Smart Health Tracker

Jeongin Yang, Hyunjoon Jeon, Anirudhha Hazarika, Jeevaha Coelho

MF Healthcare Technology Consulting

Imperial College London

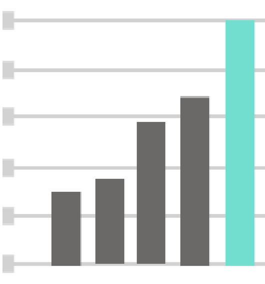
Overview

WHAT IS *SMART HEALTH TRACKER* ABOUT?

- This project makes use of the *"Multilevel Monitoring of Activity and Sleep in Healthy People"* dataset.
- Our smart health tracker aims to exploit potential correlations between physiological data and physiological markers to develop a basis for machine learning algorithms, in order to predict mental wellness markers based on physical data.
- *MMASH* dataset has been preprocessed using statistical analysis techniques to extract important features and streamline the learning process of our machine learning models. Each model is then applied to data provided by the user to predict their mental wellness and provide recommendations via an online web application stored in the cloud (AWS).

RATIONALE

- Given the rise of global stress levels due to COVID-19, our project aims to allow users to better understand how their daily routines and habits affect their mental well-being and highlight areas of improvement help manage their stress.



Global stress levels

Future Work

ADOPTING WEARABLE BIOSENSORS

- Wearable devices like *Fitbit Charge 2* can be used to measure accelerometer values, and *Polar H7 Bluetooth 4.0 Heart Rate Sensor* can be used to measure R-R intervals. The data collected can be exported to our web application, where our machine learning model will detect stress level based on R-R intervals, with the aid of our Activity Tracking model.
- Prevalent use of wearable sensors today can easily open doors for cost-effective adaptation of wearable devices to conveniently collect data needed for our Smart Health Tracker.

IMPROVING ACCURACY FOR GREATER DATASET

- Greater dataset will be available for training, with more data collected by the users via wearable devices which will be stored on our web application database.
- Deep learning
  - Due to limited training data size, traditional algorithms have been adopted. With greater dataset, adopting LSTM architecture could be considered to maintain or improve accuracy further.

SLEEP DISTURBANCE DETECTION

- Various other studies have found a meaningful correlation between Heart Rate Variability and Sleep quality.
- With various HRV features and Pittsburgh scale, which is an indicator for sleep quality, our machine learning algorithm can be extended further to detect one's sleep quality.

Methodology

Data Analysis & Preprocessing → Activity and Stress Tracking with Machine Learning → Access via AWS web application

Data Analysis

ABOUT THE *MMASH* DATASET

- *MMASH* dataset contains physiological data, such as beat-to-beat heart data, accelerometer data, sleep quality and physical activity, as well as the physiological markers such as anxiety, stress and emotional events.

ACTIVITY ANALYSIS

- The activities of each user have been categorised to 3 categories – *mental stressor*, *physical stressor*, and *rest*.
- They have been categorised based on the effects each activity poses on one's Central Nervous System, and the hormones they produce.
- This helps us distinguish and identify different types of stress. This categorisation has been used to improve accuracy of our stress detection model.

Activity.csv →
- Mental Stress: 8) Small screen usage, 9) Large screen usage 10) Caffeine consumption, 11) Smoking 12) Alcohol consumption
- Physical stress: 4) Light, 5) Medium, 6) Heavy movements
- Rest: 1) Sleeping, 2) Laying down 3) Sitting, 7) Eating

ACTIGRAPH ANALYSIS

- Accelerometer, heart rate, steps and inclinometer readings were taken every second for 2 days for each participant.
- By matching timestamps between the *Actigraph* data and the Activity questionnaire we can label each datapoint with an activity, providing a basis for a supervised learning model.

HEART RATE VARIABILITY (HRV) ANALYSIS

- Raw RR beat-to-beat data has been cleaned by removing outliers, NaN values, and ectopic beats. Then it was interpolated to produce Normal beat-to-beat data (NN intervals), ready for HRV feature extraction.
- Different HRV features have been extracted – time domain and frequency domain features.

Heart Rate Variability (HRV) calculation

Raw R-R intervals → Preprocessing (- Removing outliers - Removing NaN values - Interpolation) → N-N intervals (normal-to-normal beats) →
- Time Domain Features: MeanNN, SDNN, RMSSD, SDSD, pnn50, pnn20
- Frequency Domain Features: Low Frequency (LF), High Frequency(HF), LF/HF ratio

- Short-term HRV values for 5-minute time frame have been calculated for each subject. Then, mean HRV values have been calculated and have been compared between rest condition and mentally stressful condition. Differences between HRV values have been observed between these two conditions.
- *Time-domain feature analysis* : It has been found that a greater stress level is related to a lower pNN50 (a time-domain HRV feature).
- *Frequency-domain feature analysis* : Studies have found that LF is related to sympathetic (stressful) activities, while HF is related to parasympathetic (restoring, relaxing) activities. A correlation between high LF/HF ratio and high stress level has been observed.

PSYCHOLOGICAL PARAMETER ANALYSIS

- 2 indicators for mental wellness, *State-Trait Anxiety Inventory (STAI)* and *Daily Stress* have been selected and analysed for exploration.
- Thresholds have been calculated to distinguish the *STAI* values and *Daily Stress* values into different classes – low, and high levels.
- *STAI* index has been selected for stress detector, as it showed greater correlation with the HRV features than the *Daily Stress* values.

Machine Learning

ACTIVITY CLASSIFIER

- *Aim* : Classify Activities to 2 different classes – rest and physically stressful activity.
- *Input* : Actigraph values
- *Output* : Activity class

Input: Actigraph values (x, y, z movements, vector magnitude, heart rate, number of steps per second, inclinometer values) → Random forest Classifier → Output (Rest, Physically stressful activity)

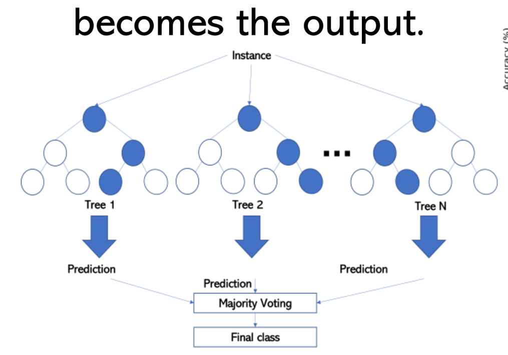
RESULTS

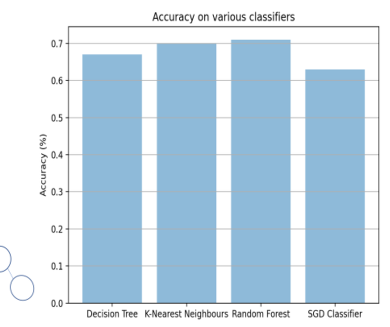
- *Final accuracy* : 71%
- Classification Report :

	Precision	Recall	F1-score	Count
Physical Stress	0.69	0.77	0.73	33960
Rest	0.74	0.66	0.70	33671

RANDOM FOREST CLASSIFIER

- Consists of multiple decision trees.
- Each independently trained decision tree makes a prediction of an output class for the given input.
- The output class which receives the most votes from the trees becomes the output.
- Random Forest Classifier has been selected as it showed greatest accuracy out of different classifiers tested.





DISCUSSION

- Out of 12 output activities, 5 activities which have been categorized as 'mentally stressful' have been removed. Remaining 7 activities have been categorized to two categories – physically stressful activities and rest. With these adaptations, the accuracy of the classifier has increased significantly, from *40%* to *70%*.
- Due to the imbalance in the dataset, undersampling was done to make sure the dataset is balanced for training.

CONCLUSION

- This activity tracker will get rid of the inconvenience that users may experience while filling in forms manually to indicate the times they had done physical activities.

STRESS DETECTOR

- *Aim* : Predict *STAI* index between the two classes – low and high levels.
- *Input* : Time-domain and Frequency-domain HRV features
- *Output* : Anxiety Level (Low, High)

Input: HRV features (*Time-domain features:* meanNN, RMSSD, SDSD, SDNN, pnn50, pnn20; *Frequency-domain features:* LF, HF, LF/HF) → SVM Classifier → Output (Low, High levels of STAI)

SVM CLASSIFIER

- Due to small dataset size, a low-complexity algorithm should be adopted in order to avoid overfitting.
- Linear SVM classifier has been selected to avoid overfitting and maintain high accuracy.

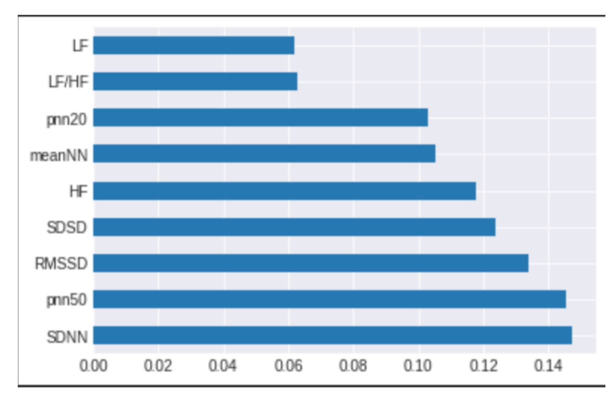
RESULTS

- *Final accuracy* : 88%
- Classification Report :

	Precision	Recall	F1-score	Count
Low	0.88	1.00	0.93	7
High	1.00	0.75	0.86	4

DISCUSSION

- With *Activity Classifier* to identify physical activities, mental stress could be differentiated from stress caused by physical movements, so that the changes in HRV features are only caused by mental stress. HRV values at times when physical stress was present were filtered.
- Due to the imbalance in the dataset, sampling had to be done. Since the dataset is small, oversampling was done to achieve balanced dataset for training.
- Feature Importance map has been produced for feature selection, so that important features can be selected for the simplification of the model, while maintaining its accuracy.

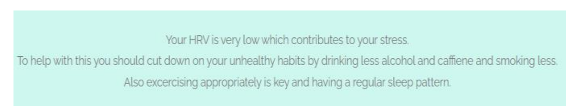


AWS Web Application

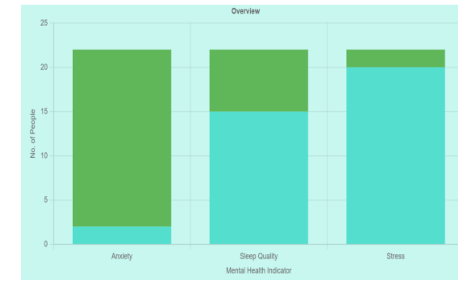
***SMART HEALTH TRACKER* WEB APPLICATION**

STRESS METER

- A stress meter predicts the anxiety/ stress levels based on STAI indicator using our Stress Detecting Model, after collecting physiological data provided.
- 2 different levels detected by the stress detector – low, and high will be indicated on the meter.



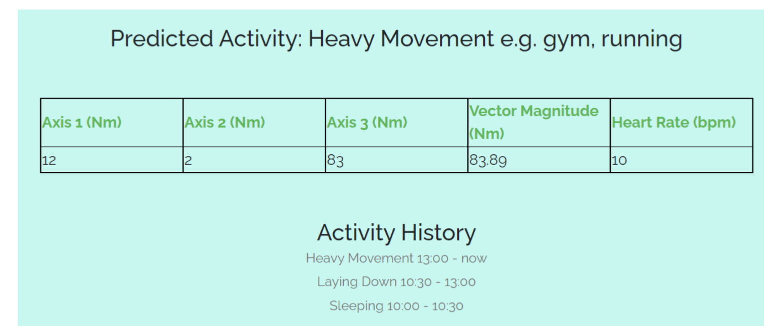
- If stress levels are high, possible interventions are recommended to improve stress levels.



DASHBOARD

- Data of 22 subjects are given and correlations between data points are exploited to show how stress and anxiety can be predicted.
- Individual subjects are compared across all subjects.
- This can help each person to analyse their health status better.

ACTIVITY TRACKING



- Data from the *Activity Tracker* is sent to the website, and this is displayed clearly to the user. Activity history for the day is also provided.

CLIENT-SERVER COMMUNICATION

- Data from the wearable device can be sent via MQTT to the database stored on the server.
- The website runs data analysis and Machine Learning models on the data and displays the outputs on the website.

DATABASE

- An SQL relational database is stored on the AWS server and can be frequently updated with data from the wearable devices. Each relation contains information about each user and their activity, and the website requests data needed from the database, as necessary.