[1]

Student Stress Prediction Using Mobile Sensing Data

SIDDHARTH SHUKLA, University of California, San Diego, USA

Students in college are subjected to an immense amount of stress and the contributing factors range from academic courseload to maintaining a healthy social life. While some students find their way of managing stress, others might struggle to cope with challenging and stressful environments. With the growing popularity of wearable devices and the ubiquity of smartphones the ability to utilize physiological data collected from such devices to predict the user's mental state such as mood and stress has become more approachable. In this report, I try to explore the possibility of predicting students' stress conditions based on various remote sensed data along with Ecological Momentary Assessment (EMA) surveys sampled throughout the academic quarter.

The dataset used in this report is from 48 Dartmouth students over 10 weeks term to assess their mental health which is famously known as the StudentLife dataset. Three different periods, namely 12 hours, 24 hours, and 48 hours of aggregation windows were experimented with different classifiers such as *Logistic Regression, Random Forest, Support Vector Machines,* and *Light Gradient Boosting Machines*. The light Gradient Boosting Machine classifier achieved the highest **AUC score** of **0.76** for the 24-hour aggregation window for sensor data combined with Ecological Momentary Assessment (EMA) surveys. The above modeling approach can be implemented through various mobile applications which can predict a student's stress and suggest various intervention methods like meditation etc.

Additional Key Words and Phrases: Mobile Sensing, Stress prediction, Light Gradient Boosting Machine, Machine Learning

1 MOTIVATION

A student's university experience can be stressful as well as fun and exciting. In psychology, stress is a feeling of emotional strain and pressure. One may feel stressed about starting university, exams, getting along with new people, coursework deadlines, or thinking about the future. A little bit of stress is natural and biologically it is designed to help cope with challenging situations. For example, it pushes a student to work harder and do his best during exams or enhances decision-making skills, especially when working against tight deadlines. The problem arises when stress surpasses manageable levels and turns into serious problems like depression or anxiety. It leads to many negative impacts such as finding it harder to focus or get things done. This has also been shown in various national studies where stress leads to other serious emotional health challenges like anxiety, sleep deprivation, and depression thus leading to impediments to academic success. Research also indicates that during overwhelming stress a student not only feels demotivated to complete his work it also leads to other negative coping mechanisms like substance abuse, lashing out at others, and taking sleeping pills and other anti-anxiety pills.

When students are exposed to stressors or things that provoke stress, it leads to an array of physical, emotional, behavioral, and cognitive reactions. This affects different students in different ways. Below are some of the different dimensions into which stress manifests in people:

- `Physical symptoms`: These could be sweating, increased heart rate or blood pressure, vertigo, shortness of breath, muscle tension, headaches, stomachaches, and fatigue
- `Emotional symptoms`: Hostility, irritability, mood changes increased worrying. It could also lead to feelings of helplessness or loneliness.
- `Behavioral symptoms`: Binge or reduced eating, drug or alcohol misuse, decreased sex drive, erratic sleep habits.

[1] code available at: https://github.com/slickFix/Student-Stress-Prediction

Author's address: Siddharth Shukla, sishukla@ucsd.edu, University of California, San Diego, USA.

- `Cognitive Symptoms`: Memory loss, loss of concentration, negative outlook, dissociation such as disconnection from thoughts, feelings, and identity.

Prolonged exposure to a stressful environment has many harmful effects on human health such as cardiovascular disease, and alterations of brain structures involved in cognition [11]. Acute stressors (lasting minutes) are associated with potentially adaptive upregulation of some parameters of natural immunity and downregulation of some functions of specific immunity [18]. Whereas chronic stressors are associated with the suppression of both cellular and humoral measures. Chronic stress can increase inflammation and alter protective immune responses, and thereby may increase susceptibility to certain types of cancer by suppressing type 1 cytokines and protective T cells, and increasing regulatory/suppressor T cell function [6]. Thus unchecked stress can lead to adverse health impacts and figuring out what situations might cause stress is only half the battle for college students.

Stress can take one of three forms:

- `Acute Stress`: This is the most common form of stress and it is a result of day-to-day stressors, such as running to class, receiving a bad grade, and waking up late. The good news is this form of stress fades away quickly and has a little physical or mental impact.
- `Episodic Acute Stress`: This form of stress, as the name suggests develops when a student experiences acute stress multiple times over an extended period. It shows up with certain symptoms like migraines and tension headaches.
- `Chronic Acute Stress`: This is a more serious form than the other two where a student can't avoid a long-term stressful situation. For example, a student struggling to consistently perform academically in a credit-heavy course may develop into chronic acute stress which could also lead to weight gain, sleep deprivation, and anxiety.

If we dig into the statistics of stress, a survey by American College Health Association (2015) found that stress become the most serious academic impediment among students at over a hundred colleges and universities across the US. Nearly 40% of college students in the US admit to feeling inadequately-rested for five out of seven days a week. More than 45% of college students in America claimed to have more than average stress whereas 33% of students reported average stress and 12.7% claimed tremendous stress. Around 80% of university students in the UK reported stress in school or college. The recent coronavirus pandemic has made things worse for students as the closure of campus led to less social life, transitioning to online learning environments and loss of paid jobs in the middle of the semester led to excessive stress. A study by [8] on students in a public research university in Kentucky during an early phase of COVID-19 showed that 88% of students experienced moderate to severe stress, with 44% of students showing moderate to severe anxiety and 36% of students having moderate to severe depression.

Various efforts have been made by researchers at different institutions using varied technologies for detecting stress. Some of the noticeable ones have been using heart rate and rate variability by [24], cortisol level by [7] and skin conductance by [20]. Other techniques may not depend on the sensor and can simply try to extract information from users through self-reporting tools [16] and surveys like the Perceived Stress Scale [4]. The use of passive sensing data from mobile phones and other ubiquitous devices like wearable wristbands has started to find its place in the application of mental health observation.

Since devices like mobile phones and wearable bands have an array of embedded sensors which can passively monitor numerous physical parameters such as heart rate, luminosity, motion, and location they form an excellent source of information to infer different mental states. However, these sensors can not directly capture a user's cognitive context such as mood and well-being. Correlations between sensor data and mental health conditions have been found by many researchers which form the basis of using them to predict the mental states of students.

As there is no direct link between any sensor data with mental health, Machine Learning algorithms can be used to uncover relations and hidden patterns between multiple sensors and mental well-being.

In this report, I try to investigate the potential of using machine learning algorithms to detect stress levels using passively sensed data from mobile phones of 48 college students of Dartmouth college during a 10-week long academic term [25]. The dataset contains rich and multi-dimensional data collected in the form of passively sensed using mobile phones along with Ecological Momentary Assessment surveys throughout the academic term with pre and post-term mental health surveys. Experiments of using the passively sensed mobile data and its combination with Ecological Momentary Assessment for different aggregation time windows such as 12 hours, 24 hours and 48 hours have been performed using Machine Learning algorithms such as Light Gradient Boosting Machines, Logistic Regression, Support Vector Machines, and Random Forest.

2 RELATED WORK

Various studies have been carried out to understand the relationship between mobile phone usage and its use to predict a user's cognitive context such as the mood of a person. As mobile phones are part of the everyday life of billions of people, they provide a suitable tool for psychological experiments unobtrusively. The authors of [15] deployed a system to study participant's emotions using mobile phones and cross-validated their results using questionnaires filled by the users demonstrating the possibility of detecting speakers and participants' emotions using a locally running classifier on off-the-shelf mobile phones. Another paper [13] tried to understand the causal links between emotional states and their interaction with mobile phones. As most of the papers are based on correlation analysis between emotional states and mobile phone interaction, this paper shows the causal impact of user's emotion on different aspects of mobile phone interaction validated on 5,118 mood reports from 28 users over a period of 20 days.

The authors of [19] present a longitudinal in-the-wild study of mood through smartphones by using an Android app to collect participants' self-reported moods and system-triggered experience sampling data which passively measures the physical activity, sociability, and mobility with the help of mobile's sensors which is collected from 18,000 users for about three years. The paper demonstrates the usage of physical and software sensors in mobile phones to automatically and accurately identify routines and demonstrate a strong correlation between these routines to predict a user's mood with an accuracy of 70%. This paper [10] reports a software system that infers the mood of users based on how the mobile phone is used. Most smartphone sensors measure physical properties such as acceleration, light, and others whereas, MoodScore is a sensor that measures the mental state of the user. The results were tested on 32 participants over two months using a formative statistical mood study with smartphone-logged data achieving an initial accuracy of 66% which got improved to 93% after a two-month personalized training period.

One of the other studies exploiting mobile sensing data to understand well-being states is [17] where the authors have established the correlation between mobile phone sensors and depressive symptom severity. The study recruited 40 adult participants for two weeks to explore the detection of the daily-life behavioral marker using phone GPS and usage sensors for identifying depressive symptom severity. This paper used PHQ-9 self-reported depression survey for validation and concluded by suggesting phone sensors' utility for detecting at-risk populations. The paper [12] lists down various sensing modalities that are already explored such as location, call records, SMS, and overall application usage logs to infer the depressive state of users and presents the initial results of multi-modal sensing via smartphones to demonstrate the association of depressive states with smartphone interaction features. Another paper is [3] which uses mobile sensing to predict changes in the depressive mood of an individual by analyzing their movements. This paper seeks to answer the question of only using the mobility patterns of a user from GPS traces by unobtrusively monitoring individuals affected by depressive mood disorders and predicting changes in mood. [22] paper investigates the idea of using data collected from smartphone of users to predict the stress level of a user by conducting a study that collects smartphone data and stress data as measured on the Perceived Stress Scale, seven times a day for two weeks. The paper reports a strong correlation between stress levels and mobile sensor data making them suitable for stress prediction.

Many studies and research papers have attempted to understand the stress of a user by machine learning algorithms and mobile-sensed data. [5] describes a stress detection system based on a combination of two physiological signals: Galvanic Skin Response and Heart Rate. The proposed approach of the authors is unique in terms of using individual stress templates for gathering the behavior of individuals under different degrees of stress than providing a global stress classification approach. The proposed system achieved a 99.5% rate on 80 individuals by using well-known machine learning algorithms like SVM, kNN, and Linear Discriminant Analysis. DeepMood by [23] is another attempt to use a machine learning algorithm to predict severely depressed moods based on self-reported histories. It uses a recurrent neural network algorithm which was trained on 2,382 self-declared depressed people showing performance improvement on additional long-term historical information and performing better than the SVM algorithm. citesiddique2021machine is a paper where the authors have tried to use the google forms dataset that is filled by students, to predict the stress levels of student duet o online education by using machine learning algorithm and have reported an accuracy of 73.91% using random forest algorithm. Authors of [9] used deep neural networks to analyze physiological data collected from chest-worn and wrist-worn sensors to first perform a binary classification of stress and then classify emotion using multi-class classification.

Some of the previous works on the StudentLife dataset [25] involve predicting students' GPAs based on passively sensed data [26]. The authors proposed new methods for better understanding study duration and social behavior such as partying of a group of undergraduate students and used a simple model based on linear regression with lasso regularization that could accurately predict cumulative GPA. [14] investigates the effectiveness of neural network models for predicting user's stress levels by using location information collected from smartphones that are present in the student life dataset. The authors of [21] present a general platform for personalized predictive modeling of behavioral states like students' level of stress using Auto encoders and Multitask learning and achieving an improvement of 45.6% in F1 score on the student life dataset. Paper by [1] demonstrates the usage of LSTM, CNN, and CNN-LSTM algorithms, which accept sequence data as input from the student life dataset for predicting mental stress.

3 PROJECT AIM

The aim of the report involves testing the feasibility of predicting students' stress using passively sensed mobile data by exploring different machine learning algorithms. In addition to this, augment the passively sensed data with additional interaction data and report the improvement in the performance of metrics. The first step in this process involves identifying a dataset that presents the opportunity of predicting stress by providing ground truth information as well as passively sensed mobile data records. The student life dataset which is collected from 48 Dartmouth students over 10 weeks term to assess their mental health contains mobile sensed data capturing various information such as location, activity, etc as well as interaction data in the form of Ecological Momentary Assessment surveys. Post the dataset finalization step, the data exploration step begins which involves gaining familiarization with the data structure and its relation to the real world. This step also tries to explore various interdependencies between different forms of data and identifies the presence of ground truth if any. Once these steps are complete then the data modeling exercise of predicting the stress using various predictors and experimenting with different setups can start. Predicting students' stress would help in ideating various stress mitigation steps that can be implemented in mobile phones such as pop-ups reminding them to meditate or nudging them for evening walks to lower their stress levels.

Below is a bullet point representation of the complete involved process:

- `Data identification`: The process of selecting the dataset had criteria of the presence of passively sensed mobile data along with interaction data with reliable ground truths. Since collecting this form of the dataset would be challenging, a deidentified open-source dataset was used.
- `Data exploration`: After the finalization of a dataset, gaining familiarization with the dataset by exploring forms the next crucial step. It involves understanding the context of the data collection its limitation and its relevance to the problem at hand. Usually, the dataset is present in raw form and in different modalities which need to be brought at the same level for modeling purposes.
- `Data cleaning`: the process of data cleaning involves preparing the data according to the limitations of the model such as removing outliers and scaling the features.
- `Data modeling`: the prepared data is fed into different machine learning algorithms for training and its performance on the validation set is noted for further tuning of hyperparameters. A separate test set is maintained to evaluate the model's performance for new data.
- `Feature experimentation`: This step involves improving the model performance by crafting new features or using different data setups.

4 METHODOLOGY

This section deals with the informed explanation of the dataset, its finding and all the steps that were taken to perform the modeling steps and reporting of final metrics.

4.1 Dataset

In this section, a detailed description of the dataset is provided. Studentlife dataset is one of a kind study that uses passive and automatic sensing data from the phone of a class of 48 Dartmouth students over a 10-week term to assess their mental health such as depression, loneliness, stress, and academic performance in terms of grades across all their classes, term GPA and cumulative GPA and behavioral trends like degree of stress, sleep hours, visits to the gym, change in response to college workload concerning the assignment, midterms, and finals as their term progress.

The data collection used computational methods and machine learning algorithms on the phone to assess sensor data to make higher-level inferences such as sleep, sociability, activity, etc. The data is collected from 48 undergrads and grad students at Dartmouth over 10 weeks of spring term containing continuous data and 320000 self-reports and pre-post surveys. The StudentLife application that ran on students' phones automatically measure the following human behaviors 24/7 without any interference:

- `bed time, wake up time and sleep duration`
- `the number of conversations and duration of each conversation per day`
- `physical activity (walking, sitting, running, standing)`
- `where they were located and how long they stayed there (i.e., dorm, class, party, gym)`
- `in-situ comments on campus and national events: dimension protest, cancelled classes; Boston bombing.`
- `the number of people around a student through the day`
- `outdoor and indoor (in campus buildings) mobility`
- `stress level through the day, across the week and term`
- `positive affect (how good they felt about themselves)`
- `eating habits (where and when they ate)`
- `app usage`

Along with this, 10 automatic sensing data corresponding to physical activity, audio inferences, conversation inferences, BlueTooth scan, light sensor, GPS, phone charge, phone lock, WiFi, and WiFi location were present in 10 different folders in a CSV file for each user. All the questions asked in Ecological Momentary Assessment surveys (EMA) were present in EMA_definition.json as a data dictionary and the EMA data for each group such as Activity, Behavior, Cancelled Classes, exercise, mood, stress, etc was present in separate folders with each user response information stored in JSON file. Among all these survey data, 9 EMA data out of which 7 were used as predictors and 2 were used as ground truth labels and among the 10 remote sensing data, 7 were used for modeling purposes. Below are the list of 7 remote sensing features and 7 EMA features used in machine learning algorithms:

- **Remote Sensing**: App Usage, **EMA**: Social - number of social interactions
- **Remote Sensing**: Activity, **EMA**: Class - class experience and assignment due
- **Remote Sensing**: Audio, **EMA**: Sleep - hours slept and sleep quality rating
- **Remote Sensing**: Conversation, **EMA**: Activity - working percentage, relaxing percentage
- **Remote Sensing**: Time spent in darkness, **EMA**: Exercise - exercise duration, done or not
- **Remote Sensing**: Phone charge duration, **EMA**: Mood - happy or not
- **Remote Sensing**: Phone lock duration, **EMA**: Class 2 - class is challenging or not, expected grades

Two EMA data used as ground truth are :

- **Stress EMA** (2132 data points) – Option 1 to 3 (stressed) and Option 4,5 (feeling normal)
- **Mood 2 EMA** (414 data points) – Option 1 (happy), Option 2 (tired) and Option 2 (stressed)

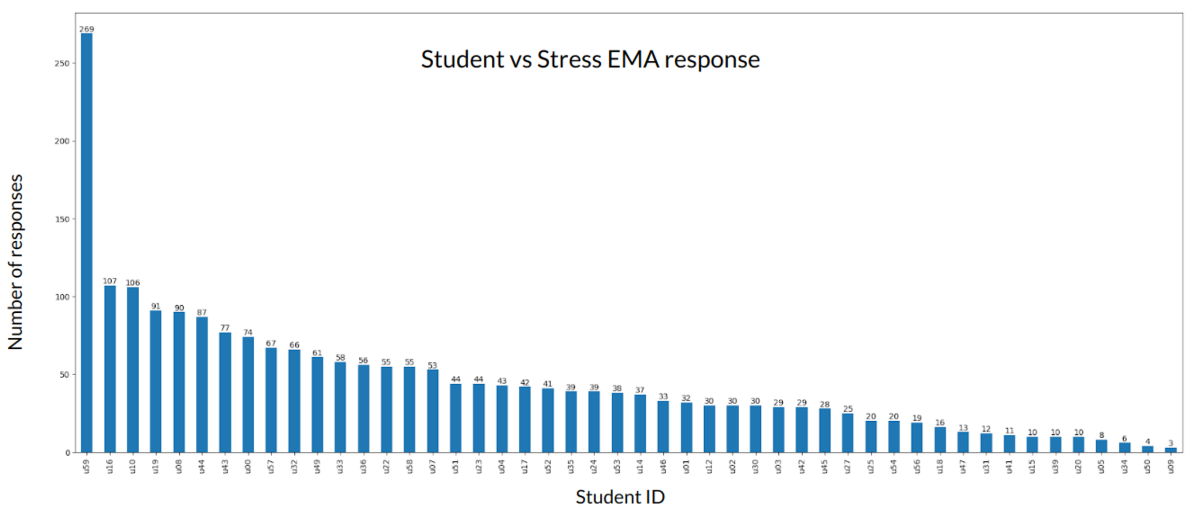

Fig. 1. Number of Stress EMA response by each student

4.2 Data cleaning

Stress EMA and Mood 2 EMA data were present in JSON files for each user, so the first step was to combine the data from both folders into one file in a format that would data visualization easier. A custom function for reading the JSON file for each user and converting it into a pandas data frame was written for easier access and plotting of data points. Duplicate records that were present were dropped from the data frame by keeping only the latest timestamp record of the EMA survey response. This might have occurred due to failure to capture the response in the central database or multiple entries by the user. After removing the duplicated records, all the EMA responses were checked for being the academic calendar period as for a few users some of the records

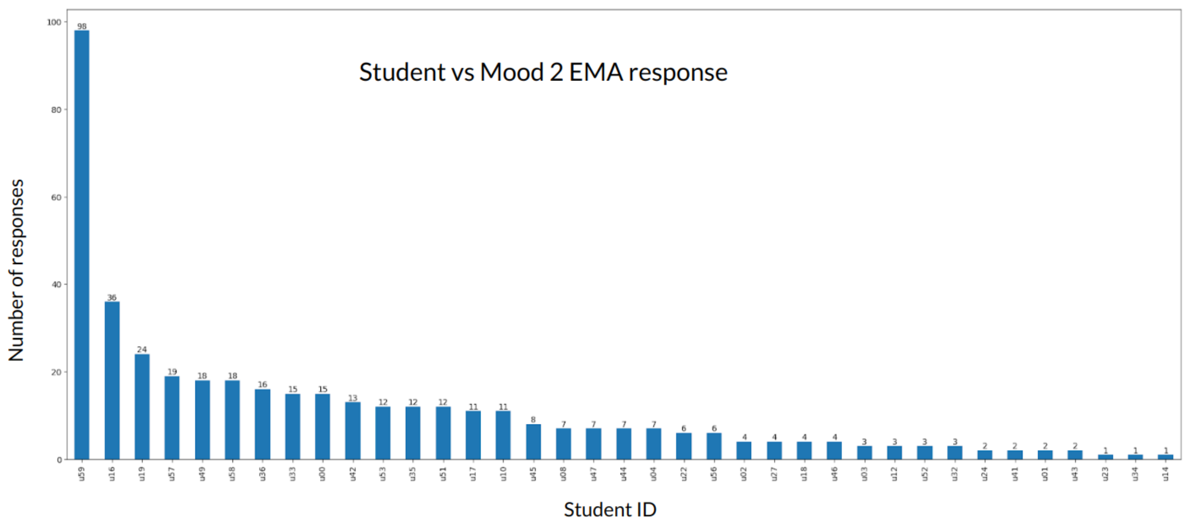

Fig. 2. Number of Mood 2 EMA response by each student

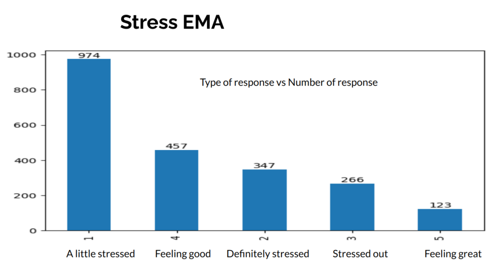

Fig. 3. Number of response for each Stress EMA options

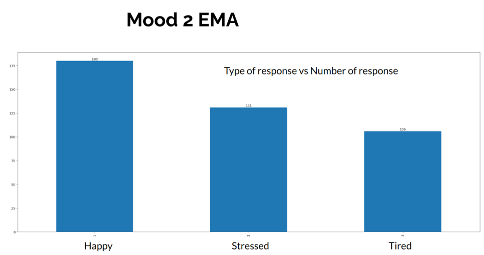

Fig. 4. Number of response for each Mood 2 EMA options

were present outside this period. After completion of this step, the data was ready for combining with predictors which were extracted by different custom functions.

4.3 Data Aggregation

After the records for ground truths and predictor variables are extracted and cleaned, combining the two into one data frame is the next step so that the data is ready for the modeling exercise. Since the ground truth data that is our Stress EMA and Mood 2 EMA responses are at different timestamps than our independent variables a custom logic for aggregating independent variables was written. This function would pick one record from ground truth and based on user input group independent features by summing or averaging based on the record type for the input period. For example, if the dependent variable which is our ground truth is a record at 3'O clock in the afternoon and the aggregation period is 1 day then it would look at the predictor variable records which are present between yesterday's 3'O clock and today's 3'O clock and if the predictor variable was the number of social interaction then they all would be summed up to form one record for model's input. For the experimentation purpose, 12 hours, 24 hours, and 48 hours of aggregation window were used.

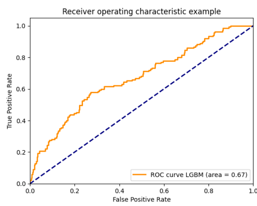

Fig. 5. LGBM performance on sensing data (period = 12 hrs)

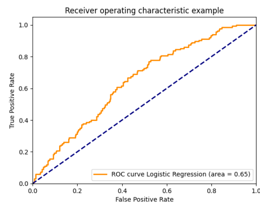

Fig. 6. Log reg performance on sensing data (period = 12 hrs)

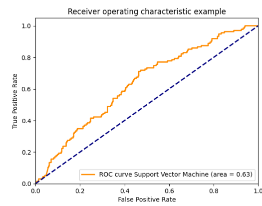

Fig. 7. SVM performance on sensing data (period = 12 hrs)

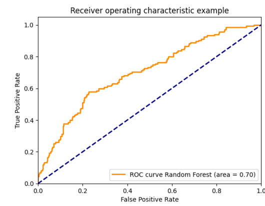

Fig. 8. Random Forest performance on sensing data (period = 12 hrs)

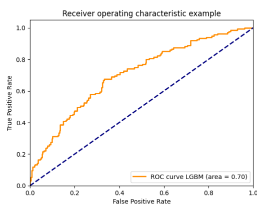

Fig. 9. LGBM performance on sensing and EMA data (period = 12 hrs)

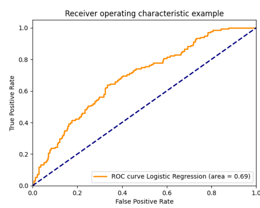

Fig. 10. Log Reg performance on sensing and EMA data (period = 12 hrs)

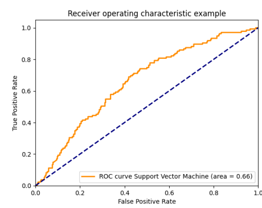

Fig. 11. SVM performance on sensing and EMA data (period = 12 hrs)

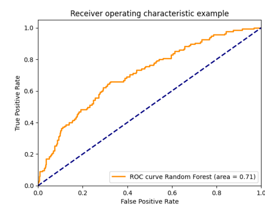

Fig. 12. Random forest performance on sensing and EMA data (period = 12 hrs)

4.4 Modeling

This is the final and most important step of the process. The aggregated data from the previous step is spitted into three sets namely train, validation, and test set. A total of 2454 records were present after cleaning and 60% of the total that is 1472 records were used for training different models and 20% of the total was used for hyperparameter tuning on the validation set and 20% for reporting mode's performance on the test set. After the data split, train data was label encoded for categorical columns and all the features were standardized using standard scalar for logistic regression and Support Vector Machines for faster convergence. For tuning the hyperparameters optuna [2] framework was used which is a hyperparameter optimization framework. The above sets of data records were created for two settings:

- **First**: Only mobile sensing data is used as independent variable
- **Second**: Combination of mobile sensing data and EMA survey data is used as independent variable

For both of the above settings and different aggregation periods 4 models were trained:

- `Light Gradient Boosting Machine classifier`
- `Logistic Regression`
- `Support Vector Machine classifier`
- `Random Forest classifier`

5 RESULTS

The dependent variable is a binary flag suggesting if a particular user is stressed or not. So to measure the performance of each classifier different classification metrics such as accuracy, precision, recall, and F1 score were obtained. For selecting the best classifier AUC metric was used as captures the overall performance of

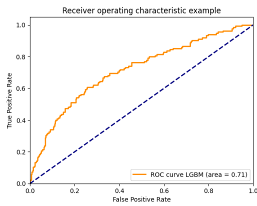

Fig. 13. LGBM performance on sensing data (period = 24 hrs)

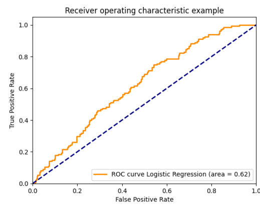

Fig. 14. Log reg performance on sensing data (period = 24 hrs)

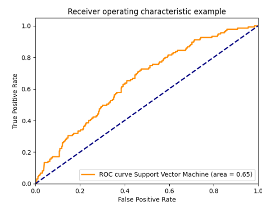

Fig. 15. SVM performance on sensing data (period = 24 hrs)

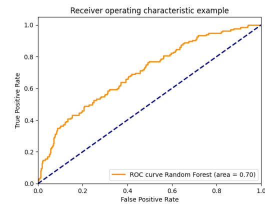

Fig. 16. Random forest performance on sensing data (period = 24 hrs)

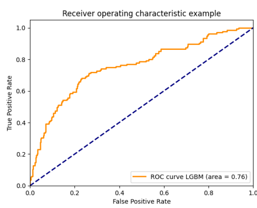

Fig. 17. LGBM performance on sensing and EMA data (period = 24 hrs)

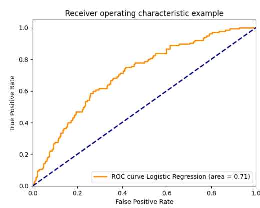

Fig. 18. Log Reg performance on sensing and EMA data (period = 24 hrs)

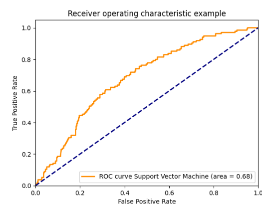

Fig. 19. SVM performance on sensing and EMA data (period = 24 hrs)

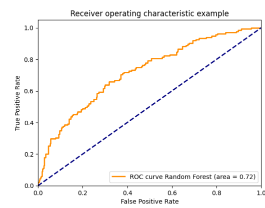

Fig. 20. Random forest performance on sensing and EMA data (period = 24 hrs)

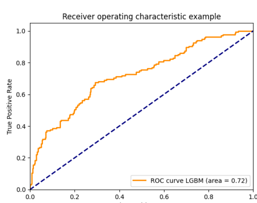

Fig. 21. LGBM performance on sensing data (period = 48 hrs)

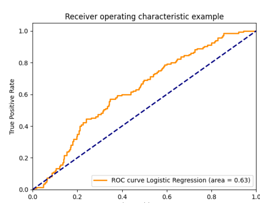

Fig. 22. Log reg performance on sensing data (period = 48 hrs)

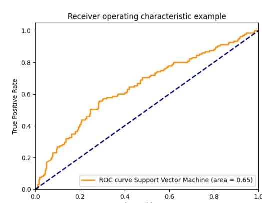

Fig. 23. SVM performance on sensing data (period = 48 hrs)

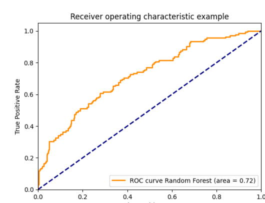

Fig. 24. Random forest performance on sensing data (period = 48 hrs)

the classifiers. Plots of different AUC metrics of each of the classifiers under different settings and aggregation periods are provided in 5 6 7 8 9 10 11 12 13 14 15 17 18 19 20 21 22 23 24 25 26 27 28

6 DISCUSSION

Table 1 and 2 present the consolidated view of all the experimentations performed for different data setup and aggregation time frames. The light Gradient Boosting Machine (LGBM) algorithm performs the best with an AUC of 0.76 using sensing and EMA data for a 24-hour aggregation window. A general trend for improvement in all model performance is seen with using additional EMA survey data over only sensing data which makes sense as EMA survey data has some of the mobile interaction features that are usually strong predictors of stress such as quality of sleep, mood, number of social interactions, etc. For most of the models, there is an improvement in score with an additional aggregation period except for logistic regression in the case of using only sensing data.

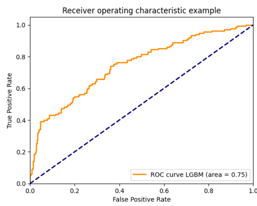

Fig. 25. LGBM performance on sensing and EMA data (period = 48 hrs)

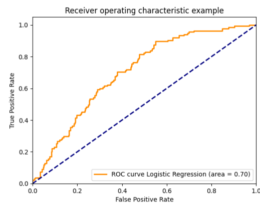

Fig. 26. Log Reg performance on sensing and EMA data (period = 48 hrs)

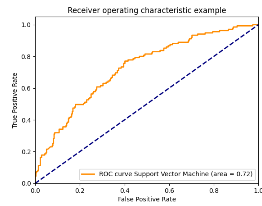

Fig. 27. SVM performance on sensing and EMA data (period = 48 hrs)

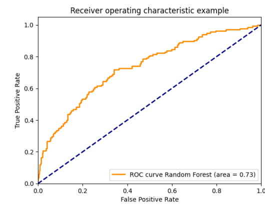

Fig. 28. Random forest performance on sensing and EMA data (period = 48 hrs)

Table 1. AUC metric of different models using only **sensing data**

Model	12 Hour	24 Hour	48 Hour
LGBM	0.67	0.71	0.72
Logistic Regression	0.65	0.72	0.63
Support Vector Machines	0.63	0.65	0.65
Random Forest	0.70	0.70	0.72

Table 2. AUC metric of different models using only **sensing** and **EMA data**

Model	12 Hour	24 Hour	48 Hour
LGBM	0.70	**0.76**	0.75
Logistic Regression	0.69	0.71	0.70
Support Vector Machines	0.66	0.68	0.72
Random Forest	0.71	0.72	0.73

In general additional data due to extra time might be useful for complex nonlinear models but might contain noise and confuse linear models such as logistic regression.

Additional features from other EMA surveys like the lab, and study space information may help all the models to improve the score further. In addition to this, manual handcrafted features like time left for the next assignment deadline, if the day of stress prediction is a holiday or not, and time spent on social media etc. may lead to further boosting of the score but due to time and data limitations could not be explored. Another promising method for boosting the score could involve model personalization for each user as all the presented models are trained for the population than the individuals.

7 CONCLUSION

A Student's experience at a university could have a range of stressors such as doing well in academics to maintaining a healthy social life. Stress in manageable forms is useful as it keeps the student motivated to work hard but becomes a problem in case of chronic stress which has many harmful effects on the body ranging from cardiovascular diseases to changing cognitive structure of the brain leading to memory loss etc. Prolonged stress could also lead to a negative coping mechanism like taking anti-anxiety pills or substance abuse. Many statistics show that most college students are stressed and any form of intervention would be helpful to alleviate stress levels.

The ubiquity of mobile sensors presents us the opportunity to not only passively record various facets of daily life such as the number of human interactions, and time spent talking to others in real life vs on social media platforms in addition to smartphone interaction data such as a popup requesting to report the happiness level but also provide nudging for various behaviors like meditation or going for an evening walk after a prediction of stress state to reduce stress levels.

In this report, machine learning algorithms such as Light Gradient Boosting Machines (LGBM), Random forest, and others are applied to StudentLife dataset to show the accuracy of prediction for Stress state by modeling the problem through a binary variable which is captured in the form of Stress EMA survey. Different data setup such as using only sensing data and a combination of sensing and EMA data with various aggregation windows has been explored. The Light Gradient Boosting machines' performance of 0.76 AUC on the combination of sensing and EMA survey data for a 24-hour aggregation window demonstrates the feasibility of accurately predicting stress and taking a step further for nudging behaviors to effectively manage stress.

REFERENCES

[1] Yasin Acikmese and S Emre Alptekin. 2019. Prediction of stress levels with LSTM and passive mobile sensors. *Procedia Computer Science* 159 (2019), 658–667.

[2] Takuya Akiba, Shotaro Sano, Toshihiko Yanase, Takeru Ohta, and Masanori Koyama. 2019. Optuna: A Next-generation Hyperparameter Optimization Framework. In *Proceedings of the 25rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*.

[3] Luca Canzian and Mirco Musolesi. 2015. 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*. 1293–1304.

[4] Sheldon Cohen, Tom Kamarck, and Robin Mermelstein. 1983. A global measure of perceived stress. *Journal of health and social behavior* (1983), 385–396.

[5] Alberto de Santos Sierra, Carmen Sánchez Ávila, Gonzalo Bailador Del Pozo, and Javier Guerra Casanova. 2011. Stress detection by means of stress physiological template. In *2011 third world congress on nature and biologically inspired computing*. IEEE, 131–136.

[6] Firdaus S Dhabhar. 2014. Effects of stress on immune function: the good, the bad, and the beautiful. *Immunologic research* 58, 2 (2014), 193–210.

[7] Sally S Dickerson and Margaret E Kemeny. 2004. Acute stressors and cortisol responses: a theoretical integration and synthesis of laboratory research. *Psychological bulletin* 130, 3 (2004), 355.

[8] Jungmin Lee, Hyun Ju Jeong, and Sujin Kim. 2021. Stress, anxiety, and depression among undergraduate students during the COVID-19 pandemic and their use of mental health services. *Innovative higher education* 46, 5 (2021), 519–538.

[9] Russell Li and Zhandong Liu. 2020. Stress detection using deep neural networks. *BMC Medical Informatics and Decision Making* 20, 11 (2020), 1–10.

[10] Robert LiKamWa, Yunxin Liu, Nicholas D Lane, and Lin Zhong. 2013. Moodscope: Building a mood sensor from smartphone usage patterns. In *Proceeding of the 11th annual international conference on Mobile systems, applications, and services*. 389–402.

[11] Sonia J Lupien, Bruce S McEwen, Megan R Gunnar, and Christine Heim. 2009. Effects of stress throughout the lifespan on the brain, behaviour and cognition. *Nature reviews neuroscience* 10, 6 (2009), 434–445.

[12] Abhinav Mehrotra, Robert Hendley, and Mirco Musolesi. 2016. Towards multi-modal anticipatory monitoring of depressive states through the analysis of human-smartphone interaction. In *Proceedings of the 2016 ACM international joint conference on pervasive and ubiquitous computing: adjunct*. 1132–1138.

[13] Abhinav Mehrotra, Fani Tsapeli, Robert Hendley, and Mirco Musolesi. 2017. MyTraces: Investigating correlation and causation between users' emotional states and mobile phone interaction. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 3 (2017), 1–21.

[14] Gatis Mikelsons, Matthew Smith, Abhinav Mehrotra, and Mirco Musolesi. 2017. Towards deep learning models for psychological state prediction using smartphone data: Challenges and opportunities. *arXiv preprint arXiv:1711.06350* (2017).

[15] Kiran K Rachuri, Mirco Musolesi, Cecilia Mascolo, Peter J Rentfrow, Chris Longworth, and Andrius Aucinas. 2010. EmotionSense: a mobile phones based adaptive platform for experimental social psychology research. In *Proceedings of the 12th ACM international conference on Ubiquitous computing*. 281–290.

[16] Tauhidur Rahman, Mi Zhang, Stephen Voida, and Tanzeem Choudhury. 2014. Towards accurate non-intrusive recollection of stress levels using mobile sensing and contextual recall. In *Proceedings of the 8th International Conference on Pervasive Computing Technologies*

for Healthcare. 166–169.

[17] Sohrab Saeb, Mi Zhang, Christopher J Karr, Stephen M Schueller, Marya E Corden, Konrad P Kording, David C Mohr, et al. 2015. Mobile phone sensor correlates of depressive symptom severity in daily-life behavior: an exploratory study. *Journal of medical Internet research* 17, 7 (2015), e4273.

[18] Suzanne C Segerstrom and Gregory E Miller. 2004. Psychological stress and the human immune system: a meta-analytic study of 30 years of inquiry. *Psychological bulletin* 130, 4 (2004), 601.

[19] Sandra Servia-Rodríguez, Kiran K Rachuri, Cecilia Mascolo, Peter J Rentfrow, Neal Lathia, and Gillian M Sandstrom. 2017. Mobile sensing at the service of mental well-being: a large-scale longitudinal study. In *Proceedings of the 26th international conference on world wide web*. 103–112.

[20] Cornelia Setz, Bert Arnrich, Johannes Schumm, Roberto La Marca, Gerhard Tröster, and Ulrike Ehlert. 2009. Discriminating stress from cognitive load using a wearable EDA device. *IEEE Transactions on information technology in biomedicine* 14, 2 (2009), 410–417.

[21] Abhinav Shaw, Natcha Simsiri, Iman Deznaby, Madalina Fiterau, and Tauhidur Rahaman. 2019. Personalized student stress prediction with deep multitask network. *arXiv preprint arXiv:1906.11356* (2019).

[22] Thomas Stütz, Thomas Kowar, Michael Kager, Martin Tiefengrabner, Markus Stuppner, Jens Blechert, Frank H Wilhelm, and Simon Ginzinger. 2015. Smartphone based stress prediction. In *International conference on user modeling, adaptation, and personalization*. Springer, 240–251.

[23] Yoshihiko Suhara, Yinzhan Xu, and Alex'Sandy' Pentland. 2017. Deepmood: Forecasting depressed mood based on self-reported histories via recurrent neural networks. In *Proceedings of the 26th International Conference on World Wide Web*. 715–724.

[24] Tanja GM Vrijkotte, Lorenz JP Van Doornen, and Eco JC De Geus. 2000. Effects of work stress on ambulatory blood pressure, heart rate, and heart rate variability. *Hypertension* 35, 4 (2000), 880–886.

[25] Rui Wang, Fanglin Chen, Zhenyu Chen, Tianxing Li, Gabriella Harari, Stefanie Tignor, Xia Zhou, Dror Ben-Zeev, and Andrew T Campbell. 2014. 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*. 3–14.

[26] Rui Wang, Gabriella Harari, Peilin Hao, Xia Zhou, and Andrew T Campbell. 2015. SmartGPA: how smartphones can assess and predict academic performance of college students. In *Proceedings of the 2015 ACM international joint conference on pervasive and ubiquitous computing*. 295–306.