7-2019

StressMon: Large scale detection of stress and depression in campus environment using passive coarse-grained location data

Camellia ZAKARIA
Singapore Management University

StressMon: Large Scale Detection of Stress and Depression in Campus Environment using Passive Coarse-grained Location Data

CAMELLIA ZAKARIA

SINGAPORE MANAGEMENT UNIVERSITY

2019

StressMon: Large Scale Detection of Stress and Depression in Campus Environment using Passive Coarse-grained Location Data

by

Camellia Zakaria

Submitted to School of Information Systems in partial fulfilment of the requirements for the Degree of Doctor of Philosophy in Information Systems

Dissertation Committee:

Rajesh BALAN(Supervisor/Chair)
Associate Professor of Information Systems
Singapore Management University

Youngki LEE(Co-supervisor)
Assistant Professor of Computer Science and Engineering
Seoul National University

Kenneth GOH
Assistant Professor of Strategic Management
Singapore Management University

Archan MISRA
Professor of Information Systems
Singapore Management University

Julie KIENTZ
Associate Professor of Human Centered Design & Engineering
University of Washington

Singapore Management University
2019

I hereby declare that this PhD dissertation is my original work and it has been written by me in its entirety. I have duly acknowledged all the sources of information which have been used in this dissertation.

This PhD dissertation has also not been submitted for any degree in any university previously.

Camellia Zakaria

17th July 2019

StressMon: Large Scale Detection of Stress and Depression in Campus Environment using Passive Coarse-grained Location Data

Camellia Zakaria

ABSTRACT

The rising mental health illnesses of severe stress and depression is of increasing concern worldwide. Often associated by similarities in symptoms, severe stress can take a toll on a person's productivity and result in depression if the stress is left unmanaged. Unfortunately, depression can occur without any feelings of stress. With depression growing as a leading cause of disability in economic productivity, there has been a sharp rise in mental health initiatives to improve stress and depression management. To offer such services conveniently and discreetly, recent efforts have focused on using mobile technologies. However, these initiatives usually require users to install dedicated apps or use a variety of sensors, making such solutions hard to scale. Moreover, they emphasise sensing individual factors and overlook 'physical social interaction' that plays a significant role in influencing stress and depression. This thesis presents *StressMon*, a monitoring system that can easily scale across entire campuses by passively sensing location information directly from the WiFi infrastructure.

This dissertation explores how, by using only single-attribute location information, mobility features can be comprehensively extracted to represent individual behaviours to detect stress and depression accurately; it is important to note that this is without requiring explicit user actions or software installation on client devices. To overcome the low-dimensional data, *StressMon* additionally infers physical group interaction patterns from a group detector system. First, I investigate how mobility features can be exploited to better capture the dynamism of natural human behaviours indicative of stress and depression. Then, I present the framework to detect stress and depression accurately, albeit separately. In a supplementary effort, I demonstrate how optimising *StressMon* with group-based mobility features greatly enhances the performance of stress detection, and conversely, individual-based features improve depression detection. To extensively validate the system, I conducted three different semester-long longitudinal studies with different groups of undergraduate students at separate times, totalling up to 108 participants. Finally, this dissertation documents the differences learned in understanding stress and depression from a qualitative perspective.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENT

Iqbal, you taught me to be selfless and to touch people in the most positive way. To my family, you nurtured me to what I am today. Tim, you listened even when nothing I said was important. Raj, you made my campus our chillout place because I was always too busy. I appreciate and love all of you.

I would like to express my sincerest gratitude to my advisors, Rajesh Balan and Youngki Lee, for being tremendous mentors. You found me lost, and pushed me to my limits and beyond. I have so much to improve in my skills and my self. I thank you for your patience and guidance. I would also like to thank my committee members, Kenneth Goh, Julie Kientz and Archan Misra for their time and invaluable advice. Archan, thank you for grudgingly participating in my defense. :')

A special mention to Richard Davis, Zachary Walker, William Tov, Jennifer Mankoff, Quentin Roy, Aniket Pujara and Ritesh Kumar. Thank you for the opportunity to work with and learn from you. Meera and Kenny, thank you for always looking out for me. To all my university buddies, I cherish our friendship. Finally, I would like to thank the SIS Research Center, LiveLabs and Living Analytics Research Centers for supporting my education here at Singapore Management University and providing opportunities at Carnegie Mellon University and Seoul National University.

It has been a long journey, and I take from the friendship of all to finally say these words out loud, "Stress no more!"

CHAPTER 1

INTRODUCTION

Stress and depression are mental health illnesses most commonly associated with challenges in everyday life. Stress is often underestimated as a normal reaction to daily pressures with at least 74% of people reportedly struggling to cope with stress [1]. Depression is a global phenomenon affecting more than 300 million people of all ages [2]. Much research has progressed towards understanding the relationship between stress and depression [3]. Stress, depending on its severity – ranging from acute to chronic – can result in a particular set of consequences on one's body, mood and behaviour. For example, acute stress typically causes irritability, but chronic stress causes aggressiveness leading to social isolation [4, 5]. In fact, chronic stress can lead to more serious health consequences of depression [6–8]. While depression is generally preceded by chronic stress, depression can occur without an individual feeling stressed at all [3]. Hence, depression remains an independent health condition, and its treatment must be addressed directly [9, 10]. Recently, depression is reported by the World Health Organisation (WHO) as one of the leading causes of lost economic productivity estimated to cost the global economy US $1 trillion each year [11], originating from reasons such as absenteeism from unmanaged stress [4]. To reduce this impact, it has been shown that treating symptoms of depression can reduce work absenteeism[1] [11]. With work increasingly structured via teams to capitalise on various skill sets and experiences, an

[1]Note: Without losing its authentic meaning, I refer to 'work' as activities that require skills, time and effort to complete.

individual's stress could be as a result of social factors [12, 13]. Indeed, the underlying mechanisms causing stress are much more complex to understand as individuals are frequently working in groups where group dynamics and social interactions can greatly influence the stress levels of group members [12]. For example, being in supportive groups allows individuals to receive peer support which greatly helps reduce their stress levels [14]. Having a strong support system allows one to develop a more positive emotional response towards the group in stressful work situations [15]. The serious impact of stress and depression have raised much need for mental health support to be provided in workplaces [4, 11, 13, 16]. For instance, the inability of individuals to cope with stress is commonly found to increase work absenteeism [4], while the treatment of depression reduces work absenteeism [11]. Prior studies have shown that recovering from stress to a normal state is much easier, in terms of the length of time and treatment required, compared to depression [3] and that treating depression early can result in earlier relative recovery times. Thus, there is great merit in detecting individuals' stress and depression early. It is noteworthy that this dissertation aims to address the binary classification problem of detecting individuals likely to experience (1) severe stress over 6-days interval and (2) depression over 15-days interval in a work setting; however, it does not reveal the underlying causes of those conditions[2].

Much work in psychology, small group and systems research has focused on monitoring stress and depression in work environments [14–16, 18–23]. The dominant paper-pencil stress and depression scales must be validated with different group and work-related measures to thoroughly assess these mental illnesses in workplaces. Constructs for each measure are bounded by fixed questions; responses are biased to socially desired norms and highly sensitive to the timing of assessment [24, 25]. While these limitations have spurred initiatives to automatically as-

[2]I use the term "severe stress" to refer to individuals showing signs suggestive of chronic stress and "depression" to refer to individuals showing significant signs of depression. It should be highlighted that the stress assessment scale used in this study cannot clearly distinguish between acute to chronic stress, but a high score on the assessment is suggestive of chronic stress [17]. While this study utilised a clinically validated depression scale, it is not a diagnosis tool for depression.

sess mental health through mobile and wearable applications [26–29], existing solutions do not consider the notion of group interaction in the physical environment and only social interaction through applications [30–32]. Using various combinations of sensors to appropriately find physiological and behavioural indicators [30, 33–37] related to stress and depression increases overall power and privacy burdens to users [38]. Finally, these methods require explicit user interaction of installing a dedicated app, providing a much less extensible mechanism to conveniently accommodate data gathering from tens of thousands of users feeling stressed or depressed at a given time. Recent research has recommended the use of location information as single-sourced data to produce strong indicators for depression and schizophrenia among many other mental illnesses [39–42]. Using location information generally involves making direct observations on active behaviours, or the lack thereof, in human mobility. The ability to infer users' activities and engagements from location information help associate these traces with social interactions in the same environment [43, 44] (Figure 1.1 summarises how my thesis differentiates from prior work).

A feasible approach to meet these challenges could be the use of a standalone wearable device such as *sociometer* [45]. *Sociometer* was constructed to collect very fine granularity location information from GPS and Bluetooth sensors. However, measuring group interaction is through speech emotion analysis from built-in microphone, which is rather intrusive [28]. Further, using a standalone device introduces a strong bias to users who own such devices. In contrast, leveraging publicly available sensors at community-wide scale can support sensing everyone and bypasses connections to personal devices. Passively sensing at scale is already enabled by technologies such as video surveillance for public safety [46] and WiFi for location-independent network access [47]. Live video can be coupled with emotion tracking analysis to detect stress [30], but this method poses significant privacy threat from direct exposure of a user's identity [48]. Deriving mobility patterns from location proximities of WiFi signals [49] is less accurate but relatively unobtrusive.

While utilising WiFi infrastructures means monitoring capabilities are localised to within an indoor vicinity, high concentrations of people spend approximately 70%-80% of their time indoors [50]. Everyone in the environment can be monitored as anonymously as possible without additional devices or cost. Advantageously, WiFi infrastructure is a predominant solution deployed in most public buildings, providing users with free connection and public services.

Inspired by the aforementioned prior works, in this dissertation I build on existing WiFi passive-sensing components which provide single-sourced location information to easily and accurately detect stress and depression across the entire population in a work environment. We overcome data sparsity by (1) using group interaction features and (2) measuring the changes in an individual's behaviour including group interaction habits by comparing to *their own* and *population's* past periods to enhance the detection of stress and depression. Therefore, my contribution to this dissertation lies in ***(1) synthesising only location information into mobility features that sufficiently represent individual behaviours and group interaction habits, (2) statistically analysing these features to build a high-performing machine learning detection module for stress and depression, and (3) examining several individual and group measures as optimisation features***. I demonstrate the feasibility of our approach through three rounds of IRB-approved long-term user studies. The studies ran from Fall AY2017 to the end of Fall AY2018, running for a maximum of 81 days. 108 (out of 140) students actively participated in a standard protocol (of experience sampling every 3 days, retrospective assessment approximately every 2 weeks, and semistructured interview sessions) to capture their self-evaluations on mental well being.

In the rest of the chapter, I clarify the boundaries to the topics of this dissertation. Then, I briefly discuss existing technological approaches to stress and depression monitoring. Accordingly, I describe our proposed solution called *StressMon*, and present my thesis statement and validation plan. The chapter closes with an organising overview of this dissertation.

1.1 Defining Stress and Depression

Often conceptualised as being a psychological state, stress embodies the dynamic interaction of individuals with their environment, including people [51]. Specifically, Lazarus and Folkman describe psychological stress as a perception of how an individual appraises a situation as harmful, and then makes a secondary appraisal of how to best cope with the case; either through reacting rationally or emotionally [51–53]. While stress is an everyday occurrence, understanding the most common types of stress and knowing how to identify them can be challenging. There are several levels of stress – acute stress, episodic stress and chronic stress – characterised by Lazarus [5] to result in different symptoms and requiring different treatment approaches. It is noteworthy that this thesis does not address other types of stresses, which are critically influenced by life-threatening events (e.g., violence, death) or exposed at a very young age – traumatic and toxic stress [54, 55].

Acute stress is commonly observed as a natural response to everyday demands and pressures; associated wit both negative and positive outcomes. For example, an individual experiencing acute stress might be motivated to work, but be emotionally and physically distressed over a short period of time. In that distress, they could feel frustrated and fatigued after experiencing difficulties concentrating [5]. However, such stress is highly treatable with treatments generally focusing on mindfulness and cognitive behavioural therapy [56]. Moving up the scale is episodic stress, which describes acute stress over more frequent occurrences. An individual's emotional distress is more severe, as it tends to exhibit aggressiveness and low tolerance; commonly affecting interpersonal relationships. Unfortunately, episodic stress puts individuals at risk of heart disease, chest pain and persistent headaches [5]. Thus, treatment typically involves professional help, which may take many months [57]. Finally, chronic stress constitutes a wide variation of stressors over long periods of time. More often than not, chronic stress is a result of work-related strains [58]. When left untreated, such stress can cause serious disabilities such as major depressive disorders, high blood pressure and serious behavioural changes including insomnia and social isolation [5].

Significant research suggests that depression is the most likely outcome of exposure to psychological stress [7, 12, 59, 60]. Melchior and Siegrist also found depression resulting from chronic work-related stress is more often associated with individuals whose current work performances are affected by stresses of their previous jobs [58, 60]. Factors such as high work demand, poor social support and working relationships, and limited control over situations are stressors that commonly predict depressive symptoms [7, 60]. As a result, substantial evidence observes severely stressed individuals making changes in their normal behaviours [61–63]; for example, withdrawing from others and the inability to rest [64, 65]. Some of these behavioural symptoms overlap with depression [10, 66]. Further, the struggle to reorient or adapt can bring about more serious consequences [67]. These findings highlight the importance of recovering early from severe stress to a normal state, compared to when more serious conditions (i.e., depression) have manifested [3, 68].

On the other hand, depression, as recognised by the Diagnostic and Statistical Manual of Mental Disorders Fifth Edition (DSM-5), is a diagnosis which requires an individual to display at least five specific symptoms in the period of two weeks. The primary criteria is either being in a depressive mood or anhedonia (i.e., losing interest to pleasurable stimuli), while secondary criteria include diet changes, sleep abnormalities, reduced physical movements, fatigue, feelings of worthlessness and suicidal ideation [69]. Much research has investigated the relationship of stress and depression. One may argue that a depressive mood could be as a result of a stressor, thus a compound factor to a stressful situation [70]. However, anhedonia is an affective response, highly associated with hereditary or substance disorders (also from taking taking medications or treatments). It is also a complex process associated with personality characteristics [71–74]. For this reason, depression can occur without an individual feeling stressed [3], and stress and depression should be treated as two separate entities [9, 69, 70].

This dissertation narrows its scope of assessing the psychological stress to experiences of chronic stress. The overlapping symptoms of chronic stress in depression make depression an essential aspect of this investigation, and an independent mental health issue. It should be noted that the user studies in this work did not include any formal clinical diagnosis. Hence, this dissertation is not aimed to develop a diagnosis tool for stress and depression – rather, it aims to detect individuals showing signs suggestive of chronic stress (i.e., severe stress) and significant levels of depression. In comparison to prior work, this dissertation explores the use of location information to generate strong behavioural indicators of stress and depression on two levels; treating location traces of each individual as a singleton and including the physical group interactions with others in the same environment. Further, the value of location data is maximised from measuring the changes in an individual's behaviour, including group interaction habits by comparing to *their own* and *their (work) population's* past periods.

1.1.1 Factoring Group Interactions

To successfully manage stress at work, Lazarus argues that the individual, collectives (e.g., colleagues, social friends), and workplace must be treated as one unit, not independently [52]. Separately, Cox *et al.* characterise work-related stress by *work content* (e.g., task, load), but in no small degree, driven by *work context* (e.g., interpersonal relationships) [12, 13]. These findings show overwhelming evidence that group interaction is a critical factor influencing stressful work environments.

With more organisations structuring work within teams to capitalise on skills and perspectives, much research in Organisational Behaviour (OB) has been dedicated to understanding several characteristics of group interaction in team processes. These efforts include social identification [22], social cohesion [75], social loafing [76], and group potency [77], among many others. Positive effects of group interaction can bring about individual motivation and support for members, but there are also negative implications of group influence including work stress [21, 23, 78, 79]. Recent research has suggested *social identification*, the degree to which an individual psychologically associates themselves with a group, is a strong indicator of loss productivity [80], stress [81, 82] and depression [18]. Since its inception, social identification has been researched in pursuance of defining and assessing varied aspects of the interaction between an individual and the group [83–85] including conflict [86, 87]. Conflict is inevitable in groups whose members have diverse goals, opinions and attitudes, which consequently influences within-group behaviour [88]. Studies have found individuals with high social identification towards their (work) group are more likely to receive peer support, be committed to work [14], and as a result, develop a more positive emotional response in stressful work situations [15].

To understand how the inner dynamics of groups can affect individuals' overall mental wellness and productivity, this dissertation additionally examines if social identification can similarly be determined from mobility features which are representative of group interactions, and the implication for detecting stress.

1.2 Existing Technologies for Mental Health Monitoring

There has been substantial prior work in developing stress and depression monitoring applications. Some examples of stress applications include *UStress* [26], *StressSense* [28], and *AutoSense* [33], among many others. These context-aware applications make use of physiological readings from electrodermal activity (EDA) and electrocardiogram (ECG) [34,35,89], and device activity data [90,91] from mobile and wearable devices to detect stress in real-life environments. *StudentLife* [29] from Wang *et al.* analysed behavioural changes related to stress and the same authors also analysed symptoms features to predict depression scores [92]. However, these solutions typically require: (1) the orchestration of multiple mobile and/or wearable sensors, which place privacy and power burdens on users and their devices – resulting in lower user participation rates and increased attrition rates [38] and (2) explicit user interaction of installing a dedicated application on their mobile device, which may limit the resource to only proactive users who would be willing to install an app – unfortunately, many studies found highly stressed and depressed people (in greatest need of help) tend to behave passively and helplessly in finding a solution [93, 94].

More recently, the understanding of mobility patterns from sensing individuals' location has represented some of the most promising research areas including mental health and social behaviour. For example, much work has discovered strong relationships between social isolation and mental disorders such as depression [39, 41], schizophrenia [42] and happiness [40]. Brown *et al.* [95] used wearable RFID tags to collect indoor location traces of employees interacting with colleagues in different building spaces. Ware *et al.* used location data collected from the WiFi infrastructure to detect depression [96]. In a similar fashion, Zhou *et al.* [97] used WiFi indoor localisation data to learn about student behaviour. [96, 97] are conceptually the closest to our approach. However, these works neither monitored stress nor inferred group interaction factors in detecting depression. In the next section, I describe our goals and proposed solution. [92]

1.3 Solution: *StressMon*

Our goal is to develop a stress and depression monitoring solution, *StressMon*, as a first-level safety net whose benefits can be distributed across an entire population of users in a work environment. It complements fine-grained sensing solutions that require installing dedicated mobile applications where timely interventions can be made in a more personalised context. Key characteristics of *StressMon* are:

1. ***No application installation or additional device required*** to accommodate monitoring tens to thousands of users, at any given time. To enable ease of

deployment, our approach leverages a passive, WiFi-based location detection to track all users in a work setting.

2. ***Utilises single-sourced location data***, obtained from RSSI values reported directly by the WiFi access points (APs). Using just the location information, individual routine behaviours are derived from inferring different user activities in the environment.

3. ***Considers physical group interaction patterns***, inferred from grouping location traces of individuals in the same vicinity. It utilises a group detection system, which clusters devices into logical groups from the same location traces.

4. ***Ascertains changes in individual and group behaviours***, which compares individuals' normal behavioural patterns to past periods of their own, called *absolute change*, and their work population's, called *relative change*.

5. ***Leverages additional work-related information as features.*** Specifically, an individual's personality traits and social identification towards their workgroup, which were not measured passively from location data, were combined with passive location-based features to improve the detection of stress and depression.

6. ***Accurately detects stress and depression over different time windows***. Fundamentally using the same sets of mobility features, albeit with sufficient differences to separate the detection models, an individual with severe stress can be detected at 6-days interval and an individual with depression can be detected at 15-days interval. The addition of work-related information features to the stress and depression models leads to more accurate detection results.

	Detection Capabilities	User Study	Sensing/Model	Features (Individual behaviour)	Social Behaviours	Validation	Performance
StressMon, 2019	Detects (1) a person with signs of chronic stress over 6-days, and (2) a person with clinically significant depression over 15-days.	Long-term, campus (individual and team setting), 3 phases between 36-81 days, 108 active participants √	WiFi-based location system, Random Forest (binary classification) √	Location features of individual routines and changes in behaviours, personality (big-5 score) √	Location features of physical group interactions and changes in group interactions √	PHQ-8 for depression,PSS-4 for stress, FISI for social identification as ground truth √	Stress + social identification:98.93% TPR, 69.49% TNR Depression + personality: 90.21% TPR, 69.45% TNR √
Choi et al., 2009	Detects occurrence of stressful events in real-time (unspecified form of stress).	Short-term, laboratory setting (4 tasks over 6 minutes), 3 subjects	Heart rate monitor (wearable), Fisher's linear discriminant analysis (binary classification)	Individual heart rate √	N.A.	Did not use stress scales – not reported	83% accuracy Other measures such as TPR, TNR and AUC were not provided.
AutoSense, Ertin et al., 2011	Detects (1) a minute of sensor data is a physiological response to a stressor, and (2) a person perceives stress during a particular minute (unspecified form of stress).	Short-term, laboratory setting + 2 full days in the field, 17 subjects	ECG, RIP, GSR, accelerometer (wearable), Hamilton-Tompkins QRS detection	Various individual physiological characteristics	N.A.	Did not use stress scales; self-reports but did not specify questions.	90% accuracy on physiological classifier, and 0.72 with self-reported rating of stress √
StressSense, Lu et al., 2012	Detects stressed speech and neutral speech (unspecified form of stress).	Short term, marketing job and job interview setting, 14 subjects (over several – unspecified – days)	Smartphone speakers, Gaussian mixture model (binary classification)	Subject's voice √	N.A.	Did not use stress scales, used Affectiva GSR indicative of stress as ground truth	81% and 76% accuracy for indoor and outdoor environments
UStress, Egilmez et al., 2017	Detects occurrence of stressful events in real-time (unspecified form of stress).	Short-term, laboratory setting (9 tasks over several minutes), 9 students	GSR, accelerometer, heart rate; chest band and smartwatch, Random Forest (binary classification)	Various individual physiological and physical characteristics	N.A.	Did not use stress scales – activities were predetermined to be stress or no stress.	approx. 80% TPR, approx. 80% TNR (AUC is not provided)
StudentLife, Wang et al., 2014	Did not solve a detection problem, presented significant correlations between the sensor data and mental health and educational outcomes	Long-term, campus (individual), 1 phase for 70 days, 48 active participants √	Accelerometer, colocation and application usage	Various individual physiological and physical characteristics	Social engagement, Speech and conversational interaction √	EMA as ground truth – PHQ-9 for depression, PSS for stress √	N.A.
Canzian et al., 2015	Predicted depression (mean+1std-split) at 14-days interval, presented significant correlations between the GPS location and depression	Long-term, in-the-field, 1 phase for 71 days, 28 active participants √	GPS, SVM classifier with a Gaussian RBF kernel (binary classification)	Location features of individual routines √	N.A.	PHQ-8 for depression as ground truth √	60% TPR, approx. 80% TNR
EDUM, Zhou et al., 2016	Did not solve a detection problem, only characterised classroom behaviours such as attendance and punctuality	Long-term, campus, 1 phase for 63 days, 700 participants √	WiFi-based location system + mobile app – individual routines (colocation), phone activity states	Location features of individual routines, mobile activity characteristics √	N.A.	No validation of any mental health outcomes	N.A.
Wang et al., 2018	Predict (1) depression score, and (2) predict depressed state (mean-split) both on a weekly basis	Long-term, in-the-field/campus, 2 phase for 63 days, 83 active participants √	Accelerometer, colocation and application usage, Lasso regularized linear regression model, logistic regression model (binary classification)	Various individual physiological and physical characteristics	Social engagement, Speech and conversational interaction √	PHQ-4 (once a week) and PHQ-8 (every two weeks) for depression as ground truth √	81.5% recall and 69.1% precision to detect depressed state √
Ware et al., 2018	Detect clinically significant depression over 13-days interval	Long-term, campus, 2 phase for a total of 19 months, 182 participants √	WiFi-based location system, SVM classifier, (binary classification) √	Location features of individual routines √	N.A.	PHQ-9 for depression as ground truth, clinician √	77.00% TPR, 62.50% TNR for day-time monitoring (excluding participants living on campus)

Figure 1.1: Summary differences from prior work; rows highlighted in yellow are related to stress detection, blue for depression detection, and green are significant pieces which did not solve detection problems. A tick on "User Study" indicates user studies of similar scale, "Sensing/Model" indicates similar sensing mechanisms (without the need for direct sensing and/or a dedicated mobile application), "Features" are extracted from a single sensor data. "Social behaviours" account for features representing interaction patterns, "Validation" adopts similar methodology for validating results, and "Performance" indicates works with best results, comparable to ours.

Figure 1.1 summarises the differences of *StressMon* from prior work. *StressMon* makes several clear distinctions. First, it is the only work which builds detection models for stress and depression as one solution. Unlike prior work on stress detection, the user study for *StressMon* was conducted over a longer period of time and is first work in using location data. *StressMon* does not detect stress in real-time but in 6-days – for the reason that signs suggestive of an individual experiencing severe stress is due to emotional pressures suffered over a prolonged period of time [17]. It detects depression in 15-days, typically from following symptoms over a two-week period [98]. Next, *StressMon* is a first work in utilising physical social interaction [12, 13, 53] and changes in behaviour [61–65] as key features in its detection model compared to [39, 96]. Compared to [92], features representative of social patterns were not extracted from a different data source – but from a single-attribute location data. Finally, using location-based input and optimisation features, it produces results far different from prior work which utilised fine-grained data [92].

1.3.1 Testbed

As a start to this endeavour, I studied different groups of undergraduate students to demonstrate the application of *StressMon* as a campus-wide resource for our university. The university campus is an excellent testbed for *StressMon* for two main reasons: First, considerable evidence suggests students in universities and higher education are most vulnerable to severe amounts of stress [99, 100], embodying a trend of campuses as high-stress environments [16, 101, 102]. Second, as with many other universities and organizations worldwide [103], our university is deployed with a campus-wide Wi-Fi system (the LiveLabs) that scales to tens of thousands of clients and has been running stably since 2013.

It should be noted the long term goal of *StressMon*, as a fully functional community-wide health monitoring system, must be principally designed with appropriate policies and procedures; a deployment of *StressMon* in a professional work environment would likely raise concerns among employees fearing negative reviews and discrimination from being frequently detected to experience severe stress or depression. Would such assessment impact a user's career in any way? Until such privacy ensuring policies are established, it is viable to use students as test users since they are not paid employees. Note: The implementation of privacy policies appropriate for an end-to-end solution as *StressMon* is not part of this dissertation, however its implications will be discussed in Chapter 7.

1.3.2 *StressMon*'s Ethical Practice

With increasing progress in enabling technologies for large-scale behavioural research, ethical concerns remain a challenge within the research community. Since *StressMon* considers the influence of social relationships to detect severe cases of mental health issues, the process of collecting and deriving behavioural patterns in groups of user data must abide by compelling ethical principles. Referencing the CSCW guidelines for social computing [104], we argue that the mechanisms operationalising *StressMon* comply with principles of the Belmont Report [105] in the following ways:

1. Respect for Persons: Having an informed consent form (with IRB-approval) is the most straightforward way of respecting and protecting users from harm. As *StressMon* conveniently collects its data through the WiFi access points, it bypasses the direct communication with the user's personal device for data collection; thus minimises privacy threats. It should be highlighted that the localisation mechanism used in *StressMon* [106, 107] maintains anonymity by applying a one-way hash function, which prevents a user's device from being easily identified. *StressMon* itself does not provide anonymous protocol capabilities. Therefore, the system requires additional scrutiny by organisational officials responsible for approving data security controls to maintain confidentiality of (all) users utilising the WiFi.

2. Beneficence: Beneficence is weighing the risks over the benefits of any research. The risks of exposing users' identity is minimised because we are making no direct communication with users' devices to collect location information and no requirement for personally identifiable information to conduct behavioural analyses. At this stage, *StressMon* operates by detecting individuals who display signs of severe stress and depression (the minority class), and involves the risk of misclassification. Ideally, *StressMon* should achieve high sensitivity (True Positive Rate) and high specificity (True Negative Rate) in real-world scenarios. As the cost of misclassifying a minority-class is substantially greater than the cost of misclassifying a majority-class, *StressMon* seeks to maximise sensitivity while ensuring moderate specificity. While the implication of such performance may possibly lead to more students being misclassified as severely stressed, *StressMon* could integrate an early identification step to verify the severity of stress and depression for borderline cases. Despite these risks, we believe *StressMon* provides greater social benefit for both individual and collective levels. With approximately 66% of college students suffering from either depression or stress [108], and campus service providers facing resource crises [109], *StressMon* can be deployed as a campus-wide "safety net" for those in greatest need of help. *StressMon* can be an enabler for students to receive help via external methods such as interventions moderated by counsellors. Another strength of *StressMon* is in its ability to detect stress and depression derived from rudimentary behavioural analysis of different groups of anonymised users within a population – enabling group-level health interventions, which are less-intrusive to targeted individuals.

3. Justice: Justice requires fair user participation. This is true for *StressMon*, as its data collection is not influenced by factors such as the socioeconomic status or technical experience of the user. Instead, *StressMon* leverages Wi-Fi, which is readily available in public spaces (e.g., offices, campuses and shopping malls) and commodity devices (e.g., laptops and mobile phones). In fact, a majority of mobile devices today are designed to prioritise Wi-Fi. *StressMon* does not require any explicit user interaction of installing/running a dedicated application on their phone. Thus, the resource is readily available to all users in the environment having access the bare minimum of a smartphone that can connect to wireless networks via Wi-Fi.

The evolution of social sensing enables the measuring of large-scale human behaviour. Technologies, such as *StressMon*, provide foundational mechanisms for interdisciplinary research communities to explore new methods of facilitating mental health benefits/interventions and studying the natural processes of small group phenomena, while ensuring such studies stay within the boundaries of ethical practices.

1.4 Motivating Scenario

Ultimately, the usefulness of *StressMon* must depend on the value it brings in real-world situations among a broader group of users. The following scenarios illustrate the application of *StressMon* for two different sets of users.

Scenario #1

A Team of Undergraduate Student Developers. Alice, Bob, and Charlie are developing a game app for their final year project, and have coordinated among themselves to work closely together on campus. The campus integrates a WiFi indoor localisation system that can track the location of devices using data collected directly from the WiFi infrastructure itself – i.e., without needing any input or apps installed on client devices. Recently, the university deployed *StressMon* as a campus-wide "safety net" to automatically detect students and groups who are displaying signs suggestive of chronic stress, intending to help them via external methods such as interventions by student counsellors.

Alice, Bob, and Charlie are predominantly connected to the campus WiFi on all their smart devices. *StressMon* determines, from location traces, that Alice predominantly works with Bob. They are both responsible for front-end development. In the same way, *StressMon* detects that Charlie likes working from various workspaces; albeit by himself. In actuality, Charlie is struggling to deliver his tasks. *StressMon* detects that Charlie is under severe stress. An external system (that is beyond the scope of this paper), using interventions prescribed by student counsellors, promptly provides Charlie with stress management tips. At the same time, the external system notifies Alice and Bob that their team member is experiencing high stress and provides them with concrete steps to take to maintain good work balance and harmony within their group. Alice and Bob are stunned at this revelation and meet up with Charlie, asap, and utilise the suggestions provided to them. This intervention has an immediate impact, Charlie's stress levels decrease, and he can complete his portion of the task.

Scenario #2

New Nurse Preceptorship Program. Amy is a new graduate nurse who joins the med-surgery unit and is assigned to a nurse preceptor, Carol. The hospital employs a Bluetooth indoor positioning system to help staff find their way to departments or wards. Recently, the hospital deployed *StressMon* to similarly detect new nurses who are showing significant signs of depression with the intention of supporting by notifying their preceptors.

Three weeks into the programme, *StressMon* determines Amy making significantly different behavioural changes and workgroup interactions. For example, Alice takes shorter break times and deviates further away from common locations she previously frequented during these times. *StressMon* detects Amy as depressed. In actuality, Amy struggles with taking up more caseload and is emotionally affected from experiencing verbal abuse by patients. At the same time, she hesitates to make a report to her supervisor-in-charge for fear of being appraised as incompetent. Using interventions prescribed by nurse counsellors (beyond the scope of this dissertation), *StressMon* promptly provides Amy with self-care tips. At the same time, *StressMon* notifies Carol (Amy's preceptor) and the system provides her with concrete steps to guide Amy with knowledge and skills that are required for the depressive situations that Amy is struggling with. This intervention helps Amy gain better learning experiences and her depression levels decrease.

These scenarios highlight our overall vision wherein *StressMon* is a key part of an overall health monitoring and intervention solution that can provide organisation-wide coverage. This solution has use cases beyond academia, and is of high value in hospitals where are under significant daily stress leading to depression; many studies have linked the severe shortages of medical professionals, especially nurses, to the highly stressful work conditions [110, 111].

1.5 Thesis

My thesis statement is as follows:

> *It is possible to easily and accurately detect stress and depression across entire school campuses using location data while overcoming the limitations of low-dimensional data by incorporating inferred individual and group features.*

To restate the main points of this dissertation, *StressMon* addresses the binary classification problem of detecting individuals likely to experience (1) severe stress over 6-days interval and (2) depression over 15-days interval in a work setting; however, it does not reveal the underlying reasons for those conditions. In the rest of the dissertation, I use the term “severe stress” to refer to individuals showing signs suggestive of chronic stress and “depression” to refer to individuals showing significant signs of depression. This dissertation answers the following research questions:

1. Can we detect severe stress in a person using only location data?
2. What effects do group-related features have on stress detection?
3. Can we detect depression in a person using only location data?
4. What effects do individual-related features have on depression detection?
5. What can be learned from the findings of severe stress and depression?

1.5.1 Validation Plan

This dissertation establishes the thesis via the following steps:

1. It presents our solution, *StressMon*, which leverages a key sensing apparatus comprised of the *LiveLabs* and *Grumon* systems, to collect location data passively across all users in the entire campus environment and build a detection module to accomplish its goals.

2. It characterises mobility features extracted from location data, representing individual and group interaction behaviours. Features are categorised by their function and degree of specificity; that is, some features require the verification of highly specialised references such as a work schedule.

3. It provides a detailed exploratory data analysis on mobility features to distinguish significant differences between a group of students with severe stress versus others, from comparing the changes in their behaviours and group interaction patterns.

4. Then, via a primary user study, *Study_SE*, it demonstrates how it achieves the first goal of accurately detecting stress. In addition, it shows the effects of various group-based features on the prediction accuracy of stress.

5. In the same manner, it demonstrates how it achieves the second goal of accurately detecting depression with sufficient differences to the detection model. It shows the effects of individual-based features on the prediction accuracy of depression.

6. Finally, it exhibits how *StressMon* can effectively maintain its high accuracy in detecting stress and depression via two validation studies, *Study_Valid1* and *Study_Valid2*. In particular, the stress and depression models yielded comparable performance to what was achieved in the primary study, *Study_SE*.

Section	Table/ Figure	Experiment	Achieved Results	Comparison
4.3	4.6	Severe Stress: **Individual + Group** vs. Individual features (note: at 3-days interval)	TPR: 99.33% TNR: 69.49% ACC: 72.08%	TPR: 95.70% TNR: 60.44% ACC: 63.37%
4.3	4.11	Severe Stress: **Location + optimisation** vs. Location-based features (note: at 6-days interval)	TPR: 88.08% TNR: 93.09% ACC: 92.05%	TPR: 98.93% TNR: 75.54% ACC: 77.59%
4.4	4.13	Depression: **Individual + Group** vs. Individual features (note: at 15-days interval)	TPR: 70.25% TNR: 63.53% ACC: 65.58%	TPR: 51.83% TNR: 64.72% ACC: 63.32%
4.4	4.10	Depression: **Location + optimisation** vs. Location-based features (note: at 15-days interval)	TPR: 90.21% TNR: 69.45% ACC: 77.16%	TPR: 70.25% TNR: 63.53% ACC: 65.58%

Table 1.1: Summary of experiments, central to this thesis and its forward references to the sections, respectively. Note: all results were tested on primary dataset, *Study_SE* students. Refer to sections for results of other populations.

Table 1.1 provides forward references to experiment details that are central to this thesis.

1.6 Organisation of Document

This dissertation presents seven chapters and three appendices as follows:

Chapter 2 presents the building blocks to *StressMon*. It details two (out of three) components: First is the Key Sensing Apparatus, comprised of two existing solutions that have been utilised solely for collecting location data. Second is the Location Feature Extractor, a component which I developed to apply various heuristics and map activities onto location programmatically. Then, it extracts mobility features that are representative of individual behaviours and group interactions of users in the work environment. The third block briefly describes all the detection models that make up the *StressMon* detection component.

Chapter 3 describes the long-term study I conducted in three separate phases.

First is the primary study, *Study_SE*, which specifically looked at a population of Information Systems students undertaking a highly stressful course. The second and third studies, *Study_Valid1* and *Study_Valid2*, included a broader range of students from different schools and courses to validate all the detection models. This chapter details the various assessment scales used to measure participants' stress (at a frequency of every 3 days) and depression levels (at a frequency of every 2 weeks) as ground truth to validate *StressMon*'s detection models.

Accordingly, Chapter 4 describes in detail the development of a stress model and depression model separately. The chapter comprehensively includes a clear explanation of how ground truth labels for stress and depression were processed, and then, mapped to the location data (input) as binary classification problems. The evaluation of both models adopted a standard procedure of group-fold cross validation for the primary study, *Study_SE*, and additional tests on *Study_Valid1* and *Study_Valid2* as validations.

Chapter 5 comprehensively presents my findings on participants' experiences in stress and depression through validated scales and qualitative interviews. The interviews were aimed to elicit what caused stress in participants' life (at the point of study) and how it affected their behaviours and social interactions, especially with their workgroups. It is important to note that I did not explore the experience of depression in interviews due to my clinical inexperience to handle such cases.

Chapter 6 reviews the most relevant research works related to my sociological understanding of stress, depression and social identification. Then, I expand on the assessment methods for stress, both traditionally and technologically. Finally, Chapter 7 closes this dissertation with a summary of its main contributions, a discussion that combines all findings from evaluating *StressMon*'s detection engine to inform its limitations and future work.

CHAPTER 2

BUILDING BLOCKS OF *StressMon*

Earlier, I described *StressMon* as a solution utilising a single-sourced location data. This decision is based on several reasons: First, prior work found strong associations between location traces and mental health outcomes such as depression and schizophrenia. Second, server-side location data provides the ability not only to infer individuals' behaviours but to group interaction patterns in a less intrusive manner. Further, our decision to adopt WiFi location tracking as a sensing technique is based on the technology having predominant presence in most public buildings to provide free connection and services. This chapter begins by discussing our decision's implications. The rest of the chapter describes in detail the key blocks comprised in *StressMon* and the outputs each component produced.

2.1 Understanding Location Sensing Techniques

The increasing availability of location-based services has transformed urban research by exploiting geographical position of personal devices to environmental information gathered from networks into various applications; ranging from health to business analytics [29,39,41,92,112]. Location provides a wealth of information on individual movement patterns and their physical communication patterns; for example, users can navigate both in physical and virtual space for directions, and receive updates derived from locating other individuals or routes. Then, activities or engagements inferences can be made from the venue and duration [43, 113].

The implementation of location-based services is fundamentally driven by its sensing technique to actively provide location estimates. A great deal of commercial services utilise global positioning systems (GPS) [114] but they suffer from indoor inaccuracies. Indoor-positioning can be enabled by radio-based sensing such as infrared (IR), radio frequency identification (RFID), ultra-wideband (UWB), Bluetooth low energy (BLE) and wireless fidelity (WiFi) among many non-radio based sensing techniques (e.g., camera, sound, magnetic field) [115]. However, a considerable number of these technologies require massive transceiver and infrastructure deployments for emitted signals to be picked up periodically [116]. Fortunately, in the present day, WiFi is almost universally available as a means of ubiquitous and continuous wireless network coverage. This practice is especially true in public buildings [117, 118] from campuses to offices, shopping malls to airports. WiFi is also integrated mainly in consumer communication devices.

Bluetooth might seem comparably common to WiFi as it is supported on most consumer devices, and beacons are small, inexpensive and long-lasting. Bluetooth's higher beacon density could also provide more accurate positioning than WiFi [119]. However, a user cannot be tracked continuously (with a fixed Bluetooth address) at different periods; unfortunately, Bluetooth address of a device is replaced with random values at specific time intervals [120]. One of *StressMon*'s key characteristics (see Section 1.3) is to ascertain the changes in behaviours and in group interaction individuals experience when they are stressed. Such a feature would require continuous identification of the user, albeit anonymously. WiFi uses MAC address, which is a globally unique ID for a device. One way to maintain anonymity is to use reliable one-way hash functions [106, 121]. Identifying a user from the MAC address can be challenging, but inferences could eventually be made [121]. For these reasons, WiFi indoor positioning makes a good option for *StressMon* to adopt as its sensing technique.

2.1.1 Implications of Using WiFi Indoor Localisation System

Evidently, WiFi indoor localisation system is one of the most common techniques due to the widespread deployment of Wifi infrastructures. While accuracy estimates largely depend on the signal-to-noise ratio of the received signal, these estimates can be influenced by many other environmental factors related to building structures [122]. Common hardware concerns such as access point (AP) displacement could result in inaccuracies of received signal strength indicators. Therefore, radio maps (i.e., APs in close proximity are annotated by the corresponding location information) must be constructed and updated regularly. This requirement demands time and labour, especially in large urban areas [123].

One of the widely adopted methods to determine location from WiFi signals is through inferring position from the strongest received signal strength of all APs within range and finding the location that best matches the signal from the radio map. Studies have found that this technique might be effective in academic buildings (such as ours) since fewer human movements are expected periodically compared to large public spaces such as shopping malls [106, 124]. Since *StressMon* monitors for deviations in behaviours, it is important that differences in movements are large enough to be picked up by different APs. Finally, offsite behaviours and group interactions cannot be supported by this technique. The ability to monitor outdoor movements can be extended (by *StressMon*) to use GPS, if necessary. However, substantial research has found that humans spend approximately 70%-80% of their time indoors [50].

2.2 *StressMon* System Overview

Figure 2.1 illustrates an overview of *StressMon*; comprised of three components: (1) Key sensing apparatus, (2) Location feature extractor and (3) Machine learning detection engine. It is important to note that my contribution to this dissertation is focused on enabling accurate detection of stress and depression using only location data supplied by the key sensing apparatus.

2.3 Block 1: Key Sensing Apparatus

StressMon leverages two key sub systems, which make up a crucial component to collect location and group information directly from the environment's infrastructure. Building on existing deployment, performance inaccuracies for these systems were rare and bounded, as explained below. Note that the evaluation of these systems is not part of my dissertation contribution.

2.3.1 *LiveLabs* WiFi Indoor Localisation System

StressMon adopts the *LiveLabs* WiFi indoor localisation system as a server-side solution, which uses real-time location services (RTLS) to extract Receiver Signal Strength Information (RSSI). This is the signal strength of each device connected to the AP as measured by the AP. These signal strengths decay as the device moves further away from the AP. Hence, by using RSSI observed by multiple APs, we can compute the position of each device using a method known as *reverse triangulation*. This approach uses data collected solely from the infrastructure (each WiFi AP) and thus can work across any mobile device (e.g. iOS, Android) and does not require installing any client software.

The *LiveLabs* system has been operating in our test environment (Singapore Management University) since August 2013, including several other public spaces. The system was tested accurate between 3 to 8 meters in most places; sufficient to localise a device to a specific room. Based on empirical studies, Khan *et al.* reported its performance at 87% accuracy [125]. Further, its location errors could increase to 10-20% for non-negligible periods [106].

By default, the *LiveLabs* system anonymises the MAC addresses of all connected devices using a 1-way hash function; thus, enables monitoring of users at community-level without exposing identities of tracked devices. However, personal

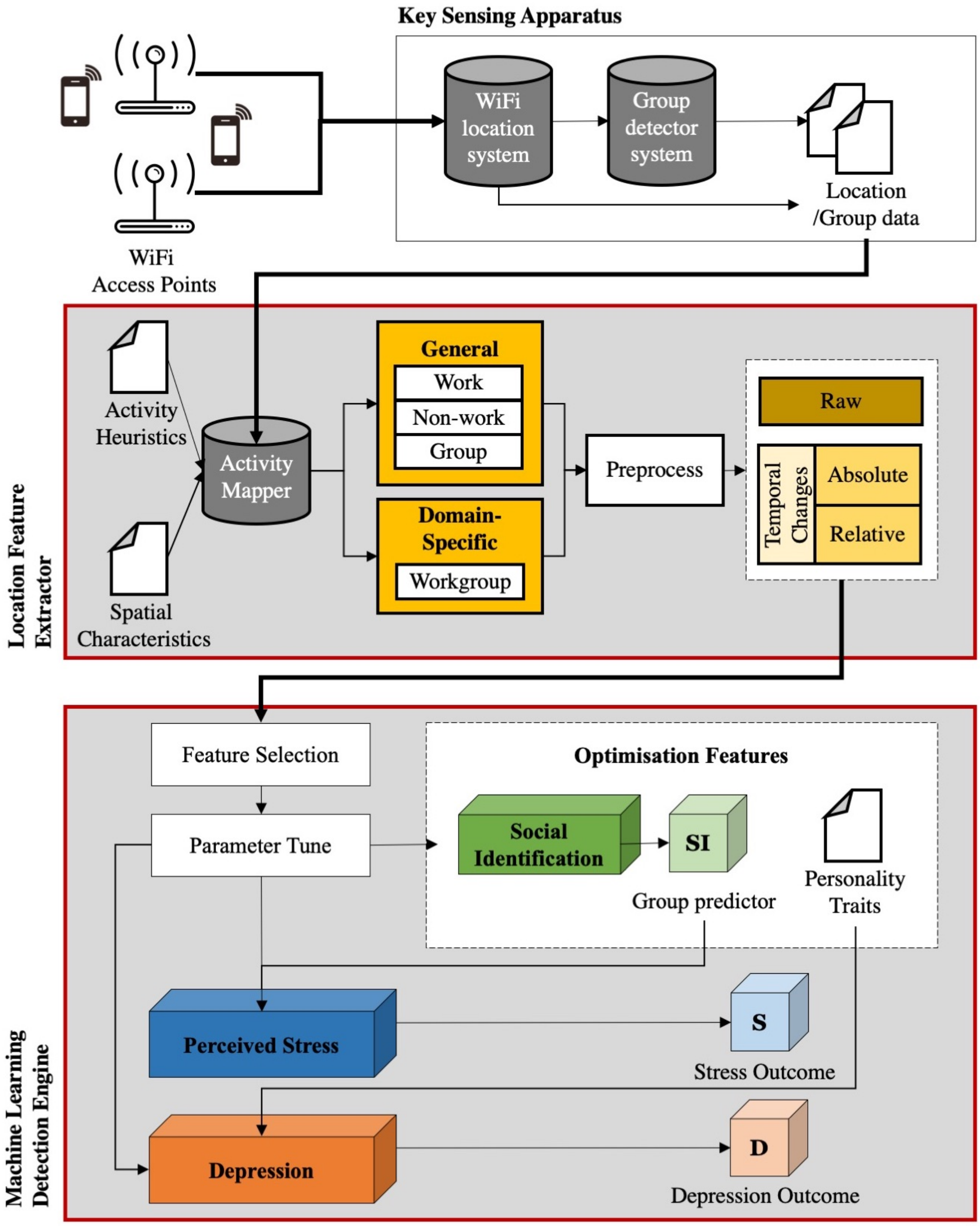

Figure 2.1: Illustration of *StressMon*, comprised of three blocks, with shaded blocks indicating my contribution to this dissertation: (1) Key sensing apparatus is made up of existing systems deployed in the test environment, (2) Location feature extractor makes decisions on the types of activities inferred from spatial characteristics and generates predictive features including features that measure the changes in behaviours and interaction patterns and (3) Machine learning detection engine, which consists of separate detection models for stress and depression.

Size	No. of devices	Interaction Type
Solo	1	Alone
Small	2 <= Device => 5	With close/work group
Medium	6 <= Device => 20	With medium-sized group
Large	Device > 20	Mass participation

Table 2.1: Group sizes are defined to extract interaction patterns on campus.

monitoring of stress can also be supported as a key use case – users (upon consent) will need to provide their device MAC address to *StressMon* so the same hash function can be applied to identify their devices from these location traces. As will be explained in Chapter 3, the studies conducted as part of this dissertation were to demonstrate *StressMon* as a "safety net" solution for a community of users. However, I collected the MAC addresses of consented participants so their location traces could be matched against their self-reports to validate the results of *StressMon*'s detection models.

2.3.2 *Grumon* Group Detector System

The second sub-system utilised for sensing is the *Grumon* group detector system, , which extracts group information from the localisation system [107]. Specifically, the group detector processes location information to cluster devices located in the same vicinity that move together using Markov Cluster algorithm (MCL). GruMon was shown to be highly accurate, detecting over 80% of the groups, with 97% precision, within 10 minutes of observing location data. Most of its errors arose in detecting large groups where a large group was defined as 7 or more individuals. It was much more accurate when detecting small and medium groups.

2.3.3 Implications of Key Sensing Apparatus

By leveraging these key sub-systems, it is clear that *StressMon* is susceptible to some compounding errors accumulating in the final prediction. For example, Balan and colleagues explained the *LiveLabs* deployment in our study environment achieved excellent accuracies. Its performance, however, highly depends on environmental factors (e.g, layout of building, number of APs installed) and (changing) density of crowds [106]. These findings support Bak's conclusion of indoor localisation systems performing better in academic buildings than other large public spaces (e.g., shopping mall, airports) [124]. Similarly for *Grumon*, Sen *et al.* found higher inaccuracies from the indoor localisation system deployed in malls. With the indoor location system (*LiveLabs*) yielding its best case performance in our university, it is assumed that the *Grumon* system maintains its reported results of a 91% precision and 82% recall. Accordingly, Jayarajah and colleagues validated this assumption by including various levels of random noise in the location data to determine its robustness, and, subsequently, did not observe any significant changes in its performance [49]. Finally, the group detector requires (next-place) transition features to achieve higher precision to detect a group more accurately [107], and thus fits well with the student population who are commonly observed to transit between classes and/or buildings [49].

For these reasons, the negative effects of both key systems' performances are expected to be rare and bounded in the current evaluation of *StressMon*. It is important to note that *StressMon*'s performance might be impacted by the inaccuracies of its sensing component, especially in public spaces with high density of people and unpredictable crowd movements [126] – I address this as part of Limitations in Chapter 7, and discuss how further studies should be focused on testing its robustness. While research to improve the accuracy of indoor positioning mechanisms is still ongoing, I believe there is sufficient progress to assume that WiFi indoor location systems will be able to accurately produce location information and, subsequently, group location information in different kinds of public spaces. In this dissertation, I am focused on identifying the types of behaviours and interaction patterns that strongly indicate individuals under severe stress and/or depression, so that these events can be detected accurately.

2.4 Block 2: Location Feature Extractor

2.4.1 Activity Mapper

The activity mapper determines the most likely activity to occur at a particular location based on the detected venue, time thresholds of students' routines calculated from population mean, and regular measurements of activity in 15 minute windows. These heuristics were developed in the following steps. First, common predetermined activities were manually assigned to each campus facility (e.g., {"location":"<building name>_<level number>_Seminar Room3.2", "activities":["seminar", "study", "transit"]}). Then, an activity is decided from the list based on simple decision making statement of time and day (e.g., lectures are conducted on fixed time slots on weekdays). Second, the activity is verified against the student's project schedule, a material collected as part of the study to assign activities specific to SE workgroup. Note: while verification was performed manually in this user study, its process can be automated by synchronising individuals' calendars. Further, these activities were confirmed based on time thresholds of students' daily activities averaged over the sampled population (note: students were surveyed of their daily campus routines as part of a demographic survey, described in Chapter 3). For example, an instructor consultation is capped at 1 hour, gym at 2 hours, transition between places at 15 minutes, eating at 30 minutes. Finally, an activity is only assigned to a nomadic device switching connections between different APs if a connection lasts for at least 15 minutes; otherwise labelled as "in transition".

time	location	mapped_location
2/10/17 12:46	1010100049	SMU-SIS-L1-FilteringArea
2/10/17 12:46	1010100049	SMU-SIS-L1-FilteringArea
2/10/17 19:01	1010200048	SMU-SIS-L2-StudyArea2
2/10/17 19:01	1010200044	SMU-SIS-L2-StudyArea2
3/10/17 10:09	1010100027	SMU-SIS-L1-ReceptionAndLobby
3/10/17 10:09	1010100027	SMU-SIS-L1-ReceptionAndLobby
3/10/17 10:09	1010100027	SMU-SIS-L1-ReceptionAndLobby
3/10/17 10:09	1010100027	SMU-SIS-L1-ReceptionAndLobby
3/10/17 10:09	1010100027	SMU-SIS-L1-ReceptionAndLobby
3/10/17 10:10	1010100029	SMU-SIS-L1-ReceptionAndLobby
3/10/17 10:10	1010100027	SMU-SIS-L1-ReceptionAndLobby

(A)

date	groupid
22/10/17	19566b5371d22a839bed11c489d9ff4c998a2851
22/10/17	43bc78399ce14fdaada37246bf406b00107625e8
22/10/17	80df141ba5a49be56aff19d877efdf5c96e63baa
22/10/17	06e1f4985ab594cde27a7f6abe52a76159c5dd68:9af0695338dd7e5750200c6657cebca0533a9cea
22/10/17	06e1f4985ab594cde27a7f6abe52a76159c5dd68:9af0695338dd7e5750200c6657cebca0533a9cea
22/10/17	623ea3a03088affe7d2f5cdf9e0c70d35a820f5d
22/10/17	42ac16a0240c88ce7559b498d1b19c7a3242c43b
22/10/17	12b2082baea3715e8b5d0658f92b8b028c88208b:560b86812f577f473c581d2d5f96f2b317d1a9c2
22/10/17	12b2082baea3715e8b5d0658f92b8b028c88208b:560b86812f577f473c581d2d5f96f2b317d1a9c2
22/10/17	ace8f3a914a8bd1d31b6b72171e351a52b28d1a2
22/10/17	4606b253e7ec8a5fc769e54db159a3267ff62a9a

(B)

•••

last_location	total_transitio	location_history	size
SMU-SIS-L5-StudyArea1	0	SMU-SIS-L5-StudyArea1:00:00:40	1
SMU-SIS-L2-StudyArea2	0	SMU-SIS-L2-StudyArea2:00:00:40	1
SMU-SIS-B1-NearLiftLobby	0	SMU-SIS-B1-NearLiftLobby:00:00:40	1
SMU-SIS-L4-MRCorridor	0	SMU-SIS-L4-MRCorridor:00:00:40	2
SMU-SIS-L4-MRCorridor	0	SMU-SIS-L4-MRCorridor:00:00:40	2
SMU-SIS-L4-OusideAcadOffice2	0	SMU-SIS-L4-OusideAcadOffice2:00:00:40	1
SMU-SIS-L2-LiftLobby	1	SMU-SIS-L2-LiftLobby:00:00:20;SMU-SIS-L2-LiftLobby:00:20:20	1
SMU-SIS-L2-StudyArea1	0	SMU-SIS-L2-StudyArea1:00:00:40	2
SMU-SIS-L2-StudyArea1	0	SMU-SIS-L2-StudyArea1:00:00:40	2
SMU-SIS-L5-StudyArea1	0	SMU-SIS-L5-StudyArea1:00:00:40	1
SMU-SIS-L5-AcadOffice	0	SMU-SIS-L5-AcadOffice:00:00:40	1

Figure 2.2: Sample of data inputs to the activity mapper: (A) is the location record of a specific device, (B) is the group record of the entire population. Note that 'groupid' in (B) is a concatenation of multiple hashed MAC addresses if group size is more than 1. Unless the actual MAC address is provided, a user cannot be identified as being part of the group.

2.4.2 Pre-processing Noise and Missing Values

Figure 2.2 illustrates a portion of the location data and group data, inputs to the activity mapper. Note that (A) is the location record of a specific device (MAC address), while (B) is the entire *Grumon* record of the population. To extract group data of a particular device, records specific to the MAC address needs to be queried. For example, in (B), the 'groupid' consists of hashed MAC addresses joined as part of a group. The hashed MAC address must first be decrypted by applying a 1-way hash function, before its records can be filtered. In the interest of space, group data in (B) only reflects 1 or 2 participants in the group. Having more than 2 participants would mean that the 'groupid' will be a concatenation of more than 2 hashed MAC addresses.

Notice in (A) of Figure 2.2, location data is sensed every 5 seconds. Occassionally, these records tend to bounce from one access point to another, especially when users transit to other locations with their connected devices. To resolve the issues of fluctuating signals, the final location is determined by the *mode* of a 5-minute sliding window. Another problem is that devices may not always be connected to campus WiFi, thus constitute to missing values. Fortunately, missing features were not in large time intervals (no more than a few days), hence, treated with AKIMA interpolation. AKIMA spline affects only the curve of neighbouring data points, minimising the error of its estimates [127].

2.4.3 Mobility Features

Table 2.2 summarises all the mobility features. Once activities were assigned to all location entries, I proceeded to extract features as described in the following list.

1. ***Number of unique visits per day*** records the number of different buildings visited. Our university campus is comprised of seven buildings (five storeys each) and has an underground concourse that connects most buildings.

2. ***Total time spent on <activity type> & Number of times engaged in <activity***

Type	**Set**	**Mobility Feature**
General	Non-work (Location data)	count_building count_campus, time_campus count_gym, time_gym count_food, time_food count_clinic, time_clinic count_transit, time_transit
	Work (Location data)	count_study, time_study count_seminar, time_seminar count_consultation, time_consultation
	Group (Group data)	count_group, time_group count_smallgroup, time_smallgroup count_mediumgroup, time_mediumgroup count_largegroup, time_largegroup
Domain-specific	Workgroup (Location data)	count_meetings, time_meetings count_pairprogramming, time_pairprogramming count_knowledgeshare, time_knowledgeshare count_unique, time_unique count_knowledgeshare, time_knowledgeshare
	(Group data)	count_groupmembers, time_groupmembers

Table 2.2: Summary of feature types extracted from location and group data, grouped in sets.

type> per day consist of the following activities with 15 minutes unit time per activity: on campus (sums up all activities), studying, attending lecture, group meetings, study consultation, transiting, eating, exercising, visiting the clinic. Domain-specific workgroup activities include the types of tasks declared in the project schedule such as pair-programming, knowledge sharing, application design, and milestone preparation and *unique* events.

3. ***Total time spent being in <group type> & Number of times being in <group type> per day*** consist of the various group types listed in Table 2.1.

Fundamentally, these features were based on activity types; most of these activities can be generalised to the entire student population (e.g., studying, attending a lecture, eating), however, a subset of features were highly specific to students enrolled in the Software Engineering course (i.e., domain-specific features) as they were verified against their project team schedule. Note that each schedule entry captures location, date, duration, types of task, and attendees. Off-campus and contradicting entries, that is, a detected location which did not match the logged location (for a particular time of the day), were identified as 'unique' task.

Feature Sets: Further, these mobility features can be categorised into four broad categories; *Work* (W), *Non-work* (NW), *Group* (G) and *Workgroup* (WG). *Work* features are events that take place in locations such as open-study areas, seminar rooms, and group meeting rooms. *Non-work* features are events that take place in locations such as the campus gym, dance studio and cafeterias. *Group* features capture properties in the group data. *Workgroup* features are *Work* features verified against the project schedule to represent project specific events. Except for *Workgroup* features, all features are *General* in type. *Workgroup* features are *domain-specific*.

2.4.4 Changes in Behaviours as Features

We hypothesised that ***changes in an individual's behaviour and group interaction patterns in reference to themselves and their peers are key indicators of perceived stress***. This hypothesis is based on prior research that showed how changes in behaviour occur due to stress [62,63,128], and struggles to reorient could entail serious consequences [67].

$$x^o_{i,j} = \sum_{V \in [1..N] \setminus u} x^u_{i,j}/N - 1$$

$$\hat{x}^*_{i,j} = \sum_{k=i}^{i+w} x^*_{i,j} \quad i \in [1..K-w], [w] := \{3, 6, 9, \ldots, w\}$$

$$\boldsymbol{abs}^u_{\mathbf{i,j}} = \hat{x}^u_{i+1,j} - \hat{x}^u_{i,j} \tag{1}$$

$$\boldsymbol{rel}^u_{\mathbf{i,j}} = (\hat{x}^u_{i+1,j} - \hat{x}^o_{i+1,j}) - (\hat{x}^u_{i,j} - \hat{x}^o_{i,j}) \tag{2}$$

To implement changes in behaviours as features, each feature listed in Table 2.2 is compared against its own history of an earlier period, $x^u_{i,j}$, as ***absolute change*** (***abs***), as in Equation 1. In addition, I compared the changes an individual displayed against their peer population (i.e., users who were enrolled in the same course), $x^o_{i,j}$, as ***relative change*** (***rel***), as in Equation 2. These change values are calculated over different time windows in multiples of three days (as our ground truth data was collected every three days).

Conclusively, both *General* and *Domain-specific*-typed features comprises ***raw*** (i.e., original counts), ***abs*** (i.e., absolute changes) and ***rel*** (i.e., relative changes) features. In the upcoming chapters, I will present the top features used in building each detection model using ROC curve analysis to quantify the diagnostic ability of each feature.

2.5 Block 3: Detection Engine

At its core, *StressMon* uses a standard machine learning pipeline of feature selection and classification. Note that the design and evaluation of the stress and depression models are presented as chapters of their own in Chapter 4 and 5.

Similar to existing work [39, 44], I paid particular attention to solving a binary classification problem of detecting stress and depression. As illustrated in Figure 2.1, the engine is comprised of three machine learning models; they are for stress and depression, and a social identification model whose outcome is used as additional predictor for stress. In building each model, the engine adopts a recursive feature elimination (RFE) process to select the best set of feature (*General* and *Domain-specific*-typed) and parameters.

2.5.1 The Stress Model

Following prior work [26,44], which recommends using Random Forest (RF) to detect stress, I similarly adopted RF and compared the algorithm with Logistic Regression (LR) and Support Vector Machine (SVM), as much evidence has established these algorithms to outperform for binary classification [129, 130]. As justified in the next Chapter 4, the stress model was rigorously evaluated to best perform using the RF algorithm.

2.5.1.1 The Social Identification Model

In contrast, automatically detecting social identification through machine learning techniques is the first of its kind. With no prior work suggesting which technique works best, I maintained the same choices of algorithms for detecting social identification. In addition, I sought to consider the temporal sequencing of mobility features which is inherently related to using a HMM classifier [131]. This decision is based on prominent research which argues the identity one holds to their (work) group is likely to change over time from factors such as team conflict [14, 87, 132].

2.5.2 The Depression Model

In separate works by Canzian *et al.* [39] and Ware *et al.* [96], the authors employed a generic SVM model to solve a binary classification problem of depression. Keeping in mind that stress is commonly associated with depression, I made the decision to retain the choices of algorithm for depression similar to stress. The results from evaluating different classifiers presented in Chapter 5 found RF to perform best in detecting depression.

2.5.3 Success Criteria

StressMon is a successful solution if it fulfils the following:

1. ***The changes in an individual's behaviour make the strongest predictors.*** A key differentiating factor of *StressMon* (see Chapter 1.3) is ascertaining *relative changes* and *absolute changes* in behaviours to identify stress and depression more thoroughly.

2. ***Detect stress and depression accurately.*** In this dissertation, '*accuracy*' is defined by AUC, ideally a score close to 1. However, it prioritises TPR (i.e.,

students with severe stress or depressed students) over TNR.

3. ***Maintains high accuracy on other populations.*** *StressMon* must demonstrate generalisability by achieving high accuracy in detecting stress over a period of time, sufficient to suggest the stress is likely to be unmanaged, and depression over at least 2 weeks for students in *Study_Valid1* and *Study_Valid2* population.

2.5.4 Performance Metrics

As will be discussed in Chapter 4, I obtained a highly imbalanced dataset, particularly for stress, from the three user studies. To cater to class imbalances, I defined the metrics as (1) area under the ROC curve (AUC), indicating whether true positives are ranked higher than false positives and an AUC score close to 1 indicates a perfect predictor, (2) true positive rate (TPR or sensitivity) is the frequency the model correctly identifies positive cases out of all true positives, (3) true negative rate (TNR or specificity) is the frequency of the model correctly identifying actual negatives, and (4) overall accuracy (ACC). It is noteworthy that AUC alone can lead to potentially misleading results when the classifier is tested on an imbalanced dataset – for example, Florkowski *et al.* reported an AUC of 0.91 with higher sensitivity (97%) and lower specificity (62%, accuracy is 79%) [133]. Hence, it is important to consider all different types of metrics to evaluate the models.

In the case of *StressMon*, it is more important to build a classifier with high TPR than TNR to prioritise the accuracy of predicting students with signs of severe stress or depression rather than healthy students as being healthy. As the occurrence of healthy subjects were more common than others (resulting in skewed predictions on the majority class), I combined a resampling technique (strictly on the training set) with modified classification threshold to achieve high TPR over TNR [134].

2.6 Summary

StressMon is made up of three components. Block (1), Key sensing apparatus, comprised of the *LiveLabs* WiFi indoor localisation system and the *Grumon* group detector system, is used as crucial systems to collect location and group information. It is important to note that my contribution to this dissertation is focused on enabling accurate detection of stress and depression – requiring Block (2) and (3). Block (2), the Location feature extractor, details the steps for data-preprocessing, feature extraction and feature engineering using only location data. In addition, this component establishes how changes in individuals' normal behaviours were ascertained. That is, an individual's (individual and group) behaviours are compared against its own history (absolute change) and population (relative change) of an earlier period. Finally, Block (3), the Machine learning detection engine puts together separate models, mainly for stress and depression. In addition, a social identification model whose outcome is used as additional predictor for stress leverages on the same set of mobility features. The next chapter presents the accuracy of each set of predictors for detecting stress and depression.

2.6.1 Contribution of Thesis and Acknowledgement

The overall ideation and realisation of this research project were accomplished together with my advisors, Dr. Rajesh Balan and Dr. Youngki Lee. *StressMon* was orchestrated by the works of many. I thank SMU's LiveLabs Urban Lifestyle Innovation Platform for providing the *LiveLabs* and *Grumon* systems, to act as main sensing mechanisms, and resources to conduct the user studies for *StressMon*. In Table 2.3, I summarise how various people contributed to *StressMon*.

	Camellia	**Org./Person**
Gathering of user study resources, utilisation of assessment scales, design of user study procedure		With the help of Dr. Balan and Dr. Lee, I walked through the designs of the user study with several subject-matter experts as follows: (1) Dr. Kenneth Goh (Assistant Professor of Strategic Management, SMU) (2) Dr. Wiliam Tov (Associate Professor of Psychology, SMU) (3) Dr. Ada Chung (Head, Mrs Wong Kwok Leong Student Wellness Centre) (4) Dr. Mythily Subramaniam (Director, Research, Institute of Mental Health) (5) Dr. Janhavi Vaingankar (Deputy Director, Research, Institute of Mental Health) (6) Dr. Grace Park (Assistant Professor of Psychology, SMU) Additionally, the following instructors provided me with IRB-approved course materials: (7) Ong Hong Seng and Thiang Lay Foo (Software Engineering Course Instructors)
Management of user study resources	100%	-
Conducting user study interviews and coding	50%	(1) Vinoj Jayasundara, (2) Chetan Mittal (SMU LiveLabs Intern) Both students were employed by LiveLabs to assist in various tasks pertaining to the long-term user studies, primarily in preparing online application materials, conducting and coding interviews
Block 1: Passive Sensing Mechanisms	0%	SMU LiveLabs (100%)
Block 2: Location Feature Extractor		
Database management for all data collection	50%	(1) Ritesh Kumar, (2) Aniket Pujara (SMU LiveLabs Research Engineer) Both research engineers assisted in setting up the server for storing location data. Aniket assisted in generating group data in the *Grumon* system using the *LiveLabs*'s location data.
Development of activity mapper/heuristics	70%	(1) Chetan Mittal, (2) Vinoj Jayasundara (SMU LiveLabs Student Intern) Both students assisted me in improving the pre-processing pipeline for extracting/ assigning activities to location.
Block 3: Detection Engine		
Implementation of detection models	100%	-
Analysis of results	100%	-

Table 2.3: Contributions to *StressMon*

CHAPTER 3

LONG-TERM USER STUDY

In this chapter, I describe the long-term user studies conducted across various student groups during academic year 2017 (Fall) to 2018 (Fall) at Singapore Management University (SMU). These user studies were a means through which I sought to understand (1) students' behaviours on campus and group interaction with their workgroup peers, (2) their mental well-being and team processes associated with work stress, and (3) the validations of my assumptions and model solutions. Appendix A lists the IRB-approvals to conduct the studies. Appendix B and C list the validated scales used to collect ground truth data.

3.1 Ethical User Consideration

With ethics as the moral anchor of mental health research, the user studies were carried out with a full approval from the Institutional Review Board (IRB) at SMU. Potential risks include adverse psychological and social effects on students from reflecting on stressful events. Further, the user study was vulnerable to conflicts of interest since students' academic performances (grade point average, GPA, and course grades) were collected. Hence, the following risk mitigations were taken:

1. We consulted a team of medical professionals from an established mental health institution and academic professors from other disciplinary departments (Psychology department for mindfulness and mental health, and Or-

ganisational Behaviour department for team-related processes and states) to verify scales of assessments and questions for the interviews.

2. I emphasised voluntary participation in the study recruitment advertising to avoid any form of adverse coercion among students signing up for team participation. For this reason, I worked on time arrangements with 10 Professors for more than 20 different class sections to conduct in-class advertising.

3. Participants' anonymity was guaranteed in three ways – First, all participants were issued anonymised *Participant ID* to identify themselves in assessments/interviews. Second, I maintained a protected file that holds a mapping of issued IDs and actual records. Third, I manually anonymised unintentional disclosures of identities in materials such as interview findings and prior to storing information.

4. I informed and encouraged participants to make sound health consultations provided by our university from my first-hand encounters of psychological distresses; for example, when participants broke down emotionally during the interview session.

5. I avoided a potential conflict of interest, particularly in academic grading, by collecting course grades and GPA after term results had been finalised and published to all students.

6. Finally, I worked with course instructors who were not part of this research to gather anonymised Software Engineering project team schedules.

3.2 Participants

To validate *StressMon*, I conducted user studies during Fall 2017 to end of Fall 2018, recruiting different student populations. In particular, the primary user study, *Study_SE*, was purposed to build the detection models, while the other two studies, *Study_Valid1* and *Study_Valid2* were to validate these models. Table 3.1 summarises the demographics of the three studies.

3.2.1 Main Study: *Study_SE*

I recruited a total of 76 student participants (39 M, 37 F) in Fall AY2017, over a period of 81 days. All participants were in the second year of Information Systems (IS) major, and enrolled in the Software Engineering (SE) module that requires students to work in pre-assigned groups to build semester-long projects. SE groups are made up of 4 to 5 pre-assigned members, with no bias in gender, grade and nationality. Additionally, students work with people they do not know, and who have varying capabilities, personalities, and work styles. Project requirements and team groupings for Software Engineering are typically announced on the first day of class, and each team has the semester to build a fully working software artefact while following the various processes required. These processes required team members to spend equal amounts of time coding, then rotate programming pairs, and deliver in multiple releases. We chose this course as it is reported anecdotally as a highly stressful IS core module, mostly due to the pressures of having to work closely with people whom you do not know, and juggling a multitude of technical and management tasks.

3.2.2 Validation Study: *Study_Valid1*

Study_Valid1 is a smaller scale study in the Spring AY2017. This study ran on the second half of the Spring AY2018 semester for 36 days, typically a time perceived to be most stressful due to increasing deadlines and preparation for exams. 13 participants (3 M, 10 F) with a mix of Business Management and Social Science majors were enrolled in a Business elective module, Social Entrepreneurship. Students were in their Sophomore, Junior and Senior years. As part of the curriculum, students worked in their own groups to complete a business plan presentation and written report during the last few weeks of the semester.

	Study_SE (primary)	***Study_Valid1*** (validation 1)	***Study_Valid2*** (validation 2)
Period	Fall AY2017, 81 days	Spring AY2017, 36 days	Fall AY2018, 81 days
Total	76 students (39 M, 37 F)	13 students (3 M, 10 F)	51 students (24 M, 27 F)
Active	62 students (34 M, 28 F)	11 students (3 M, 8 F)	35 students (15 M, 20 F)
Team	50 students	0 students	25 students
Individual	12 students	11 students	10 students
Age	19 - 25 (22 med)	20 - 24 (22 med)	19 - 26 (22 med)
GPA	1.64 - 3.84 (2.85 med)	2.90 - 3.99 (3.33 med)	0 - 3.78 (2.63 med)
Major	Information Systems (62)	Finance (1) Business management (9) Social Sciences (1)	Information Systems (31) Business Management (1) Economics (2) Accountancy (1)
Study year	Sophomore (62)	Sophomore (3) Junior (3) Senior (5)	Sophomore (12) Junior (9) Senior (1) Freshman (13)
Course	Software Engineering (62)	Social Entrepreneurship (11)	Software Project Management (8) Interaction Design & Prototyping (13) Computational Thinking (1) Information Systems & Innovation (7) Programme in Writing & Reasoning (6)

Table 3.1: Demographics summary of participants from 1 primary (*Study_SE*) and 2 validation studies (*Study_Valid1* and *Study_Valid2*). GPA ranges from 0-4, 0 due to Freshmen with no GPA.

3.2.3 Validation study: *Study_Valid2*

In Fall AY2018, 51 students (24 M, 27 F), across different majors and year of study, participated. There was no specific target module; I approached instructors for five different courses being Software Project Management (SPM, IS core), Interaction Design & Prototyping (IDP, IS core), Information Systems & Innovation (ISI, IS core), Computational Thinking (CT, IS core), and Programme for Writing and Reasoning (PWR, University core). In all these modules, students must form their own groups to work either on a semester-long project (SPM and IDP) or short-term project (ISI, CT and PWR). None of these courses required the intensive juggling of technical and management practice, as is expected of SE.

3.2.4 Retention at Study End

Table 3.1 summarises details of all active participants, that is, students who contributed at least 80% of all survey types and attended at least one interview session. As with other long-term studies, managing user retention is critical yet inherently challenging. At the end of study, 62 (out of 72) participants remained active in *Study_SE*. 35 (out of 51) students, and 11 (out of 13) students were active for *Study_Valid1* and *Study_Valid2* respectively. Out of the 11 teams who signed up for *Study_SE*, 10 teams participated throughout the study, 5 teams enrolled and participated for *Study_Valid2*. *Study_Valid1* did not receive any team participation. I validated from the interview responses that none of the inactive participants had reported experiencing a highly stressful semester.

3.3 Study Procedure

Table 3.2 summarises all data collected for the user study. Each participant filled out a pre-study questionnaire outlining their campus routines (e.g., meal breaks, sports, frequented campus workspaces), academic (current GPA, modules enrolled), personal and social backgrounds (e.g., records of part-time jobs, sleeping habit, condi-

Type	Frequency	Data Collected
Demographic	Once	Household, Social, Education, Campus-related activities routines, Big-5 [135], Gratitude [136] Meaning in Life [137], Satisfaction with Life [138]
Experience Sampling	every 3 days	Sleep duration, PSS-4 [139]
Self Assessment	every 2 weeks	PHQ-8 [98], PANAS [140], Social Identification [141], Social Loafing [76], Social Cohesion [142] Group Potency [143]
Interview	mid term, term end	major sources of stress, accounts of significant group experiences (pos/neg), ways of managing team conflict
Post Interview Assessment (M)	term end	source of stress [144] American Time Use Survey on weekday activities [145]
MAC address of user's mobile phone	end of interview 1 end of interview 2	location traces from Wi-Fi localisation system

Table 3.2: Summary of different types of data collected under four different procedures; demographic survey, experience sampling, self-assessment and semi-structured interview sessions. Note: Post interview assessment was added as part of a modification procedure only for students from *Study_Valid2*, indicated with (M).

tion of living). Further, students self-assessed their personality traits (Big-5 [135]) and several other social and life indicators [136–138]. During the study, students reported their stress levels using the PSS-4 [139], which is commonly adopted among student and employee populations to evaluate their stresses [139, 146], every three days. Approximately every two weeks, corresponding to the timeline provided in Table 3.3, students participated in retrospective assessment for various measures such as depression [98] and social identification [141]. Note that the practising psychiatrists who evaluated our entire study strongly encouraged using PHQ-8 and not PHQ-9, to avoid the ninth question about suicidal thoughts – our research team was not trained to handle a positive answer to that specific question. As I wanted to strike a balance between frequency of surveying and reducing user burdens, administering surveys every three days allowed me to collect sufficient samples for every day of the week. Retrospective assessments were also timed before/after a critical

Event Description	Period (in Day)
Collect assessment #1	3
M1: Release of proj. specs.	04 - 08
Collect assessment #2	15
M2: Team Goal	25 - 29
Collect assessment #3	36
1-Week Recess	39 – 43
Conduct interview #1	39 – 43
M3: User Acceptance Test (UAT)	53 – 56
Collect assessment #4	57
M4: Final deliverable	74-78
Collect assessment #5	75
Conduct interview #2	77-81

Table 3.3: Data collection periods were timed before and after critical SE milestones (shaded rows and indicated as M#).

SE project milestone.

Additionally, students attended two semi-structured interview sessions at the midpoint and study's end. Sessions were guided by the questions stated below, with follow-up questions to better understand how their team experiences affected different aspects of their perceptions of being part of the team or how these experiences were part of their major stressors.

1. What is your main source of stress and experiences of critical (positive or negative) team events? Elaborate.

2. Did any of these events change the dynamics of the team, and if so, how did it emotionally affect you?

3. If applicable, were problems in the team solved and how did the group communicate?

As part of the interview, students also provided their mobile phone MAC address, allowing their location traces to be queried from the *LiveLabs* and *Grumon* systems. Finally, *Study_SE* participants provided access to their SE project schedule (i.e., a graded document maintained by all teams to keep track of project plans). These records included information about meeting dates, duration, and location.

Note that all surveys collected were purposed as ground truth labels to validate *StressMon*'s detection models - once the system is operational, it evaluates stress solely based on passive location data.

3.3.1 Modification to Study Procedure

Surveying the main sources of stresses through open-ended means (semi-structured interview) among students from *Study_SE* and *Study_Valid1* proved inadequate to understand the nuances and details of stresses students typically face, and thus led me to revise a small portion of the second interview session with a survey adapted from Yumba *et al.* [144]. The survey clearly defined the types of stress: academic, personal, financial and environmental factors. I adopted the American Time Use survey [145] to clarify the amount of time students spent on different activities on a weekday.

3.3.2 Survey Portal

To enhance the process of data collection, I administered all surveys using Qualtrics [147], embedded in a custom online portal driven by October-CMS open-source PHP platform [148]. This portal, as illustrated in Figure 3.1, served as the primary platform to facilitate in sending survey reminders and compensation updates. This portal can be accessed at http://apollo.smu.edu.sg/mh_portal/ index.php

3.3.3 Compensation

Students were compensated with a maximum amount of USD 30 in two ways, (1) for entering the study, and (2) for remaining active respectively. The compensation amount varies depending on how active they were at providing self-reported surveys. Active participants were entered into a lucky draw to win a USD 76 cash prize. In addition, a USD 37 bonus was offered to participants whose entire project group joined and completed the study. All compensations were generously funded

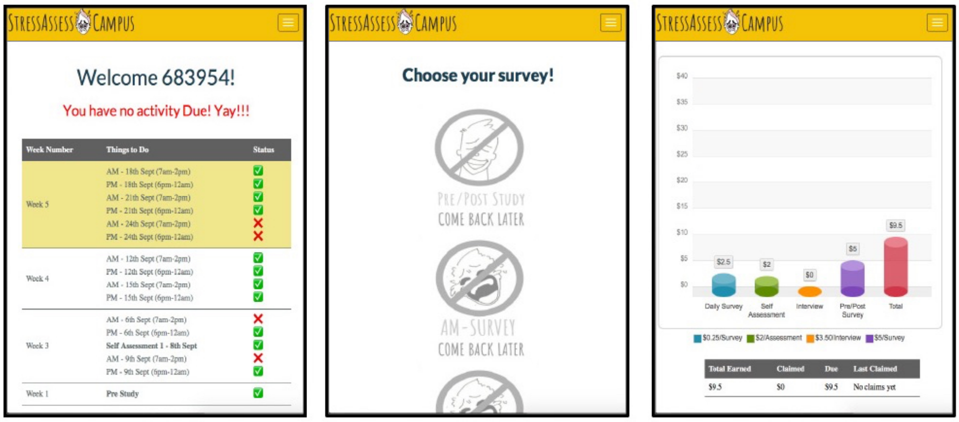

Figure 3.1: *StressMon* survey portal for data collection. This portal was developed using October CMS and Qualtrics, for students to participate in surveys and check monetary information.

by National Research Foundation (NRF), Prime Minister's Office, Singapore under its IDM Futures Funding Initiative.

3.3.4 Description of Assessment Scales

Despite a wide range of survey data being collected in these user studies, not all of these assessments were used in this dissertation. Instead, they contribute to additional studies to potentially expand the capabilities of *StressMon*. In this section, I describe the scales used to primarily validate *StressMon*.

Perceived Stress Scale (PSS-4): PSS-4 is a well-established scale, ranging from 0 to 16, used extensively by students and employees [146, 149]. PSS-4 accounts for negative and positive typed stresses through scoring. Having a score close to 16 suggests severe amounts of negatively perceived stress [139]. Severe negative stress has been found to result in adverse cognitive and emotional consequences and in vulnerability to depression [150]. It is important to note that PSS-4 is not a diagnostic tool. Instead, the scale is only used to compare between people in the same population.

Patient Health Questionnaire (PHQ-8): The use of PHQ-8 is more straightforward as the scale is a diagnostic tool with clear cutoffs – 0-4 (no/minimal depression), 5-9 (mild depression), 10-14 (moderate depression), 15-19 (moderate-severe depression), and 20-24 (severe depression). In fact, a PHQ-8 score above 10 can yield categorical diagnosis of clinically significant depression [98].

Four-item Social Identification (FISI): The FISI item is the rating on a 7 point Likert scale of ones agreement on the following statements; *"I identify with [In-group]", "I feel committed to [In-group]", "I am glad to be [In-group]", "Being [In-group] is an important part of how I see myself"*. A higher score indicates high level of social identification. The assessment needs to be conducted with more than one group in order to determine if an individual highly identifies themselves with their group. However, an alternative way is to conduct longitudinal studies on how differently they assess themselves relative to the group over time [151].

Big Five Inventory (Big-5): The Big-5 is a test for personality traits on five different dimensions; *Extraversion*, *Agreeableness*, *Conscientiousness*, *Neuroticism* and *Openness to experience*. For each of these personality dimension, the scale has 6 sub traits for each trait to be assessed independently. In this study, I am particularly interested in participants' *Neuroticism* scoring. A high score indicates an individual experiencing negative feeling such as anxiety, anger, or depression.

It is important to note that with the exception of Big-5, all scales were used as validation of *StressMon*; as presented in Chapter 5, *Neuroticism*, one of the five personality traits of Big-5, was used as a predictor for *StressMon*'s depression detection model. Further, While these scales used as ground truth, the scores were not treated as a diagnosis for any mental health condition. Instead, scores for each assessment were processed into binary categories – for example, for stress, it is **0:** ***normal stress*** and **1:** ***severe stress*** . Process of creating labels will be addressed in the following chapters.

3.3.5 Location and Group Data

The bulk of my data collection comprised of WiFi signal sensed directly from every AP to generate location and group information. I extracted three months worth of WiFi signal data for *Study_SE* and *Study_Valid2* and one-month for *Study_Valid1*;

this accumulated to an average of 5,000 preprocessed location data points (equivalent to 7 hours of location data) and 820 preprocessed group data points (2 hours of group interaction data) per participant each day. Students were also detected to have visited, on average, 96 unique locations on campus each month.

3.3.5.1 Description of Location Data

Each WiFi location entry corresponds to a connection made from the mobile phone to an AP every 5 seconds. Each tuple consists of [d_i, u_i, l_i, a_i], where d is datetime stamp, u is the hashed MAC address of connected devices (representing the users), l is the location code at which the device is localised, a is the accuracy of the localisation and i is the number of entries in the dataset. Each location code, l_i, is mapped to a location name (in the format of <building name>_<level>_<room name>), the location's maximum capacity and current occupancy. Thus, each tuple informs us of the amount of time a user is detected to be at a room-level location and how 'busy' the location is between 97-99% accuracy.

3.3.5.2 Description of Group Data

Each group data extends a location entry with [d_i, g_i, ct_i, ll_i, tt_i, lh_i, s_i], where d is datetime stamp, g is a concatenation of hashed MAC addresses connected to the same AP over a period of time, ct is the last datetime the devices were detected as a group, ll is the last location code the devices were detected as a group, tt is the total time detected as a group, lh is the location history which provides a concatenation of location and the detected time, and s is size of group or number of devices concatenated in g. Consequently, group data gives information pertaining to users who make up the group, and the various locations and amount of time spent at different locations over a period of time.

3.4 Summary

Three rounds of long-term user studies were conducted as part of this dissertation. *Study_SE* was a primary study purposed to build the detection models in *StressMon*, while *Study_Valid1* and *Study_Valid2* were purposed to validate these models. I recruited a total of 140 undergraduate participants. The studies ran between Fall AY2017 to the end of Fall AY2018, running for a maximum of 81 days. 108 (out of 140) students actively participated in a standard protocol (of experience sampling every 3 days, retrospective assessment approximately every 2 weeks, and semi-structured interview sessions) to capture their self-evaluations on mental well being and workgroup processes, as they naturally occur in their everyday life. While a wide range of survey data was collected, in this dissertation, I narrowed my scope of investigation to stress and depression and several strong indicators associated with the mental health issues.

CHAPTER 4

DETECTION STRESS & DEPRESSION

In this chapter, I address three research questions: First, *"Can we detect stress in a person using only location data?"* and, subsequently, *"What effects do group-related features have on stress detection?"*. With substantial evidence of individuals experiencing depression from severe stress, I will address the third questions: *"Can we detect depression in a person using only location data?"* and, subsequently, *"What effects do personalised features have on depression detection?"* This chapter describes in detail the processes followed in building a stress detection model and the outputs it produced. In addition, the chapter includes the development of a social identification model to produce an optimisation output for better stress detection results.

4.1 Preprocessing Labels Procedure

Recall in Chapter 3, students self-reported their stress levels every three days using the PSS-4 as ground truth (see Table 3.2). First, I describe how these stress scores were mapped into their binary equivalent and the way missing survey data was handled. Then, I discuss the technique employed to deal with a highly imbalanced dataset from converting the scores to binary labels.

4.1.1 Labelling Stress

In all three user studies, PSS-4 scores have the following distribution: min=1, max=16, median=8, mean=7.66, SD=2.35. I divided the scores into two groups: ***severe stress*** (1, positive class) for the scores of 12 and above, two standard deviations away from the mean, otherwise ***normal stress*** (0, negative class). Since PSS-4 is not designed as a diagnostic tool for severe stress, the threshold was referenced from a work by Wartig *et al.* that provides norms for a sample with various ethnicity (White, Mixed, Black African, and Asian, N>1500) [146].

Treating Missing Data: Note that all experiments in this dissertation utilised data strictly from active participants, who self-reported at least 80% for each assessment. That is, out of 27 stress samples collected, some students would have missed a maximum of 6 data points. I treated the missing data by performing multiple imputation, particularly using multivariate imputation by chained equations (MICE) [152]. Unlike more conventional methods (e.g., taking the mean or last observation carried forward), MICE creates multiple predictions for each missing data from (generally) 10 cycles to create a complete dataset as the coefficients in the regression models converge to stability. When the existing data is highly predictive of missing values, the imputation will result in small but accurate standard errors [153].

	Study_SE	***Study_Valid1***	***Study_Valid2***
no. of participants	62	11	35
severe stress	145 (9%)	3 (2%)	1 (1%)
normal stress	1529 (91%)	129 (98%)	944 (99%)

Table 4.1: Distribution of stress labels for all studies; 27 samples per *Study_SE* and *Study_Valid2* participants, and 12 samples per *Study_Valid1* participants.

Label Data Distribution

Table 4.1 lists the distribution of labels. The PSS-4 conversion resulted in a distribution of more than 90% *normal stress* labels for all studies. Prior work [154] suggests that the label imbalance is to be expected as individuals overwhelmed by stress tend to be outliers. Such skewed datasets can lead to poor prediction performance if not corrected [154].

Treating Imbalanced Dataset:

I addressed the problem of imbalanced data-set by applying SMOTE [155] to synthetically oversampled training set data in the *severe stress* classes. SMOTE is widely applied in similar dataset problems as my own, and has shown to improve over other re-sampling techniques including modifying loss ratio and class weights [156]. Instead of duplicating original observations (likely to result in overfitting), SMOTE works by first segmenting the data and generating synthetic observations that closely represent the original data. That is, a synthetic sample is created between an original data point and its nearest neighbour. By creating new examples inferred from existing ones, SMOTE avoids the problem of overfitting (from randomly copying existing examples), but it introduces noise. Through SMOTE, the handling of imbalanced dataset is achieved independently of learning algorithm. Past experiments have found classifiers benefiting from SMOTE if data are low- dimensional (variables are not more than 100), and SMOTE reduces the bias towards the classification in the majority class for SVM, RF and CART [157]. Note that up-sampling was strictly contained in the training set; thus, the presented results reflect the true performance of an unaltered imbalanced test set.

4.1.2 Labelling Depression

Based on related work [96, 158], assessments with PHQ-8 score >= 10 are treated as a significant level of depression. Accordingly, scores of 10 and above (min=0, max=24, median=8, mean=8.23, SD=4.77) were grouped as ***depressed*** (1, positive class), otherwise ***non-depressed*** (0, negative class).

	Study_SE	***Study_Valid1***	***Study_Valid2***
no. of participants	62	11	35
depressed	534 (32%)	28 (21%)	330 (35%)
non-depressed	1140 (68%)	104 (79%)	615 (65%)

Table 4.2: Distribution of depression labels for all studies; 27 samples per *Study_SE* and *Study_Valid2* participants, and 12 samples per *Study_Valid1* participants.

Label Data Distribution

Table 4.2 lists the distribution of labels. The PHQ-8 conversion resulted in approximately 30% depressed students for all studies, matching published statistics of about 4 million college students in the US [159]. Similarly, the data imbalance was treated with SMOTE sampling.

4.2 Hypothesis Testing

Using the transactional model of stress and coping [53] as a way to understand how individuals cope with activities, we know that coping is treated as the transactions between the individual and environmental factors, their perceptions of stress. The effectiveness of one's coping strategies [53] determines if an event is indeed perceived as stressful. As part of coping, for instance, it has been suggested in past works, when experiencing stressful situations, individuals are likely to change their behaviour, environment or the way they evaluate the situation. Miller *et al.* found in their study that the change in behaviour was reported a common reaction in coping with stress. Specifically, individuals reporting higher levels of stress were more likely to change their activity entirely [160].

To explore the feasibility of detecting stress using location data, I first developed a set of hypotheses on mobility-driven features. I characteristically determined the factors contributing to the presumed effect of stress based on common interview responses by *Study_SE* students (to be discussed in Chapter 6), including how students mainly felt stressed from spending the majority of their time on study-related matters and how Software Engineering group project required much of their time.

Hypothesis testing was performed to statistically validate features that can differentiate students experiencing *severe stress*. For the ground-truth labels, I placed the students in *Study_SE* into *severe stress* (n=4) and *normal stress* (n=58) categories, based on their total average PSS-4 scores.

Methodology: The analysis was conducted in the following order: First, I visually examined the changes in mobility features over time, between the two groups by averaging features every three days, and plotting them over the study duration of 81 days. I defined *Time Point, T_x* as a sample made every 3 days – i.e., T_{24} = 24 * 3 = day 72 of the study. A secondary axis was added to plot the average stress levels (measured by PSS-4 score) between groups. Then, I performed one-way MANOVA to investigate the significance of the multivariate mean effects on different features, and ran individual t-tests with Bonferroni correction to check for specific mean differences across time periods.

4.2.1 Campus Routines

Based on prior research [161] which found high correlations between high perceived stress of students and absenteeism (in classes and work-related activities), I formulated hypotheses beginning with a conjecture that students with *severe stress* are more likely to reduce their (physical) presence and interactions with working peers on campus.

H1: Students with *severe stress* spend fewer hours on campus.

Overall, I observed students with *normal stress* incrementally spent more time on campus, especially towards the second half of the semester. This is an expected trend because students generally spend more time working on projects and preparing for examinations as the semester ends. Yet, students with *severe stress* were found to spend significantly less time on campus (p=0.04), specifically on T_{24} and T_{26} (see Figure 4.1). Interestingly, students with *severe stress* began to show decline participation during the same time their stress level peaked on Day 21.

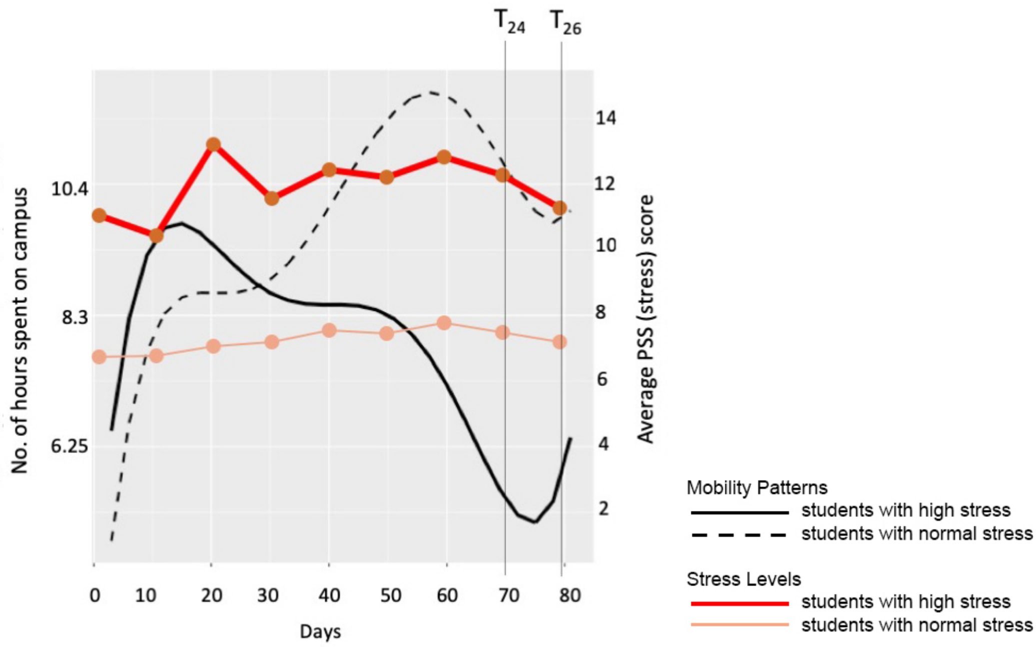

Figure 4.1: Mean plot on total time spent on campus between *normal stress* and *severe stress* student groups - T24 and T26 are highlighted time points where the two distributions were statistically different (p=0.04).

H2: Students with *severe stress* participate less in work activities on campus.

Students with *severe stress* were significantly less involved ($p<0.01$) in work-related activities (i.e., seminar attendance, self-study and group project activities) than students with *normal stress* (see Figure 4.2). A more interesting observation is how severely stressed students began the semester displaying more participation in these activities, but decreased over time. Further, two large dips occurred around the recess week (T_{12}-T_{14}) and the end of the semester (T_{23}-T_{25}). These time points closely corresponded to important SE project milestones and were significantly different between both groups (see Table 3.3 for critical SE milestones). As with the time spent on campus (Figure 4.1), students with *severe stress* demonstrated declining participation in their study activities at the start of Day 21, with more significant time points spread out throughout the semester.

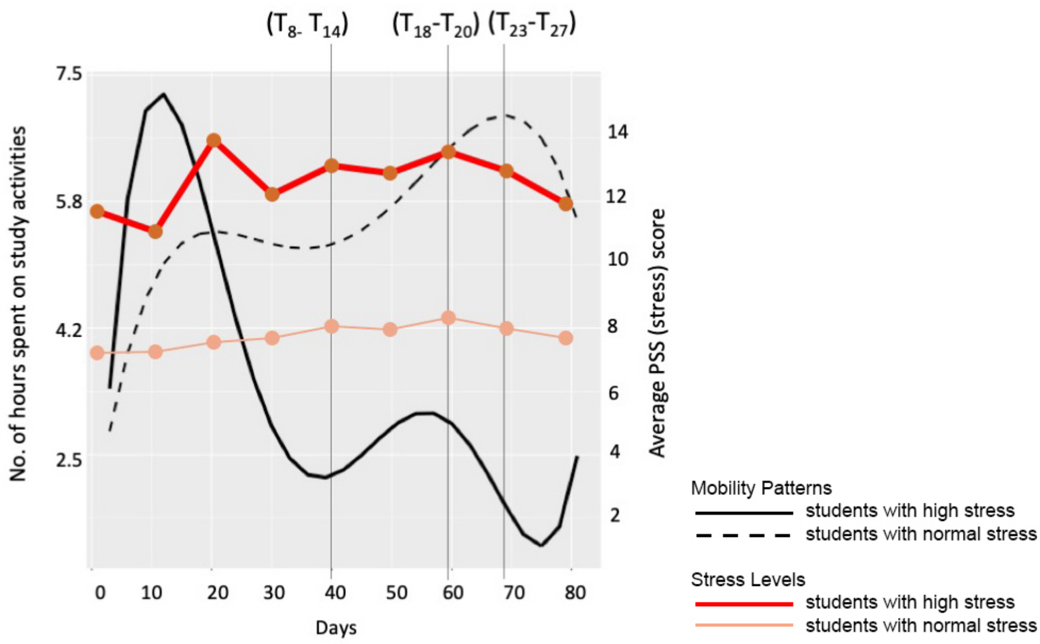

Figure 4.2: Mean plot on total time spent on work-related activities between *normal stress* and *severe stress* student groups - multiple time points are highlighted where the two distributions were statistically different (p<0.01).

4.2.2 Group Interaction

The interview findings identified almost half (42%) of the *Study_SE* students who expressed Software Engineering as the primary stressor, and attributed the overwhelming pressure to relationship tension. With group interaction proving as crucial indicators of stress, I hypothesised:

H3: Students with *severe stress* participate more in SE-related activities.

Figure 4.3 illustrates trends, specifically for SE sessions among our students. Having SE (anecdotally) as one of the most stressful modules among *Study_SE* students, I expected students with *severe stress* would spend significantly more time on SE. This is true, however, specifically for project management matters (p=0.04). Note: students with *severe stress* spent less time on programming-related tasks (on campus) but the difference is not significant. In the beginning, their involvement appeared to be significantly lower than the rest at T_4, gradually peaked during the recess week (Day 39-43), and were significantly different at various time points. As

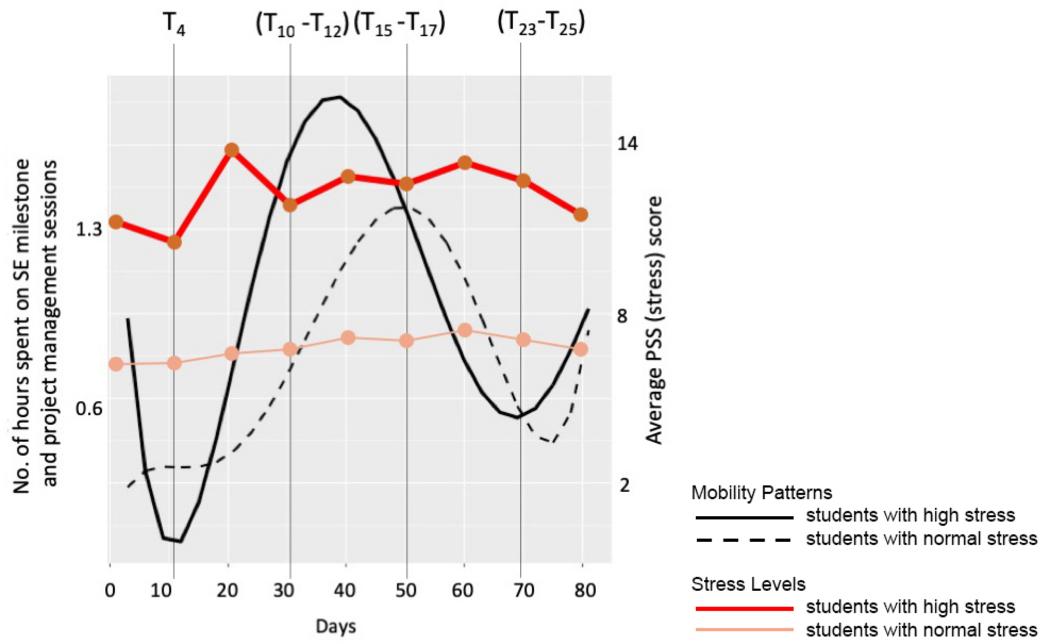

Figure 4.3: Mean plot on total time spent on Software Engineering project management and knowledge sharing sessions between *normal stress* and *severe stress* student groups - multiple time points are highlighted where the two distributions were statistically different (p=0.04).

explained, this is a typical time among students to ramp up their SE project for two reasons – (1) class breaks allowed for more project time and (2) it is right before a major project milestone (user testing). A similar pattern was also displayed between Day 70 to 80, which corresponds to the final project milestone.

H4: Students with *severe stress* are generally likely to be more involved in group activities on campus.

From Figure 4.4, I found students with *severe stress* spent significantly more time in groups on work-related activities than *normal stress* students (p=0.04). For example, at T_7, severely stressed students spent approximately 5 out of their total 9 hours on campus involved in group-related activities. Note, this record does not distinguish between work and non-work group activities. However, it was noted from the interviews that students spent most of their group projects on Software Engineering. Their group involvement dipped during the term break when group projects were most expected. Despite showing higher stress levels towards the end of the semester, group participation among students with *severe stress* continued to peak on Day 70.

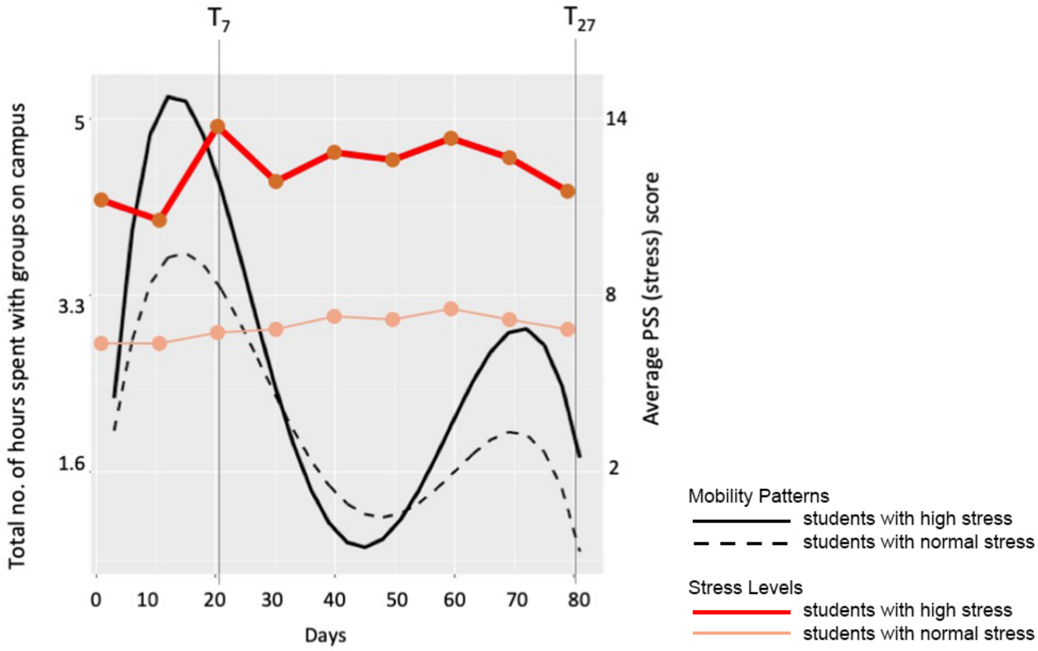

Figure 4.4: Mean plot on total time spent with groups on campus between *normal stress* and *severe stress* student groups - T7 and T27 are highlighted time points where the two distributions were statistically different (p=0.04).

4.2.3 Theoretical Argument for Hypothesis Results

To date, cognitive psychology has led researchers to promising leads on the effects of stress. Assuming that unmanaged stress will, at some point, change one's behaviour, this hypothesis testing sheds light on the activities that were most significantly affected by stress. Note that all features were tested but not all yielded significance.

The transactional model of stress predicts that how an individual interprets a stressful event influences their reaction to the situation. Additionally, Zohar *et al.* found that while individuals are likely to experience various stressors at once, the most salient stressor takes controls over how the person appraises and respond to a situation. In this case, I determined that severely stressed students exhibited significantly different behaviours in their overall involvement on campus and with their

Software Engineering group. Separately, I determined through the interview sessions that Software Engineering was their main stressor throughout the semester. For three severely stressed student, workgroup situations were particularly threatening as they perceived themselves as not possessing the skills required to fulfil group tasks compared to others in the team. In contrast, one student expressed he did not receive the support of other members to deliver the project up to his standard. The concept of coping with stress was originally focused on the individual aspect in terms of personality variables [162]. However, ongoing research efforts today are defining stress as per [52], where *individuals, collectives and place must be treated as one unit, not independently*. These findings led me to conjecture that a measure sufficiently representing the social dynamics of individuals within their workgroup is a key indicator of stress.

4.3 Evaluation of Stress Model

Table 4.3 summarises the results with different sets of features used for the stress models, ***Model$_{S_SE}$*** and ***Model$_S$***. ***Model$_{S_SE}$*** is specific to *Study_SE* student population as it used *Domain-specific* (SE-related) features. ***Model$_S$*** is a generalised version which excluded all *Domain-specific* features.

			Feature Settings			**Performance**	
Sec.	**Model**	**Study**	**Type (Set)**	***raw/abs/rel***	**Interval**	**AUC**	**TPR** (%)
4.3.1	***Model$_{S_SE}$***	*Study_SE*	*General (All) + Domain-specific (All)*	***rel+abs***	6-days	0.96	98.93
4.3.2	***Model$_S$***	All	*General (All)*	***rel+abs***	6-days	0.97	96.01
4.3.3.2	***Model$_{S_SE+SI}$***	*Study_SE*	*General (All) + Domain-specific (All) + SI*	***rel+abs***	6-days	0.96	88.08

Table 4.3: Summary of stress model configurations achieving best performances. Sections 4.3.1, 4.3.2 and 4.3.3.2 provide detailed results for each experiment. ***Model$_{S_SE}$*** is a highly specific stress model that uses *Domain-specific* (SE-related) features, while ***Model$_S$*** excludes all *Domain-specific* features. ***Model$_{S_SE+SI}$*** adds social identification binary outcome as additional feature to detect stress.

Methodology: I conducted the evaluation in three parts: First, I performed a *Group K-fold* cross-validation (CV), whereby 12-13 distinct students from *Study_SE* made up each test fold to determine various model settings for $\boldsymbol{Model_{S\_SE}}$. Next, I conducted *Train-Test* by training on the whole *Study_SE* dataset, and validated individually on *Study_Valid1* and *Study_Valid2* . Finally, I built an *All-population* model by combining all users from three populations and performed a Group K-fold CV. A Leave-one-out validation was not used due to the highly imbalanced dataset which might result in no one *severe stress* sample in a user. Across 12-13 students each contributing 27 samples, at least 1 sample in each group reported *severe stress*. Note that input to model and performance metrics was defined in Chapter 2.5.

Input and Output: Datasets to build and evaluate the stress model are made up of location and group features. Over an 81 day study period, these mobility features were aggregated every three days, where time periods correspond to the frequency of self-reports collected as ground truth labels. Accordingly, the labels are mapped to each feature vector. The output for the stress model is a binary outcome of **0:** ***normal stress*** and **1:** ***severe stress***.

4.3.1 Cross-Validation Experiment: *Study_SE* – $\boldsymbol{Model_{S\_SE}}$

Classifier: First, I determined the use of algorithms, comparing Support Vector Machine (SVM) and Logistic Regression (LR) against Random Forest (RF) and all features, as the base classifier. Tuning all classifiers to achieve good performance on the positive class (high AUC), I empirically determine the cutoff of the classifier, which is typically set at 0.5 to 0.45 (thus, at the cost of a high false negatives rate). As shown in Table 4.4, RF yielded significance with AUC=0.97 (at p=0.01 level) than LR and SVM (0.57 and 0.86 respectively) and, subsequently, was retained as the choice algorithm.

Features: Next, I investigated the hypothesis that change features make the strongest predictors of stress (see Section 2.4.4). I achieved the highest AUC score of 0.97 using a combination of raw and change features (***raw+rel+abs***). However,

	Logistic Regression (LR)	**Support Vector Machine (SVM)**	**Random Forest (RF)**
AUC	0.57	0.86	0.97
TPR (%)	69.77	79.41	99.60
TNR (%)	35.35	68.27	72.00
ACC (%)	37.35	69.27	74.44 (*)

Table 4.4: Results from using all features on different algorithms. (*) indicates significance at p=0.01 level.

	raw+rel+abs	***raw***	***rel+abs***
AUC	0.97	0.91	0.95
TPR (%)	99.60	98.93	99.33
TNR (%)	72.00	59.65	69.49
ACC (%)	74.44	63.18	72.08

Table 4.5: Results from using different combination of feature types on chosen Random Forest algorithm; All (***raw+rel+abs***), Raw (***raw***) and Change (consists of relative change, ***rel*** and absolute change, ***abs***).

	All Change ***rel+abs***	Individual (location data)	Group (group data)
AUC	0.95	0.85	0.69
TPR (%)	99.33	95.7	97.99
TNR (%)	69.49	60.44	22.03
ACC (%)	72.08	63.37	28.38

Table 4.6: Results separating the changes in group-related features from individual (routine) features; Change (***rel+abs***), Individual (change features extracted from location data) and Group (change features extracted from group data) to detect stress at 3-days interval.

the addition of ***raw*** set did not lead to significantly better performance. Hence, I retained only the change set as a smaller set of features to avoid developing an over-fitted or computationally expensive model. Using change features, I was able to achieve an AUC of 0.95 (see Table 4.5). In addition, I used recursive feature elimination (RFE), with a backward elimination of step size=1, on all change features. However, no change features were completely redundant and the performance of the RF classifier peaked with all change features considered. Note: this includes *Domain-specific* change features.

Individual and Group Interaction Features: To better understand the best set of features used in the stress detection model, my next step was to compare each

Description	Type	varImp
Number of times engaged in *studying*	***abs*** (*Work*)	100.0
Number of times being in *solo group*	***abs*** (*Group*)	48.57
Number of times being in *solo group*	***rel*** (*Group*)	40.22
Number of times engaged in *eating*	***abs*** (*Non-work*)	33.68
Number of times engaged in *studying*	***rel*** (*Work*)	32.38
Number of times engaged in *exercising*	***rel*** (*Non-work*)	31.83
Number of unique building visits	***abs*** (*Non-work*)	29.60
Number of times engaged in *transiting*	***abs*** (*Non-work*)	29.21
Total time spent with *all groups*	***abs*** (*Group*)	25.34
Number of times engaged in *attending lectures*	***rel*** (*Work*)	24.67

Table 4.7: Top 10 features for detecting *severe stress*, using ROC curve analysis, and sorted by variable importance (varImp).

model performance using only individual routine features (features extracted from location data, see Table 2.2) and social interaction features (features extracted from group data). As summarised in Table 4.6, the use of group-related features alone did not yield high performance. Instead, a large portion of its inaccuracies was attributed by the signification reduction in TNR. In contrast, the changes in individual routines make stronger predictors from correctly classifying more negative cases. However, the combination of all change features proved to significantly improve overall accuracy to 72.08% from 63.37%. Table 4.7 lists the top 10 features sorted in the order of variable importance, that is, the ROC curve analysis conducted on each predictor was used as the measure of importance. Accordingly, I retained all change features.

Time Window Experiment: Finally, I sought to determine the time at which the stress model would detect *severe stress* most accurately. With gradual time increase every 3 days (corresponding to the frequency of PSS-4 samples collected), I observed the highest AUC=0.96 on a 6-days interval. The reduced misclassification rate of 22.41% with 6-days interval is a significant improvement (at p=0.1 level) compared to the 3-days interval (see Table 4.8). As the interval increases to 9-days, both TPR and TNR achieved comparable results, leading to better overall accuracy of 85.48%.

At this point, it is important to consider time as key factor of intervention. That

	3-days	6-days	9-days	12-days	15-days	18-days
AUC	0.95	0.96	0.89	0.82	0.89	0.83
TPR (%)	99.33	98.93	88.55	77.91	67.35	56.93
TNR (%)	69.49	75.54	84.62	87.86	90.27	91.52
ACC (%)	72.08	77.59 (*)	85.48 (*)	87.93	89.41	89.83

Table 4.8: Results from calculating chosen Change type (***rel+abs***) features on different time intervals; from 3 to 18-days. (*) indicates significance at p=0.1 level.

is, while stress is regarded as an everyday experience and chronic stress evolves over a longer period of time, prolonging detection of severe stress by more than a week might result in students missing out on vital help, leading to depression. Since timeliness should be prioritised over accuracy, similarly, prioritising true positives over true negatives, I concluded the best model settings for detecting severe stress using Random Forest (RF) algorithm, all change set features (***rel+ abs***) calculated at a 6-days interval –$\boldsymbol{Model_{S\_SE}}$.

4.3.2 Additional Validation: *Study_Valid1 & 2 – $Model_S$*

Using all change features included *Domain-specific* (SE) features for $\boldsymbol{Model_{S\_SE}}$; thus, it is highly tailored to SE students in the *Study_SE* sample. To build a generalised model for other populations, I excluded all *Domain-specific* features as $\boldsymbol{Model_S}$. First, I trained on *Study_SE* sample and tested on different populations. Then, I performed a Group 5-fold CV on all three populations. Table 4.9 lists our results in detail.

Method	Train	Test	AUC	TPR (%)	TNR (%)	ACC (%)
Train-Test	*Study_SE*	*Study_Valid1*	0.91	66.67	90.70	99.17
	Study_SE	*Study_Valid2*	0.94	100.0	81.25	81.27
Group5-fold (All population)	Folds 2-5	Fold 1	0.98	94.44	86.81	87.06
	Folds 1,3-5	Fold 2	0.96	88.88	84.26	84.34
	Folds 1-2,4,5	Fold 3	0.97	98.64	80.88	82.09
	Folds 1-3,5	Fold 4	0.96	98.07	75.46	77.65
	Folds 1-4	Fold 5	0.96	100.0	76.38	77.65
		Average	0.97	96.01	80.76	81.76

Table 4.9: Summarised results for stress model, $\boldsymbol{Model_S}$, on three different validations. $\boldsymbol{Model_S}$ is a generalised stress model that excludes all *Domain-specific* features.

This solution achieved a reduced AUC=0.91 and 66.67% TPR for *Study_Valid1*. That is, out of 3 *severe stress* reported by 2 students, 1 was misclassified. Approximately 10% misclassification (90.70% TNR) was as a result of 6 students, 2 of whom did not report *severe stress* but felt *depressed*. The same test on *Study_Valid2* students successfully yielded an AUC=0.94, 100% TPR as it correctly detected 1 *severe stress* instance. While misclassification rate dropped to 18.73% (81.25% TNR), the false detection, unfortunately, affected most students (31) in the sample. However, out of 35 students, 14 had reported feeling *depressed* despite not experiencing *severe stress*.

The final step combined all students from three user studies to build an all-population stress model, evaluated using a Group 5-fold CV. I achieved an average AUC=0.97 and 96.01% TPR (4 out of 149 *severe stress* instances would go unnoticed). Unfortunately, the overall accuracy of 81.76% continued to affect most participants by identifying them as severely stressed at some point in the study. Further, at least 19 (out of 108) students who did not report any accounts of *severe stress* were misclassified multiple times during the study.

4.3.3 Optimising Stress Model

To the best of my knowledge, there has been no Systems-related work which examines social dynamics specific to workgroups in detecting stress. Recall in Section 4.3.1, I did not fully capitalise on the *Domain-specific* features (representing workgroup interactions), which only achieved an overall accuracy for 77.59% and was eventually left out to build a generalised model for other populations. In this section, I demonstrate how *Domain-specific* features can be further exploited to separate workgroup patterns between groups of students. Prior work has found social identification to influence the time individuals dedicate themselves to working with the team [163]. Separately, it was concluded to be a strong indicator of stress from reasons such as working long hours and burnout [15]. Accordingly, I experimented on generating a collective indicator for individuals of their workgroups using mobility features – specifically, by artificially categorising the measure of social identification. The experiments presented in this section only examined *Study_SE* population for *Domain-specific* features to be used. Unfortunately, generating this set of features cannot be generalised to other populations as it requires an additional validation of a specialised resource such as a project schedule. It is important to note this dissertation did not investigate the temporal dynamics of social identification and its relationship to stress – instead, it sought to easily differentiate both workgroup-specific behaviours and interaction patterns of individuals from the process of distinguishing two groups of students with contrasting identification. However, I discuss several findings of social identification in relation to stress by conducting my qualitative assessments in Chapter 5, and present future extensions to this work in Chapter 7.

4.3.3.1 Labelling Social Identification

To determine a student's social identification at study's end, I performed a thorough inspection on all self-reports related to social identification, specifically by verifying 5 samples of Four-Item Social Identification (FISI) assessment [83, 141, 164] with 2 interview responses. Recall, the assessments were intentionally timed close to critical SE milestones (listed in Table 3.3), as I expected to observe changing identifiers during stressful situations of meeting project milestones. Missing data were similarly treated with MICE (see Section 4.1.1). A final social identification outcome at study's end, ***SI**$_{final}$*, was achieved in three-folds: (1) FISI scores were binarised based on a median-split, (2) labels (FISI #1 to #5) were compared with the interview responses; specifically, FISI #1-3 with Interview #1, and FISI #4 and #5 with Interview #2, and (3) I manually assigned a concluding social identification label, annotated by ***SI**$_{final}$*, based on participant's interview response in session #2.

I utilised the Four-Item Social Identification (FISI) scale, which is anchored on a Likert scale from 1 (strongly disagree) to 7 (strongly agree) – a high score indicates a high identifier. Figure 4.5 illustrates the distribution of FISI scores by *Study_SE* students throughout the study (N=62, mean=5.21, SD=0.99, median=5.5, min=1, max=7). To simplify the interpretation of social identification, I performed a median split to transform a continuous variable into a categorical variable – a score below median (5.5) places an individual as having **0: low social identification**, otherwise **1: high social identification**. Unfortunately, categorising continuous data weakens the observed relations between variables, with studies showing reduction in correlations between variables by 20.2% [165]. Humphreys *et al.* argued against using artificial categorised data to perform analysis of variance (ANOVA) to test influences of an outcome variable [166]. Despite numerous rebukes, Farrington proposed this method as one way of handling data with highly skewed distribution while understanding a variable is not linearly related to an outcome [167]. This study did not address the temporal dynamics of social identification and its relationship to stress – instead, it sought to easily differentiate workgroup-specific behaviours and interaction patterns of individuals from distinguishing two groups of students with different identification.

Coding Interview Response: To ensure stability, accuracy and reproducibility, the same two coders were maintained for all interviews, and both coders used a standard coding scheme to reflect key categorisations such as the main source

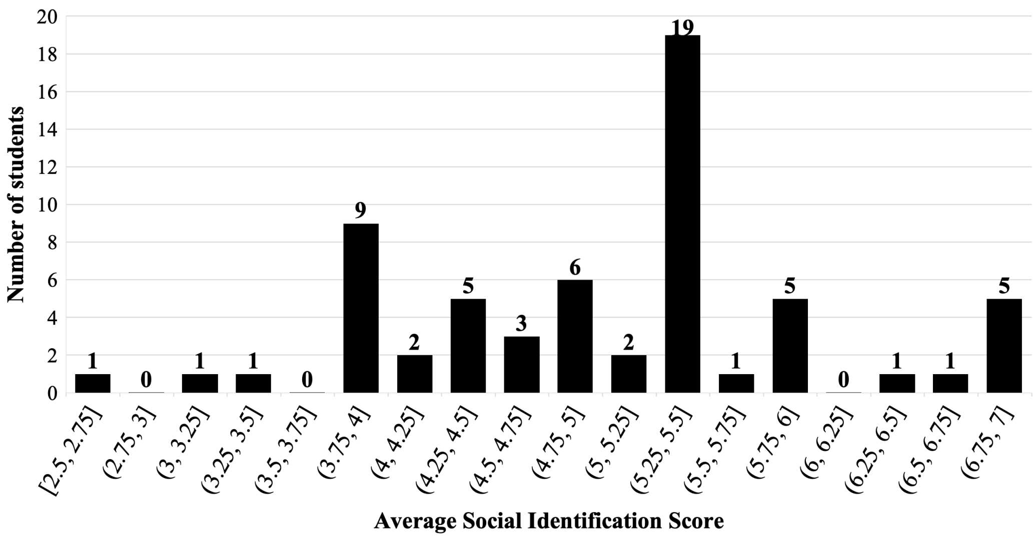

Figure 4.5: Social identification: Distribution of scores for *Study_SE* students on their Software Engineering workgroup throughout the study.

of stress, types of and reasons for critical events, changes in team involvement, types of emotional reaction from critical team events, and main strategy for group communication. Rousseau *et al.* concluded that showing care and support for team members [168] through means of creating opportunities to learn, providing tangible assistance and communicating with more positive emotional engagement, all help reinforce a stronger sense of identification in an individual. Hence, responses that checked negatively on critical events, less engagement with the team, negative emotional reaction and declining communication would all indicate students did not identify with their team (**0: low social identification**), otherwise, a **1: high social identification**.

Label Data Distribution: Grouping participants based on their $\boldsymbol{SI\_{final}}$ resulted in 36 students with **high social identification** and 26 students as **low social identification**. It is important to note that contradicting labels were determined for at least 7 students who consecutively reported FISI labels not matching their interviews; for example, FISI #1-3=[0-0-0] but Interview #1=[1] and FISI #4-5=[0-0] but Interview #2=[1]. Unfortunately, through this method, I was not able to precisely identify untrue scores. Hence, these samples were retained.

4.3.3.2 Evaluation of Model – $\boldsymbol{Model_{S\_SE+SI}}$

This experiment aimed to investigate if it is possible to detect patterns that characterise an individual as a high or low identifier through inferring their workgroup behaviour over a period of time. Here, I defined model performance by three measures – accuracy, precision, recall and AUC score (i.e., value of 1 indicates perfect classification). I maintained a group 5-fold CV, splitting participants at 80-20 (%) for training and testing; that is, 50 participants exclusively for training and 12 participants for testing.

Input and Output: Input dataset to build and evaluate the model is made up of features extracted from location and group data collected from the WiFi infrastructure. Over an 81 day study period, features were aggregated at critical time points corresponding to Table 3.3 (Day 3, 15, 36, 57 and 75). Each tuple consists of [$x_{a\_t1...tn}$,$x_{b\_t1...tn}$, $x_{c\_t1...tn}$, y], where x_a to x_c represent different mobility features at time point 1 to 5, and y is an individual's social identification label, **0: low social identification** or **1: high social identification**, as described in Section 4.3.3.1.

Features: Figure 4.6 summarises the results for detecting social identification at study's end. I employed a wrapper-based feature selection method which trains a RF classifier, resulting in a small subset of important *Group (G)* and *Workgroup (WG)* typed ***raw*** features (all important features selected were based on setting a threshold of more than 50% using ROC curve analysis). To avoid developing an overfitted or computationally expensive model, I limited the use of features to only (G)+(WG) sets with feature threshold above 80%.

Description	**Type (Set)**	**varImp**
Total time spent with *all groups*	General	100.0
Number of times being in *solo group*	(*Group-G*)	86.28
Number of times being in *small group*		81.81
Number of times being in *medium groups*		81.69
Number of times engaged in *SE*	Domain-specific	80.68
Number of times did not engaged in *SE on campus*	(*Workgroup-WG*)	63.06
Number of times engaged with *SE* members		57.47
Number of times engaged in *SE* *knowledge sharing sessions*		54.06
Number of times engaged in *SE meetings*		52.73

Table 4.10: Top 9 features based on a threshold > 50% using ROC curve analysis, and sorted by variable importance (varImp).

HMM-based Classification: Detecting individuals in two separate groups is a binary classification problem, which is typically solved using algorithms such as RF or SVM. However, these algoritms do not consider the temporal sequencing between observations, which can be achieved from using HMM. I utilised the HMMWeka package [169] for this implementation. The objective is to determine from a set of observations over a period of 81 days whether an individual is in either one of two states – **high social identification** or **low social identification** – at every two weeks. Finally, I classify an individual as a high or low identifier based on the majority of the sequenced outcomes. Given the model, λ, and (G)+(WG) features as sequence of observations VT, an HMM consists of an initial state probability, t, a matrix T of transition probabilities between states, and emission distribution $e(s)$ – the maximum likelihood training for these parameters were calculated by counting the occurrence of the observations and the hidden states (ground truth labels of artificially categorised FISI assessment, see Section 4.3.3.1). Then, given a new sequence of observations, o, the forward algorithm is used to find the posterior probability of observing that sequence of states given the model, P(o|λ). Finally, I determine the final outcome of an individual's identification by employing a majority rule from the series of outcomes, and measure its accuracy against the ground truth, $\boldsymbol{SI}_{\_final}$.

As charted in Figure 4.6, the average accuracy of the classification is 75.90% (79% precision, 76% recall, 0.83 AUC). That is, 15 out of 62 participants' social identification were misclassified at study's end. This experiment also compared the performance of classification with features treated independently using SVM and RF as common supervised learning methods (see Chapter 2.5). The HMM-based classification maintained a significantly better performance (at p=0.01 level).

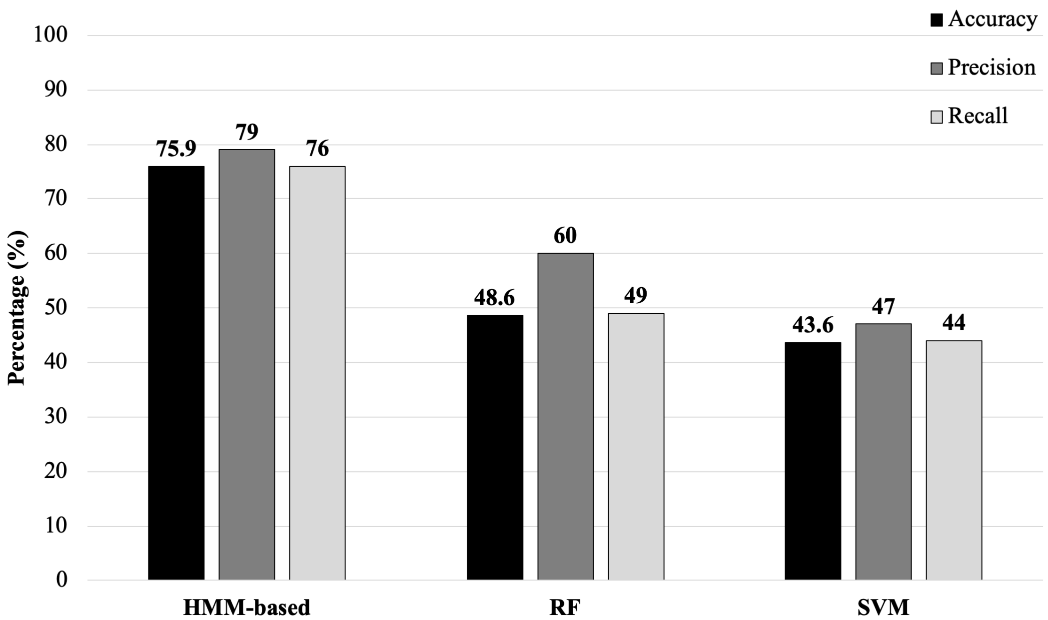

Figure 4.6: Results from using Group and Workgroup set features on three different algorithms; HMM-based classification which takes into account the temporal sequence of features, and Support Vector Machine (SVM) and Random Forest (RF).

	Model$_{S_SE}$ **All-typed features**	**incl. manually-assigned social identification label**	**incl. mobility-driven social identification outcome**
AUC	0.96	0.96	0.96
TPR (%)	98.93	90.04	88.08
TNR (%)	75.54	92.43	93.09
ACC (%)	77.59	91.61(*)	92.05 (*)

Table 4.11: Summarised results compared ***Model$_{S_SE}$*** with mobility-drive predicted outcomes, ***Model$_{S_SE+SI}$*** and manually-assigned labels for social identification. (*) indicates significance at p=0.01 level.

Another concern is deriving poor quality labels; since the final label relied on interview coding, which may not accurately represent an individual's identification. For instance, I found at least 7 participants who provided contrasting interview responses from their FISI assessments (see Section 4.3.3.1). The removal of these dubious instances from the training set improved precision by 3%, correctly classifying two more participants (79.10% accuracy, 82% precision, 79% recall and 0.85 AUC score).

Including SI-outcome as a Feature: Table 4.11 summarises the difference between using a mobility-driven social identification feature and manually-assigned social identification label (described in Section 4.3.3.1). I continued the optimisation experiment by adding the predicted outcomes from the social identification model as an additional feature to the stress model, noted as ***Model$_{S_SE+SI}$*** . Interestingly, the use of social identification outcome would lead to improvements in detecting true negative cases, while maintaining AUC at 0.96 (difference is not significant). However, TPR reduced by 8%. Despite the inaccuracies of misclassifying 13 participants using a social identification model, the use of a mobility-driven feature helps maintain true negative cases with a significantly higher accuracy above 90% (at p=0.01 level). While true positive rate is decreased by approximately 10% for ***Model$_{S_SE+SI}$*** compared to ***Model$_{S_SE}$***, the AUC score is retained at 0.96.

4.3.3.3 Workgroup Features Analysis

To better understand how mobility patterns relate to social identification in workgroup (SE) practices, I analysed the *domain-specific Workgroup (WG)* ***raw*** features, as described in Table 2.2. Specifically, I aimed to investigate if an individual's social identification [170] could be explained by observing mobility variations. For instance, Bos *et al.* argue that physical proximity can lead to emergence of groups [171]. As part of the SE course, students are graded for their project management practices (e.g., students must engage in pair programming, project managerial role and alternate programming partners). For this reason, I began with a conjecture that:

> ***Students with high social identification demonstrate more physical proximity to their workgroup-related activities than students with low social identification.***

Methodology: For this analysis, students were grouped based on their ***SI***$_{\textit{final}}$: **high social identification** (n=36) and **low social identification** (n=26) and their mobility patterns on SE-related activities, totalled for each given period, were compared at different assessment time points (corresponding to Table 3.3). I conducted a Mann-Whitney U non-parametric test and used the median differences between groups as they had similar distribution patterns. Note that this analysis included all (WG) features; however, only presents findings that indicated trends towards significance.

Results: Figure 4.7 illustrates the different trends of mobility patterns observed for each group on several SE-related group tasks, most likely to take place on campus. Overall, the maximum time spent on project-related tasks was lower for students with low social identification than for high identifiers. On assessment period 3 (Day 39 to 43), where contributions among members were expected to be the most (to complete development tasks during recess week and meet UAT milestone, see Table 3.3), low identifiers spent approximately 1.5 hours less on meetings ($p=0.11$), 5 hours less on pair-programming (p=0.11) and 6 hours less on knowledge sharing session (p=0.05) than others. 14 out of 17 low identifiers who reported primarily being burdened by SE throughout the semester, expressed in the interview they received very little team support from team members for various reasons. For example, one participant was frustrated for not receiving the guidance needed for technical tasks. One student admitted that he chose to exclude himself from the team after failing to deliver components for UAT, one of the key milestones that is graded in the course.

Students with low social identification were observed to start the semester's project with fewer off-campus and/or discrepant tasks (SEunique_count) between assessment periods 1 and 2. This trend incrementally increased over the semester. Despite observing close to significant differences for other SE activities (p=0.11), the contrasting behaviours supported my qualitative findings from the interview sessions that low identifiers spent less time working together with their group members. For example, one student said that members were often busy (with other project commitment) and could not dedicate themselves to the project as much as he did. These students led the technical tasks and quite often juggled project management responsibilities. Some students believed their technical incompetencies held the team back, but surprisingly maintained high identification from the support of other members – in these case, students who found themselves less competent were more likely to be in conflict with the most (technically) competent in the group. For example, one participant reported that she was blamed for making errors in the code. Later, it became much easier for her not to participate in decision-making processes, as how other members remained passive. Interestingly, there were more cases of low identifiers who were technically inclined expressing reluctance to work with the same group again. Unfortunately, students such as these were the most salient and emotionally charged with low social identification. Note that when asked to describe positive critical events such as a celebratory meal for a team member's birthday or milestone completion, there was no mention such events had been shared

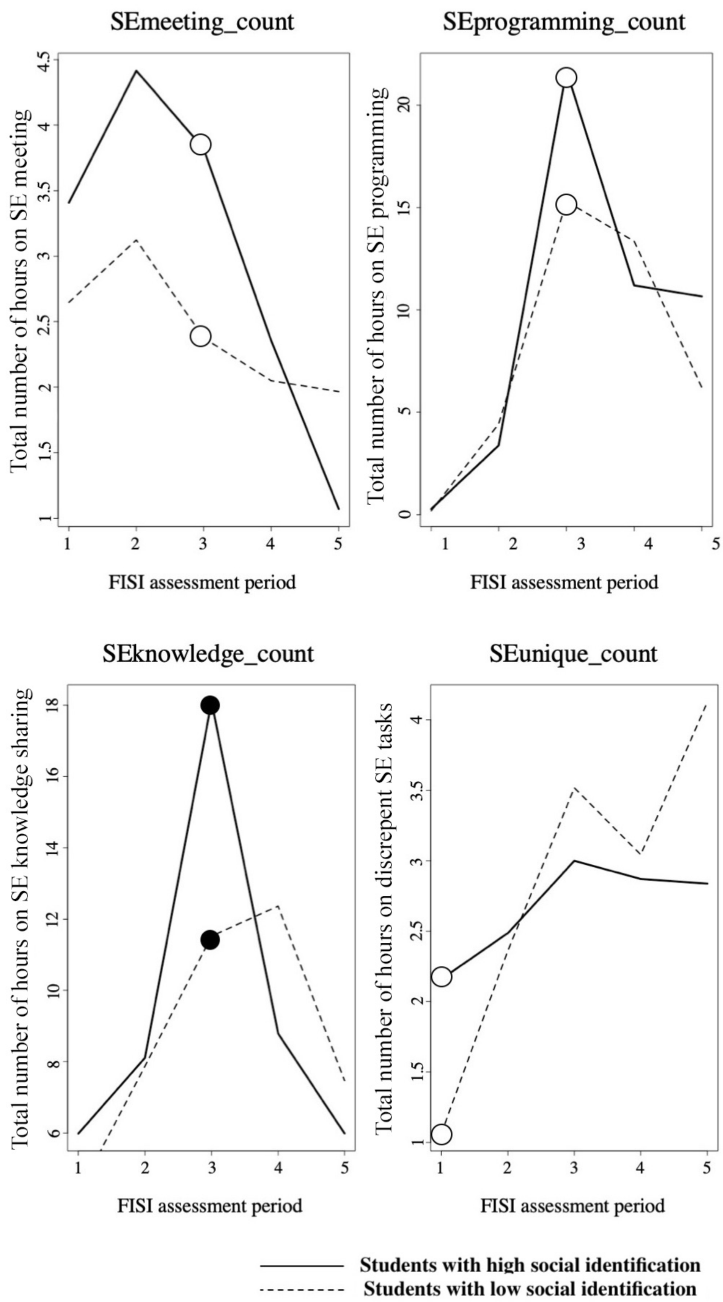

Figure 4.7: Graphs chart median differences for SEmeeting_count, SEknowledge_count, and SEunique_count, and SEprogramming_count that correspond with assessment periods. Shaded circle on graph indicates significant differences (p=0.05), and unshaded circle indicates close to approaching significance (p=0.11).

within the team.

Conclusively, these experiments demonstrate the potential of separating different behavioural modes by observing an individual's identification. Nonetheless, it must be pointed out that the findings related to social identification require in-depth analysis, particularly in its relation to stress. In Chapter 5, I report several comple-

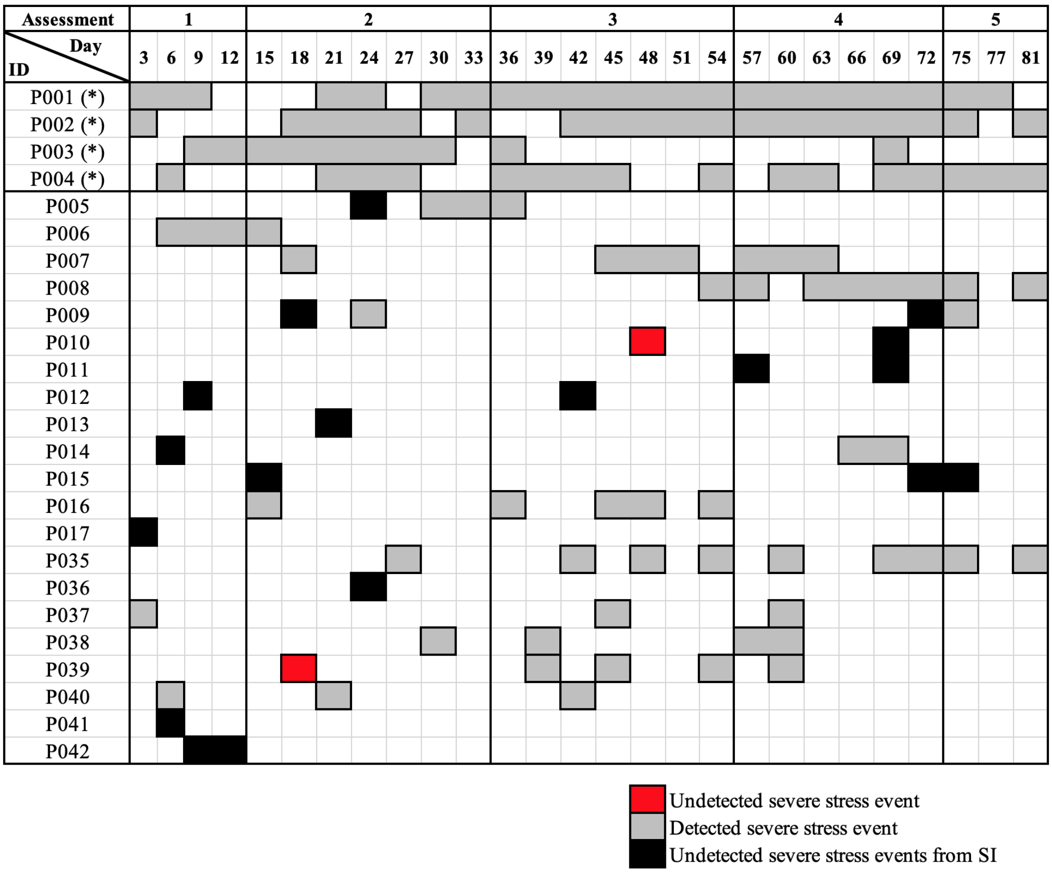

Figure 4.8: Detection results of 28 participants from *Study_SE*; Note that four *Study_SE* students who scored an average of 12 in their PSS-4 assessment throughout the study are indicated by (*). They were the most critical cases of *severe stress* .

mentary insights for understanding how social identification affects stress, and vice versa.

4.3.4 Summary of Stress Detection

This chapter presented the results of detecting stress in three groups of students, as summarised in Table 4.3. Overall, the evaluation demonstrated the strength of *StressMon* in two different ways. First, this approach did not require training a new stress model to detect *severe stress* in different groups of students. For example, even when using a model trained solely from students enrolled in Software Engineering (*Study_SE*), the ***Model***$_{S}$ still achieved a high 0.91 and 0.94 AUC score when used on two different populations, *Study_Valid1* and *Study_Valid2*, respectively. Second, the removal of *Domain-specific* features (i.e., features related to

Software Engineering project) increased the TNR for our all-population model. I demonstrated how *StressMon* was able to achieve and prioritise high TPR, which improves the likelihood of *StressMon* detecting the small number of *severe stress* cases even at the cost of misclassifying more cases of no stress as *normal stress*. The tradeoff which prioritises detecting the more important cases correctly is appropriate for an early warning solution. Third, improving the detection of true negative cases can be achieved from using *social identification* as additional feature. Unlike the baseline stress model, ***Model$_S$***, which uses strictly change features (***rel+abs***), the social identification model is built from utilising ***raw*** mobility features to generate a predicted outcome at study's end. That is, the model did not account for changes in behavioural patterns in individuals for predicting social identification. The downsides of using social identification predicted outcome are (1) predicting such outcome, which require the use of *Domain-specific* features generated from workgroup activities, and (2) the feature is generated after approximately 2.5 months of behavioural analysis; ***Model$_{S_SE+SI}$*** retrospectively improves detection of true negative cases in severely stressed students. As shown in Figure 4.8, using social identification outcome as an additional feature resulted in the misclassification of *severe stress* for 12 students. Overall, however, the reports of *severe stress* for these students were sparse and not what should be considered a critical case for *StressMon* to flag.

4.4 Evaluation of Depression Model

In this section, I sought to investigate if the features (and setting) used to build the stress model, ***Model$_S$*** could accurately detect *depressed* users. Using a Random Forest (RF) algorithm, *General* change set features (rel+ abs) calculated at a 6-days interval, the model's predictions were at random chance (TPR: 59%, TNR: 50%) in detecting individuals showing significant signs of depression; the results were unsurprising. While stress and depression often bear similarities in behavioural

			Feature Settings			Performance	
Sec.	Model	Study	Type (Set)	*raw/abs/rel*	Interval	AUC	TPR (%)
4.4.1	$\boldsymbol{Model_D}$	*Study_SE*	*General (All)*	***rel+abs***	15-days	0.72	70.25
4.4.2	$\boldsymbol{Model_D}$	All	*General (All)*	***rel+abs***	15-days	0.78	77.96
4.4.3	$\boldsymbol{Model_{D+}}$	All	*General (All)* + *N*euroticism	***rel+abs***	15-days	0.88	91.21

Table 4.12: Summary of depression model configurations achieving best performances. Sections 4.4.1, 4.4.2 and 4.4.3 provide detailed results for each experiment. $\boldsymbol{Model_{D+}}$ is an optimised depression model that uses personality (neuroticism score) as addition feature, while $\boldsymbol{Model_D}$ operates exclusively on mobility-driven features.

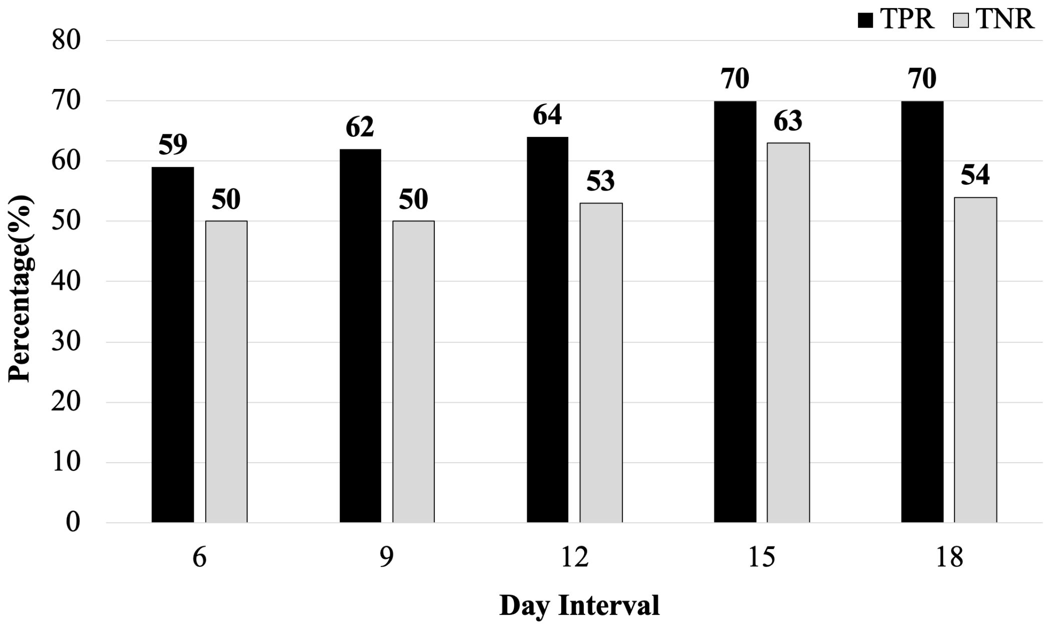

Figure 4.9: Results from calculating Change type (rel+abs) features on different time intervals; from 6 to 18-days.

symptoms, depression evolves over a much longer period of time. For example, typical assessment of depression using validated scale is measured at 2 weeks [98].

4.4.1 Cross-Validation Experiment: *Study_SE* – $\boldsymbol{Model_D}$

Time Window Experiment: Building on prior work, I repeated the time window experiment by calculating the change in all *General* features between 3 days up to 30 days. Through this experiment, I was able to empirically determine the optimal window size for calculating changes in behaviours as 15-days to best detect depression (differences in time intervals are not significant). Figure 4.9 charts the performances (TPR and TNR) from day 6 to day 18. Note that performances from day 18 onwards continued to decline on both metrics.

Individual and Group Interaction Features: Depression detection using RSSI values has been previously proposed by Ware *et al.* [96], but the authors did not consider group-related features. Note that their experiments ran on two phases, resulting in an average TPR of 77.00% and TNR of 62.50% (Phase 1, TPR:80% and TNR: 62%. Phase 2, TPR: 73% and TNR: 63% for day-time monitoring). Accordingly, I investigated the difference in model performance from using group-related features and (individual) location features. To achieve this, I built the depression model using different sets of change features, calculated at a 15-days interval. As summarised in Table 4.13, I was able to quickly determine that by removing group features, the model was correctly classifying the positive cases at random chance. In comparison, group-related features achieved 65.24% TPR at the expense of correctly classifying non-depressed students. However, these features alone did not yield high performance. Table 4.14 lists the top 10 most important features being used.

Conclusively, these results shed light on two small but interesting differences in the detection of depression versus stress. First, changes in behavioural features must be calculated at a longer time interval of 15-days, compared to stress at 6-days. Second, the group-related features rose as the strong predictors for depression, compared to individual (routine) features for stress. I concluded the best model settings for detecting depression using Random Forest (RF) algorithm, *General* change set

	General Change ***rel+abs***	Individual (location data)	Group (group data)
AUC	0.72	0.63	0.67
TPR (%)	70.25	51.83	65.24
TNR (%)	63.53	64.72	54.88
ACC (%)	65.58	63.32	56.91

Table 4.13: Results separating the changes in group-related features from individual (routine) features; Change (***rel+abs***), Individual (change features extracted from location data) and Group (change features extracted from group data) to detect depression at 15-days interval.

Description	Type	varImp
Total time spent with *all groups*	***abs*** (*Group*)	100.0
Number of times being in *solo group*	***abs*** (*Group*)	96.66
Number of times being in *all groups*	***abs*** (*Group*)	91.31
Number of times engaged in *studying*	***rel*** (*Work*)	85.66
Number of times engaged in *transiting*	***abs*** (*Non-work*)	82.51
Number of times engaged in *attending lectures*	***abs*** (*Work*)	66.44
Number of times being in *small group*	***abs*** (*Group*)	65.64
Number of unique building visits	***rel*** (*Non-work*)	64.88
Number of times engaged in *studying*	***abs*** (*Work*)	64.75
Number of times engaged in *all non-work activities*	***rel*** (*Non-work*)	64.28

Table 4.14: Top 10 features for detecting *depression*, using ROC curve analysis, and sorted by variable importance (varImp)

Method	Train	Test	AUC	TPR (%)	TNR (%)	ACC (%)
Train-Test	*Study_SE*	*Study_Valid1*	0.70	71.43	62.50	64.39
	Study_SE	*Study_Valid2*	0.69	63.03	66.02	64.97
Group5-fold (All population)	Folds 2-5	Fold 1	0.76	71.98	69.75	70.57
	Folds 1,3-5	Fold 2	0.78	88.99	53.41	60.47
	Folds 1-2,4,5	Fold 3	0.81	82.14	59.91	65.43
	Folds 1-3,5	Fold 4	0.80	82.21	65.65	72.07
	Folds 1-4	Fold 5	0.73	64.47	67.31	66.11
		Average	0.78	77.96	63.21	66.93

Table 4.15: Summarised results for depression model, $\boldsymbol{Model_D}$, on three different validations. $\boldsymbol{Model_D}$ is a generalised depression model that excludes all *Domain-specific* features.

features (***rel***+ ***abs***) calculated at a 15-days interval as $\boldsymbol{Model_D}$.

4.4.2 Validation Experiment: *Study_Valid1 & 2 – Model$_D$*

I replicated the steps in Section 4.3.2 to validate $\boldsymbol{Model_D}$ on different population sets. Additionally, I combined all students from the three user studies to build an all-population depression model, evaluated using a Group 5-fold CV. Table 4.15 summarises these results in detail.

Overall, my approach maintained an accuracy of approximately 65% for all studies. Similarly, I combined all students from three user studies to build an all-population depression model, evaluated using a Group 5-fold CV. It is worth noting that the performance of $\boldsymbol{Model_D}$ is comparable to that in [96], achieving 77.96% TPR and 63.21% TNR. As these numbers were achieved in two different environments with different population groups, it is possible that accuracy numbers would change significantly in different environments. That being said, using only individual features (as baseline comparison in this environment, see Table 4.13), ***Model$_D$*** yielded between 18-26% improvement in TPR.

4.4.3 Optimising Depression Model

Considering depression is defined by stricter ethical boundaries, my user studies did not include subjective assessments of users' accounts of depression; doing so might best be moderated by health professionals. Alternatively, I manually analysed the demographic information of students such as their gender, academic year, GPA and Big-5 personality assessment [135]. The analysis of *Study_SE* revealed only 1 case of depression by a student who scored low on neuroticism (score <= 2.25 out of 5). The most significant portion of depression reports was by students whose neuroticism scores were 3.75 and 4. Indeed, many studies draw correlations between high neuroticism scores and depression [73,74]. Thus, I revised the model to include personality traits (i.e., Openness, Conscientiousness, Extroversion, Agreeableness and Neuroticism). Additionally, I experimented with the social identification outcome previously predicted (in Section 4.3.3.2) for stress as a possible feature.

4.4.3.1 Cross-Validation Experiment: *Study_SE* – *Model$_{D+}$*

Table 4.10 charts the results of detecting depression with different types of optimisation features. The inclusion of social identification outcome as a feature did not improve overall accuracy. In contrast, the addition of all personality traits as features improved overall accuracy by approximately 10% (difference is significant at p=0.05 level). A brief examination on each dimension revealed 'Neuroticism' as the strongest predictor (achieving 100% variable importance, while the next best-ranking trait, 'Openness', yielded a 10% importance). The exclusion of all other traits resulted in a higher TPR of 90.21%, however, this difference is not significant from using all Big-5 dimensions.

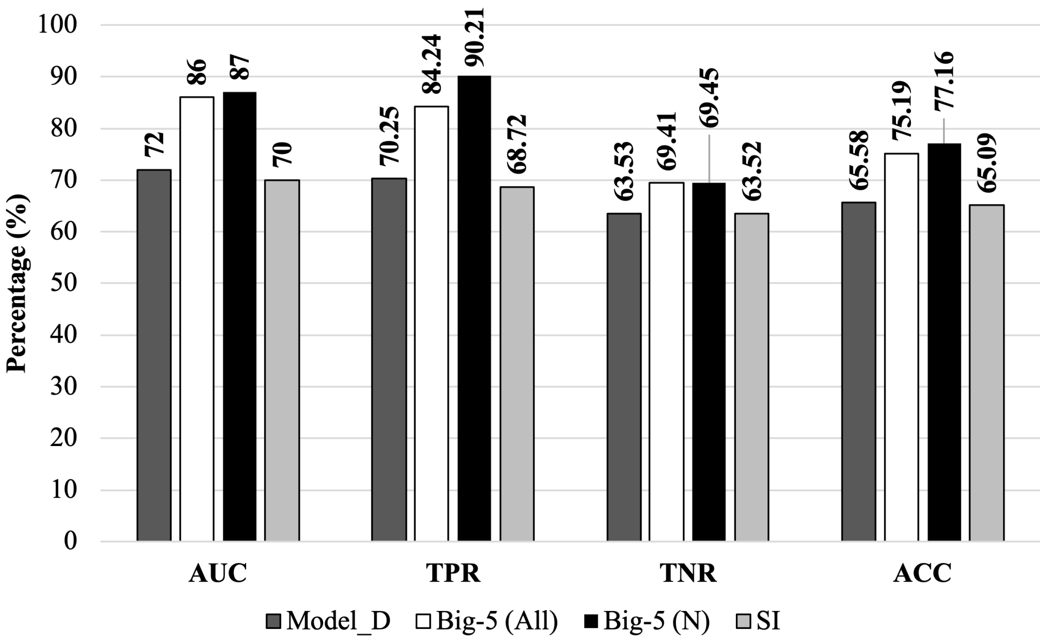

Figure 4.10: Results from adding different optimisation features; SI is the social identification outcome, Big-5 (All) considers all 5 personality traits and Big-5 (N) particularly uses Neuroticism score as feature.

Method	**Train**	**Test**	**AUC**	**TPR** (%)	**TNR** (%)	**ACC** (%)
Train-Test	*Study_SE*	*Study_Valid1*	0.41	57.14	47.12	49.24
	Study_SE	*Study_Valid2*	0.86	81.52	72.68	75.77
Group5-fold (All population)	Folds 2-5	Fold 1	0.86	91.3	47.62	63.65
	Folds 1,3-5	Fold 2	0.93	92.66	76.82	79.96
	Folds 1-2,4,5	Fold 3	0.93	95.71	76.65	81.38
	Folds 1-3,5	Fold 4	0.88	90.87	68.69	77.28
	Folds 1-4	Fold 5	0.84	85.53	63.75	73.00
		Average	0.88	91.21	66.71	75.06

Table 4.16: Summarised results for depression model with Neuroticism as optimisation feature, $\boldsymbol{Model_{D+}}$, on three different validations.

At this point, it is important to consider the practicality of using Big-5 assessment. That is, while the Big-5 personality assessment is often deemed too lengthy (assessment comprises 44 questions), using only a subset of this information could increase model accuracy (Neuroticism has 8 questions). Since using only one personality dimension reduces the user burdens of sampling, I concluded $\boldsymbol{Model_{D+}}$ as including Big-5 ‘Neuroticism’ score as an optimisation feature.

4.4.3.2 Additional Validation: *Study_Valid1* & *2* – $\boldsymbol{Model_{D+}}$

Validation results of $\boldsymbol{Model_{D+}}$ (trained on *Study_SE* and tested on *Study_Valid1* and *Study_Valid2*) are summarised in Table 4.16. While overall accuracy of 49.24% and TPR of 57.14% seemed significantly lower for *Study_Valid1* population, the model was successful in detecting two out of three participants who reported depression. Finally, I built an all-population depression model and evaluated it using a Group 5-fold CV. This model achieved an average AUC score of 0.88, maintaining more than 85% TPR across all folds; a total of 9 out of 55 students who reported depression had several instances of depression misclassified. Unfortunately for 1 student was completely missed by the model.

The Big-5 assessment is a widely accepted instrument of measuring the most common aspects of a person's personality. It has since been rigorously validated across cultures [172] and encouraged to facilitate open science practices [173]. Unlike the social identification feature (utilised as a retrospective feature for stress), this scale could potentially be sampled as demographic variable to improve the performance of *StressMon* in detecting depression.

4.4.4 Summary of Depression Detection

The evaluation of $\boldsymbol{Model_{D}}$, using a Random Forest algorithm and *General* change (***rel***+***abs***), demonstrated the possibility for *StressMon* to detect individuals showing signs of depression. However, two small but critical factors to different the models from detecting depression from stress are (1) the time interval used to calculate changes in features – from 3-days (stress) to 15-days (depression) – typically from following symptoms over a two-week period [98], and (2) the inclusion of group-related features (extracted from group data) as they make much stronger predictors than individual features. I have demonstrated that *StressMon* achieved results that are comparable to [96]. However, using features from our location data, *StressMon* only yielded 51.83% TPR (it is possible that accuracy numbers would change significantly in different environments). An error analysis on *Study_SE* students for ***Model$_D$*** presented a clear trend of students who frequently reported depression scoring high on Neuroticism as one of their personality traits; that is, an individual is more likely to experience negative emotions than those who are emotionally stable [135]. Results from my experiments support prior work which establishes the strong association between this personality trait and depression. Building on this finding, I included students' Big-5 personality score by particularly using their scoring for Neuroticism attribute, as additional feature to the model ***Model$_{D+}$***. The addition of this feature resulted in a higher AUC=0.88 (91.21% TPR and 66.71% TNR) on all studies. Unfortunately, however, this information would require individuals to participate in a one-time demographic survey as the information cannot be generated through analysing mobility features. As shown in Figure 4.11, the addition of Neuroticism as an optimisation feature led to an accurate model for depression for most students. Unfortunately, 2 students (P103 and P217) were left completely unnoticed by *StressMon* .

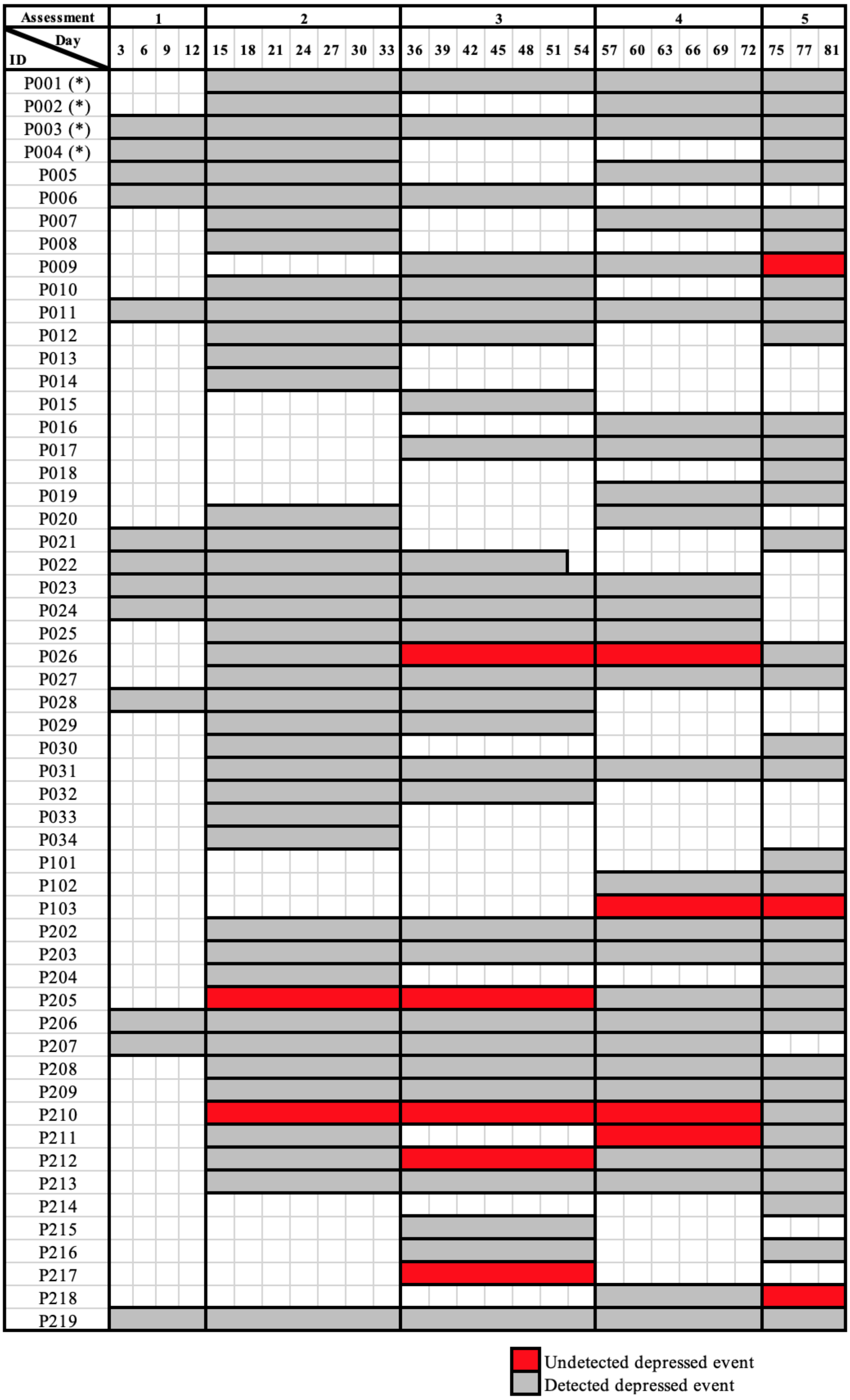

Figure 4.11: Detection results of 55 participants from all studies; Note that four *Study_SE* students who scored an average of 10 or more in their PHQ-8 assessment throughout the study are indicated by (*). They were the most critical cases of *depressed* and *severe stress* at the same time.

CHAPTER 5

UNDERSTANDING STRESS & DEPRESSION

This chapter addresses the final research question: *What can be learned from the findings of stress and depression?* This chapter begins by identifying main causes of stress in students. In addition, I discuss how individual behaviours in workgroup factors such as roles in leadership, team support and misconducts can be applied to several existing theoretical perspectives, which are key in explaining why some students are more stressed than others from working in teams. Finally, I discuss how the differences in stress and depression influenced our decisions in developing *StressMon*.

5.1 Main Stressors

58 out of 62 *Study_SE* students identified *academic stress* as the primary cause of stress (note: these students might not necessarily be overwhelmed by the stress). Approximately 50% (33) of the population recognised the Software Engineering group project as their primary stressor, while three others agreed it was due to another course they were taking at the same time. Four students attributed their primary stressor (due to personal business or poor health). Of the 33 students, 14 students attributed the "Software Engineering" stress to relationship tension with

their team members; for example, not receiving a reciprocal commitment from team members. In addition, 5 out of 11 *Study_Valid1* students reported their main stressor was for academic reasons; specifically, 2 students faced interpersonal relationship strains with their group members. 3 students were mainly stressed from their job search as they were in their last semester. 1 *Study_Valid1* student was primarily stressed from settling back into a competitive school environment after spending the previous semester on exchange studies.

With these findings, however, I was not able to understand the nuances and details of stresses students typically face. As explained in Chapter 3.3.1, I later mod-

Cat	**Cause**	**Score**
Academic	Increased class workload	4.21
	Group projects	4.21
	Many hours of group projects	4.04
	Examinations	3.96
	Many hours of studies	3.88
	Lower grade	3.77
Environmental	Lack of vacations/breaks	3.63
Personal	Change in sleeping habits	3.44
Interpersonal	Difficult personalities of (school-related) group members	3.40
	School-related misunderstanding	3.27
	Work with people (school-related group members) you don't know	3.15
Personal	Change in eating habits	3.15
Interpersonal	Change in relation with school mates	2.98
Personal	Financial difficulties	2.9
	Combining job and studies	2.88
Academic	Lack of university support	2.88
Personal	Personal relationship issues	2.85
	Health issues	2.75
	Family issues	2.69
Environmental	Computer problems	2.67
Academic	Language difficulties	2.58
Environmental	Unfamiliar educational environment	2.56
Academic	Missing seminars/classes	2.52
Environmental	Bad living conditions	2.44
	Quit jobs	2.23
	Moving to a new city	2.04

Table 5.1: Sources of stress among *Study_Valid2* students ranked from highest to lowest score. With the exception of "Environmental" and "Personal" reasons, the top ranked reasons can be categorised into *work content* and *work context*.

ified the user study procedure for *Study_Valid2* students to include two additional assessments: One is the *sources of stress*, adapted by Yumba *et al.* [144]. Out of a 5-point scale, students ranked 'increasing workload' (academic, 4.21), and 'having to work on group project' (academic, 4.21) as the top two stressors. With the exception of 'lack of vacations' and 'change in sleeping habits', students' major sources of stress were related to academic and interpersonal-related linked to working with their peers. These findings support prior work of students under considerable stress, especially from high amounts of workload (i.e., *work content*) and working with people (i.e., *work context*), as described by Cox *et al.* [12].

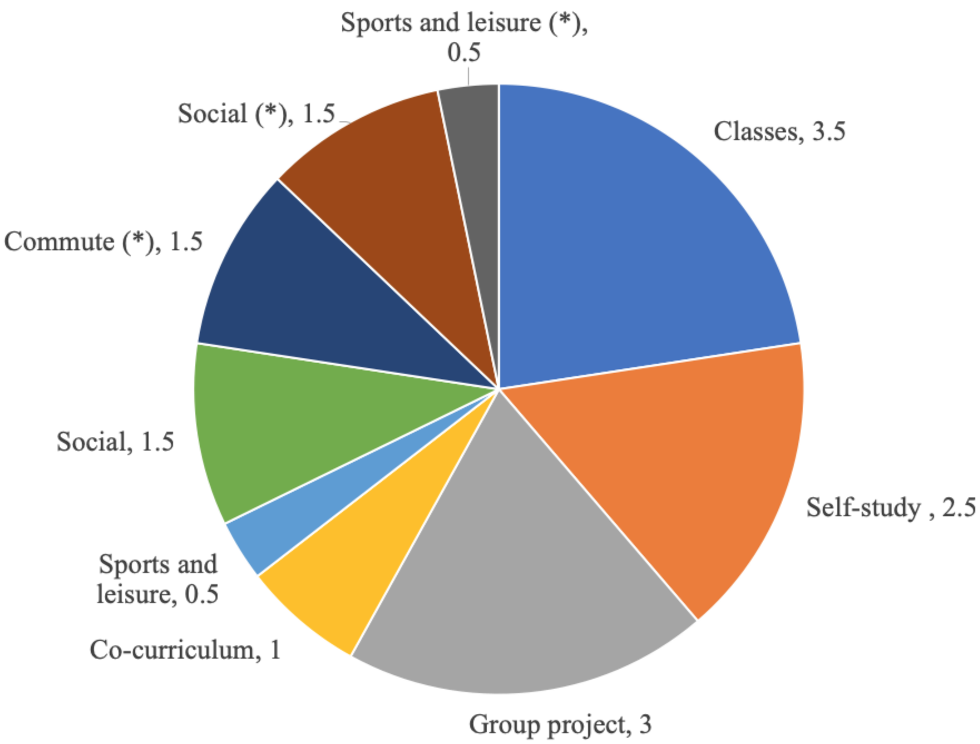

Figure 5.1: Diagram charts the average amount of time spent (in hours) on different common activities among *Study_Valid2* students, totalling to 12 hours on campus each day. Unless indicated with (*), all activities took place on campus.

Additionally, *Study_Valid2* students estimated the time spent on campus using a survey adapted from the American Time Use Survey [145]. Interestingly, 35 students reported an average of 12 hours each day on various activities on campus (see Figure 5.1); 3.5 hours spent on attending seminars, 3 working on group projects, 2.5 hours on self-study and the remaining hours spread across sports, leisure and social activities. Over a long 12 hour period on campus, it is no wonder that academic-typed stress has imprinted itself as a major stressor.

5.2 Stress and Social Identification

In practice, it is hard to distinctly associate social identification with stress. For example, Researchers have claimed that being saliently identified to a group increases an individual's sense of obligation to work harder and for longer hours, which may possibly lead to mental strains and burnout [15]. In contrast, an individual who is less identified from lack of support of coworkers, may experience less (work-related) stresses as a result of dedicating less time to work. Comparing the differences in social identification between students with *severe stress* and *normal stress*, as in 4.2, I determined that severely stressed students generally identified themselves more with the team (see Figure 5.2).

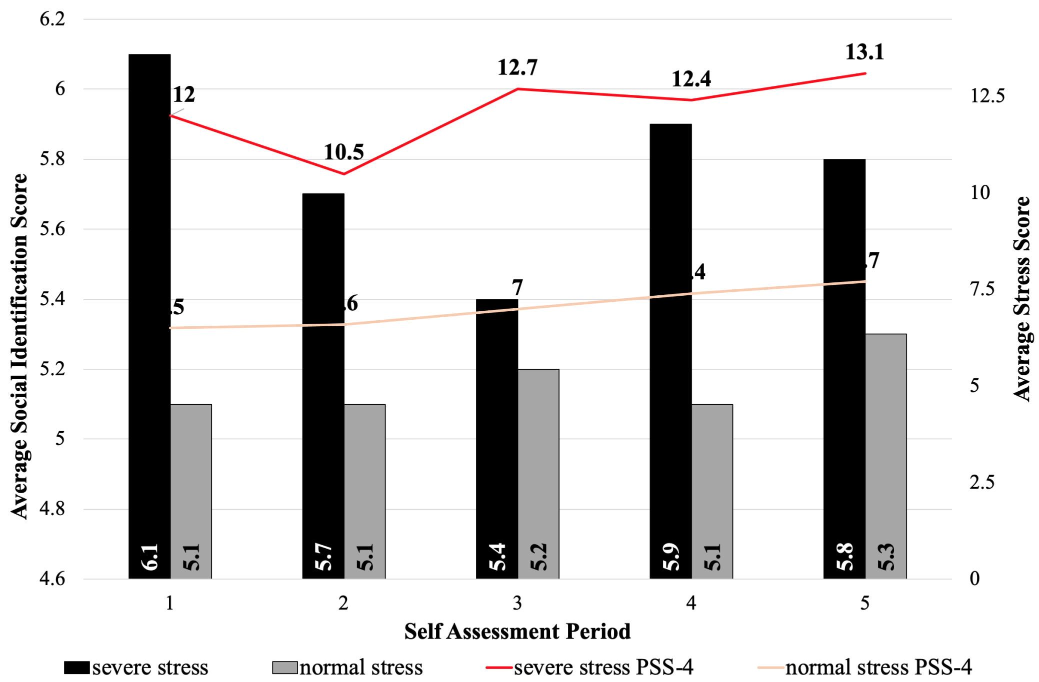

Figure 5.2: Social identification: Distribution of scores for *Study_SE* students with *severe stress* vs. *normal stress* throughout the study.

Overall, students with *severe stress* scored above the mean and displayed more variance in their scores. In contrast, students with *normal stress* were fairly consistent with their social identification throughout the semester. 3 (out of 4) of the severely stressed students indicated facing challenges in fulfilling project tasks due to their incompetencies in programming. Nonetheless, students generally maintained good working relationships with their team members, who actively encouraged and assisted them in various tasks. Interestingly, the average score for social identification was at its lowest (5.4), corresponding to the time students were typically engaged in development work (see Table 3.3). For 1 low identifier with *severe stress*, he expressed feeling overwhelmed from taking the lead in development work.

To better understand how social relationships between and among workgroup members negatively influence an individual's stress, I critically evaluated the interview findings by *Study_SE* students, as their workgroup procedures were intensive and extensively documented. Accordingly, I made several inferential connections between students' behavioural characteristics and theoretical perspectives related to social identification (i.e., social identity), as summarised by Seering *et al.* [170].

> ***"The identities we tend to embody are those that are the most accessible and have the best 'fit' within a given situation."***

Individuals are quick to associate themselves with an accessible social identity, which is most likely to be suited to their current goal. I quickly learned that students collectively categorised themselves as *A-coder*, *B-coder*, and *C-coder* (from the interviews). These categories are based on their abilities which manifested through a prerequisite course and inherently created social comparison between and among team members. The notion of *social comparison* based on (academic) abilities, unfortunately, is judgmental and punishing to those who were evaluated less favourably [174, 175]. I learned that 5 *C-coders* felt they were less valued for their opinions and being assigned less challenging types of work. One student reported being blamed for her task decisions. Another student expressed, "the *A-coder* [in his team] values someone of higher IQ than himself." To the *A-coder*'s defence, he said, "[I couldn't] trust the quality of work by my members throughout the whole semester." The implication of this behavioural existence is an uneven work distribution where an *A-coder* was more likely to take on lead responsibilities of the programming aspects, despite fair work distribution being a project management requirement (this may explain the discrepancies I found in their schedule logs). There were more instances where non *A-coders* expressed low identification, however, these were not among those who reported severe stress.

> ***"A person's identity can be defined at various distinct levels, with the most common differentiation being between one's personal and social identities."***

Fundamentally, an individual possesses a multitude of social identities for the groups they belong to. Similarly, in this study, I found cases where students described themselves as possessing multiple identities in different subgroups (within the workgroup). For example, one student (a high identifier) believed he mostly acted as the 'middle person' for the group. While not the most technically inclined (he also identified himself as a *B-coder*), he was well relied upon by the *A-coders* for fulfilling his technical duties and was able to communicate with the *C-coders* (as the *A-coders* were very explicit in showing their dislike of the C-coders). Nonetheless, he would identify himself more with the *A-coders* since they contributed more to the success of their project. Despite being randomly assigned into groups, one student was teamed up with a good friend, who, regrettably, was in many conflicts with the rest of the team. Adopting multiple identities as a friend to one and an *A-coder* to others, the student eventually felt tired of being part of the team and participated less in group meetings. She did, however, maintain regular online communication with the workgroup. Note: All teams managed a SE Telegram group chat as a standard online communication channel. Forming multiple identities helps students maintain positive relationships with different members in the group. In the case of three students with *severe stress*, they eventually maintained high social identification from working with everyone else but the *A-coder*. Two students described themselves as having purely functional relationships with members of their team. While they kept healthy and respectful working practices, they would treat each other like strangers outside of work. The most prominent of these cases was an inconsistent reporting of critical group events which students reportedly experienced at midweek and semester end. For instance, two students who broke down in their first interview, expressed reconciling with their group members over disagreements. The experiences of intergroup conflict (for some escalating to relationship tensions) may explain the occurrence of changing social identification assessment scores every few weeks.

> ***"In groups, the leaders who emerge are the members who are the most prototypical of the group's norms."***

Bicchieri *et al.* described 'prototypical' as a collection of physical, mental and psychological characteristics that an individual in the group is believed to embody [176]. Our findings support prior research of high identifying *A-coders* being liked for their prototypical properties [170, 177], and as a result, became more influential than others in the group. *A-coders* reportedly exercised more control over critical decisions such as directing coding standards, determining system functionalities and task assignments, and approving completed tasks. It came as no surprise that many *A-coders* felt saliently identified to their team. However, the implications to "follow the leader" can be dangerous and destructive. Many students admitted to passive participation, fully trusting the *A-coder* and following their instructions without hesitation. At times when *A-coders* reacted aggressively under overwhelming amounts of stress, their work incivilities were tolerated more than others in regard of their expertise, and often left the weaker coders disrespected. One student said, "I am the weakest in the group. It was quite shocking the way I was being treated [by the A-coder], and I do not dare to voice out my opinion." Another student clarified, "Scolding became [A-coder's] common practice on Telegram group chat. Text messages were sent in caps and [he used] coarse language." Despite having trouble dealing with their rude behaviour, group members chose to maintain a subservient relationship with the *A-coder*, suggesting: "As long as we stay on [our] path, the group work will be fine. So I stayed passive," when being asked if support was offered to help group members who were targeted.

5.3 Reports of Stress and Depression

Figure 5.3 charts the percentage of our students in each of three user studies reporting *severe stress* over a 3-days interval amounting to 27 samples.

7% of the *Study_SE* students reported *severe stress* at the beginning of the semester and peaked at 17% on Day 69 before the final project deliverable. Only one student (3%) from *Study_Valid2* reported *severe stress* on Day 45 when the semester resumed. I sampled *Study_Valid1* students from the second half of the semester (Day 45 onwards) and received the first reports of *severe stress* on Day 51 and towards semester end. In contrast, a higher percentage of students reported feeling depressed (see Figure 5.4). Further, the analysis revealed a concerning trend of 40 student participants who reported feeling *depressed* continuously for approximately four weeks (for all studies). Among these 40 were four *Study_SE* students, who simultaneously experienced frequent *severe stress* from SE (see Figure 5.5, participants are indicated with (*)).

Figure 5.6 charts the stress and depression episodes for *Study_Valid1* and *Study_Valid2* students in detail. In real-world operation, students who are concurrently *depressed* and *severely stressed* or frequently *depressed* but not severely stressed are those that *StressMon* detects as "red-flags" so that interventions can take place as early as possible. The higher percentage of students who reported *depression* as compared to *severe stress* raised a serious need for our technique to successfully detect *depressed* students separately from detecting students with *severe stress*. While models for stress ($\boldsymbol{Model_S}$) and depression ($\boldsymbol{Model_{D+}}$) were fundamentally built on similar sets of mobility features, their key differences lie in the time window for calculating changes in individual behaviour and group interaction.

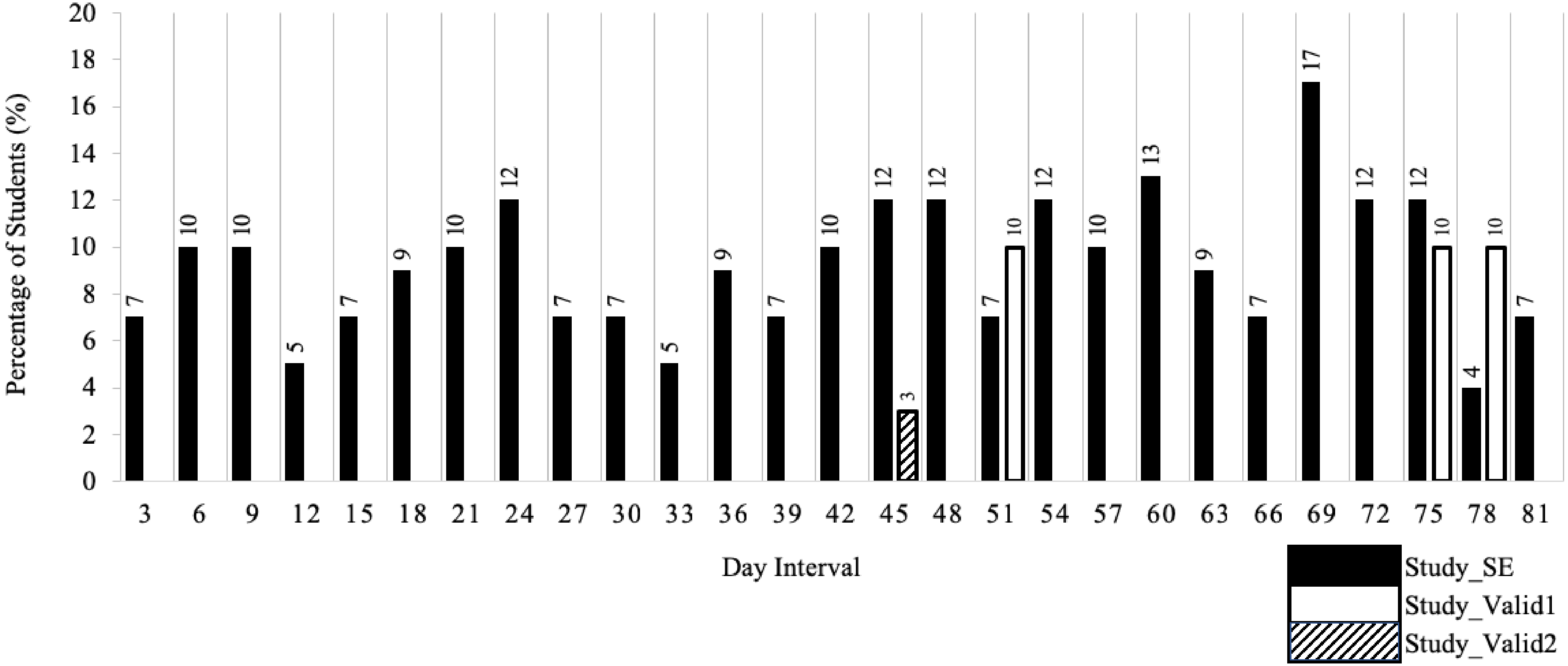

Figure 5.3: Histogram of percentage of students from different user studies reporting *severe stress* (PSS-4 score more than 12) every 3 days. Samples for *Study_Valid1* students were only collected from day 45 onwards.

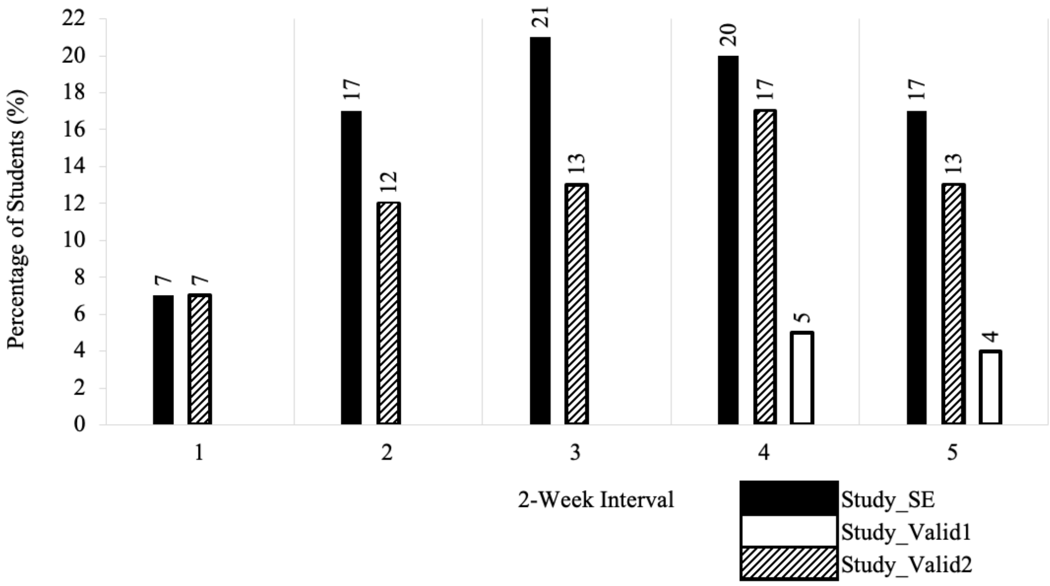

Figure 5.4: Histogram of percentage of students from different user studies reporting feeling *depressed* (PHQ-8 score more than 9) approximately every 2 weeks. Samples for *Study_Valid1* students were only collected from day 45 onwards, corresponding to sample 4 and 5.

5.4 Summary

In summary, these findings corroborate prior studies that college students mainly experience academic stress: A significant percentage of students, for example, 93% of *Study_SE* students, believed their primary stressor was related to various academic matters, many of whom confirmed working on the group project as one of the most stressful involvements. Further, this study documented evidences of students facing tremendous difficulty in working with unfamiliar people and personalities such that interpersonal aspects of workgroup stress became most familiar to them. As summarised by Seering *et al.* [170], individual behaviours can be explained by key social identity perspectives to understand the roles of leadership, team support, collaborations and misconducts between and among individuals working in a group having different social identification. It is, however, interesting to note a low percentage of *severe stress* was reported (26%, 28 out of 108 students). One of the most common remarks made during the interviews implied strong acceptance of stress being a normal reaction throughout the semester. On the other hand, a more

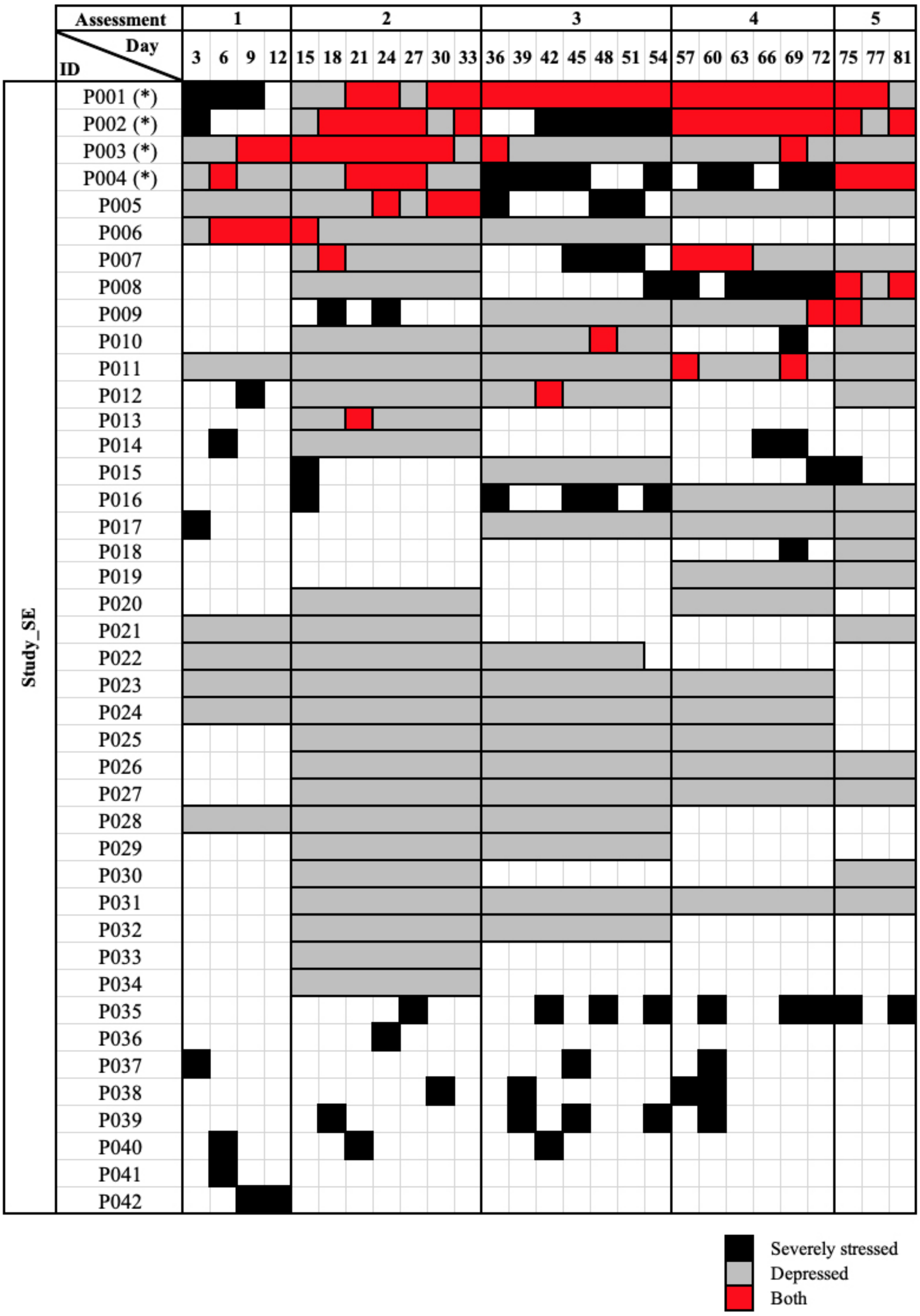

Figure 5.5: Reports of *severe stress* and/or *depression* of 42 students from *Study_SE* charted to illustrate two different patterns; (1) frequent instances of *severe stress* and *depression* (P001-P004), (2) occasional *severe stress* and *depression* (P005-P018), (3) only *depression* (P019-P034), and (4) only *severe stress* (P035-(042). The remaining 20 students did not report such events.

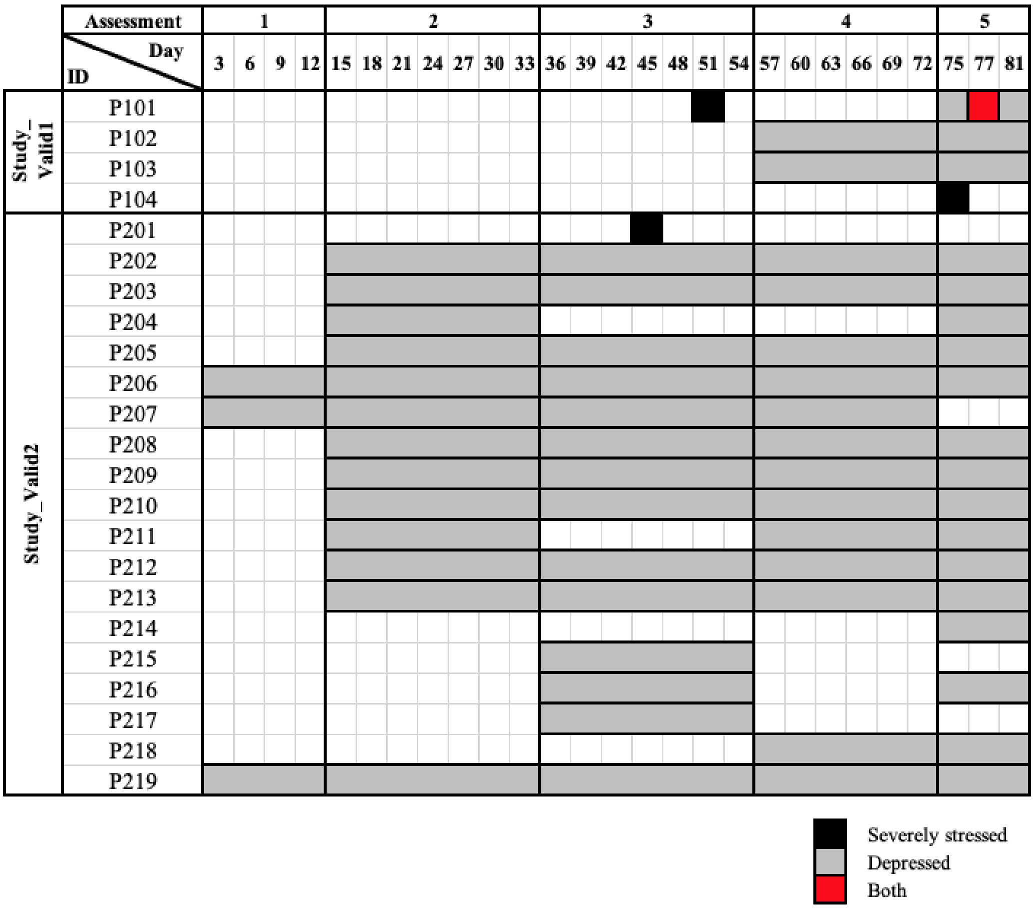

Figure 5.6: Reports of *severe stress* and/or *depression* of 4 students from *Study_Valid1* and 19 students from *Study_Valid2* charted to illustrate three different patterns; (1) occasional *severe stress* and *depression* (P101,P201), (2) only *depression* (P102,P103,P202-P219), and (3) only *severe stress* (P104). The remaining 24 students (6 *Study_Valid1* , 16 *Study_Valid2*) did not report such events.

concerning takeaway was of students reported feeling more depressed (38%) than severely stressed. The studies shed light on different patterns of depression and severe stress. In particular, I found cases where students who reported feeling depressed did not report feeling stressed. Four students, P001-P004, reported multiple consecutive severely stressed and depressed periods. Eight students reported multiple experiences of *severe stress* but did not feel depressed. These findings validated a key design decision of *StressMon* to use separate detection models for stress and depression. While stress is commonly associated with depression, depression remains an independent mental health condition [9, 10].

CHAPTER 6

LITERATURE REVIEW

In reviewing prior work, my first objective was to establish a comprehensive clinical perspective on the associations of stress and depression. Subsequently, I extended my research to look into group factors related to stress and individual factors related to depression. I thoroughly reviewed the assessment methods for stress and depression in two folds. First, I researched validated scales that have been principally used by different populations. Second, I concentrated on technology-driven applications for stress and depression monitoring. In addition, this investigation included systems-related research related to large-scale sensing and utilising location information for stress, depression and group interactions.

6.1 Associations of Stress and Depression

Individuals, ranging from children to working adults [60,99,144,149,178] and holding different professions [8,179,180], experiencing stressful life events are strongly associated with being at risk for depression [81, 150, 181–184]. The relationship between both conditions, however, was often described to be unidirectional. Much of the early research focused on independent stressful life events. That is, these events are beyond one's control such as death of a loved one. More recently, the investigation on dependent events has gained recognition. Dependent events are occurrences influenced by characteristics of another individual; for example, receiving little peer support or having a conflict. Unfortunately, dependent stress episodes play a more significant role in depression or increase the likelihood of subsequent depression [181]. These findings verified the importance of group interaction as a key factor in monitoring stressful events and confirmed our decision to include group interaction as part of stress.

However, it is important to note that while these works suggest the association of depression with higher likelihood of stress [185, 186], depression can be influenced by anhedonia, which is an affective response linked to an individual's hereditary or personality characteristics [71, 72]. Harkness and colleagues discovered that cognitive-affective symptoms such as pessimism and sadness are strong predictors of pressures around interpersonal relationships. Further, researchers found that personality characteristics, specifically neuroticism, highly influence depression [71–74].

6.1.1 Social Identification in Stress

Expanding the assessment of stress to include group interaction factors, many studies have investigated several characteristics of group interaction in team processes; for example, social identification [22], social cohesion [75], social loafing [76], and group potency [77]. Among these factors, *social identification* has been suggested to be a strong indicator of stress [81, 82] and depression [18].

Since the development of social identity theory (SIT) by Henry Tajfel [187], social identity has climbed to be of importance in social psychology research – Our sense of 'who we are', defined by 'which group we belong to' and 'how different are we from other members in the group', affects our sense of self-worth, thus influences self-esteem and well being [22, 23]. Substantial research has found social identification as a strong indicator of mental health such as stress and depression [18, 81, 82]. Some of this research highlights increased commitment resulting from social identification with an organisation. On the other hand, developing a high sense of work commitment from increased identification may also result in long working hours, eventually affecting physical health [15]. One of the most common misconceptions regarding social identity is its association with social identification. While much work has used these terms interchangeably, the constructs of social identity and social identification are different. Social identity describes a range of characteristics to which a person identifies with the group; this, for example, could be age-related or professionally-related [141, 188]. In contrast, social identification describes the extent to which a person feels identified to a group (regardless of the nature); that is, the affective aspect towards the group [141]. With more organisations structuring work within teams to capitalise on skills and perspectives, much research in Organisational Behaviour (OB) has been dedicated to understanding social identification [22] and its implications for workgroups [14, 21, 23, 78, 79]; for example, social identification is negatively correlated with workplace bullying [189] and positively correlated with work commitment [15].

6.2 Assessment Methods

Here, I describe my considerations in using a specific assessment scale. The applicability of an instrument used to measure these conditions depends largely on how I intended to operationalise each construct in this study. Then, I provide the most prominent examples of measuring stress and depression through technological means.

6.2.1 Validated Scales

Stress: The Wheaton measure of chronic stress [190] and UCLA Life Stress Interview [178, 181, 182] investigate the construct of chronic stressor for major life events and the Stressful Life Events Screening Questionnaire (SLESQ) focuses on traumatic events [183]. Job Stress Survey and the Occupational Stress Indicator [58, 179] are specific to employment. While work stress is one of the widespread psychological stressors among people, the operationalisation of its construct, in this study, is not bound to work stress but extends to how participants judge any life circumstances as stressful. Among the most common methods of assessment is the Perceived Stress Scale (PSS), which has been used extensively in stress-related studies investigating factors at workplaces [180, 191, 192] and schools [146, 193, 194] related to physical health and mental well being [184,195,196]. The scale has been adapted to over 25 languages to be applied widely for different cultures and social norms [139,180,191,192,194]. Although PSS is available in various versions (e.g., 14-items, 10-items), the shortened version of 4-items (PSS-4) has demonstrated adequate internal consistency and reliability [195]. Evidently, the PSS is efficacious for overall comparison with existing and future studies. Nonetheless, the scale is not designed as a diagnosis tool and its results should only be interpreted as a screening tool to judge those in need of further support. That is, the service of a mental health professional is essential to make formal diagnosis for chronic conditions. PSS does not determine an individual's primary stressor.

Depression: The assessment for depression is encouraged by Psychiatry research to follow a diagnostic evaluation of criteria according to DSM and ICD [197, 198]. These criteria require an individual experiencing multiple symptoms indicating substantial functional impairment (including decrease or increase in appetite, feeling of worthlessness, recurrent suicidal ideation) within the same 2 week period [70]. The aim is clearly to avoid misdiagnosis of depression or trivialisation of its concept. Many depression rating scales are available. For example, the Hamilton Depression Rating Scale (HAM-D) [199], historically a common tool, is, however, not intended for diagnosis [200]. Other scales include the Montgomery-Asberg Depression Rating Scale (MADRS) [201], Beck Depression Inventory (BDI) [202] and Patient Health Questionnaire (PHQ) [98]. While studies have found reasonably strong correlation between these scales [203,204], PHQ may indicate an advantage over other scales as it has been frequently adopted by similar studies (as ours) in Systems research [39,96]. Further, the screener provides comparable results across multiple countries and cultures [205, 206]. In this study, I used

PHQ-8, an 8-item questionnaire. Advised by colleagues from our behavioural psychology department as well as practicing psychiatrists from a local mental health hospital, using PHQ-9 was strongly discouraged to avoid the ninth question – it asks the tough question of suicidal thoughts and our research team was not trained to handle a positive answer to that specific question.

Social Identification: Several scales were developed to measure social identification on multiple dimensions of self-esteem, self-categorisation and commitment to group [207–210]. However, the multiple dimensions in these scales are excessively elaborated and complex for users [141]. In contrast, the development of Single-Item Social Identification (SISI) [141] was found to strongly correlate with [209] and has since been preferred [18, 189, 211]. The extension of SISI, a Four-Item Social Identification (FISI) [83, 141, 164] was found in many studies to demonstrate much stronger reliability [212].

Conclusively, these prior works validated our decision to measure social identification with FISI. Despite being established, measuring social identification, as with any other survey measurements, has its limitations. First, survey questions restrict how we conceptualise the construct. Also, surveys do not cater to the sensitive timing of assessment. Over time, surveys become cumbersome to participants, and exacerbate the problem of missing data [24]. In the context of longitudinal studies, attrition (of participants) and non-responses can have severe implications on analyses [213].

6.2.2 Technology-driven Health Monitoring Systems

Research in developing systems for mental health continuously offers new capabilities to monitor psychological states and mental conditions in real-time [29, 41, 42, 97]. The early works of *StudentLife*, spearheaded by Wang *et al.* assesses mental health and performance of university students using fine-grained sensor data collected directly from mobile devices [29]. Most recently, the same authors looked at symptoms features to predict depression scores [92]. Other works specific to stress includes *UStress* [26], *cStress* [27], *StressSense* [28], and many more [33, 214, 215]. All these applications utilise electrodermal activity (EDA), electrocardiogram (ECG) from wearable sensors and/or smartphones to detect stress in real-life environments. Unfortunately, these applications are prone to fall foul from active sensing and incompatibility between systems/hardware requirements. Simultaneously, collecting multiple sources of information poses higher risk to one's privacy.

One form of lightweight sensing technique is the use of location data. Canzian *et al.* explored the correlation between GPS-based location features and depressions [39]. While informative, relying on GPS suffers from insufficient indoor precision; as a result, disregarding a substantial amount of behavioural information from humans spending 70%-80% of the time indoors [50]. Furthermore, group interaction habits were not considered as part of their investigation. To make up for indoor imprecisions, Brown *et al.* utilised wearable RFID tags to collect indoor location traces of employees interacting with colleagues in different building spaces [95]. However, this study was not intended for studying stress or depression. Moreover, while the technique is feasible to support indoor monitoring, it might pose impracticality problems as RFID tags are not (part of) commodity devices.

6.2.2.1 Large-scale Sensing Solutions

There has also been a lot of research devoted to developing sensing applications that scale from individuals to entire communities [216]. These applications, however, are mostly in the areas of urban planning [217] and security [46]; for example, using community-wide video surveillance for purposes of public safety. Specific to the area of mental health, Ware *et al.* [96] utilised WiFi association data from the university's WiFi infrastructure to detect depression. In a similar fashion, Zhou *et al.* [97] used WiFi indoor localisation data to learn about student behaviour. These works are conceptually the closest to our approach, however, they neither monitored stress nor inferred group interaction factors. *StressMon* uses a similar method to [96] but differs in the following ways: 1) it is a full system that works in real-time by pulling data directly from the WiFi infrastructure while prior work used data provided in an offline fashion from campus IT, 2) it incorporates group features into its models which prior work did not, 3) it achieves a significantly higher detection of individuals with depression at 91.21% true positive rate. (Note: these numbers were achieved in two different environments with two different population groups. It is possible that performance results would change significantly in different environments).

6.2.3 Sensing Group Interaction Patterns

Motivated by a desire to investigate how the assessment of social identification can evolve from its paper-pencil traditions, a promising step is to extend the aforementioned sensing technique, which had proven successful in capturing individual and collective behaviours [49]. Further, location information had previously shown favourable results in deriving social behaviours from knowing a person's activities or engagements [43, 113].

6.2.3.1 Workgroup Interactions And Location

Nemeth and Staw argue that workgroups tend to develop norms as a collective in their work practices [218]. This claim is supported by Jetten *et al.* who described highly (socially) identified individuals are more likely to incorporate salient in-group norms as a guide for behaviour [219]. However, a sudden change from these norms could be indicative of team friction or emotional upset within the group [220]; for example, members with high conflict tend to avoid each other [86, 87]. In fact, low identifiers might react against group norms by engaging in the opposite of group norms [219]. Ultimately, these prior works suggest that in-group participation varies to the degree with which individuals perceive their social identification [219, 221, 222]. We build on these findings to investigate if the location can be used to distinguish individual differences from their group norms, and accordingly, detect social identification. This question remains unclear in present literature.

6.2.3.2 Social Identification in Systems

Recently, Seering *et al.* enriched our knowledge on a set of ACM-related work [170] which referenced the scientific work of Henri Tajfel [187]. Most of these works are related to identifying key aspects of productivity in distributed groups through online collaborative spaces such as WikiProjects [80, 223], GitHub [224, 225], online multi-player games [171] and chatrooms [226]. The closest to our area of exploration is Min *et al.* – using call and message logs and applying SVM machine-learning technique, the authors sought to classify the relationships of people in an individual's contact list into groups of family, work and social [227]. While this work presented social identity theory as part of its background work to differentiate social roles, they did not directly investigate the constructs of social identity. In [170], Seering *et al.* argue that technologies could define various typical forms of groups and how high identifiers are more likely to be influenced by the group's norms. Building on these findings, we were determined to investigate if individual behaviours measured by technologies could similarly present distinct patterns in their groups. However, we did not use social identity in our investigation – as explained in Section 6.1.1, we utilised the measure of social identification.

6.3 Summary

These bodies of prior work inspired *StressMon*, and it differs in two distinct ways. First, we use passively collected coarse-grained location data, computed from RSSI signal strengths reported by the environment's WiFi infrastructure, as our primary source of data and show how even this coarse single-attribute data can be effectively used to generate a variety of location and group interaction features. Second, we use these features to calculate changes in work routines and group interactions and then to measure and detect stress and depression effectively. The use of these features to detect stress and depression has until now been unexplored. Overall, *StressMon* is designed to be a first-level safety net that provides mental health information about the entire population. It thus nicely complements more fine-grained solutions that require installing active stress trackers etc. that can be used to better understand the health of specific individuals who desire closer monitoring.

CHAPTER 7

CONCLUSION & FUTURE WORK

Ultimately, the long term goal of *StressMon* is to serve as an end-to-end solution in work environments, professional or academic, as an extra factor providing effective interventions after timely detections. This dissertation has detailed a solution to the problem of detecting stress and depression, separately using different models but fundamentally built on the same set of mobility features. In this chapter, I summarise the main contributions of this dissertation. Then, I address its limitations and discuss the deeper implications of deploying a solution such as *StressMon* that can automatically detect stress and depression across entire campuses without explicit user involvement.

7.1 Thesis Conclusion

In this dissertation, I demonstrated the possibility of *easily and accurately detecting stress and depression across entire school campuses using location data while overcoming the limitations of low-dimensional data by incorporating inferred individual and group features.* This is *easily* achieved through leveraging two key sensing mechanisms, the *LiveLabs* WiFi indoor location system and *Grumon* group detector system – where single-attribute RSSI values were directly sensed from the WiFi APs – to enable the passive collection of location and group data for any mobile device, without installing any client software.

To *overcome the limitations of low-dimensional data*, individual routine behaviours and group interaction patterns were inferred based on various activity heuristics. Then, changes in behaviours in comparison to an individual's past period and population were ascertained every 6-days and 15-days as features for detecting stress and depression, respectively. Detection models for stress and depression prioritise the accuracy of predicting individuals with signs of severe stress or depression rather than healthy individuals as being healthy – higher TPR at the expense of TNR. Using only mobility-driven features, ***Model$_S$*** *accurately* detects individuals showing signs of severe stress at 0.97 AUC score, 96.01% TPR, 80.76% TNR and overall accuracy of 81.76% for three student populations (see Section 4.3.2). The inclusion of a social identification [141] predicted outcome for ***Model$_{S_SE+SI}$*** , similarly from building a detection model with *Domain-specific* mobility features, helps to increase overall accuracy to 92.05% (see Section 4.3.3). However, this feature requires a specialised resource such as a project schedule to extract workgroup-related features. The detection of individuals showing significant signs of depression using only location data, unfortunately, yielded a reduced performance of 0.78 AUC, 77.96% TPR, 63.21% TNR and overall accuracy of 66.93% (see Section 4.4.2). However, the addition of *neuroticism* (one of five personality dimensions in the Big-five assessment [135]) successfully achieved improved performance of 0.88 AUC score, 91.21% TPR, 66.71% TNR and overall accuracy of 75.06% (see Section 4.4.3).

7.2 Summary Contributions

This dissertation makes contributions in three folds. First, we present *StressMon*, which is conceptually novel in that it is able to accommodate monitoring stress and depression to tens to thousands of users, at any given time. Second, this dissertation produced a set of mobility-based features, which comprehensively capture individuals' behaviours and physical group interaction patterns in their work environment.

The final contribution is the evaluation and longitudinal user studies which validate the detection models for stress and depression, making up the core engine of *StressMon*.

***StressMon* Solution**: In this dissertation, I illustrated the development of *StressMon*, which we envisioned as a first-level safety net to monitor stress and depression for an entire population of users. We adopted the *LiveLabs* indoor localisation system to passively collect single-attribute location data, obtained from RSSI values reported directly by the WiFi access points (APs). In doing so, we were able to bypass the installation of dedicated mobile applications and direct connections with individual devices for data collection. Further, location data does not directly expose user identities, thus minimises privacy risks. To make up for the lack of data, we proposed including physical group interactions and employed the *Grumon* group detector system, which clusters the location traces into logical groups. Altogether, the location traces allowed us to extract user behaviours and group interaction patterns. Finally, we tested our hypothesis on capturing changes in individuals' behaviours and interactions to maximise the value of location data. Expanding feature characteristics from just a single attribute (location) significantly reduces the pipeline of collecting and processing fine-grained mobile and wearable inputs, established in prior works.

System Artefacts: In developing an exhaustive set of mobility-based features, this dissertation has produced several system artefacts. First is the activity mapper, which is a fully functional data processing pipeline whose input is location information, provided in a .CSV format. I programmed the activity mapper to procedurally clean the location data in a 5-min time window, and treat missing data with AKIMA interpolation. Then, activities were programmatically determined based on a consolidation of the population's routined activities from demographic survey and project schedules. Second is the feature extractor, which extracts individuals' behaviour and interaction patterns; they can be generalised to the entire student population (*general* features), and a subset of features are highly specific to a certain course (*domain-specific* features). In addition, I calculated the changes in an individual's behaviour and interaction patterns by comparing each feature against its own history of an earlier period (***absolute change***, ***abs***), and against their peer population (***relative change***, ***rel***) over a 6-day time window for stress and 15-day time window for depression.

Evaluation of Detection Engine: The evaluation results of *StressMon* detection engines reported several important findings. I conducted a thorough evaluation on three different populations of students, degrees and year of study, over long periods of time (between 36 to 81 days); *Study_SE* (primary), *Study_Valid1* and *Study_Valid2* (validations). The first evaluation demonstrated the high performance of *StressMon*'s stress model in detecting individuals with *severe stress* at every 6 days; *severe stress* is equivalent to an individual whose stress score is 12 or more, out of a validated 16 points PSS-4 scale. Specifically, $\boldsymbol{Model_S}$ achieved a high 0.97 AUC score and 96% TPR. I conducted additional experiments to build a *social identification* detection model, specific to *Study_SE* students based on their *domain-specific* mobility features. Then, the output of their *social identification* was used as optimisation feature, which significantly improved detection of true negative cases, thus increased overall accuracy from 81.76% to 92.05%. The second evaluation demonstrated high performance of *StressMon*'s depression model in detecting *depressed* individuals at a slightly longer interval of every 15 days; *depressed* is equivalent to an individual whose depression score is 10 or more, out of a 24 points PHQ-8 scale. The depression model takes in an additional feature based on individual's personality trait, *neuroticism*, to achieve an AUC score of 0.88 and 91.21% TPR. The requirement of this additional feature, however, would mean that individuals must be surveyed for their personality at the start of the assessment.

7.3 Practical Limitations

7.3.1 Indoor Location SubSystem Requirement

StressMon fundamentally requires the availability of an indoor positioning system that can generate location information for every device in the environment using data collected solely from the infrastructure. This data is then processed by our software to generate group information and predictions. Currently, we use WiFi as it is the predominant solution deployed and used on our campus and we believe it is the predominant solution deployed on most campuses worldwide as well. If the WiFi deployment in a particular environment is sparse, then the accuracy of the location tracking will decrease and this could affect the performance of *StressMon*. The indoor location solution used by *StressMon* currently works with WiFi networks that use equipment from Aruba [228], Cisco [229], Zebra [230], or Ubiquiti [231]. *StressMon* can leverage other techniques such as Bluetooth if it is deployed generally; for example, at hospitals to help staff find their way to departments or wards [232]. In the future, if new technologies such as 5G replace WiFi in indoor environments, *StressMon* will be modified to use these technologies for its base sensing needs.

7.3.2 Beyond an Academic Setting

In this dissertation, I demonstrated how *StressMon* can accurately detect stress and depression amongst students in a university campus setting. *StressMon* would work on other campuses as well as there is nothing in our solution that is tied specifically to our campus. But how easy is it to deploy *StressMon* in other work environments? Fundamentally, *StressMon* uses deviations in work routines and interactions to produce its output and thus it will not work well in highly regimented work environments where the location of an individual does not change significantly across time – for example, factories where each worker is assigned to a dedicated point in the assembly process and stays there the entire day with minimal interaction with their peers (except during brief breaks) or elementary level education where students are in the same classroom the entire day. Instead, *StressMon* works best in work environments where monitoring deviations in work schedules and collaborative practices is possible; for example, on university campuses, hospitals, or military bases where students, nurses, and service personnel frequently move, daily, to different parts of the environment, and have ample opportunities to interact with different people. Moreover, offsite work behaviours and online work collaborations are not yet supported. It is important to note that *StressMon* primarily operationalises in work settings. Stressors come in varying forms (e.g., death of a loved one, divorce/separation, losing a job etc.) and are prioritised differently across the population. Recent surveys in the US and UK found those nations increasingly concerned about money (62% US, 43% UK (female), 30% UK (male)) and work (61% US, 41% UK (male)) [233, 234]. *Work* characteristically reflects numerous factors of an individual's well-being and effects on their well-being.

Above all, I have learned that depression is a condition, which may occur entirely independently of stress; for example, due to an experience of mood disorder. While my experiments have shown that stress is more likely to affect one's campus routines, the same cannot be said for depression. The strongest predictors of depression are related to social interactions, which might not have been comprehensively captured within campus; students spent at least 3.5 hours off-campus on leisure, socialising and commuting (see Figure 5.1). As in [39], Canzian *et al.* considered the mobility patterns of users to their homes and workplaces by collecting their GPS data. This study did not go beyond the academic setting. However, I believe *StressMon*'s technique can easily extend to use more sensors, such as GPS, if necessary. Scalable features which could be extracted from GPS and incorporated in *StressMon* include the time taken to commute from campus to home, number of places visited or transitions made between campus to home, and the types of places visited. However, adding more sensors reduces the scalability (as these sensors will require apps or other mechanisms) and increases the privacy concerns.

7.3.3 Latency of Predictions

StressMon currently detects stress every 6 days and depression every 15 days. Thus, it is not real-time even though it collects and processes real-time data. There are health monitoring solutions that offer real-time stress analysis [235] and then suggest interventions. However, we have designed *StressMon* to detect large and significant swings in mental health and this requires a sufficiently long measurement period – for example, to differentiate a truly stressed individual from one that is just instantaneously stressed and then recovers. In addition, detecting depression requires a longer observation window as this is a fairly fundamental change in mental health which needs to be carefully assessed. As stated previously, our goal was to design a first level safety net that flagged dramatic changes in mental health *at scale* and we believe *StressMon* succeeds at that – indeed its detection abilities are much faster than any competing solution at the scale in which it was designed to operate.

7.3.4 Other Limitations

I have shown that changes in an individual's routine and their group interactions, extracted from coarse-grained location data, make useful features in detecting *severe stress* and *depression*. It should be highlighted that this dissertation did not evaluate the sensitivity/robustness of *StressMon* , particularly to the data collected of its sensing mechanisms (*LiveLabs* and *Grumon*). It is possible that the accuracy numbers would change significantly in different environments. These experiments were tested on three different and separate student populations, sampled at different times. However, it is possible that students in other environment/cultures might adopt different routines on campus. Hence, further studies will be required to determine the efficacy of *StressMon* in other work settings (scholastic and professional).

7.4 Future Work

7.4.1 Providing Appropriate Interventions

One of the more interesting takeaways from our studies was a validation of prior findings that stress and depression are inconsistently correlated [236]. In particular, I found cases, see Figure 5.5 and 5.6, where students who were detected as depressed (and who indicated as such on their PHQ-8 surveys) did not report being stressed. I also found cases of highly stressed students, who reported being highly stressed over multiple consecutive reporting periods, who do not report themselves as being depressed. Overall, this reinforced my decision to use separate models for stress and depression. While the models share many features, sufficient differences make them unique and distinct. This highlights the impact of (solutions such as) *StressMon* as it accurately detected students who were depressed even though they were not stressed – these students would be almost impossible for peers or faculty to identify.

7.4.1.1 Individual-level Interventions

This dissertation has only addressed the detection of severe stress and depression without determining the underlying reasons for those conditions. For effective prevention and treatment to take place, greater attention is needed to understand the main stressor an individual is exposed to. Our goal is to create an end-to-end solution that also provides effective and timely interventions, which needs to be done quite carefully. For example, *StressMon* is a probabilistic system and thus it will make errors and sending interventions to individuals who do not need them could be problematic. Even more importantly, sending interventions to individuals *with* problems needs to be even more carefully monitored – to avoid the intervention accidentally worsening the condition. We are currently working with our psychology colleagues, our student counsellors, and with practicing psychiatrists from our primary mental health hospital to design and evaluate various interventions that can be sent by a system that uses triggers generated by *StressMon*. This is a very important and exciting area of future research as it raises questions such as "Should interventions be sent to individuals or to groups? Sending to groups would minimise the impact of an incorrect detection but might also result in less effective interventions", "When should interventions be sent and with what frequency?", "When should technology assisted (e.g. using an app) interventions be sent versus interventions provided by a human (e.g. a student counsellor)", "How can technology be used to send novel and potentially more effective interventions?", and "What are the appropriate privacy and ethical policies to ensure that no individual feels unfairly targeted or discriminated against while ensuring that anyone who needs help (even if they are not aware of it) receives it?".

7.4.1.2 Group-level Interventions

I learned that a common practice among teams was to maintain a computer-mediated communication (CMC) such as a SE workgroup Telegram chat. We found evidence to support Lampinen's claim that users perform self-censorship to manage adverse group situations even through communicative platforms [237]. For example, some of the top programmers would remove themselves from the chat groups of their teams as they did not identify with their groups and did not want to "stay in touch". However, cutting off the team's most convenient form of communication only led to more tenuous relationships between the students. Further, many students reportedly allowed this type of anti-social behaviour and remained passive as they feared provoking the top programmer who was leading their group. One possible use of our detection model, with the goal to build more supportive and productive groups, could be as an intervention mechanism where changing social identification scores could be revealed to the group early so that corrective actions could be taken before the situation became irreparable.

7.4.1.3 Evaluate Mental Wellness Programs

With increasing awareness in mental wellness, many organisations, academic institutions and governments [11] are actively developing actionable plans to improve the overall well-being of individuals, communities, and nations. Concurrently, it is important that these programs are objectively measured for the mental health outcomes. Hence, much work has gone into developing surveys that systematically evaluate one's career, social, financial, physical, and community from enrolling in such mental health programs. We believe *StressMon* can contribute to research in evaluating mental wellness programs. At its current state, *StressMon* can provide continuous behavioural monitoring, complementing the manual participation in surveys to evaluate individuals' work performances and mental progresses. As mentioned in Section 7.3, its sensing mechanism can be extended to include more fine-grained sensors for measuring off-work behaviours.

7.4.2 Dynamics of Social Identification

This dissertation only scratched the surface of investigating whether binary levels of social identification can be detected using mobility patterns. This dissertation did not provide a synthesis on the temporal dynamics of social identification. One way to take this analysis forward is to build on the preliminary findings in Chapter 5.2, where I noted students with severe stress tend to display more variance in their social identification compared to students with normal stress. In reality, social identification can be influenced by many factors other than stress; for example social comparison, leadership or even personality. Understanding such dynamics will require connecting individual-level dynamics (of their behaviours) with social identification, which is generally difficult to observe on a longitudinal basis using traditional assessments (i.e., interview, observations). Instead, *StressMon* can be leveraged to offer explanatory breadth of behavioural insights.

7.4.3 Privacy Policies

In the three longitudinal user studies, it was possible to obtain IRB approval to deploy *StressMon*, as university students are not paid employees, and detecting students with *severe stress* and *depression* has no lasting negative implications. Indeed the university has numerous mechanisms in place to help students with issues that do not impact their careers in any way. In contrast, a deployment of *StressMon* in a professional work environment, including its use to monitor university employees, would likely raise concerns among employees fearing negative reviews and discrimination. Thus, *StressMon* must have appropriate policies and mechanisms protecting employee rights before it can be widely deployed. This is especially important as *StressMon* is likely to be used, due to its inherent mechanisms, without explicit user consent. This is a rich area for future research, especially with the rise of community-wide sensing systems, and we are currently working with experts in Privacy and Ethics Law to develop appropriate policies and procedures for community-wide health monitoring systems such as *StressMon* that balance the privacy of individuals with the ability to provide help to those who most need it (and may not realise it).

7.5 Concluding Remarks

Stress and depression are increasingly prevalent mental illnesses all around the world. We see these topics progressively gaining research partnerships in different areas from Psychology to Organisational Behaviour to various Computing disciplines. With mobile devices becoming a commodity, mental health resources have evolved and transformed greatly from patient-clinical services to personalised mobile assessments, encouraging individuals to be more proactive in keeping up with their wellness plans. While promising, existing approaches exhibit limitations by posing higher privacy risk and power demands relative to collecting and transmitting data. Furthermore, they introduce a strong self-bias where only users who are interested in getting help would install the application on their phone. With more and more people in need of help to manage their stress and depression, it is not only critical the aforementioned limitations are addressed, but such resource should be made easily available and accessible to users in large scale. Moreover, the design of these apps is highly personalised and may unintentionally ignore the unique forms of collective factors that can result in experiencing stress and depression.

This dissertation aims to provide a feasible solution to overcome the three limitations. We envisioned providing a campus-wide “safety net” solution to automatically and non-intrusively detect individuals exhibiting signs of excessive stress or depression. Leveraging on the *LiveLabs* indoor localisation system and the *Grumon* group detector system as the sensing apparatus allows *StressMon* to easily scale across entire campuses as it does not require any dedicated app or explicit user interaction. In addition, the sensing apparatus supports grouping devices into logical groups based on location alone. The development and evaluation of both systems are not part of this dissertation. This dissertation is focused on synthesising location information from the sensing apparatus into mobility features that adequately represent individual behaviours and group interaction habits. I statistically analysed these features to build a high-performing machine learning detection engine for stress and depression, and examining how individual and group characteristics improved the performance of identifying the onset of stress or depression more thoroughly. The design of *StressMon*’s detection engine are fundamentally similar in terms of the classification algorithm (Random Forest) and mobility features (features measuring changes in behaviours and interaction patterns) being used. However, both models are sufficiently different to distinguish a stressed and depressed individual from longer observation windows, 6 to 15 days. The models competently flag extreme changes in behaviours, supporting evidences of behavioural symptoms for stress and depression. With the evaluation of our engine based on several user populations, we believe *StressMon* succeeds at its detection abilities to provide mental health assessment *easily*, *effectively* and *at scale*.

Appendices

APPENDIX A

INSTITUTIONAL REVIEW BOARD APPROVAL

1. IRB-17-079-A087(817) was approved for *Study_SE* , which ran for 81 days for Software Engineering students.

2. Modification IRB-17-079-A087-M2(218) was approved for *Study_Valid1* study, which lasted 36 days for Social Entrepreneurship course

3. Modification IRB-17-079-A087-M5(1118) was approved for *Study_Valid2* study, which lasted 81 days for multiple courses and included two additional scales in Appendix B.6 and B.7.

SMU
SINGAPORE MANAGEMENT
UNIVERSITY

21 August 2017

Nur Camellia Binte Zakaria
SMU Student
School of Information Systems

Dear Camellia,

IRB APPROVAL OF RESEARCH
CATEGORY 2A: Expedited Review
Title of Research: Understanding the Relation of Time Management to Stress among College Students
SMU-IRB Approval Number: IRB-17-079-A087(817)

Thank you for your IRB application for the above research that we received the latest revised application on 21 Aug 2017.

I am pleased to let you know that, based on the description of the research in your IRB application, the IRB has determined that your research falls under Category 2 and has approved your application.

Please note the following:

1. Indicate the above SMU-IRB approval number in all your correspondence with the IRB on this research.

2. If any adverse events or unanticipated problems involving human subjects occur during the course of the research project, you must complete in full the SMU-IRB Unanticipated Problem/Adverse Events Report Form (see SMU-IRB website) and submit it to the SMU-IRB within 24 hours of the event.

3. If you plan to modify your original protocol that was approved by the SMU-IRB, you must complete in full the SMU-IRB Protocol Modification Request Form (see SMU-IRB website) and submit it to the SMU-IRB to seek approval before implementing any modified protocol.

4. This IRB approval for your research is valid for one year (12 months) from the date of this letter. If you plan to extend your research project beyond one year from the date of the IRB approval, you must submit a request to renew the research protocol using the Continuation Review Form (see SMU-IRB website) or Protocol Modification Request Form ***prior to the IRB approval expiry date***.

5. Please be reminded to be compliant with Singapore's Personal Data Protection laws in carrying out your research activities.

If you have any queries, please contact the IRB Secretariat at irb@smu.edu.sg or telephone +65 6828-1925.

Yours Sincerely,

Li Jing
Committee Member
Institutional Review Board

Administration Building 81 Victoria Street Singapore 188065
Tel: +65 6828 0100 Fax: +65 6828 0101 www.smu.edu.sg Reg. No. 200000267Z

SMU Classification: Restricted

SMU
SINGAPORE MANAGEMENT UNIVERSITY

21 February 2018

Nur Camellia Binte ZAKARIA
SMU Student
School of Information Systems

Dear Camellia,

IRB PROTOCOL MODIFICATION REQUEST APPROVAL
CATEGORY 2A: EXPEDITED REVIEW
Title of Research: Understanding the Relation of Time Management to Stress among College Students
SMU-IRB Exemption/Approval Number: IRB-17-079-A087(817)
SMU-IRB Modification Number: IRB-17-079-A087-M2(218)

Thank you for your IRB Protocol Modification Request application for the above research in which we received the latest revised copy on 21 February 2018.

I am pleased to let you know that, the IRB has approved your application for the modification based on the description of modified research protocol stated in your Modification Request form.

Please note the following:

1. Indicate the above SMU-IRB approval number and SMU-IRB modification number in all your correspondence with the IRB on this research.

2. If any adverse events or unanticipated problems involving human subjects occur during the course of the research project, you must complete in full the SMU-IRB Unanticipated Problem/Adverse Events Report Form (see SMU-IRB website) and submit it to the SMU-IRB within 24 hours of the event.

3. If you plan to modify your original protocol that was approved by the SMU-IRB, you must complete in full the SMU-IRB Protocol Modification Request Form (see SMU-IRB website) and submit it to the SMU-IRB to seek approval before implementing any modified protocol.

4. This IRB approval for your modified protocol is valid one year from the date of this letter. If you plan to extend your research project beyond one year from the date of the IRB approval, you must submit a request to renew the research protocol using the Continuing Review Form (see SMU-IRB website) or Protocol Modification Request Form ***prior to the IRB approval expiry date***.

5. Please be reminded to be compliant with Singapore's Personal Data Protection laws in carrying out your research activities.

If you have any queries, please contact the IRB Secretariat at irb@smu.edu.sg or telephone +65 6828-1925.

Yours Sincerely,

Christopher Chen
Deputy Chair
Institutional Review Board

Administration Building 81 Victoria Street Singapore 188065
Tel: +65 6828 0100 Fax: +65 6828 0101 www.smu.edu.sg Reg. No. 200000267Z

SMU Classification: Restricted

SMU
SINGAPORE MANAGEMENT
UNIVERSITY

2 November 2018

Nur Camellia Binte ZAKARIA
SMU Student
School of Information Systems

Dear Camellia,

IRB PROTOCOL MODIFICATION REQUEST APPROVAL
CATEGORY 2A: EXPEDITED REVIEW
Title of Research: Understanding the Relation of Time Management to Stress among College Students
SMU-IRB Exemption/Approval Number: IRB-17-079-A087(817)
SMU-IRB Modification Number: IRB-17-079-A087-M5(1118)

Thank you for your IRB Protocol Modification Request application for the above research in which we received the latest revised copy on 1 November 2018.

I am pleased to let you know that, the IRB has approved your application for the modification based on the description of modified research protocol stated in your Modification Request form.

Please note the following:

1. Indicate the above SMU-IRB approval number and SMU-IRB modification number in all your correspondence with the IRB on this research.

2. If any adverse events or unanticipated problems involving human subjects occur during the course of the research project, you must complete in full the SMU-IRB Unanticipated Problem/Adverse Events Report Form (see SMU-IRB website) and submit it to the SMU-IRB within 24 hours of the event.

3. If you plan to modify your original protocol that was approved by the SMU-IRB, you must complete in full the SMU-IRB Protocol Modification Request Form (see SMU-IRB website) and submit it to the SMU-IRB to seek approval before implementing any modified protocol.

4. This IRB approval for your modified protocol is valid one year from the date of this letter. For Expedited Review applications, if you plan to extend your research project beyond one year from the date of the IRB approval, you must submit a request to renew the research protocol using the Continuing Review Form (see SMU-IRB website) or Protocol Modification Request Form ***prior to the IRB approval expiry date***. Please note that for Full Review applications, continuing review applications must be submitted and approved until the research study is closed (i.e., at least one research paper has been published or presented).

5. Please be reminded to be compliant with Singapore's Personal Data Protection laws in carrying out your research activities.

If you have any queries, please contact the IRB Secretariat at irb@smu.edu.sg or telephone +65 6828-1925.

Yours Sincerely,

Forrest Zhang
Chair
Institutional Review Board

Administration Building 81 Victoria Street Singapore 188065
Tel: +65 6828 0100 Fax: +65 6828 0101 www.smu.edu.sg

Reg. No. 200000267Z

Appendix B

StressMon-related Assessments

B.1 Demographics Questions

Instructions: Please answer the following questions. Your answers will be kept confidential. However, feel free to skip any questions you might find uncomfortable in addressing.

Note: Questions indicated with (*) are only for students taking Software Engineering course.

1. What is your Participation ID? ________________
2. What is your mobile phone MAC Address? ________________
3. What is your age? ________________
4. What is your gender? ________________
5. What is your citizenship? ________________
6. What is your marital status? ________________
7. Are you pursuing a Double major? ________________
8. Are you working alongside your studies to support yourself? And if so, why? ________________
9. What is the highest education level you would like to achieve? ________________
10. What is the highest education level you are expected to achieve by your family? ________________

11.	How many credit units (CUs) are you taking this semester?	________________
(*) 12.	Are you taking PMSB module this semester	________________
13.	What type of housing are you currently living in? (select one)	Dorm rented room rented home owned public home owned private home
14.	Are you paying for your accommodation?	Yes / No
15.	How many people are living with you at present?	___ Grandparents ___ Parents ___ Siblings ___ Relatives ___ Friends ___ Others:_______

Eating/Sleeping/Physical Activities Patterns

16.	In general, how many meals do you eat in a day?	________________
17.	In general, how many times do you go to the campus eateries in a day? (e.g Koufu, 1983 etc.)	________________
18.	In general, what are your usual meal hours when on campus? (You can tick more than one option)	___ 07:00-08:00 ___ 08:00-09:00 ___ 09:00-10:00 ___ 10:00-11:00 ___ 11:00-12:00 ___ 12:00-13:00 ___ 13:00-14:00 ___ 14:00-15:00 ___ 15:00-16:00 ___ 16:00-17:00 ___ 17:00-18:00 ___ 18:00-19:00 ___ 19:00-20:00 ___ 20:00-21:00
19.	In general, what are the top 3 most frequented locations to study when on campus?	(1)______________ (2)______________ (3)______________
20.	In general, how many hours of sleep do you get in a day?	________________
21.	While on campus, do you take naps in general? If so, how many hours approximately? And which location are you most likely to go to?	________________
22.	How often do you exercise in a week?	________________
23.	Do you frequent any of these facilities when on-campus? Select where applicable	___ campus gym ___ swimming pool ___ sports hall

B.2 Perceived Stress Scale 4, PSS-4

Referenced Cohen *et al.* [139]

Note: The questions in this scale ask you about your feelings and thoughts during the last 3 days.

1. In the last 3 days, how often have you felt that you were unable to control the important things in your life?

2. In the last 3 days, how often have you felt confident about your ability to handle your personal problems?

3. In the last 3 days, how often have you felt that things were going your way?

4. In the last 3 days, how often have you felt difficulties were piling up so high that you could not overcome them?

Scoring for the PSS-4:

Questions 1 and 4: 0 = Never, 1 = Almost Never, 2 = Sometimes, 3 = Fairly Often, 4 = Very Often

Questions 2 and 3: 4 = Never, 3 = Almost Never, 2 = Sometimes, 1 = Fairly Often, 0 = Very Often

Lowest score: 0, Highest score: 16 – Higher scores are correlated to more stress.

B.3 Patient Health Questionnaire, PHQ-8

Referenced Kroenke *et al.* [98]

Note: Over the last 2 weeks, how often have you been bothered by any of the following problems?

	Not at all	**Several Days**	**More than half the days**	**Nearly every day**
Little interest or pleasure in doing things	0	1	2	3
Feeling down, depressed, or hopeless	0	1	2	3
Trouble falling or staying asleep, or sleeping too much	0	1	2	3
Feeling tired or having little energy	0	1	2	3
Poor appetite or overeating	0	1	2	3
Feeling bad about yourself, or that you are a failure, or have let yourself or your family down	0	1	2	3
Trouble concentrating on things, such as reading the newspaper or watching television	0	1	2	3
Moving or speaking so slowly that other people could have noticed. Or the opposite – being so fidgety or restless that you have been moving around a lot more than usual	0	1	2	3

Scoring for PHQ-8:

Total score and severity of depression: 0–4 None, 5–9 Mild depression, 10–14 Moderate depression, 15–19 moderately severe depression, 20–24 severe depression.

B.4 Four-Item Social Identification, FISI

Referenced Postmes *et al.* [141]. Adapted from Doosie [83, 164].

Note: [In-group] refers to the workgroup you are assigned to in the course you have registered as part of this study. For example, "Software Engineering (SE)", "Social Entrepreneurship (ScE)" etc.

1. I identify with [In-group].
2. I feel committed to [In-group].
3. I am glad to be [In-group].
4. Being [In-group] is an important part of how I see myself.

Scoring for FISI: On a scale from 1 (strongly disagree) to 7 (strongly agree).

B.5 The Big Five Inventory, Big-5

Referenced John *et al.* [135]

Note: Please write a number next to each statement to indicate the extent to which you agree or disagree with that statement.

I see myself as someone who...

01.	Is talkative	23.	Tends to be lazy
02.	Tends to find fault with others	24.	Is emotionally stable, not easily upset
03.	Does a thorough job	25.	Is inventive
04.	Is depressed, blue	26.	Has an assertive personality
05.	Is original, comes up with new ideas	27.	Can be cold and aloof
06.	Is reserved	28.	Perseveres until the task is finished
07.	Is helpful and unselfish with others	29.	Can be moody
08.	Can be somewhat careless	30.	Values artistic, aesthetic experiences
09.	Is relaxed, handles stress well	31.	Is sometimes shy, inhibited
10.	Is curious about many different things	32.	Is considerate and kind to almost everyone
11.	Is full of energy	33.	Does things efficiently
12.	Starts quarrels with others	34.	Remains calm in tense situations
13.	Is a reliable worker	35.	Prefers work that is routine
14.	Can be tense	36.	Is outgoing, sociable
15.	Is ingenious, a deep thinker	37.	Is sometimes rude to others
16.	Generates a lot of enthusiasm	38.	Makes plans and follows through with them
17.	Has a forgiving nature	39.	Gets nervous easily
18.	Tends to be disorganised	40.	Likes to reflect, play with ideas
19.	Worries a lot	41.	Has few artistic interests
20.	Has an active imagination	42.	Likes to cooperate with others
21.	Tends to be quiet	43.	Is easily distracted
22.	Is generally trusting	44.	Is sophisticated in art, music, or literature

Scoring for Big-5: On a scale from 1 (strongly disagree) to 5 (strongly agree). "R" denotes reverse-scored items

Extraversion: 1, 6R, 11, 16, 21R, 26, 31R, 36

Agreeableness: 2R, 7, 12R, 17, 22, 27R, 32, 37R, 42

Conscientiousness: 3, 8R, 13, 18R, 23R, 28, 33, 38, 43R

Neuroticism: 4, 9R, 14, 19, 24R, 29, 34R, 39

Openness: 5, 10, 15, 20, 25, 30, 35R, 40, 41R, 44

B.6 Sources of Stress

Adapted from Yumba *et al.* [144]

Please kindly reply to the following questions:

During your studies you may some how experience stress. How would you rate these sources of stress that may cause stressful situations at any time during your studies? What do you think is causing stress during your studies?

Strongly disagree: 1 to 2, **Agree:** 3, **Strongly agree:** 4 to 5, **Don't know:** 6

A. INTERPERSONAL FACTORS						
1. Change in relation with school mates	1	2	3	4	5	6
2. Work with people (school-related group members) you don't know	1	2	3	4	5	6
3. School-related misunderstandings	1	2	3	4	5	6
4. Difficult personalities of (school-related) group members	1	2	3	4	5	6
B. PERSONAL FACTORS						
1. Change in sleeping habits	1	2	3	4	5	6
2. Change in eating habits	1	2	3	4	5	6
3. Financial difficulties	1	2	3	4	5	6
4. Combining job with studies	1	2	3	4	5	6
5. Personal relationship issues	1	2	3	4	5	6
6. Family issues	1	2	3	4	5	6
7. Health issues	1	2	3	4	5	6
C. ACADEMIC FACTORS						
1. Increased class workload	1	2	3	4	5	6
2. Lower grade	1	2	3	4	5	6
3. Many hours of studies	1	2	3	4	5	6
4. Language difficulties	1	2	3	4	5	6
5. Lack of university support	1	2	3	4	5	6
6. Examinations	1	2	3	4	5	6
7. Missing lectures	1	2	3	4	5	6
8. Group projects	1	2	3	4	5	6
9. Many hours of group projects	1	2	3	4	5	6
D. ENVIRONMENT FACTORS						
1. Lack of vacations/breaks	1	2	3	4	5	6
2. Computer problem	1	2	3	4	5	6
3. Bad living conditions	1	2	3	4	5	6
4. Quit job	1	2	3	4	5	6
5. Unfamiliar educational environment	1	2	3	4	5	6
6. Moving to a new city	1	2	3	4	5	6

B.7 Time on Education

Adapted from the American Time Use Survey [145]

Estimate the average amount of time IN HOURS spent EACH DAY on the following activities:

1. Classes (seminars, lab sessions, extra-classes etc.)
2. Self-study ON campus
3. Group project ON campus
4. Co-curriculum activities ON campus
5. Leisure and sports such as gym and swimming ON campus (exclude hours spent for sports-related CCA)
6. Social hangouts ON campus
7. Leisure and sports OFF campus
8. Social hangouts OFF campus
9. Commute to campus (both ways)

APPENDIX C

ADDITIONAL DATA COLLECTION

C.1 Gratitude

Referenced McCullough *et al.* [136]

Using the scale below as a guide, indicate how much you agree with the following statements:

Strongly disagree: 1, **Disagree:** 2, **Slightly disagree:** 3, **Neutral:** 4, **Slightly agree:** 5, **Agree:** 6, **Strongly agree:** 7

1.	I have so much in life to be thankful for	1	2	3	4	5	6	7
2.	If I had to list everything that I felt grateful for, it would be a very long list	1	2	3	4	5	6	7
3(*).	When I look at the world, I don't see much to be grateful for	1	2	3	4	5	6	7
4.	I am grateful to a wide variety of people	1	2	3	4	5	6	7
5.	As I get older I find myself more able to appreciate the people, events, and situations that have been part of my life story	1	2	3	4	5	6	7
6(*).	Long amounts of time can go by before I feel grateful to to something or someone	1	2	3	4	5	6	7

Scoring instructions: Add up the scores for items 1, 2, 4, and 5. Then, reverse the scores for items 3 and 6. That is, if a question is scored "7," it is "1," if the question is scored a "6," it is "2," etc. Finally, add the reversed scores for items 3 and 6 to the total from Step 1. The score for GQ-6 is a number between 6 and 42.

C.2 Meaning in Life

Referenced Steger *et al.* [137]

Please take a moment to think about what makes your life and existence feel important and significant to you. Please respond to the following statements as truthfully and accurately as you can, and also please remember that these are very subjective questions and that there are no right or wrong answers. Please answer according to the scale below:

Absolutely untrue: 1, **Mostly untrue:** 2, **Somewhat untrue:** 3, **Can't say true or false:** 4, **Somewhat true:** 5, **Mostly true:** 6, **Absolutely true:** 7

1.	I understand my life's meaning.	1	2	3	4	5	6	7
2.	I am looking for something that makes my life feel meaningful.	1	2	3	4	5	6	7
3.	I am always looking to find my life's purpose.	1	2	3	4	5	6	7
4.	My life has a clear sense of purpose.	1	2	3	4	5	6	7
5.	I have a good sense of what makes my life meaningful.	1	2	3	4	5	6	7
6.	I have discovered a satisfying life purpose.	1	2	3	4	5	6	7
7.	I am always searching for something that makes my life feel significant.	1	2	3	4	5	6	7
8.	I am seeking a purpose or mission for my life.	1	2	3	4	5	6	7
9.	My life has no clear purpose.	1	2	3	4	5	6	7
10.	I am searching for meaning in my life.	1	2	3	4	5	6	7

Scoring instructions:

Presence = 1, 4, 5, 6, & 9-reverse-coded

Search = 2, 3, 7, 8, & 10

C.3 Satisfaction with Life

Referenced Diener *et al.* [138]

Below are five statements that you may agree or disagree with. Using the 1 – 7 scale below, please indicate your agreement with each item.

Strongly disagree: 1, **Disagree:** 2, **Slightly disagree:** 3, **Neither agree nor disagree:** 4, **Slightly agree:** 5, **Agree:** 6, **Strongly agree:** 7

1.	In most ways my life is close to my ideal	1	2	3	4	5	6	7
2.	The conditions of my life are excellent	1	2	3	4	5	6	7
3.	I am satisfied with my life.	1	2	3	4	5	6	7
4.	So far I have gotten the important things I want in life.	1	2	3	4	5	6	7
5.	If I could live my life over, I would change almost nothing.	1	2	3	4	5	6	7

Scoring instructions:

The SWLS is a 7-point Likert style response scale. The possible range of scores is 5-35, with a score of 20 representing a neutral point on the scale. Scores between 5-9 indicate the respondent is extremely dissatisfied with life, whereas scores between 31-35 indicate the respondent is extremely satisfied.

C.4 Positive and Negative Affect Schedule, PANAS

Referenced Watson *et al.* [140]

Rate your feelings and emotions in the past 2 weeks:

Very slightly/not at all: 1, **A little:** 2,

Moderately: 3, **Quite a bit:** 4, **Extremely:** 5

#	Score	Feelings/Emotions	#	Score	Feelings/Emotions
1		Interested	11		Irritable
2		Distressed	12		Alert
3		Excited	13		Ashamed
4		Upset	14		Inspired
5		Strong	15		Nervous
6		Guilty	16		Determined
7		Scared	17		Attentive
8		Hostile	18		Jittery
9		Enthusiastic	19		Active
10		Proud	20		Afraid

Scoring instructions

Positive Affect Score: Add the scores on items 1, 3, 5, 9, 10, 12, 14, 16, 17 & 19. Scores can range between 10 – 50. Higher scores represent higher levels of positive affect. Mean scores: momentary = 29.7 and weekly = 33.3.

Negative Affect Score: Add the scores on items 2, 4, 6, 7, 8, 11, 13, 15, 18 & 20. Scores can range between 10 – 50. Higher scores represent higher levels of negative affect. Mean scores: momentary = 14.8 and weekly = 17.4.

C.5 Social Loafing

Adapted from a 10-item measure of George *et al.* [76]

Not at all: 1, **A little:** 2, **Somewhat:** 3, **Moderately:** 4, **Quite a bite:** 5, **A great deal:** 6

During the [workgroup] project, I…

1.	Deferred responsibilities I should have assumed to other people.	1	2	3	4	5	6
2.	Did not do my share of the work.	1	2	3	4	5	6
3.	Put forth less effort than other members of my project team.	1	2	3	4	5	6
4.	Took it easy if other teammates were around to do the work.	1	2	3	4	5	6

C.6 Social Cohesion

Referenced Kozlowski *et al.* [142]

Not at all: 1, **A little:** 2, **Somewhat:** 3, **Moderately:** 4, **Quite a bite:** 5, **A great deal:** 6

Members of my project team . . .

1.	get along well together.	1	2	3	4	5	6
2.	enjoy spending time together.	1	2	3	4	5	6
3.	have good relationships with each other.	1	2	3	4	5	6
4.	like to socialise together.	1	2	3	4	5	6
5.	are friends with each other.	1	2	3	4	5	6

C.7 Group Potency

Referenced Guzzo *et al.* [143]

Note: [In-group] refers to the workgroup you are assigned to in the course you have registered as part of this study. For example, "Software Engineering (SE)", "Social Entrepreneurship (ScE)" etc.

Not at all: 1, **A little:** 2, **Somewhat:** 3, **Moderately:** 4, **Quite a bite:** 5, **A great deal:** 6

1.	My project team has confidence in itself.	1	2	3	4	5	6
2.	My team expects to be known as a high-performing team.	1	2	3	4	5	6
3.	My team feels it can solve any problem it encounters	1	2	3	4	5	6
3.	during the [In-group] project.						
4.	My team believes it can be very productive.	1	2	3	4	5	6

Bibliography

[1] M. H. Foundation, *Mental health statistics: stress*, 2019, https://www.mentalhealth.org.uk/statistics/mental-health-statistics-stress.

[2] W. H. Organisation, *Depression*, 2019, https://www.who.int/news-room/fact-sheets/detail/depression.

[3] C. Hammen, "Stress and depression," *Annu. Rev. Clin. Psychol.*, vol. 1, pp. 293–319, 2005.

[4] G. Harnois, P. Gabriel, W. H. Organization *et al.*, "Mental health and work: impact, issues and good practices," 2000.

[5] J. Lazarus, *Stress relief & relaxation techniques*. Keats Pub., 2000.

[6] A. Bartolomucci and R. Leopardi, "Stress and depression: preclinical research and clinical implications," *PloS one*, vol. 4, no. 1, p. e4265, 2009.

[7] C. Tennant, "Work-related stress and depressive disorders," *Journal of psychosomatic research*, vol. 51, no. 5, pp. 697–704, 2001.

[8] K. Wada, T. Sairenchi, Y. Haruyama, H. Taneichi, Y. Ishikawa, and T. Muto, "Relationship between the onset of depression and stress response measured by the brief job stress questionnaire among japanese employees: a cohort study," *PLoS One*, vol. 8, no. 2, p. e56319, 2013.

[9] H. M. S. Harvard Health Publishing, *What causes depression?*, 2017, https://www.health.harvard.edu/mind-and-mood/what-causes-depression.

[10] A. P. Association, *What is Depression?*, 2017, https://www.psychiatry.org/patients-families/depression/what-is-depression.

[11] W. H. Organisation, *Mental health in the workplace*, 2017, https://www.who.int/mental_health/in_the_workplace/en/.

[12] T. Cox, *Stress research and stress management: Putting theory to work*. Hse Books Sudbury, 1993, vol. 61.

[13] T. Cox, A. Griffiths, and E. Rial-González, *Research on work-related stress*. European Communities, 2000.

[14] O. Herrbach, "A matter of feeling? the affective tone of organizational commitment and identification," *Journal of Organizational Behavior: The International Journal of Industrial, Occupational and Organizational Psychology and Behavior*, vol. 27, no. 5, pp. 629–643, 2006.

[15] T. W. Ng and D. C. Feldman, "Long work hours: A social identity perspective on meta-analysis data," *Journal of Organizational Behavior: The International Journal of Industrial, Occupational and Organizational Psychology and Behavior*, vol. 29, no. 7, pp. 853–880, 2008.

[16] T. A. I. of Stress, *Workplace Stress*, 2019, https://www.stress.org/workplace-stress.

[17] E. S. Epel, E. H. Blackburn, J. Lin, F. S. Dhabhar, N. E. Adler, J. D. Morrow, and R. M. Cawthon, "Accelerated telomere shortening in response to life stress," *Proceedings of the National Academy of Sciences*, vol. 101, no. 49, pp. 17 312–17 315, 2004.

[18] T. Cruwys, S. A. Haslam, G. A. Dingle, C. Haslam, and J. Jetten, "Depression and social identity: An integrative review," *Personality and Social Psychology Review*, vol. 18, no. 3, pp. 215–238, 2014.

[19] J. I. C. on Occupational Health, *Psychosocial factors at work: Recognition and control*. International Labour Organisation, 1986, no. 56.

[20] P. Dewe, "The transactional model of stress: some implications for stress management programs," *Asia Pacific Journal of Human Resources*, vol. 35, no. 2, pp. 41–51, 1997.

[21] S. A. Haslam, A. O'brien, J. Jetten, K. Vormedal, and S. Penna, "Taking the strain: Social identity, social support, and the experience of stress," *British Journal of Social Psychology*, vol. 44, no. 3, pp. 355–370, 2005.

[22] S. A. Haslam, J. Jetten, T. Postmes, and C. Haslam, "Social identity, health and well-being: An emerging agenda for applied psychology," *Applied psychology: an international review*, vol. 58, no. 1, pp. 1–23, 2009.

[23] S. A. Haslam and R. Van Dick, "A social identity approach to workplace stress," *Social psychology and organizations*, pp. 325–352, 2011.

[24] M. A. Cronin, L. R. Weingart, and G. Todorova, "Dynamics in groups: Are we there yet?" *The academy of management annals*, vol. 5, no. 1, pp. 571–612, 2011.

[25] T. Razavi, "Self-report measures: An overview of concerns and limitations of questionnaire use in occupational stress research," 2001.

[26] B. Egilmez, E. Poyraz, W. Zhou, G. Memik, P. Dinda, and N. Alshurafa, "Ustress: Understanding college student subjective stress using wrist-based passive sensing," in *2017 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*. IEEE, 2017, pp. 673–678.

[27] K. Hovsepian, M. al'Absi, E. Ertin, T. Kamarck, M. Nakajima, and S. Kumar, "cstress: towards a gold standard for continuous stress assessment in the mobile environment," in *Proceedings of the 2015 ACM international joint*

conference on pervasive and ubiquitous computing. ACM, 2015, pp. 493–504.

[28] H. Lu, D. Frauendorfer, M. Rabbi, M. S. Mast, G. T. Chittaranjan, A. T. Campbell, D. Gatica-Perez, and T. Choudhury, “Stresssense: Detecting stress in unconstrained acoustic environments using smartphones,” in *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*. ACM, 2012, pp. 351–360.

[29] 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*. ACM, 2014, pp. 3–14.

[30] J. Aigrain, A. Dapogny, K. Bailly, S. Dubuisson, M. Detyniecki, and M. Chetouani, “On leveraging crowdsourced data for automatic perceived stress detection,” in *Proceedings of the 18th ACM International Conference on Multimodal Interaction*. ACM, 2016, pp. 113–120.

[31] Y. Ayzenberg, J. Hernandez Rivera, and R. Picard, “Feel: frequent eda and event logging–a mobile social interaction stress monitoring system,” in *CHI’12 extended abstracts on human factors in computing systems*. ACM, 2012, pp. 2357–2362.

[32] R. Ferdous, V. Osmani, and O. Mayora, “Smartphone app usage as a predictor of perceived stress levels at workplace,” in *2015 9th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth)*. IEEE, 2015, pp. 225–228.

[33] E. Ertin, N. Stohs, S. Kumar, A. Raij, M. al’Absi, and S. Shah, “Autosense: unobtrusively wearable sensor suite for inferring the onset, causality, and

consequences of stress in the field," in *Proceedings of the 9th ACM Conference on Embedded Networked Sensor Systems*. ACM, 2011, pp. 274–287.

[34] T. Hao, K. N. Walter, M. J. Ball, H.-Y. Chang, S. Sun, and X. Zhu, "Stresshacker: Towards practical stress monitoring in the wild with smartwatches," in *AMIA Annual Symposium Proceedings*, vol. 2017. American Medical Informatics Association, 2017, p. 830.

[35] N. Maeda, Y. Hirabe, Y. Arakawa, and K. Yasumoto, "Cosms: unconscious stress monitoring system for office worker," in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct*. ACM, 2016, pp. 329–332.

[36] V. Mishra, G. Pope, S. Lord, S. Lewia, B. Lowens, K. Caine, S. Sen, R. Halter, and D. Kotz, "The case for a commodity hardware solution for stress detection," in *Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers*. ACM, 2018, pp. 1717–1728.

[37] E. Rey, A. Jain, S. Abdullah, T. Choudhury, and D. Erickson, "Personalized stress monitoring: a smartphone-enabled system for quantification of salivary cortisol," *Personal and Ubiquitous Computing*, vol. 22, no. 4, pp. 867–877, 2018.

[38] A. Pantelopoulos and N. G. Bourbakis, "A survey on wearable sensor-based systems for health monitoring and prognosis," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 40, no. 1, pp. 1–12, 2010.

[39] L. Canzian and M. Musolesi, "Trajectories of depression: unobtrusive monitoring of depressive states by means of smartphone mobility traces analysis," in *Proceedings of the 2015 ACM international joint conference on pervasive and ubiquitous computing*. ACM, 2015, pp. 1293–1304.

[40] N. Jaques, S. Taylor, A. Azaria, A. Ghandeharioun, A. Sano, and R. Picard, "Predicting students' happiness from physiology, phone, mobility, and behavioral data," in *Affective computing and intelligent interaction (ACII), 2015 international conference on*. IEEE, 2015, pp. 222–228.

[41] S. Saeb, M. Zhang, C. J. Karr, S. M. Schueller, M. E. Corden, K. P. Kording, and D. C. Mohr, "Mobile phone sensor correlates of depressive symptom severity in daily-life behavior: an exploratory study," *Journal of medical Internet research*, vol. 17, no. 7, 2015.

[42] R. Wang, M. S. Aung, S. Abdullah, R. Brian, A. T. Campbell, T. Choudhury, M. Hauser, J. Kane, M. Merrill, E. A. Scherer *et al.*, "Crosscheck: toward passive sensing and detection of mental health changes in people with schizophrenia," in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 2016, pp. 886–897.

[43] E. Cho, S. A. Myers, and J. Leskovec, "Friendship and mobility: user movement in location-based social networks," in *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2011, pp. 1082–1090.

[44] A. Exler, M. Braith, K. Mincheva, A. Schankin, and M. Beigl, "Smartphone-based estimation of a user being in company or alone based on place, time, and activity," in *International Conference on Mobile Computing, Applications, and Services*. Springer, 2018, pp. 74–89.

[45] T. Choudhury and A. Pentland, "Sensing and modeling human networks using the sociometer," in *null*. IEEE, 2003, p. 216.

[46] D. Zhang, B. Guo, and Z. Yu, "The emergence of social and community intelligence," *Computer*, vol. 44, no. 7, pp. 21–28, 2011.

[47] M. Singson *et al.*, "Awareness and use of wi-fi infrastructure in student's community: a case study pondicherry university on campus students," 2011.

[48] T. Stolovy, Y. Melamed, and A. Afek, "Video surveillance in mental health facilities: is it ethical?" *The Israel Medical Association journal: IMAJ*, vol. 17, no. 5, pp. 274–276, 2015.

[49] K. Jayarajah, Y. Lee, A. Misra, and R. K. Balan, "Need accurate user behaviour?: pay attention to groups!" in *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 2015, pp. 855–866.

[50] N. E. Klepeis, W. C. Nelson, W. R. Ott, J. P. Robinson, A. M. Tsang, P. Switzer, J. V. Behar, S. C. Hern, and W. H. Engelmann, "The national human activity pattern survey (nhaps): a resource for assessing exposure to environmental pollutants," *Journal of Exposure Science and Environmental Epidemiology*, vol. 11, no. 3, p. 231, 2001.

[51] S. Folkman, R. S. Lazarus, R. J. Gruen, and A. DeLongis, "Appraisal, coping, health status, and psychological symptoms." *Journal of personality and social psychology*, vol. 50, no. 3, p. 571, 1986.

[52] R. S. Lazarus, "Psychological stress in the workplace," *Occupational stress: A handbook*, vol. 1, pp. 3–14, 1995.

[53] R. S. Lazarus and S. Folkman, *Stress, appraisal, and coping*. Springer publishing company, 1984.

[54] B. A. Van der Kolk, A. C. McFarlane, and L. Weisæth, *Traumatic stress: The effects of overwhelming experience on mind, body, and society*. Guilford Press, 2012.

[55] H. Franke, "Toxic stress: effects, prevention and treatment," *Children*, vol. 1, no. 3, pp. 390–402, 2014.

[56] J. P. Jamieson, W. B. Mendes, and M. K. Nock, "Improving acute stress responses: The power of reappraisal," *Current Directions in Psychological Science*, vol. 22, no. 1, pp. 51–56, 2013.

[57] S. J. Lepore, H. J. Miles, and J. S. Levy, "Relation of chronic and episodic stressors to psychological distress, reactivity, and health problems," *International Journal of Behavioral Medicine*, vol. 4, no. 1, pp. 39–59, 1997.

[58] J. Siegrist, "Chronic psychosocial stress at work and risk of depression: evidence from prospective studies," *European archives of psychiatry and clinical neuroscience*, vol. 258, no. 5, p. 115, 2008.

[59] H. Pikhart, M. Bobak, A. Pajak, S. Malyutina, R. Kubinova, R. Topor, H. Sebakova, Y. Nikitin, and M. Marmot, "Psychosocial factors at work and depression in three countries of central and eastern europe," *Social science & medicine*, vol. 58, no. 8, pp. 1475–1482, 2004.

[60] M. Melchior, A. Caspi, B. J. Milne, A. Danese, R. Poulton, and T. E. Moffitt, "Work stress precipitates depression and anxiety in young, working women and men," *Psychological medicine*, vol. 37, no. 8, pp. 1119–1129, 2007.

[61] G. S. Shields, L. L. Toussaint, and G. M. Slavich, "Stress-related changes in personality: A longitudinal study of perceived stress and trait pessimism," *Journal of research in personality*, vol. 64, pp. 61–68, 2016.

[62] B. Roshanaei-Moghaddam, W. J. Katon, and J. Russo, "The longitudinal effects of depression on physical activity," *General hospital psychiatry*, vol. 31, no. 4, pp. 306–315, 2009.

[63] J. G. Van Uffelen, Y. R. van Gellecum, N. W. Burton, G. Peeters, K. C. Heesch, and W. J. Brown, "Sitting-time, physical activity, and depressive symptoms in mid-aged women," *American journal of preventive medicine*, vol. 45, no. 3, pp. 276–281, 2013.

[64] N. Schneiderman, G. Ironson, and S. D. Siegel, "Stress and health: psychological, behavioral, and biological determinants," *Annu. Rev. Clin. Psychol.*, vol. 1, pp. 607–628, 2005.

[65] R. A. Sherman, J. F. Rauthmann, N. A. Brown, D. G. Serfass, and A. B. Jones, "The independent effects of personality and situations on real-time expressions of behavior and emotion." *Journal of Personality and Social Psychology*, vol. 109, no. 5, p. 872, 2015.

[66] Anxiety and D. A. of America, *Symptoms*, 2019, https://adaa.org/understanding-anxiety/depression/symptoms.

[67] M. J. Schabracq and C. L. Cooper, "The changing nature of work and stress," *Journal of Managerial Psychology*, vol. 15, no. 3, pp. 227–241, 2000.

[68] U. D. of Health and H. Services, *Common Signs & Signals of a Stress Reaction*, 2019, https://foh.psc.gov/NYCU/StressReaction.asp.

[69] A. P. Association *et al.*, "Diagnostic and statistical manual of mental disorders, 5th edn (dsm-5)," *Washington, DC: APA*, 2013.

[70] S. L. Schmidt and J. C. Tolentino, "Dsm-5 criteria and depression severity: implications for clinical practice." *Frontiers in psychiatry*, vol. 9, p. 450, 2018.

[71] S. J. Rizvi, D. A. Pizzagalli, B. A. Sproule, and S. H. Kennedy, "Assessing anhedonia in depression: potentials and pitfalls," *Neuroscience & Biobehavioral Reviews*, vol. 65, pp. 21–35, 2016.

[72] M. T. Treadway and D. H. Zald, "Reconsidering anhedonia in depression: lessons from translational neuroscience," *Neuroscience & Biobehavioral Reviews*, vol. 35, no. 3, pp. 537–555, 2011.

[73] A. Griens, K. Jonker, P. Spinhoven, and M. Blom, "The influence of depressive state features on trait measurement," *Journal of Affective Disorders*, vol. 70, no. 1, pp. 95–99, 2002.

[74] J. Ormel, A. J. Oldehinkel, and W. Vollebergh, “Vulnerability before, during, and after a major depressive episode: a 3-wave population-based study,” *Archives of general psychiatry*, vol. 61, no. 10, pp. 990–996, 2004.

[75] S. A. Carless and C. De Paola, “The measurement of cohesion in work teams,” *Small group research*, vol. 31, no. 1, pp. 71–88, 2000.

[76] J. M. George, “Extrinsic and intrinsic origins of perceived social loafing in organizations,” *Academy of Management Journal*, vol. 35, no. 1, pp. 191–202, 1992.

[77] A. D. Stajkovic, D. Lee, and A. J. Nyberg, “Collective efficacy, group potency, and group performance: Meta-analyses of their relationships, and test of a mediation model.” *Journal of applied psychology*, vol. 94, no. 3, p. 814, 2009.

[78] H. H. Moerbeek and A. Need, “Enemies at work: can they hinder your career?” *Social Networks*, vol. 25, no. 1, pp. 67–82, 2003.

[79] J. Dillard and J. Fritz, “Interpersonal relationships and organizational performance,” in *Annual Convention of the Western Speech Communication Association, Portland OR*, 1995.

[80] H. Zhu, R. Kraut, and A. Kittur, “Organizing without formal organization: group identification, goal setting and social modeling in directing online production,” in *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*. ACM, 2012, pp. 935–944.

[81] A. Isaksson, P. Martin, J. Kaufmehl, M. Heinrichs, G. Domes, and N. Rüsch, “Social identity shapes stress appraisals in people with a history of depression,” *Psychiatry research*, vol. 254, pp. 12–17, 2017.

[82] N. R. Branscombe, M. T. Schmitt, and R. D. Harvey, “Perceiving pervasive discrimination among african americans: Implications for group identifica-

tion and well-being." *Journal of personality and social psychology*, vol. 77, no. 1, p. 135, 1999.

[83] B. Doosje, N. Ellemers, and R. Spears, "Perceived intragroup variability as a function of group status and identification," *Journal of experimental social psychology*, vol. 31, no. 5, pp. 410–436, 1995.

[84] J. E. Dutton, J. M. Dukerich, and C. V. Harquail, "Organizational images and member identification," *Administrative science quarterly*, pp. 239–263, 1994.

[85] M. A. Hogg and S. C. Hains, "Intergroup relations and group solidarity: Effects of group identification and social beliefs on depersonalized attraction." *Journal of Personality and Social Psychology*, vol. 70, no. 2, p. 295, 1996.

[86] J. W. Jackson, "Intergroup attitudes as a function of different dimensions of group identification and perceived intergroup conflict," *Self and identity*, vol. 1, no. 1, pp. 11–33, 2002.

[87] N. Nair, "Towards understanding the role of emotions in conflict: a review and future directions," *International Journal of Conflict Management*, vol. 19, no. 4, pp. 359–381, 2008.

[88] A. N. Radford, B. Majolo, and F. Aureli, "Within-group behavioural consequences of between-group conflict: a prospective review," *Proceedings of the Royal Society B: Biological Sciences*, vol. 283, no. 1843, p. 20161567, 2016.

[89] D. MacLean, A. Roseway, and M. Czerwinski, "Moodwings: a wearable biofeedback device for real-time stress intervention," in *Proceedings of the 6th international conference on PErvasive Technologies Related to Assistive Environments*. ACM, 2013, p. 66.

[90] M. Gjoreski, M. Luštrek, M. Gams, and H. Gjoreski, "Monitoring stress with a wrist device using context," *Journal of biomedical informatics*, vol. 73, pp. 159–170, 2017.

[91] B. Cvetković, M. Gjoreski, J. Šorn, P. Maslov, and M. Luštrek, "Monitoring physical activity and mental stress using wrist-worn device and a smartphone," in *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer, 2017, pp. 414–418.

[92] R. Wang, W. Wang, A. daSilva, J. F. Huckins, W. M. Kelley, T. F. Heatherton, and A. T. Campbell, "Tracking depression dynamics in college students using mobile phone and wearable sensing," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 2, no. 1, p. 43, 2018.

[93] L. J. Barney, K. M. Griffiths, H. Christensen, and A. F. Jorm, "Exploring the nature of stigmatising beliefs about depression and help-seeking: implications for reducing stigma," *BMC public health*, vol. 9, no. 1, p. 61, 2009.

[94] C. Lauber, C. Nordt, L. Falcato, and W. Rössler, "Lay recommendations on how to treat mental disorders," *Social psychiatry and psychiatric epidemiology*, vol. 36, no. 11, pp. 553–556, 2001.

[95] C. Brown, C. Efstratiou, I. Leontiadis, D. Quercia, C. Mascolo, J. Scott, and P. Key, "The architecture of innovation: Tracking face-to-face interactions with ubicomp technologies," in *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 2014, pp. 811–822.

[96] S. Ware, C. Yue, R. Morillo, J. Lu, C. Shang, J. Kamath, A. Bamis, J. Bi, A. Russell, and B. Wang, "Large-scale automatic depression screening using meta-data from wifi infrastructure," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 2, no. 4, p. 195, 2018.

[97] M. Zhou, M. Ma, Y. Zhang, K. SuiA, D. Pei, and T. Moscibroda, "Edum: classroom education measurements via large-scale wifi networks," in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 2016, pp. 316–327.

[98] K. Kroenke, T. W. Strine, R. L. Spitzer, J. B. Williams, J. T. Berry, and A. H. Mokdad, “The phq-8 as a measure of current depression in the general population,” *Journal of affective disorders*, vol. 114, no. 1, pp. 163–173, 2009.

[99] M. Pariat, M. Rynjah, M. Joplin, and M. Kharjana, “Stress levels of college students: interrelationship between stressors and coping strategies,” *IOSR J. Humanit. Soc. Sci*, vol. 19, pp. 40–45, 2014.

[100] D. Eisenberg, M. F. Downs, E. Golberstein, and K. Zivin, “Stigma and help seeking for mental health among college students,” *Medical Care Research and Review*, vol. 66, no. 5, pp. 522–541, 2009.

[101] R. H. Rahe, “Stressful life events—their nature and effects,” *Psychosomatic Medicine*, vol. 39, no. 1, pp. 64–65, 1977.

[102] C. T. Mowbray, D. Megivern, J. M. Mandiberg, S. Strauss, C. H. Stein, K. Collins, S. Kopels, C. Curlin, and R. Lett, “Campus mental health services: recommendations for change.” *American Journal of Orthopsychiatry*, vol. 76, no. 2, p. 226, 2006.

[103] A. Networks, *The Whys and Hows of Deploying Large-Scale Campus-wide Wi-Fi Networks*, 2012, https://www.arubanetworks.com/pdf/technology/whitepapers/wp_Large-scale_campus-wide_wifi.pdf.

[104] J. Vitak, K. Shilton, and Z. Ashktorab, “Beyond the belmont principles: Ethical challenges, practices, and beliefs in the online data research community,” in *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*. ACM, 2016, pp. 941–953.

[105] E. Department of Health *et al.*, “The belmont report. ethical principles and guidelines for the protection of human subjects of research.” *The Journal of the American College of Dentists*, vol. 81, no. 3, p. 4, 2014.

[106] R. K. Balan, A. Misra, and Y. Lee, "Livelabs: Building an in-situ real-time mobile experimentation testbed," in *Proceedings of the 15th Workshop on Mobile Computing Systems and Applications*. ACM, 2014, p. 14.

[107] R. Sen, Y. Lee, K. Jayarajah, A. Misra, and R. K. Balan, "Grumon: Fast and accurate group monitoring for heterogeneous urban spaces," in *Proceedings of the 12th ACM Conference on Embedded Network Sensor Systems*. ACM, 2014, pp. 46–60.

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

[109] L. Goodman, "Mental health on university campuses and the needs of students they seek to serve," *Building Healthy Academic Communities Journal*, vol. 1, no. 2, pp. 31–44, 2017.

[110] L. S. Cowin and C. Hengstberger-Sims, "New graduate nurse self-concept and retention: A longitudinal survey," *International journal of nursing studies*, vol. 43, no. 1, pp. 59–70, 2006.

[111] N. Recruitment, "Nurse recruitment and retention committee: final report," *Melbourne: Policy and Strategic Projects Division, Victorian Government Department of Human Services*, 2001.

[112] K. D'Silva, K. Jayarajah, A. Noulas, C. Mascolo, and A. Misra, "The role of urban mobility in retail business survival," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 2, no. 3, p. 100, 2018.

[113] J. Krumm and E. Horvitz, "Predestination: Inferring destinations from partial trajectories," in *International Conference on Ubiquitous Computing*. Springer, 2006, pp. 243–260.

[114] N. National Coordination Office for Space-Based Positioning and Timing, *The Global Positioning System*, 2017, https://www.gps.gov/systems/gps/.

[115] A. Khalajmehrabadi, N. Gatsis, and D. Akopian, "Modern wlan fingerprinting indoor positioning methods and deployment challenges," *IEEE Communications Surveys & Tutorials*, vol. 19, no. 3, pp. 1974–2002, 2017.

[116] V. Bahl and V. Padmanabhan, "Enhancements to the radar user location and tracking system," 2000.

[117] A. Farshad, M. K. Marina, and F. Garcia, "Urban wifi characterization via mobile crowdsensing," in *2014 IEEE Network Operations and Management Symposium (NOMS)*. IEEE, 2014, pp. 1–9.

[118] A. Misra and R. K. Balan, "Livelabs: initial reflections on building a large-scale mobile behavioral experimentation testbed," *ACM SIGMOBILE Mobile Computing and Communications Review*, vol. 17, no. 4, pp. 47–59, 2013.

[119] A. Lindemann, B. Schnor, J. Sohre, and P. Vogel, "Indoor positioning: A comparison of wifi and bluetooth low energy for region monitoring." in *HEALTHINF*, 2016, pp. 314–321.

[120] C. J. Hansen, "Internetworking with bluetooth low energy," *GetMobile: Mobile Comp. and Comm.*, vol. 19, no. 2, pp. 34–38, Aug. 2015. [Online]. Available: http://doi.acm.org/10.1145/2817761.2817774

[121] J. I. Hong, G. Boriello, J. A. Landay, D. W. McDonald, B. N. Schilit, and J. Tygar, "Privacy and security in the location-enhanced world wide web," in *Proceedings of Fifth International Conference on Ubiquitous Computing: Ubicomp*, 2003, p. 11.

[122] P. Bahl, V. N. Padmanabhan, V. Bahl, and V. Padmanabhan, "Radar: An in-building rf-based user location and tracking system," 2000.

[123] Z. Yang, C. Wu, and Y. Liu, “Locating in fingerprint space: wireless indoor localization with little human intervention,” in *Proceedings of the 18th annual international conference on Mobile computing and networking*. ACM, 2012, pp. 269–280.

[124] S. Bak, S. Jeon, Y.-J. Suh, C. Yu, and D. Han, “Characteristics of a large-scale wifi radiomap and their implications in indoor localization,” in *2013 Fourth International Conference on the Network of the Future (NOF)*. IEEE, 2013, pp. 1–5.

[125] A. J. Khan, V. Ranjan, T.-T. Luong, R. Balan, and A. Misra, “Experiences with performance tradeoffs in practical, continuous indoor localization,” in *2013 IEEE 14th International Symposium on" A World of Wireless, Mobile and Multimedia Networks"(WoWMoM)*. IEEE, 2013, pp. 1–9.

[126] D. Jaisinghani, R. K. Balan, V. Naik, A. Misra, and Y. Lee, “Experiences & challenges with server-side wifi indoor localization using existing infrastructure,” in *Proceedings of the 15th EAI International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services*. ACM, 2018, pp. 226–235.

[127] H. Akima, A. Gebhardt, T. Petzoldt, and M. Maechler, “Akima: Interpolation of irregularly spaced data. r package version 0.5-4,” 2006.

[128] M. Baruth, D.-C. Lee, X. Sui, T. S. Church, B. H. Marcus, S. Wilcox, and S. N. Blair, “Emotional outlook on life predicts increases in physical activity among initially inactive men,” *Health Education & Behavior*, vol. 38, no. 2, pp. 150–158, 2011.

[129] R. Guinness, “Beyond where to how: A machine learning approach for sensing mobility contexts using smartphone sensors,” *Sensors*, vol. 15, no. 5, pp. 9962–9985, 2015.

[130] K. Kirasich, T. Smith, and B. Sadler, "Random forest vs logistic regression: Binary classification for heterogeneous datasets," *SMU Data Science Review*, vol. 1, no. 3, p. 9, 2018.

[131] Z. Xing, J. Pei, and E. Keogh, "A brief survey on sequence classification," *ACM Sigkdd Explorations Newsletter*, vol. 12, no. 1, pp. 40–48, 2010.

[132] J. C. Turner, M. A. Hogg, P. J. Oakes, S. D. Reicher, and M. S. Wetherell, *Rediscovering the social group: A self-categorization theory.* Basil Blackwell, 1987.

[133] C. M. Florkowski, "Sensitivity, specificity, receiver-operating characteristic (roc) curves and likelihood ratios: communicating the performance of diagnostic tests," *The Clinical Biochemist Reviews*, vol. 29, no. Suppl 1, p. S83, 2008.

[134] H. Li, P. Zou, X. Wang, and R. Xia, "A new combination sampling method for imbalanced data," in *Proceedings of 2013 Chinese Intelligent Automation Conference*. Springer, 2013, pp. 547–554.

[135] O. P. John, E. M. Donahue, and R. L. Kentle, "The big five inventory—versions 4a and 54," 1991.

[136] M. E. McCullough, R. A. Emmons, and J.-A. Tsang, "The grateful disposition: a conceptual and empirical topography." *Journal of personality and social psychology*, vol. 82, no. 1, p. 112, 2002.

[137] M. F. Steger, P. Frazier, S. Oishi, and M. Kaler, "The meaning in life questionnaire: Assessing the presence of and search for meaning in life." *Journal of counseling psychology*, vol. 53, no. 1, p. 80, 2006.

[138] E. Diener, R. A. Emmons, R. J. Larsen, and S. Griffin, "The satisfaction with life scale," *Journal of personality assessment*, vol. 49, no. 1, pp. 71–75, 1985.

[139] S. Cohen, T. Kamarck, R. Mermelstein *et al.*, “Perceived stress scale,” *Measuring stress: A guide for health and social scientists*, 1994.

[140] D. Watson, L. A. Clark, and A. Tellegen, “Development and validation of brief measures of positive and negative affect: the panas scales.” *Journal of personality and social psychology*, vol. 54, no. 6, p. 1063, 1988.

[141] T. Postmes, S. A. Haslam, and L. Jans, “A single-item measure of social identification: Reliability, validity, and utility,” *British journal of social psychology*, vol. 52, no. 4, pp. 597–617, 2013.

[142] P.-R. P. Kozlowski, S. W. J. and B. C. Lim, “The developmental role of team leaders,” *Current perspectives on leadership in collective work arrangements*, 2010.

[143] R. A. Guzzo, P. R. Yost, R. J. Campbell, and G. P. Shea, “Potency in groups: Articulating a construct,” *British journal of social psychology*, vol. 32, no. 1, pp. 87–106, 1993.

[144] W. Yumba, “Academic stress: A case of the undergraduate students,” 2010.

[145] U. B. of Labor Statistics, *American Time Use Survey*, 2016, https://www.bls.gov/tus/.

[146] S. L. Warttig, M. J. Forshaw, J. South, and A. K. White, “New, normative, english-sample data for the short form perceived stress scale (pss-4),” *Journal of health psychology*, vol. 18, no. 12, pp. 1617–1628, 2013.

[147] Qualtrics, *Qualtrics*, 2019, http://www.qualtrics.com.

[148] OctoberCMS, *2017 WORK AND WELL-BEING SURVEY*, 2016, http://octobercms.com/.

[149] P. Sacco, K. K. Bucholz, and D. Harrington, “Gender differences in stressful life events, social support, perceived stress, and alcohol use among older

adults: results from a national survey," *Substance use & misuse*, vol. 49, no. 4, pp. 456–465, 2014.

[150] D. A. Pizzagalli, R. Bogdan, K. G. Ratner, and A. L. Jahn, "Increased perceived stress is associated with blunted hedonic capacity: potential implications for depression research," *Behaviour research and therapy*, vol. 45, no. 11, pp. 2742–2753, 2007.

[151] E. Manninen, "Generalized identification: Individuals' levels of identification with groups and its relation to personality," 2016.

[152] M. J. Azur, E. A. Stuart, C. Frangakis, and P. J. Leaf, "Multiple imputation by chained equations: what is it and how does it work?" *International journal of methods in psychiatric research*, vol. 20, no. 1, pp. 40–49, 2011.

[153] S. Greenland and W. D. Finkle, "A critical look at methods for handling missing covariates in epidemiologic regression analyses," *American journal of epidemiology*, vol. 142, no. 12, pp. 1255–1264, 1995.

[154] E. Burnaev, P. Erofeev, and A. Papanov, "Influence of resampling on accuracy of imbalanced classification," in *Eighth International Conference on Machine Vision (ICMV 2015)*, vol. 9875. International Society for Optics and Photonics, 2015, p. 987521.

[155] G. E. Batista, R. C. Prati, and M. C. Monard, "A study of the behavior of several methods for balancing machine learning training data," *ACM Sigkdd Explorations Newsletter*, vol. 6, no. 1, pp. 20–29, 2004.

[156] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "Smote: synthetic minority over-sampling technique," *Journal of artificial intelligence research*, vol. 16, pp. 321–357, 2002.

[157] L. Lusa *et al.*, "Evaluation of smote for high-dimensional class-imbalanced microarray data," in *2012 11th International Conference on Machine Learning and Applications*, vol. 2. IEEE, 2012, pp. 89–94.

[158] D. J. Katzelnick, F. F. Duffy, H. Chung, D. A. Regier, D. S. Rae, and M. H. Trivedi, "Depression outcomes in psychiatric clinical practice: using a self-rated measure of depression severity," *Psychiatric services*, vol. 62, no. 8, pp. 929–935, 2011.

[159] D. R. Reetz, B. Krylowicz, and B. Mistler, "The association for university and college counseling center directors annual survey," *Aurora*, vol. 51, p. 60506, 2014.

[160] T. A. Miller and S. F. McCool, "Coping with stress in outdoor recreational settings: An application of transactional stress theory," *Leisure Sciences*, vol. 25, no. 2-3, pp. 257–275, 2003.

[161] J. J. Thomas and E. A. Borrayo, "The impact of perceived stress and psychosocial factors on missed class and work in college students," *Journal of College Counseling*, vol. 19, no. 3, pp. 246–260, 2016.

[162] P. T. Costa Jr and R. R. McCrae, "Personality: Another'hidden factor'is stress research," *Psychological Inquiry*, vol. 1, no. 1, pp. 22–24, 1990.

[163] S. G. Scott, "Social identification effects in product and process development teams," *Journal of Engineering and Technology Management*, vol. 14, no. 2, pp. 97–127, 1997.

[164] C. W. Leach, M. Van Zomeren, S. Zebel, M. L. Vliek, S. F. Pennekamp, B. Doosje, J. W. Ouwerkerk, and R. Spears, "Group-level self-definition and self-investment: a hierarchical (multicomponent) model of in-group identification." *Journal of personality and social psychology*, vol. 95, no. 1, p. 144, 2008.

[165] C. C. Peters and W. R. Van Voorhis, "Statistical procedures and their mathematical bases." 1940.

[166] L. G. Humphreys, "Research on individual differences requires correlational analysis, not anova," *Intelligence*, vol. 2, no. 1, pp. 1–5, 1978.

[167] D. P. Farrington and R. Loeber, "Some benefits of dichotomization in psychiatric and criminological research," *Criminal behaviour and mental health*, vol. 10, no. 2, pp. 100–122, 2000.

[168] D. M. Rousseau, "Why workers still identify with organizations," *Journal of Organizational Behavior: The International Journal of Industrial, Occupational and Organizational Psychology and Behavior*, vol. 19, no. 3, pp. 217–233, 1998.

[169] M. Gilles, "Hmmweka. available online," 2011.

[170] J. Seering, F. Ng, Z. Yao, and G. Kaufman, "Applications of social identity theory to research and design in computer-supported cooperative work," *Proceedings of the ACM on Human-Computer Interaction*, vol. 2, no. CSCW, p. 201, 2018.

[171] N. Bos, N. S. Shami, J. S. Olson, A. Cheshin, and N. Nan, "In-group/out-group effects in distributed teams: an experimental simulation," in *Proceedings of the 2004 ACM conference on Computer supported cooperative work*. ACM, 2004, pp. 429–436.

[172] R. A. Power and M. Pluess, "Heritability estimates of the big five personality traits based on common genetic variants," *Translational psychiatry*, vol. 5, no. 7, p. e604, 2015.

[173] J. Anglim and P. O'Connor, "Measurement and research using the big five, hexaco, and narrow traits: A primer for researchers and practitioners," *Australian Journal of Psychology*, vol. 71, no. 1, pp. 16–25, 2019.

[174] L. Festinger, "A theory of social comparison processes," *Human relations*, vol. 7, no. 2, pp. 117–140, 1954.

[175] P. Brickman and R. J. Bulman, "Pleasure and pain in social comparison," *Social comparison processes: Theoretical and empirical perspectives*, vol. 149, p. 186, 1977.

[176] C. Bicchieri, R. Muldoon, and A. Sontuoso, "Social norms," in *The Stanford Encyclopedia of Philosophy*, winter 2018 ed., E. N. Zalta, Ed. Metaphysics Research Lab, Stanford University, 2018.

[177] M. A. Hogg and E. A. Hardie, "Social attraction, personal attraction, and self-categorization-, a field study," *Personality and Social Psychology Bulletin*, vol. 17, no. 2, pp. 175–180, 1991.

[178] C. Hammen, J. Shih, T. Altman, and P. A. Brennan, "Interpersonal impairment and the prediction of depressive symptoms in adolescent children of depressed and nondepressed mothers," *Journal of the American Academy of Child & Adolescent Psychiatry*, vol. 42, no. 5, pp. 571–577, 2003.

[179] I. T. Robertson, C. L. Cooper, J. Williams, and J. Williams, "The validity of the occupational stress indicator," *Work & Stress*, vol. 4, no. 1, pp. 29–39, 1990.

[180] Z. Wang, J. Chen, J. E. Boyd, H. Zhang, X. Jia, J. Qiu, and Z. Xiao, "Psychometric properties of the chinese version of the perceived stress scale in policewomen," *PloS one*, vol. 6, no. 12, p. e28610, 2011.

[181] C. Hammen, J. H. Shih, and P. A. Brennan, "Intergenerational transmission of depression: test of an interpersonal stress model in a community sample." *Journal of consulting and clinical psychology*, vol. 72, no. 3, p. 511, 2004.

[182] C. Hammen, A. Ellicott, M. Gitlin, and K. R. Jamison, "Sociotropy/autonomy and vulnerability to specific life events in patients with unipolar depression and bipolar disorders." *Journal of Abnormal Psychology*, vol. 98, no. 2, p. 154, 1989.

[183] L. A. Goodman, C. Corcoran, K. Turner, N. Yuan, and B. L. Green, "Assessing traumatic event exposure: General issues and preliminary findings for the stressful life events screening questionnaire," *Journal of Traumatic*

Stress: Official Publication of The International Society for Traumatic Stress Studies, vol. 11, no. 3, pp. 521–542, 1998.

[184] A. M. Mitchell, P. A. Crane, and Y. Kim, "Perceived stress in survivors of suicide: psychometric properties of the perceived stress scale," *Research in nursing & health*, vol. 31, no. 6, pp. 576–585, 2008.

[185] B. S. Dohrenwend and B. P. Dohrenwend, "Socioenvironmental factors, stress, and psychopathology," *American Journal of Community Psychology*, vol. 9, no. 2, p. 123, 1981.

[186] R. A. Depue and S. M. Monroe, "Conceptualization and measurement of human disorder in life stress research: The problem of chronic disturbance." *Psychological bulletin*, vol. 99, no. 1, p. 36, 1986.

[187] H. Tajfel, J. C. Turner, W. G. Austin, and S. Worchel, "An integrative theory of intergroup conflict," *Organizational identity: A reader*, pp. 56–65, 1979.

[188] H. Zacher, L. Esser, C. Bohlmann, and C. W. Rudolph, "Age, social identity and identification, and work outcomes: A conceptual model, literature review, and future research directions," *Work, Aging and Retirement*, vol. 5, no. 1, pp. 24–43, 2018.

[189] J. Loh, S. L. D. Restubog, and T. J. Zagenczyk, "Consequences of workplace bullying on employee identification and satisfaction among australians and singaporeans," *Journal of Cross-Cultural Psychology*, vol. 41, no. 2, pp. 236–252, 2010.

[190] R. J. Turner, B. Wheaton *et al.*, "Checklist measurement of stressful life events," *Measuring stress: A guide for health and social scientists*, pp. 29–58, 1995.

[191] T. Almadi, I. Cathers, A. M. H. Mansour, and C. M. Chow, "An arabic version of the perceived stress scale: Translation and validation study," *International Journal of Nursing Studies*, vol. 49, no. 1, pp. 84–89, 2012.

[192] F.-X. Lesage, S. Berjot, and F. Deschamps, “Psychometric properties of the french versions of the perceived stress scale,” *International journal of occupational medicine and environmental health*, vol. 25, no. 2, pp. 178–184, 2012.

[193] A. Denovan, N. Dagnall, K. Dhingra, and S. Grogan, “Evaluating the perceived stress scale among uk university students: implications for stress measurement and management,” *Studies in Higher Education*, vol. 44, no. 1, pp. 120–133, 2019.

[194] C. Mimura and P. Griffiths, “A japanese version of the perceived stress scale: translation and preliminary test,” *International Journal of Nursing Studies*, vol. 41, no. 4, pp. 379–385, 2004.

[195] M. A. Vallejo, L. Vallejo-Slocker, E. G. Fernández-Abascal, and G. Mañanes, “Determining factors for stress perception assessed with the perceived stress scale (pss-4) in spanish and other european samples,” *Frontiers in psychology*, vol. 9, p. 37, 2018.

[196] D. Y. Leung, T.-h. Lam, and S. S. Chan, “Three versions of perceived stress scale: validation in a sample of chinese cardiac patients who smoke,” *BMC public health*, vol. 10, no. 1, p. 513, 2010.

[197] R. A. Bressan, A. C. Chaves, L. S. Pilowsky, I. Shirakawa, and J. J. Mari, “Depressive episodes in stable schizophrenia: critical evaluation of the dsm-iv and icd-10 diagnostic criteria,” *Psychiatry research*, vol. 117, no. 1, pp. 47–56, 2003.

[198] M. Maj, “Bereavement-related depression in the dsm-5 and icd-11,” *World Psychiatry*, vol. 11, no. 1, p. 1, 2012.

[199] P. Bech, “Rating scales in depression: limitations and pitfalls,” *Dialogues in clinical neuroscience*, vol. 8, no. 2, p. 207, 2006.

[200] M. Hamilton, "Development of a rating scale for primary depressive illness," *British journal of social and clinical psychology*, vol. 6, no. 4, pp. 278–296, 1967.

[201] S. A. Montgomery and M. Åsberg, "A new depression scale designed to be sensitive to change," *The British journal of psychiatry*, vol. 134, no. 4, pp. 382–389, 1979.

[202] A. T. Beck, R. A. Steer, and M. G. Carbin, "Psychometric properties of the beck depression inventory: Twenty-five years of evaluation," *Clinical psychology review*, vol. 8, no. 1, pp. 77–100, 1988.

[203] C. J. Hawley, T. M. Gale, P. S. J. Smith, S. Jain, A. Farag, R. Kondan, C. Avent, and J. Graham, "Equations for converting scores between depression scales (mådrs, srs, phq-9 and bdi-ii): good statistical, but weak idiographic, validity," *Human Psychopharmacology: Clinical and Experimental*, vol. 28, no. 6, pp. 544–551, 2013.

[204] N. Titov, B. F. Dear, D. McMillan, T. Anderson, J. Zou, and M. Sunderland, "Psychometric comparison of the phq-9 and bdi-ii for measuring response during treatment of depression," *Cognitive Behaviour Therapy*, vol. 40, no. 2, pp. 126–136, 2011.

[205] A. Dreher, E. Hahn, A. Diefenbacher, M. H. Nguyen, K. Böge, H. Burian, M. Dettling, R. Burian, and T. M. T. Ta, "Cultural differences in symptom representation for depression and somatization measured by the phq between vietnamese and german psychiatric outpatients," *Journal of psychosomatic research*, vol. 102, pp. 71–77, 2017.

[206] C. B. Young, D. Z. Fang, and S. Zisook, "Depression in asian–american and caucasian undergraduate students," *Journal of affective disorders*, vol. 125, no. 1-3, pp. 379–382, 2010.

[207] J. E. Cameron, "A three-factor model of social identity," *Self and identity*, vol. 3, no. 3, pp. 239–262, 2004.

[208] N. Ellemers, P. Kortekaas, and J. W. Ouwerkerk, "Self-categorisation, commitment to the group and group self-esteem as related but distinct aspects of social identity," *European journal of social psychology*, vol. 29, no. 2-3, pp. 371–389, 1999.

[209] A. V. Lovakov, E. R. Agadullina, and E. N. Osin, "A hierarchical (multicomponent) model of in-group identification: Examining in russian samples," *The Spanish Journal of Psychology*, vol. 18, 2015.

[210] S. Roccas, L. Sagiv, S. Schwartz, N. Halevy, and R. Eidelson, "Toward a unifying model of identification with groups: Integrating theoretical perspectives," *Personality and Social Psychology Review*, vol. 12, no. 3, pp. 280–306, 2008.

[211] E. R. Smith, C. R. Seger, and D. M. Mackie, "Can emotions be truly group level? evidence regarding four conceptual criteria." *Journal of personality and social psychology*, vol. 93, no. 3, p. 431, 2007.

[212] S. Reysen, I. Katzarska-Miller, S. M. Nesbit, and L. Pierce, "Further validation of a single-item measure of social identification," *European Journal of Social Psychology*, vol. 43, no. 6, pp. 463–470, 2013.

[213] T. K. Daugherty and E. J. Lane, "A longitudinal study of academic and social predictors of college attrition," *Social behavior and personality*, vol. 27, no. 4, p. 355, 1999.

[214] J. Choi and R. Gutierrez-Osuna, "Using heart rate monitors to detect mental stress," in *2009 Sixth International Workshop on Wearable and Implantable Body Sensor Networks*. IEEE, 2009, pp. 219–223.

[215] A. Sano and R. W. Picard, “Stress recognition using wearable sensors and mobile phones,” in *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction*. IEEE, 2013, pp. 671–676.

[216] N. D. Lane, E. Miluzzo, H. Lu, D. Peebles, T. Choudhury, and A. T. Campbell, “A survey of mobile phone sensing,” *IEEE Communications magazine*, vol. 48, no. 9, 2010.

[217] K. Aberer, S. Sathe, D. Chakraborty, A. Martinoli, G. Barrenetxea, B. Faltings, and L. Thiele, “Opensense: open community driven sensing of environment,” in *Proceedings of the ACM SIGSPATIAL International Workshop on GeoStreaming*. ACM, 2010, pp. 39–42.

[218] C. Staw and C. Nemeth, “The tradeoffs of social control and innovation in groups and organizations,” *Advances in experimental social psychology*, vol. 22, pp. 175–210, 1989.

[219] J. Jetten, T. Postmes, and B. J. McAuliffe, “‘we’re all individuals’: Group norms of individualism and collectivism, levels of identification and identity threat,” *European Journal of Social Psychology*, vol. 32, no. 2, pp. 189–207, 2002.

[220] D. Bar-Tal, *Group beliefs: A conception for analyzing group structure, processes, and behavior*. Springer Science & Business Media, 2012.

[221] T. E. Becker and M. C. Kernan, “Matching commitment to supervisors and organizations to in-role and extra-role performance,” *Human performance*, vol. 16, no. 4, pp. 327–348, 2003.

[222] J. P. Meyer, T. E. Becker, and C. Vandenberghe, “Employee commitment and motivation: a conceptual analysis and integrative model.” *Journal of applied psychology*, vol. 89, no. 6, p. 991, 2004.

[223] B. Yu, Y. Ren, L. Terveen, and H. Zhu, “Predicting member productivity and withdrawal from pre-joining attachments in online production groups,”

in *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*. ACM, 2017, pp. 1775–1784.

[224] B. Vasilescu, D. Posnett, B. Ray, M. G. van den Brand, A. Serebrenik, P. Devanbu, and V. Filkov, "Gender and tenure diversity in github teams," in *Proceedings of the 33rd annual ACM conference on human factors in computing systems*. ACM, 2015, pp. 3789–3798.

[225] A. Filippova and H. Cho, "The effects and antecedents of conflict in free and open source software development," in *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*. ACM, 2016, pp. 705–716.

[226] J. Seering, R. Kraut, and L. Dabbish, "Shaping pro and anti-social behavior on twitch through moderation and example-setting," in *Proceedings of the 2017 ACM conference on computer supported cooperative work and social computing*. ACM, 2017, pp. 111–125.

[227] J.-K. Min, J. Wiese, J. I. Hong, and J. Zimmerman, "Mining smartphone data to classify life-facets of social relationships," in *Proceedings of the 2013 conference on Computer supported cooperative work*. ACM, 2013, pp. 285–294.

[228] H. P. E. D. LP, *Aruba Enterprise Networking and Security Solutions*, 2019, https://www.arubanetworks.com/me/.

[229] CISCO, *Cisco Digital Network Architecture (Cisco DNA)*, 2019, https://www.cisco.com/c/en/us/solutions/enterprise-networks/index.html.

[230] Zebra, *Location Technologies*, 2019, https://www.zebra.com/ap/en/products/location-technologies.html.

[231] U. Networks, *Ubiquiti Networks - Software*, 2019, https://www.ui.com/software/.

[232] E. Mendelson, “Indoor and outdoor mapping and navigation utilizing rf bluetooth beacons,” Dec. 1 2015, uS Patent 9,204,257.

[233] A. P. Association *et al.*, “Stress in america: The state of our nation,” *Stress in AmericaTM Survey*, 2017.

[234] A. S. Index, “A survey conducted by one poll on 4000 uk adults between 28 jul-8 aug 2017,” *Retrieved on*, vol. 5, 2018.

[235] H. Sarker, M. Tyburski, M. M. Rahman, K. Hovsepian, M. Sharmin, D. H. Epstein, K. L. Preston, C. D. Furr-Holden, A. Milam, I. Nahum-Shani *et al.*, “Finding significant stress episodes in a discontinuous time series of rapidly varying mobile sensor data,” in *Proceedings of the 2016 CHI conference on human factors in computing systems*. ACM, 2016, pp. 4489–4501.

[236] C. L. Hammen and S. D. Cochran, “Cognitive correlates of life stress and depression in college students.” *Journal of Abnormal Psychology*, vol. 90, no. 1, p. 23, 1981.

[237] A. Lampinen, S. Tamminen, and A. Oulasvirta, “All my people right here, right now: Management of group co-presence on a social networking site,” in *Proceedings of the ACM 2009 international conference on Supporting group work*. ACM, 2009, pp. 281–290.