pervasiveness, implications and potential points of intervention," *Int J Advancement Counselling.*, vol. 40, pp. 193–212, 2018.

[36] S. Liu, M. Zhu, D. J. Yu, A. Rasin, and S. D. Young, "Using real-time social media technologies to monitor levels of perceived stress and emotional state in college students: a web-based questionnaire study," *JMIR Ment Health*, vol. 4, no. 1, p. e2, 2017.

[37] R. Beiter, R. Nash, M. McCrady, D. Rhoades, M. Linscomb, M. Clarahan, and et al., "The prevalence and correlates of depression, anxiety, and stress in a sample of college students," *J Affect Disord*, vol. 173, pp. 90–96, 2015.

[38] C. A. Darling, L. M. McWey, S. N. Howard, and S. B. Olmstead, "College student stress: the influence of interpersonal relationships on sense of coherence," *Stress Health*, vol. 23, no. 4, pp. 215–229, 2007.

[39] J. Lee and S. Jang, "An exploration of stress and satisfaction in college students," *Service Market Q*, vol. 36, no. 3, pp. 245–260, 2015.

[40] R. Wang, F. Chen, Z. Chen, T. Li, G. Harari, S. Tignor, X. Zhou, D. Ben-Zeev, and A. T. Campbell, "StudentLife: assessing mental health, academic performance and behavioral trends of college students using smartphones," in *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pp. 3–14, 2014.

[41] N. Halder, J. H. Setu, L. Rafid, A. Islam, and M. A. Amin, "Smartwatch-based human stress diagnosis utilizing physiological signals and lstm-driven machine intelligence," *2024 Advances in Science and Engineering Technology International Conferences (ASET)*, pp. 01–08, 2024.

[42] N. C. Jacobson and Y. J. Chung, "Passive sensing of prediction of moment-to-moment depressed mood among undergraduates with clinical levels of depression sample using smartphones," *Sensors*, vol. 20, no. 12, p. 3572, 2020.

[43] B. Reeder and A. David, "Health at hand: A systematic review of smart watch uses for health and wellness," *J. Biomed. Inform.*, vol. 63, pp. 269–276, 2016.

[44] S. Ancoli-Israel, R. Cole, C. Alessi, M. Chambers, W. Moorcroft, and C. P. Pollak, "The role of actigraphy in the study of sleep and circadian rhythms," *Sleep*, vol. 26, pp. 342–392, 2003.

[45] D. Bravata, C. Smith-Spangler, V. Sundaram, A. L. Gienger, N. Lin, R. Lewis, C. Stave, I. Olkin, and J. Sirard, "Using pedometers to increase physical activity and improve health: A systematic review," *J. Am. Med. Assoc.*, vol. 298, pp. 2296–2304, 2007.

[46] J. B. Roig, N. Gilson, A. Ribera, R. C. Espinosa, and S. Trost, "Measuring and influencing physical activity with smartphone technology: A systematic review," *Sports Med.*, vol. 44, no. 5, pp. 671–686, 2014.

[47] S. R. Steinhubl, E. D. Muse, and E. J. Topol, "Can mobile health technologies transform health care?," *J. Am. Med. Assoc.*, vol. 310, no. 22, pp. 2395–2396, 2013.

[48] W. Khan, Y. Xiang, M. Aalsalem, and Q. Arshad, "Mobile phone sensing systems: A survey," *IEEE Commun. Surv. Tutorials*, vol. 15, pp. 402–427, 2013.

[49] Pew Research Center, "The smartphone difference," 2015.

[50] D. Luxton, R. A. McCann, N. Bush, M. C. Mishkind, and G. Reger, "mhealth for mental health: Integrating smartphone technology in behavioral healthcare," *Prof. Psychol.: Res. Practice*, vol. 42, pp. 505–512, 2011.

[51] M. Neary and S. M. Schueller, "State of the field of mental health apps," *Cogn Behav Pract.*, vol. 25, pp. 531–537, 2018.

[52] B. Holtz, A. McCarroll, and K. Mitchell, "Perceptions and attitudes toward a mobile phone app for mental health for college students: qualitative focus group study," *JMIR Formative Res.*, vol. 4, p. e18347, 2020.

[53] I. Montagni, T. Cariou, T. Feuillet, E. Langlois, and C. Tzourio, "Exploring digital health use and opinions of university students: Field survey study," *JMIR Mhealth Uhealth.*, vol. 6, p. e65, 2018.

[54] I. L. Ahuvia, J. Y. Sung, M. L. Dobias, B. D. Nelson, L. L. Richmond, B. London, and et al., "College student interest in teletherapy and self-guided mental health supports during the covid-19 pandemic," *J Am Coll Health.*, pp. 1–7, 2022.

[55] J. Bautista and S. M. Schueller, "Understanding the adoption and use of digital mental health apps among college students: Secondary analyses of a national survey," *JMIR Ment Health.*, vol. 10, p. e43942, 2023.

[56] A. Kern, V. Hong, J. Song, S. K. Lipson, and D. Eisenberg, "Mental health apps in a college setting: openness, usage, and attitudes," *mHealth*, vol. 4, p. 20, 2018.

[57] J. Melcher, E. Camacho, S. Lagan, and J. Torous, "College student engagement with mental health apps: analysis of barriers to sustained use," *J Am Coll Health.*, vol. 70, pp. 1819–1825, 2022.

[58] N. Lau, A. O'Daffer, S. Colt, J. Yi-Frazier, T. Palermo, E. McCauley, and A. Rosenberg, "Android and iphone mobile apps for psychosocial wellness and stress management: Systematic search in app stores and literature review," *JMIR mHealth and uHealth*, vol. 8, 2020.

[59] S. Paganini, E. Meier, Y. Terhorst, R. Wurst, V. Hohberg, D. Schultchen, J. Strahler, M. Wursthorn, H. Baumeister, and E.-M. Messner, "Stress management apps: Systematic search and multidimensional assessment of quality and characteristics," *JMIR mHealth and uHealth*, vol. 11, 2023.

[60] S. Coulon, C. Monroe, and D. West, "A systematic, multi-domain review of mobile smartphone apps for evidence-based stress management," *American Journal of Preventive Medicine*, vol. 51, no. 1, 2016.

[61] V. Sîrbu and O. E. David, "Efficacy of app-based mobile health interventions for stress management: A systematic review and meta-analysis of self-reported, physiological, and neuroendocrine stress-related outcomes," *Clinical Psychology Review*, vol. 114, 2024.

[62] C. Coelhoso, P. Tobo, S. Lacerda, A. Lima, C. Barrichello, E. Amaro, and E. H. Kozasa, "A new mental health mobile app for well-being and stress reduction in working women: Randomized controlled trial," *Journal of Medical Internet Research*, vol. 21, 2019.

[63] D. Martín, I. De La Torre, B. Garcia-Zapirain, M. López-Coronado, and J. J. P. C. Rodrigues, "Managing and controlling stress using mhealth: Systematic search in app stores," *JMIR mHealth and uHealth*, vol. 6, 2018.

[64] W. Hwang and H. Jo, "Evaluation of the effectiveness of mobile app-based stress-management program: A randomized controlled trial," *International Journal of Environmental Research and Public Health*, vol. 16, no. 21, p. 4270.

[65] S. Demirel, Y. Roke, A. Hoogendoorn, J. Hoefakker, K. Hoeberichts, and P. Van Harten, "Assessing the effectiveness of stapp@work, a self-management mobile app, in reducing work stress and preventing burnout: Single-case experimental design study," *Journal of Medical Internet Research*, vol. 26, 2023.

[66] S. H. Lovibond, "Manual for the depression anxiety stress scales," *Sydney psychology foundation*, 1995.

[67] A. Shvetcov, J. Kupper, W. Zheng, A. Slade, J. Han, A. Whitton, M. Spoelma, L. Hoon, K. Mouzakis, R. Vasa, S. Gupta, S. Venkatesh, J. Newby, and H. Christensen, "Passive sensing data predicts stress in university students: a supervised machine learning method for digital phenotyping," *Frontiers in Psychiatry*, vol. 15, 2024. Art. no. 1422027.

[68] A. Choi, A. Ooi, and D. Lottridge, "Digital phenotyping for stress, anxiety, and mild depression: Systematic literature review," *JMIR mHealth and uHealth*, vol. 12, no. 1, p. e40689, 2022.

[69] A. DaSilva, J. Huckins, R. Wang, W. Wang, D. Wagner, and A. Campbell, "Correlates of stress in the college environment uncovered by the application of penalized generalized estimating equations to mobile sensing data," *JMIR mHealth and uHealth*, vol. 7, no. 1, p. e12084, 2018.

[70] A. Sano, S. Taylor, A. W. McHill, A. J. Phillips, L. K. Barger, E. Klerman, and R. W. Picard, "Identifying objective physiological markers and modifiable behaviors for self-reported stress and mental health status using wearable sensors and mobile phones: Observational study," *Journal of Medical Internet Research*, vol. 20, no. 6, p. e10464, 2018.

[71] G. Aalbers, A. Hendrickson, M. Vanden Abeele, and L. Keijsers, "Smartphone-tracked digital markers of momentary subjective stress in college students: Idiographic machine learning analysis," *JMIR mHealth and uHealth*, vol. 11, p. e37469, 2022.

[72] J. Melcher, R. Hays, and J. Torous, "Digital phenotyping for mental health of college students: a clinical review," *Evidence-Based Mental Health*, vol. 23, no. 4, pp. 161–166, 2020.

[73] M. Boukhechba, P. Chow, K. Fua, B. Teachman, and L. Barnes, "Predicting social anxiety from global positioning system traces of college students: Feasibility study," *JMIR Mental Health*, vol. 5, no. 3, p. e10101, 2018.

[74] R. Wang, Z. Chen, T. Li, G. Harari, S. Tignor, X. Zhou, D. Ben-Zeev, and A. Campbell, "StudentLife: Using smartphones to assess mental health and academic performance of college students," in *Pervasive Computing*, pp. 2–8, Springer, Cham, 2017.

[75] Y. Luo, I. Deznabi, A. Shaw, N. Simsiri, T. Rahman, and M. Fiterau, "Dynamic clustering via branched deep learning enhances personalization of stress prediction from mobile sensor data," *Scientific Reports*, vol. 14, no. 1, p. 6115, 2024.

[76] E. Girousse and N. Vuillerme, "The use of passive smartphone data to monitor anxiety and depression among college students in real-world settings: Protocol for a systematic review," *JMIR Research Protocols*, vol. 11, no. 5, p. e38785, 2022.

[77] N. Etienne and E. Agu, "Investigating transfer learning of smartphone-sensed stress in university populations," in *2020 IEEE International Conference on Big Data (Big Data)*, pp. 3381–3390, 2020.

[78] N. Kadri, S. Turki, A. Ellouze, and M. Ksantini, "An hybrid deep learning approach for prediction and binary classification of student's stress," in *Intelligent Systems and Pattern Recognition*, pp. 327–340, Springer, Cham, 2022.

[79] N. C. Jacobson, B. Summers, and S. Wilhelm, "Digital biomarkers of social anxiety severity: Digital phenotyping using passive smartphone sensors," *Journal of Medical Internet Research*, vol. 22, no. 5, p. e16875, 2020.

[80] B. Egilmez, E. C. Poyraz, W. Zhou, G. Memik, P. Dinda, and N. Alshurafa, "US-tress: Understanding college student subjective stress using wrist-based passive sensing," in *2017 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*, pp. 581–586, 2017.

[81] S. Kang, W. Choi, C. Park, N. Cha, A. Kim, A. H. Khandoker, L. J. Hadjileontiadis, H. Kim, Y. Jeong, and U. Lee, "K-EmoPhone: A mobile and wearable dataset with in-situ emotion, stress, and attention labels," *Scientific Data*, vol. 10, no. 1, p. 370, 2023.

[82] Y. Fukazawa, T. Itoh, T. Okimura, Y. Yamashita, T. Maeda, and J. Ota, "Predicting anxiety state using smartphone-based passive sensing," *Journal of Biomedical Informatics*, vol. 93, p. 103151, 2019.

[83] Y. Acikmese and S. E. Alptekin, "Prediction of stress levels with LSTM and passive mobile sensors," *Procedia Computer Science*, vol. 158, pp. 243–250, 2019.

[84] T. Hoang, T. Dang, and N. Trang, "Personalized stress detection for university students using wearable devices," in *2025 19th International Conference on Ubiquitous Information Management and Communication (IMCOM)*, 2025.

[85] M. Gjoreski, H. Gjoreski, M. Luštrek, and M. Gams, "Automatic detection of perceived stress in campus students using smartphones," in *2015 International Conference on Intelligent Environments*, pp. 140–143, 2015.

[86] N. Khan, S. Qadir, G. Anjum, and N. Uddin, "StresSense: Real-time detection of stress-displaying behaviors," *International Journal of Medical Informatics*, vol. 185, p. 105401, 2024.

[87] M. A. Sheikh, M. Qassem, and P. A. Kyriacou, "Wearable, environmental, and smartphone-based passive sensing for mental health monitoring," *Frontiers in Digital Health*, vol. 3, p. 662811, 2021.

[88] R. Wang and et al., "StudentLife: Assessing mental health, academic performance and behavioral trends of college students using smartphones," in *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '14)*, pp. 1117–1128, 2014.

[89] J. Ahn, I. Jeong, S. Park, J. Lee, M. Jeon, S. Lee, G. R. Do, D.-E. Jung, and J. Park, "App-based ecological momentary assessment of problematic smartphone use during examination weeks in university students: 6-week observational study," *Journal of Medical Internet Research*, vol. 27, p. e69320, 2025.

[90] A. Bradley and A. Howard, "Stress and mood associations with smartphone use in university students: A 12-week longitudinal study," *Clinical Psychological Science*, vol. 11, no. 2, pp. 297–315, 2023.

[91] E. A. Sağbaş, S. Korukoglu, and S. Balli, "Stress detection via keyboard typing behaviors by using smartphone sensors and machine learning techniques," *Journal of medical systems*, vol. 44, no. 4, p. 68, 2020.

[92] H. Rashid, S. Mendu, K. Daniel, M. Beltzer, B. Teachman, M. Boukhechba, and L. Barnes, "Predicting subjective measures of social anxiety from sparsely collected mobile sensor data," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 4, no. 3, pp. 1–23, 2020.

[93] J.-P. Onnela and S. L. Rauch, "Harnessing smartphone-based digital phenotyping to enhance behavioral and mental health," *Neuropsychopharmacology*, vol. 41, pp. 1691–1696, 2016.

[94] R. Wang and et al., "SmartGPA: Sensing study and social behaviors related to academic performance." Technical Report: Dartmouth College, 2015. Based on UbiComp Adjunct Presentation.

[95] Y. S. Yang and et al., "Ecological momentary assessment using smartphone applications: A process, feasibility, and lessons learned in a clinical trial," *JMIR Mhealth and Uhealth*, vol. 6, p. e10452, 2018.

[96] S. Mengelkoch and et al., "Using ecological momentary assessments to study how daily fluctuations in stress, affect, and coping relate to health," *Annals of Behavioral Medicine*, vol. 57, pp. 90–101, 2023.

[97] A. M. Ali, A. T. Al-Mosa, M. S. Al-Shamri, and et al., "The depression anxiety stress scale-21: Development and validation/short forms (psychometric evaluation)," *Frontiers in Psychology*, vol. 12, p. 759325, 2021.

[98] R. H. Kim and et al., "Validating constructs of the DASS-21 in guam," *BMC Psychology*, vol. 12, p. 111, 2024.

[99] S. Farmer and et al., "Mobile phone ecological momentary assessment of daily stressors," *JMIR Formative Research*, vol. 1, p. e4, 2017.

[100] S. Wong and et al., "Addressing missing data in digital phenotyping: A methodological review and proposed framework," *Computers in Biology and Medicine*, vol. 132, p. 104332, 2021.

[101] D. W. Hosmer Jr, S. Lemeshow, and R. X. Sturdivant, *Applied logistic regression*. John Wiley & Sons, 2013.

[102] L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.

[103] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.

[104] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, no. 6088, pp. 533–536, 1986.

[105] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[106] D. Gholamiangonabadi, N. Kiselov, and K. Grolinger, "Deep neural networks for human activity recognition with wearable sensors: Leave-one-subject-out cross-validation for model selection," *IEEE Access*, vol. 8, pp. 133982–133994, 2020.

[107] S. Kunjan, T. S. Grummett, K. J. Pope, D. M. Powers, S. P. Fitzgibbon, T. Bastiampillai, M. Battersby, and T. W. Lewis, "The necessity of leave one subject out (loso) cross validation for eeg disease diagnosis," in *International conference on brain informatics*, pp. 558–567, Springer, 2021.

[108] D. M. Powers, "Evaluation: from precision, recall and f-measure to roc, informedness, markedness and correlation," *arXiv preprint arXiv:2010.16061*, 2020.

[109] J. Brooke *et al.*, "Sus-a quick and dirty usability scale," *Usability evaluation in industry*, vol. 189, no. 194, pp. 4–7, 1996.

[110] J. Blattgerste, J. Behrends, and T. Pfeiffer, "A web-based analysis toolkit for the system usability scale," in *Proceedings of the 15th International Conference on PErvasive Technologies Related to Assistive Environments*, pp. 237–246, 2022.

[111] N. Masrur, N. Halder, S. Rashid, J. H. Setu, A. Islam, and T. Ahmed, "Performance analysis of ensemble and dnn models for decoding mental stress utilizing ecg-based wearable data fusion," *2024 IEEE International Black Sea Conference on Communications and Networking (BlackSeaCom)*, pp. 276–279, 2024.

[112] M. Alam, J. M. Sadik-Ul Islam Smaron, M. Tashfiq Islam, N. Halder, A. Islam, and M. A. Amin, "User-centered design and evaluation of a mobile mental health app for stress management and relaxation," *2025 5th International Conference on Emerging Smart Technologies and Applications (eSmarTA)*, pp. 1–8, 2025.

[113] N. Halder, A. Alam, J. H. Setu, A. Mahmood, A. Islam, and M. A. Amin, "Analyzing machine learning models for multiclass classification of human stress levels: Comparison between balanced and imbalanced data," *2025 17th International Conference on Knowledge and Smart Technology (KST)*, pp. 104–109, 2025.

[114] N. Halder, A. Alam, J. H. Setu, A. Islam, and M. A. Amin, "Emotion recognition in bangla text: An ensemble approach with data augmentation using banglabert and multibert," *2025 17th International Conference on Computer and Automation Engineering (ICCAE)*, pp. 6–11, 2025.

[115] N. Halder, T. P. Proma, J. H. Setu, A. Noor, A. Islam, and M. A. Amin, "Using transformers for emotion recognition in bangla text: A comparative study of multibert and banglabert with data augmentation," *2024 International Conference on Machine Learning and Applications (ICMLA)*, pp. 1657–1662, 2024.