SMARTPHONE TBI SENSING USING DEEP EMBEDDED CLUSTERING AND EXTREME BOOSTED OUTLIER DETECTION

By

Srinarayan Srikanthan

A Thesis

Submitted to the Faculty

of the

WORCESTER POLYTECHNIC INSTITUTE

in partial fulfillment of the requirements for the

Degree of Master of Science

in

Computer Science

May 4, 2021

Dr. Emmanuel O.Agu , Thesis Advisor

_____

Dr. Dimitry Korkin , Thesis Reader

_____

Dr. Craig E.Wills , Department Head

_____

ABSTRACT

Traumatic Brain Injury (TBI) can have long-lasting effects and possibly cause life-long disability of patients, creating a huge economic and social burden. Many TBI patients do not get early and adequate medical care. Sensor-rich, ubiquitously owned smartphones can now be used to passively sense a wide range of ailments, facilitating continuous monitoring of patients and high-risk groups in the real world. In this thesis, we propose a Deep learning approach for distinguishing smartphone users with TBI from health controls within 24-hours of the injury. Our method analyzes smartphone sensor data by first utilizing Deep embedded clustering to identify user clusters with similar smartphone sensed behaviors. Extreme Gradient based outlier detection is then employed on each of the identified clusters to predict users with TBI. In rigorous evaluation, our method achieves a balanced accuracy of 88% and a sensitivity of 74%. Our proposed method can flag smartphone users with TBI, enabling them to receive early medical attention and improve their prognostic outlook.

ACKNOWLEDGMENTS

I would like to express my heartfelt gratitude to Professor Emmanuel O.Agu, my thesis advisor, for his patient guidance, enthusiastic encouragement, useful critiques and providing a clear roadmap for the smooth progress of this research work. I would also like to thank Professor Dimitry Korkin, my thesis reader for his invaluable and timely feedback. I wish to extend my gratitude to Levi Schoen and the Charles River Analytics team for providing much needed support whenever I ran into any data related issues. Finally, I wish to thank Bhoomi Kalpesh Patel and Florina Asani for their technical support and my parents and friends for their encouragement and motivation throughout this work.

This material is based on research sponsored by DARPA under agreement number FA8750-18-2-0077. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation thereon." "The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of DARPA or the U.S. Government.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

Introduction

Traumatic Brain Injury (TBI) as the name suggests usually refers to a severe impact or blow to the head or body. Traumatic Brain Injury can be categorized into Mild Traumatic Brain Injury, Moderate Traumatic Brain Injury and Severe Traumatic Brain injury. Mild and Moderate Traumatic brain injury may have a temporary impact on the brain cells whereas Severe Traumatic Brain Injuries have long term impacts such as bleeding, tear or bruising to the brain, risk of hypotension, hypoxaemia, and brain swelling which could potentially result in death [4][5].

1.1 Causes and Symptoms

The most common cause of TBI are falls, which account to nearly half(47.2%) of all TBI related hospitalizations as shown in Fig 1.1. This percentage increases to 49% and 81% in TBI related visits for children below 17 and older adults above 65 respectively [6]. Some of the other causes are Motor Vehicle Crashes, Assault, Struck by/against objects and self-harm.

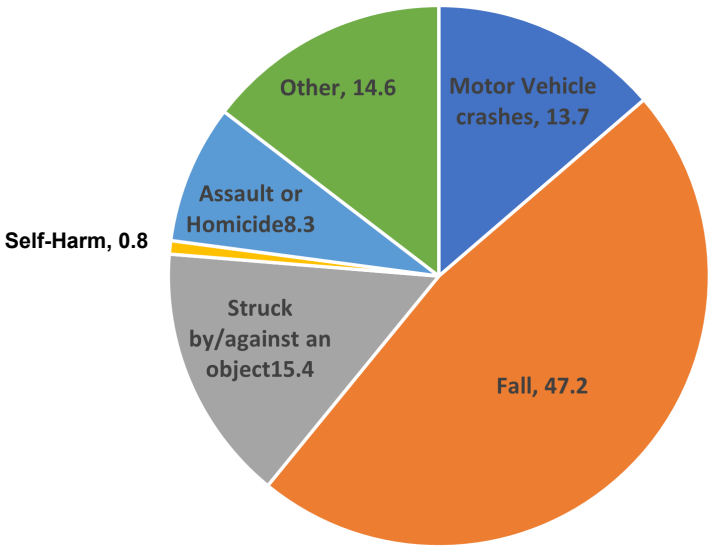

Figure 1.1: Percentage of TBI related emergency visits, hospitalizations and death by cause of injury [1].

The symptoms of TBI may appear immediately after the injury or may appear much later and can last from few minutes to lifetime depending on the severity of the injury. These symptoms can be broadly classified into four categories as listed below:

- **Thinking/ Remembering**: Difficulty thinking/concentrating/remembering and periods of unconsciousness [7][8].
- **Physical symptoms**: Nausea or Vomiting, Sensitivity to noise and light, Headache, Imbalance, Blurred vision [7][8].
- **Emotional Symptoms**: Nervousness, irritability, emotional imbalance and anxiety [7].
- **Sleep**: Improper/ Less/ More sleep than usual [7].

1.2 Prevalence of TBI

In the US alone there was around 2.8 million TBI related diagnosis in 2014 of which nearly 2.5 million TBI-related ED visits, 282,000 TBI-related hospitalizations, and 56,000 TBI-related deaths which is around 53% more than the that in 2006 [1][6]. On an average around 155 people died of injuries that include TBI everyday contributing to nearly one-third(30.5%) of all injury related deaths [6][9]. Most survivors live with significant disabilities, resulting in a huge economic burden [10]. Around $76.5 billion is spent in direct and indirect cost of TBI in the US annually of which over $1 billion is towards Pediatric TBI [11][12]. TBI has a huge impact on the United States military affecting 375,230 military personal between 2000-2017 [13]. These estimates of TBI are much higher when non-hospitalized cases are considered [14]. They go unnoticed because children below 14 and adults over 65 are more likely to sustain TBI and the symptoms of which correlate with those from old-age and playground related injuries are not often reported by children [9].

1.3 Traditional TBI measurement and detection techniques

The commonly used techniques to detect TBI can be categorized into two categories:

- **Medical Imaging** : Imaging tests like Computerized Tomography (CT) or Magnetic Resonance Imaging (MRI) are used to identify TBI, but are not frequently used due to their high costs. Other types of imaging include Inflammation Imaging, Metabolism Imaging and Tau protein Imaging [15][16][17].

- **Health Assessment questionnaires** : Glasgow Coma Scale evaluates a user and categorize them into groups based on the ability to speak, move and open eyes [15][16]. User's with a GCS score of 13 to 15 are categorized to have Mild TBI, 9 to 12 have Moderate TBI and less than 8 to have Severe TBI [18]. Another commonly used assessment is Automated Neuropsychological Assessment Metrics, which assess a users concentration, reaction time, recollection and attention [19].

1.4 The need for passive continuous TBI monitoring

Millions of people globally experience TBI, but are not diagnosed with TBI or get diagnosed much later [20]. This substantiates the need for a passive monitoring to identify TBI, with it being the leading cause of death among young adults [21]. This enforces the need for a passive surveillance system to analyze user behaviour and detect the possibility of a user to have TBI. With nearly 90% of the US population having smartphones, it presents a passive and affordable means of monitoring [22]. Most smartphones today are equipped with motion sensors(accelerometer, gyroscope, gps and proximity), light sensors, audio and video, sms and calls and application usage logs that encapsulate users social behaviour [23].They have processing capabilities in par with laptops to run high intensive computational tasks. This type of passive sensing has been employed in understanding human behaviour [24] and in identifying respiratory problems [25], influenza [26], parkinson's [27] and depression [28]. With increasing capabilities of the smartphone and its sensors, the passively collected data provides great leverage to perform modelling on and to identify underlying patterns. This can be employed to identify users with TBI using passively sensed smartphone sensor data without requiring any test that are conventionally used to detect TBI as users tend to not take some injuries seriously. This is beneficial as early detection and treatment of TBI reduces the risk and the financial burden on user's[29].

1.5 Our Approach: Passive smartphone sensing of TBI

The primary goal of this work is to identify and flag user's who could possibly be suffering from TBI as early as 24 hours from when they sustained an injury to their head. In this work we estimate if an individual sustains TBI based on mobility patterns from Accelerometer and Pedometer, movement patterns from Location and Altitude sensors, smartphone usage patterns from Gyroscope, Accessi-

bility and Battery State sensors .

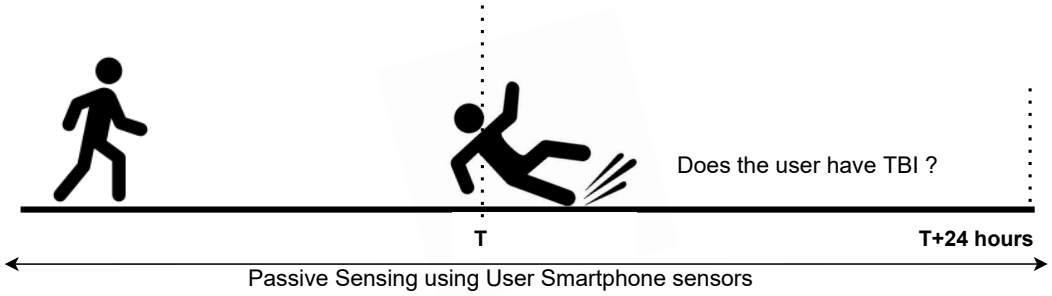

Figure 1.2: Detection Scenario : Users who have TBI are detected as soon as 24 hours from the injury.

We propose a novel framework for TBI prediction, which uses Deep Clustering to identify sub-groups of users with similar smartphone sensor patterns, for eg. users with high movement and mobility patterns are clustered together and users with low mobility and movement patterns together. Semi-Supervised Anomaly detection technique XGBOD, which uses boosting technique to identify anomalies is applied at the cluster level to identify TBI. We achieves a balanced accuracy of 88% and sensitivity of 74%.

1.6 Prior Work on TBI using smartphone

Pupil dilation based assessment of Pupillary Light Reflex (PLR) uses PupilScreen box and CNN to identify the severity of TBI in users. It uses a smartphone placed in the PupilScreen box aligned in a way that the flash of the smartphone is aligned with neutral density filter and diffuser. Our approach is passive, does not require active user involvement, additional hardware or setup, and accurately identifies TBI from passively gathered smartphone sensor data with a sensitivity of 74% balanced accuracy of 88%[2].

1.7 Contribution

These are the key contributions of our work:

1. We propose a novel framework, a deep-clustering based anomaly detection framework that precisely identifies users with possible TBI, a day after the injury using the smartphone sensor data.

2. We extract a rich set of statistical features along with hand crafted features that were identified to be useful on smartphone sensors in understanding user behaviour.

3. We deploy deep embedded clustering to segment users into homogeneous groups based on their smartphone usage. Here, we use a Auto-encoder module to perform dimensionality reduction and the loss function used is a combination of reconstruction loss and cluster assignment hardening loss.

4. We then identify users with TBI by classifying each identified cluster using XGBOD, a semi-supervised framework that employs gradient boosting based outlier detection.

5. We demonstrate through rigorous evaluations that our work attains a balance accuracy of 88% and sensitivity of 74% in detecting TBI in user's as early as 24 hours from the injury.

1.8 Thesis Roadmap

The remainder of the thesis is organized as follows. Sections II and III discuss related work and the background of the techniques used in our work. In the next section we give an overview of the dataset used and in Section V, the process of featurization from smartphone sensor data and implementation details of the our approach are discussed. In Section VI, we present the results of the experiments followed by discussion and conclusion in Section VII and Section VIII.

CHAPTER 2

Related Research

2.1 Smartphone based health sensing

With advancement in the number and the quality of sensors embedded to smartphones increasing, there have been abundance of research in understanding user behaviour [30] and predicting the health state such as influenza [31][26], zika, chikungunya and dengue virus[32], pulmonary disease[25], parkinson's [33], depression [28][34] and user moods [35][36] by analyzing the user's smartphone usage patterns. More recent research focus on calculating the possible risk of an individual to be affected by infectious disease and tracing the spread of an infectious disease in a community [37]. The Proximity Index is a risk index that identifies patients at high risk of COVID-19 infection in order to alert them by analyzing density and distances to those nearby who are infected [38]. To the best of our knowledge, our work is the first to propose a smartphone sensor's based detection of TBI among users within a day from the injury.

2.2 Mobile-health based TBI monitoring

In some prior-work researchers have tried to understand the process of recovery and quantify them among users with TBI using mobile based applications. While most of these focus on providing self-care support to users based on daily questionnaires to access a users status through sms [39] or by weekly meeting with coaches [40]. The work by J.Shannon et al tracks mood-related symptoms post TBI by a smartphone application and accessing its compliance with Ecological Monetary assessment (EMA)[41]. Though all these work track the recovery phase of TBI, our work pertains to identifying the presence of TBI passively, so that the user gets immediate assistance that is required without any specific user involvement.

2.3 PupilScreen: Smartphone based TBI detection

This work focuses on detecting the presence of TBI before reaching the ER using a smartphone application and a PupilScreen box [2]. This works by measuring the Pupillary Light reflex by directing the smartphone flashlight to the eye and recording the response through the camera. This recorded data is then processed using a fully connected Convolutional neural network to infer the

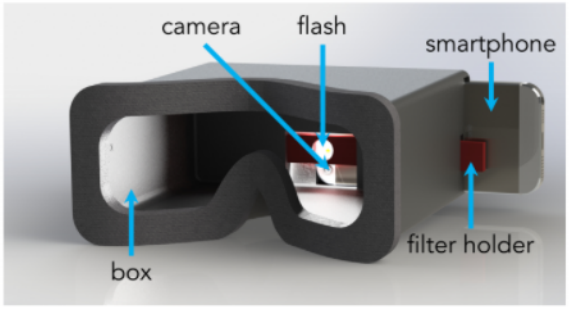

Figure 2.1: Pupil Screen box as described by Mariakakis et al [2]

diameter of the pupil. Though PupilScreen identifies the diameter of the pupil with a median error of 0.3mm, it requires a separate set up in the form of a PupilScreen box as shown in Fig. 2.1. Our work overcomes this obstacle by identifying the presence of TBI by analysing the passively sensed smartphone sensor data without any user involvement.

Other similar work were done by Lindsey [42], who identified involuntary pupil movement to correlate with TBI and Maruta et al. [43][44] tracked the gaze positional error with reference to the target and identified that for users with TBI, this variability increased.

CHAPTER 3

Background

3.1 Deep Embedded Clustering

We deploy Deep Embedded clustering to identify users with similar smartphone behaviour. By doing so, we restrict clusters to have users with similar smartphone usage behaviours, which improves the chances for the model to identify an anomalous behaviour from the regular one. We utilize Deep embedded clustering because it clusters on the latent space, which encapsulates all the information rather than the actual raw sensor data that are of extremely high dimensions. Deep embedded clustering uses a joint loss function consisting of reconstruction loss from the auto-encoder and cluster assignment hardening loss [45]. We initialize the parameters of the encoder and the decoder by pre-training using the reconstruction loss. After pre-training, the decoder module is discarded and the encoder module is fine tuned along with the KL divergence between soft cluster assignment and auxiliary distribution as shown in Eqn 5.8.

$$min \sum_i \sum_j p_{ij} log \frac{p_{ij}}{q_{ij}} \tag{3.1}$$

where, p_{ij} and q_{ij} are given by,

$$p_{ij} = \frac{\frac{q_{ij}^2}{f_j}}{\sum_{j'} \frac{q_{ij'}^2}{f_{j'}}} \tag{3.2}$$

$$q_{ij} = \frac{(1 + ||z_i - \mu_j||^2/\alpha)^{-(\alpha+1/)2}}{\sum_{j'}(1 + ||z_i - \mu'_j||^2/\alpha)^{-(\alpha+1/)2}} \tag{3.3}$$

Here, the degree of freedom α set to 1.

3.2 Extreme Boosting Based Outlier Detection

We utilize Extreme Boosting Based Outlier Detection (XGBOD) to accurately identify users with TBI. It is modelled exclusively on the different clusters of users identified by the clustering module. XGBOD is a semi-supervised outlier detection technique with unsupervised representation learning. XGBOD operates in three stages as shown in Fig 3.1 below.

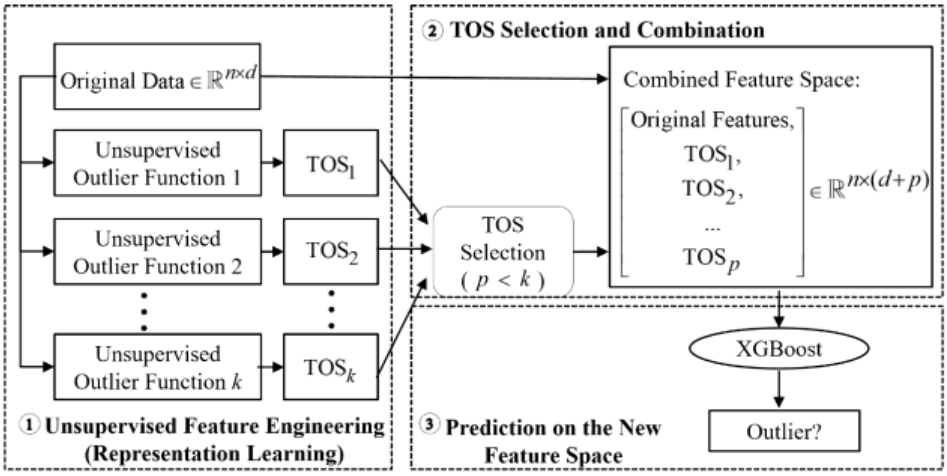

Figure 3.1: Architecture of XGBOD.[3].

In the first stage, unsupervised outlier detection techniques are used to extract learned representations of the input raw sensor data. This is done by applying outlier scoring functions to the original data as shown in Eqn. 3.4.

$$\phi(X) = [\phi_1(X)^T, ..., \phi_k(X)^T] \tag{3.4}$$

In the next stage, the augmented feature space is created by applying greedy techniques to keep useful representations from original and learned representation of data as given by Eqn. 3.5.

$$Feature\ Space_{new} = [X, \phi_1(X)] \tag{3.5}$$

In the final stage XGBoost[46], a supervised, scalable tree boosting technique is applied to identify the outliers on the new set of features.

CHAPTER 4

Dataset Description

We use the dataset gathered from the Kryptowire application. The data gathering process is ongoing with new users being recruited frequently. Uses were initially recruited through Google Ads and facebook and later through Prolific and mTurk. The inclusion and exclusion criteria for the subjects are listed below:

Inclusion criteria:

- Individuals 19 and over, English speaker, own a smart phone (iPhone or Android) with a data plan and Wi-Fi or 3G/4G capabilities, and be the primary user of the phone.

Exclusion criteria:

- Under 19 years or does not own a smart phone with Wi-Fi or 3G/4G capability
- Non-English speaker
- Unable to provide informed consent.
- Use of a VPN (virtual private network)
- Bad actors/fraudulent enrollees were identified based on failure across multiple verification procedures (e.g. VPN usage, inability to provide sensor data, CAPTCHA completion) and will be removed from the study.

	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 7	Week 8	Week 9	Week 10	Week 11	Week 12	Color Scheme
Day 1	SID1	SID8	SID15	SID22	SID29	SID36	SID43	SID50	SID57	SID64	SID71	SID78	Baseline
Day 2	SID2	SID9	SID16	SID23	SID30	SID37	SID44	SID51	SID58	SID65	SID72	SID79	Survey A
Day 3	SID3	SID10	SID17	SID24	SID31	SID38	SID45	SID52	SID59	SID66	SID73	SID80	Survey B
Day 4	SID4	SID11	SID18	SID25	SID32	SID39	SID46	SID53	SID60	SID67	SID74	SID81	Survey C
Day 5	SID5	SID12	SID19	SID26	SID33	SID40	SID47	SID54	SID61	SID68	SID75	SID82	Survey D
Day 6	SID6	SID13	SID20	SID27	SID34	SID41	SID48	SID55	SID62	SID69	SID76	SID83	Survey E
Day 7	SID7	SID14	SID21	SID28	SID35	SID42	SID49	SID56	SID63	SID70	SID77	SID84	Survey F
													Survey G
													End of Study Survey

Figure 4.1: Overview of the surveys taken by users.

Each user is a part of the data collection process for a period of 12 weeks during which smartphone sensor data is passively collected along with users responding to different surveys everyday

over the course of their participation in the study as represented by different colours in Fig. 4.1.

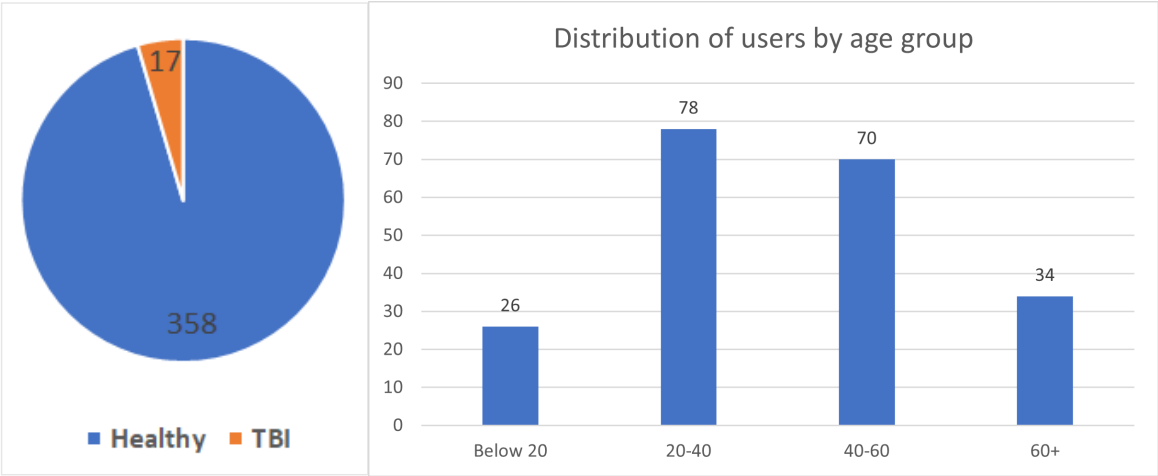

Figure 4.2: Left: Distribution of users by category. Right:Age wise distribution of users who reported their age during the start of the survey.

The subset of users considered in our work comprises of 375 users, of which around two-thirds were female and one-thirds male. The split of the users with TBI from the users considered is shown in Fig. 4.2. These users were identified based on the user responses to the survey questions highlighted in Table 4.1. The users considered in our work belonged to a variety of age groups as shown in Fig. 4.2 Participants under consideration were all iOS users and smartphone sensor data was passively collected from these devices. Smartphone sensors were polled every second and user had the option to turn off the polling.

Survey Question (as shown on mobile phone)
Were you in any accidents this week?
Did you hit your head at all this week?
Did you become unconscious due to the injury?
What day did it happen?
Did you see a doctor for your head injury?
Did the doctor say you have a concussion?

Table 4.1: Survey Questionnaire used to get ground truth labels.

The mobile sensor data collected were X-axis, Y-axis and Z-axis values along with timestamp

from Accelerometer, Gyroscope and Magnetometer. The other sensor data gathered were pressure and altitude, battery state, accessibility features(Assistive Touch, Bold Text,Closed Captioning, Darker System Colors, Voice over, Mono System Audio, Reduced Transparency, Shake to Undo, Speaking the screen, Speaking the selection and Switch control) that were turned on and the pedometer app logs.

CHAPTER 5

Methodology

The proposed framework consists of three non-overlapping, sequential components.

- **Feature Extraction**: In the first part, we compartmentalize the raw sensor data into buckets of fixed window sizes and extract statistical features and derived features over each buckets.
- **Deep Clustering**: In the next section, the featurized user smartphone sensor data is used to categorize users into disjoint groups or clusters based on their smartphone usage patterns. This is achieved through deep embedded clustering which uses a feed forward auto-encoder to perform dimensionality reduction and optimizes on the joint loss function of reconstruction loss and cluster assignment hardening loss.
- **Cluster Specific Anomaly detection**: In the last phase we perform anomaly detection on each cluster of users using Extreme Gradient Boosting Outlier detection to identify users with TBI. The overview of the workflow is shown in Fig. 5.1

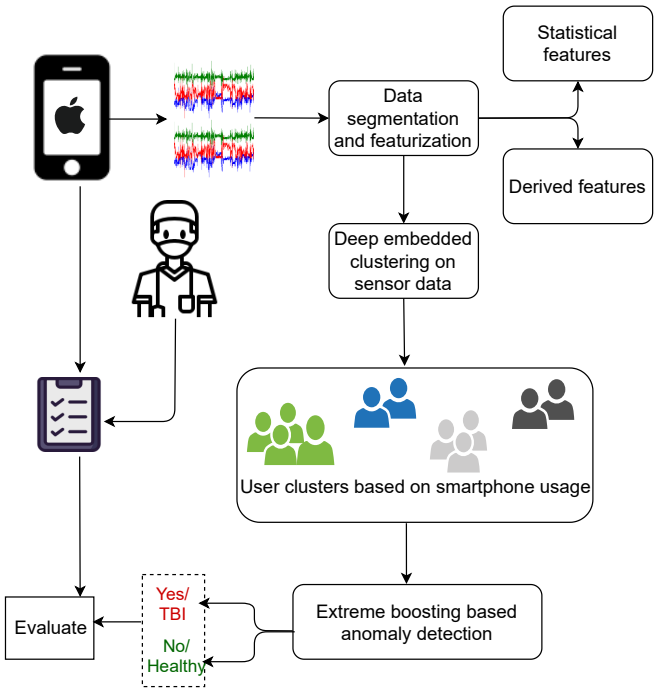

Figure 5.1: Overview of our approach.

5.1 Feature Extraction

Features of two kinds namely statistical features and derived features are extracted based on the type of sensors. These features are computed bi-hourly for 24 hours with 50% overlap between the windows. For example, each feature is extracted from 00:00-01:59(timestep 1), next 01:00-02:59(timestep 2) and next between 02:00-03:59(timestep 3) and so on until 10:00-11:59(timestep n). The process of of featurization is outlined in Fig. 5.2 below.

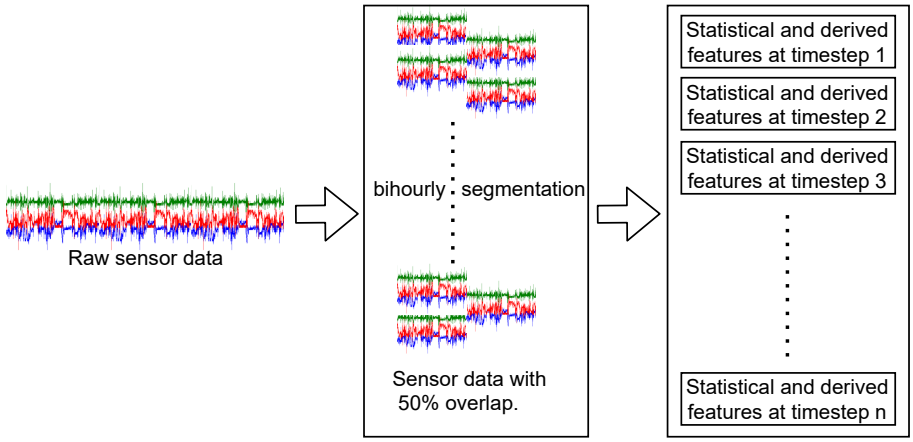

Figure 5.2: Process of data segmentation and featurization

5.1.1 Statistical Features and Derived Features

Statistical featurization provides meaningful information from the raw sensor data. They depict the raw sensor data in a high level format. Performing statistical featurization on sensors such as Accelerometer, Gyroscope and Magnetometer assist in understanding the general behaviour of the user's and the actions performed by them. The various statistical features we extract are described as follows:

- **Mean**: Mean or average is the sum of a collection divided by the number of elements in that collection.

$$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i \tag{5.1}$$

- **Variance**: Variance is the expectation of the squared deviation of a random variable from its mean.

$$\sigma^2 = \frac{1}{N}\left[\sum (x-\mu)^2\right] \tag{5.2}$$

SENSOR	FEATURE	FEATURE DESCRIPTION
Accelerometer	Mean	Mean value of X-axis, Y-axis and Z-axis separately
	Variance	Variance in data along X-axis, Y-axis and Z-axis separately
Gyroscope	Standard Deviation	Standard deviation along X-axis, Y-axis and Z-axis separately
	Skewness	Skewness in X-axis, Y-axis and Z-axis separately
Magnetometer	Kurtosis	Kurtosis of X-axis, Y-axis and Z-axis separately
	Inter-quantile range	IQR among X-axis, Y-axis and Z-axis separately
Accessibility features	Assistive Touch, Bold Text, Closed Captioning, Darker System colors, Voice over, Mono System Audio, Reduced Transparency, Shake to Undo, Speaking the screen, Speaking the selection and Switch control	These accessibility features were recorded if they were turned on-1 or off-0 *
Battery state	Battery level	The average battery level in the timeframe
	Battery Consumption	Total battery consumed in the timeframe *
Altitude	Mean	Average altitude and average pressure
	Variance	Variance in altitude and pressure

Table 5.1: List of features generated from sensor data. * represents derived features.

- **Standard Deviation**: Standard deviation is a measure of variation of a set of values.

$$\sigma = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \mu)^2} \tag{5.3}$$

- **Skewness**: Skewness is a measure of asymmetry of a probability distribution.

$$Skewness = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{x_i - \mu}{\sigma}\right)^3 \tag{5.4}$$

- **Kurtosis**: Kurtosis is a measure of how outlier prone a distribution is.

$$Kurtosis = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{x_i - \mu}{\sigma}\right)^4 \tag{5.5}$$

- **Interquartile range**: Interquartile range is a measure of the statistical dispersion of the data.

$$IQR = \frac{3}{4(n+1)}th\ term - \frac{1}{4(n+1)}th\ term \tag{5.6}$$

The derived features provide additional insights in a straightforward manner unlike raw sensors which have these information encapsulated in them providing better context.

- **Residual Step length**:

$$\text{Residual Step length} = \frac{\text{distance}}{\text{number of steps}} \tag{5.7}$$

- **Average gait velocity**:

$$\text{Average gait velocity} = \frac{\text{Average steps per sec}}{\text{step length}} \tag{5.8}$$

SENSOR	**MEASURE**	**FEATURE**	**DESCRIPTION**
Pedometer	Number Of Steps	Sum	Total number of steps
		Mean	Average number of stems
		Variance	Variance among number of steps
		Standard deviation	Standard deviation in steps taken
		Skewness	Skewness of number of steps taken
		Kurtosis	Kurtosis of steps taken
		Inter-Quantile range	IQR of steps taken
		Residual step length	Average stride length*
		Average gait velocity	Ratio of number of steps to step length*
	Active pace	Mean	Average active pace
	Distance	Sum	Total distance covered on foot
		Mean	Average distance covered on foot

Table 5.2: Pedometer features generated from raw sensor data. * represents derived features.

5.2 Deep Clustering

We perform Deep embedded clustering on the featurized data to segment users into distinct clusters based on their smartphone usage patterns as shown in Fig. 5.3. By clustering users with similar usage patterns, it is easy for anomaly detection technique to identify a change from normal behaviour. This phase consist of two modules: The auto encoder module and clustering module which internally operates in two sub modules namely, parameter initialization and parameter optimization.

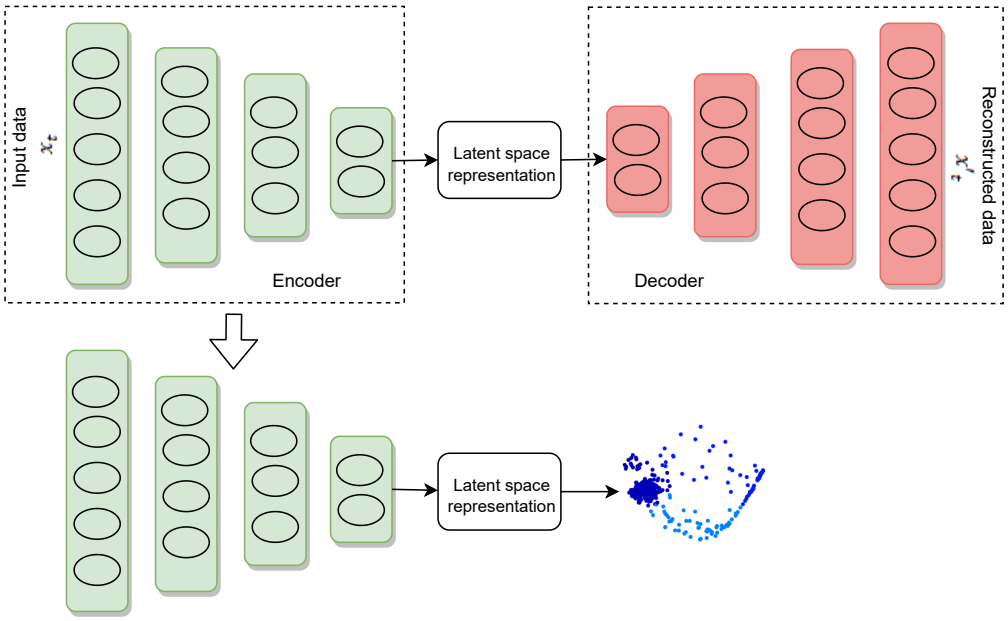

Figure 5.3: Segmentation of users by Deep embedded clustering

5.2.1 Parameter initialization

We set up the feed forward auto encoder to obtain a latent space representation with reduced dimensions. We input the data x_t at time t to the encoder with four dense layers that generate the lower dimensional latent space representation and the decoder tries to reconstruct the input from the latent space and outputs x'_t . The auto encoder uses relu activation function and at each step the encoder tries to reduce the dimension of the data from the current state to the next state, whereas the decoder tries to bring back the data to a higher dimension by reconstructing it. We stop training when we minimize the reconstruction loss which is the difference between the input data and reconstructed data and obtain the parameter θ of the trained network.

$$L_r = |x_i - x_i'|^2 \tag{5.9}$$

To initialize the cluster centroids, C_j(where j=1,2,...no of clusters) we deploy K-means algorithm on the lower dimensional latent space representation of the input data.

5.2.2 Parameter optimization

Once the autoencoder is initialized by minimizing the reconstruction loss, the decoder is discarded and the model is optimized by minimizing the KL divergence loss between the auxiliary distribution P and soft cluster assignments Q as shown in Eqn. 5.10.

$$L_c = KL(P||Q) = min \sum_i \sum_j p_{ij} log \frac{p_{ij}}{q_{ij}} \tag{5.10}$$

We then jointly optimize the cluster centroids C_j and the parameters of the autoencoder θ by applying Stochastic Gradient with momentum as a tuneable parameter γ. The final loss function is given below in Eqn. 5.11.

$$L = L_r + \gamma L_c \tag{5.11}$$

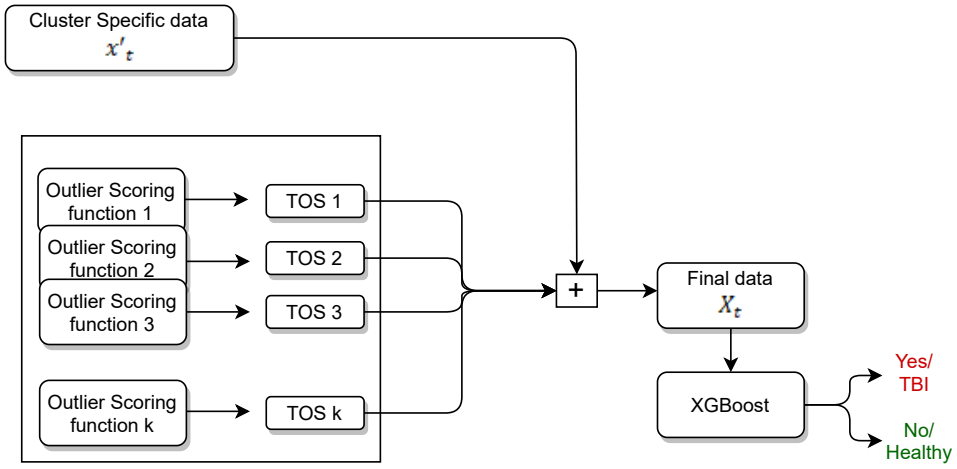

Figure 5.4: Process of data segmentation and featurization

5.3 Cluster specific Anomaly detection

Data from each clusters C_j where $j \in [0, 1, ...no\ of\ clusters]$ is fed to the Extreme Gradient Boosting based Outlier detection(XGBOD) framework. We input the learned representations from the auto

encoder x_t' where $x_t' \in C_j$. XGBOD applies outlier scoring function which maps the input data to a Transformed Outlier Score(TOS). We have k Transformed outlier scores, where k is the number of scoring functions used. We then perform greedy selection on these features and combine with x_t' to get the final data X_t. On X_t we apply XGBoost to perform binary classification as shown in Fig. 5.4[46].

CHAPTER 6

Results and Evaluation

6.1 Evaluation Metrics

We evaluated our approach using the following metrics:

- Balanced accuracy (BAC), calculated as $BAC = \frac{TP/(TP+FN)+TN/(TN+FP)}{2}$.
- Sensitivity or True Positive Rate (TPR) , calculated as $TP = \frac{TP}{TP+FN}$.
- Specificity or True Negative Rate (TNR) , calculated as $TN = \frac{TN}{TN+FP}$.

6.2 Experiments

We implement our proposed framework using passively sensed sensor data to identify users with TBI on data collected by Kryptowire and evaluate them on Balanced accuracy, Sensitivity or True positive rate and Specificity or True negative rate. The experiments are set up in a way that we use baseline models to identify the optimal window sizes and compare them against an overlapping window of 50%. We then experiment on variety of clustering techniques and evaluate the clusters and the performance of the models with and without clustering.

6.2.1 Experiment 1: Determining optimal window size

We analyze the performance of classifiers across different window sizes such as bi-hourly, six-hourly, twelve-hourly and twenty four-hourly. From Table 6.1 it is evident that Bi-hourly outperformed all the other windows. Twelve and Twenty four hour windows performed poorly predicting the majority class.

Window size	BAC	TPR	TNR	FPR	FNR
Bi-hourly(MLP)	**0.7**	**0.5**	**0.956**	**0.041**	**0.5**
6-hourly(XGBOD)	0.616	0.333	0.898	0.102	0.667
12-hourly(XGBOD)	0.5	0	1	0	1
24-hourly(XGBOD)	0.5	0	1	0	1

Table 6.1: Comparison of different window sizes.

6.2.2 Experiment 2: Determining optimal overlap

We compare the performance of classifiers with different overlap percentages using the best window size(bi-hourly) identified in our previous experiments. It is inferred that an overlap percentage of 50% outperforms the classifiers modelled on data with 33% overlap as shown in Table 6.2.

Window size	BAC	TPR	TNR	FPR	FNR
Overlap 33%	0.7374	0.5	0.9748	0.0252	0.5
Overlap 50%	**0.6276**	**0.28**	**0.9667**	**0.033**	**0.72**

Table 6.2: Comparison of different overlap percentage.

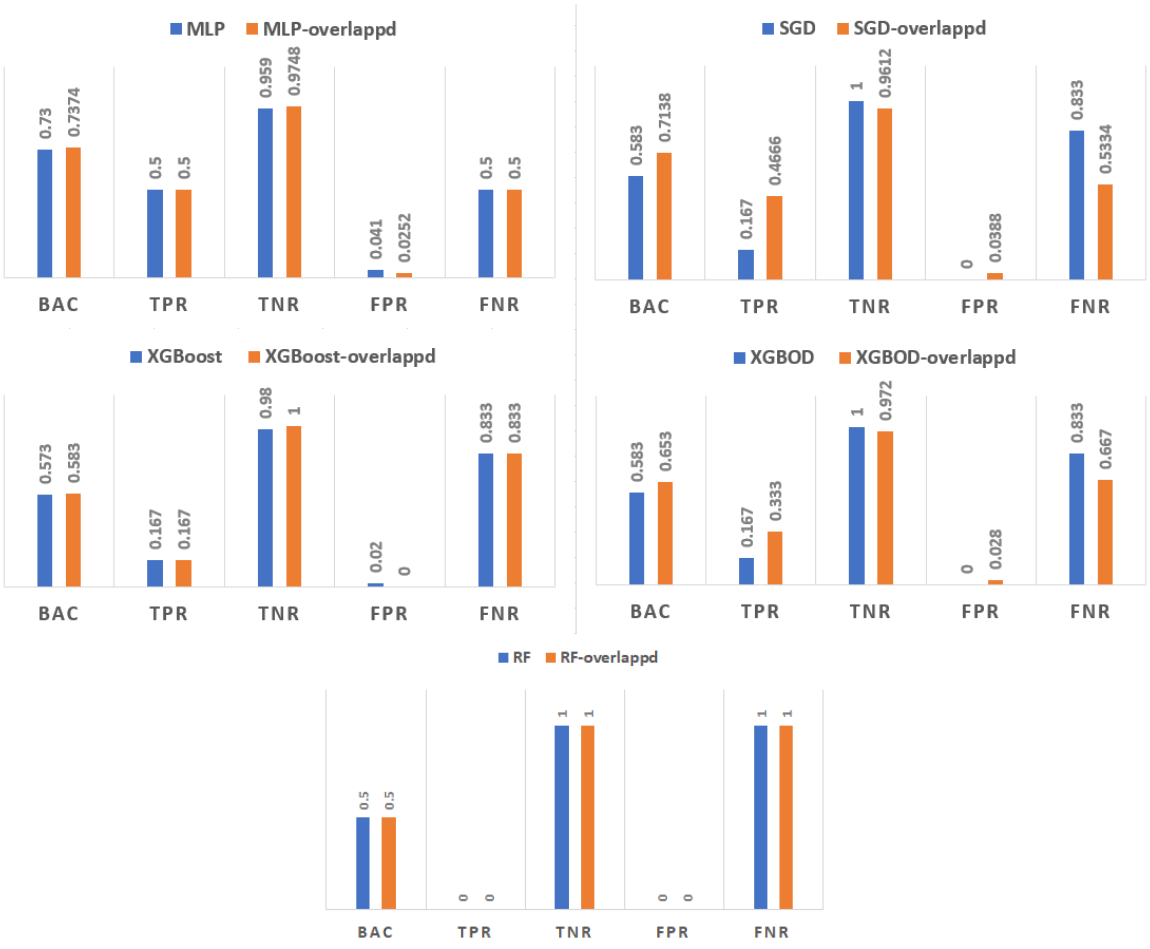

Figure 6.1: Comparison of models with non-overlapping and overlapping windows.

6.2.3 Experiment 3: Non-overlapping window vs Overlapping window

When we compare the performance of models with the best window size(bi-hourly) identified from the above experiment against overlapping windows with 50% overlap, we observed that overlapping window with 50% overlap performs better than non-overlapping windows as shown in Fig. 6.1. For all the classifiers used the balanced accuracy and the sensitivity of models with overlapping windows outperformed the models without overlap. When we compare against specificity the models with overlapping window were on par or better than the once without overlap.

6.2.4 Experiment 3: Impact of Clustering

We perform the remainder of the experiments with an overlap of 50% between windows. We experiment without clustering to set up baselines to compare against the models with clustering. The Table 6.3 below shows the performance of models in the absence of any clustering techniques.

Classifier	BAC	TPR	TNR	FPR	FNR
MLP	**0.7374**	**0.5**	**0.9748**	**0.0252**	**0.5**
RF	0.5	0	1	0	1
SGD	0.5	0	1	0	1
XGBOD	0.653	0.333	0.972	0.028	0.667
XGBoost	0.583	0.167	1	0	0.833

Table 6.3: Performance of machine learning models without clustering.

We experimented with different clustering techniques, by clustering the users into non-overlapping subsets based on their smartphone sensor data and performing modelling on them. We infer from the Fig. 6.2 below that the clusters formed are tight and distinct while projected down to 2-dimensions.

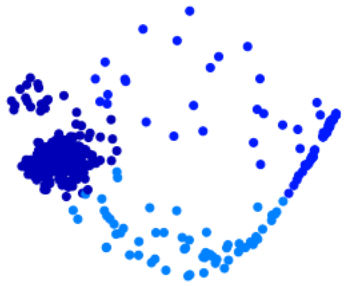

Figure 6.2: t-sne visualization of the clusters formed.

We compare K-means clustering, Spectral clustering, MeanShift clustering and Deep Embedded clustering with different classifiers using the same seed, so that we can evaluate the performance of different classifier given a clustering algorithm and the performance of different clustering algorithms given a classifier. We observed that Deep Embedded clustering had the best results across all classifiers when compared against balanced accuracy and sensitivity or True positive rate.

Classifier	Cluster	BAC	TPR	TNR	FPR	FNR
MLP	DEC	0.7962	0.6	0.9923	0.0077	0.4
	kmeans	0.733	0.4834	0.983	0.017	0.5166
	meanshift	0.7534	0.5334	0.9732	0.0268	0.4666
	spectral	0.7297	0.3833	0.9761	0.0239	0.6167
RF	DEC	0.55	0.1	1	0	0.9
	kmeans	0.525	0.005	1	0	0.95
	meanshift	0.5334	0.0666	1	0	0.9334
	spectral	0.55	0	1	0	1
SGD	DEC	0.7798	0.65	0.9096	0.0904	0.35
	kmeans	0.6538	0.3332	0.9744	0.0256	0.6668
	meanshift	0.6918	0.4	0.9832	0.0168	0.6
	spectral	0.6427	0.2166	0.9689	0.0311	0.7834
XGBOD	**DEC**	**0.8765**	**0.7333**	**0.9532**	**0.0468**	**0.2667**
	kmeans	0.667	0.333	1	0	0.667
	meanshift	0.611	0.25	0.972	0.028	0.75
	spectral	0.658	0.333	0.984	0.016	0.667
XGBoost	DEC	0.775	0.55	1	0	0.45
	kmeans	0.618	0.25	0.986	0.014	0.75
	meanshift	0.5934	0.1998	0.9866	0.0134	0.8002
	spectral	0.5825	0.0999	0.9793	0.0207	0.9001

Table 6.4: Comparison of models with different clustering techniques.

The best performing model used Deep embedded clustering with XGBOD to identify TBI in-

stances with a sensitivity of 74% and balanced accuracy of 88%. Deep embedded clustering uses a batch size of 32 and an SGD optimizer whereas XGBOD uses a learning rate of 0.001 with max depth of 20 and uses L2 regularization.

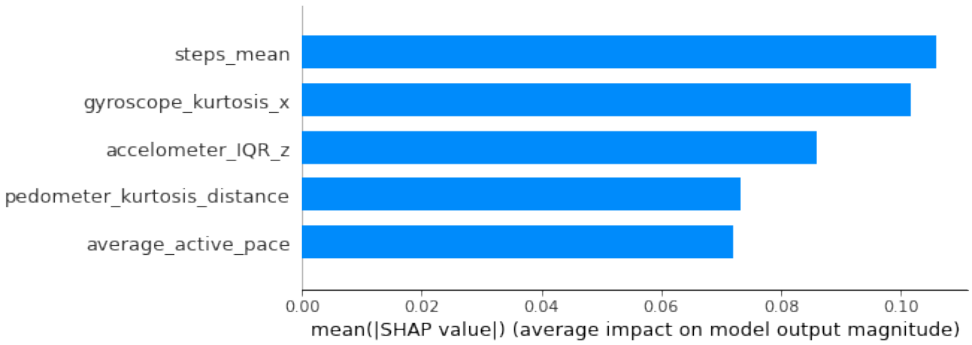

Figure 6.3: Shap value of top 5 important features.

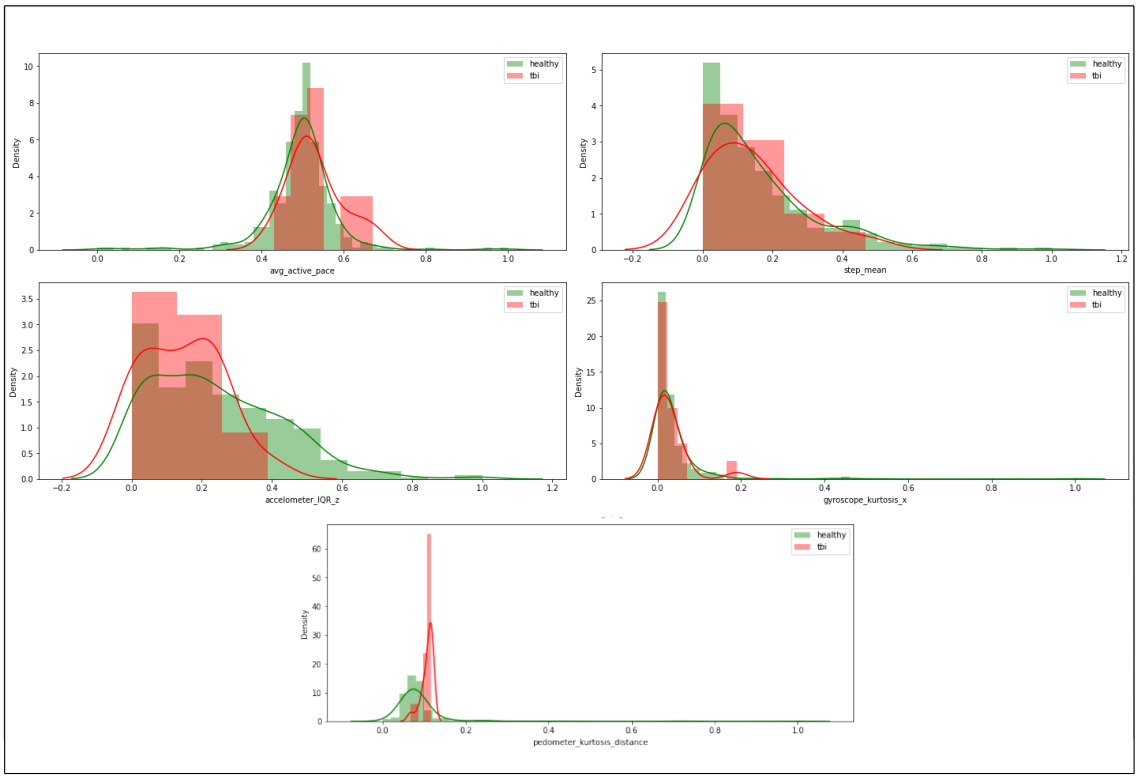

Figure 6.4: Data distribution of top 5 important features.

When the shapley values of different features across timesteps were aggregated, it was perceived that the mobility feature were the most important with three pedometer features in the top five as shown in Fig. 6.3. We analyze the distribution of these features(Fig. 6.4) in detail by comparing the

data points of TBI, healthy users it was observed evident that they had highly varying distributions with IQR of accelerometer along z axis having maximum variance. Features from Altitude sensors and two accessibility features(Gray Scale and Switch control) had 0 shap value or no influence on the predictions as shown in Fig. 6.5.

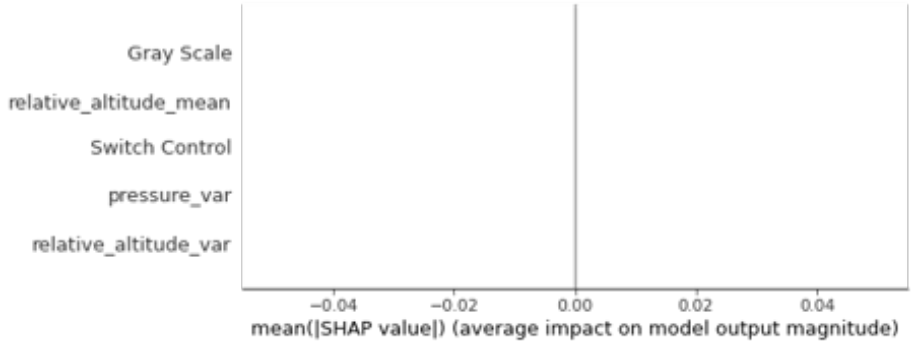

Figure 6.5: Shap value of least important features.

CHAPTER 7

Discussion

Most predictive features: The most important features that contribute largely to the predictions are shown in Fig. 6.4. Average active pace, Average number of steps, accelerometer IQR along z-axis, gyroscope kurtosis along x-axis and kurtosis of distance are the top 5 significant features when the data is prepared bi-hourly. This lines up with the fact that walking patterns of users are affected with TBI. Kurtosis determines the vulnerability of a distribution to outliers. With user patters changing post-TBI, two features of Kurtosis being in most predictive features is understandable. The least predictive sensors were features from altitude sensors. Though two accessibility features namely, gray scale and switch control had 0 shap values, other features contributed towards the predictions.

Effect of longer history of sensor data: We observed that by increasing the history of the sensor data provided to our work from 24 hours to 48 hours and 72 hours, it was observed that the model was performing better in identifying non-TBI users but had a flip side effect on identifying TBI instance as shown in Fig. 7.1. It can also be seen that from 24 hour history to 48 hour history, there was only a minor drop in sensitivity, but when we increase it to 72 hour history there was a comparatively higher drop.

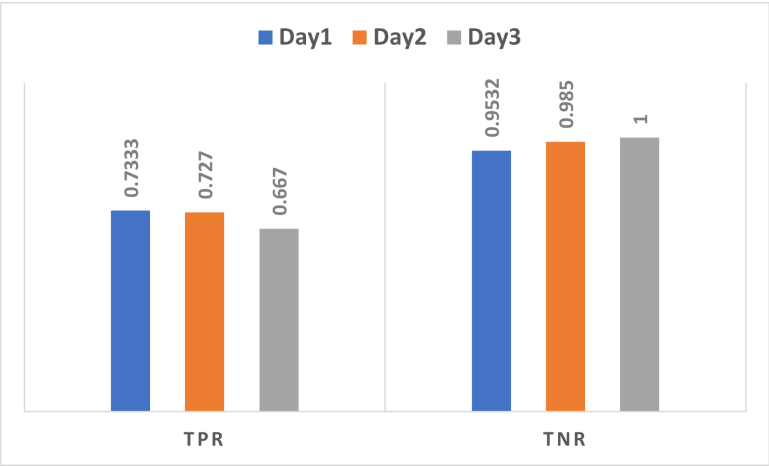

Figure 7.1: Comparison of performance with increased history.

Limitations of our work: Limitations of our work are primarily pertaining to the quality of the sensor data from different devices. In our work we consider data collected from iOS devices only. Though most of the features are not any device specific, accessibility features are iOS specific features.

Future work: As future work, we plan to include additional types of sensors across different platforms, unlike only iOS devices under consideration. We also plan to expand the data as the data collection process is ongoing and recruit more diverse population into the study, as currently only people in US are considered. With more diverse users under the study, we can unravel additional hidden patterns. We also plan to include user recovery monitoring, when a user is diagnosed with TBI.

CHAPTER 8

Conclusion

With increasing numbers of TBI-related emergency room visits and with many cases not being treated, prompt identification of possible TBI in users is gaining importance. This facilitates the need for accurate and prompt identification of TBI in users after injury.

We propose a framework which utilizes passively sensed smartphone data such as accelerometer, gyroscope, magnetometer, accessibility features, altitude and pedometer to precisely identify and flag users who have TBI as soon as a day from the injury. It uses deep embedded clustering based anomaly detection to achieve a sensitivity of 74% and a balanced accuracy of 88%. It can be primarily used by young adults who are prone to high impact falls on the playground and by elderly people, as they associate unintentional falls to their age and not get medically treated.

References

[1] C. A. Taylor, J. M. Bell, M. J. Breiding, and L. Xu, "Traumatic brain injury–related emergency department visits, hospitalizations, and deaths—united states, 2007 and 2013," *MMWR Surveillance Summaries*, vol. 66, no. 9, p. 1, 2017.

[2] A. Mariakakis, J. Baudin, E. Whitmire, V. Mehta, M. A. Banks, A. Law, L. Mcgrath, and S. N. Patel, "Pupilscreen: using smartphones to assess traumatic brain injury," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 1, no. 3, pp. 1–27, 2017.

[3] Y. Zhao and M. K. Hryniewicki, "Xgbod: improving supervised outlier detection with unsupervised representation learning," in *2018 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–8, IEEE, 2018.

[4] MayoClinic, "Traumatic brain injury." https://www.mayoclinic.org/diseases-conditions/traumatic-brain-injury/symptoms-causes/syc-20378557. Accessed: 2021-01-28.

[5] J. Ghajar, "Traumatic brain injury," *The Lancet*, vol. 356, no. 9233, pp. 923–929, 2000.

[6] CDC, "Tbi: Get the facts." https://www.cdc.gov/masstrauma/resources/gcs.pdf. Accessed: 2021-02-02.

[7] CDC, "Symptoms of traumatic brain injury (tbi)." https://www.cdc.gov/traumaticbraininjury/symptoms.html. Accessed: 2021-02-06.

[8] NIH, "What are common symptoms of traumatic brain injury (tbi)?." https://www.nichd.nih.gov/health/topics/tbi/conditioninfo/symptoms. Accessed: 2020-12-04.

[9] M. Faul, M. M. Wald, L. Xu, and V. G. Coronado, "Traumatic brain injury in the united states; emergency department visits, hospitalizations, and deaths, 2002-2006," 2010.

[10] E. Finkelstein, P. S. Corso, and T. R. Miller, *The incidence and economic burden of injuries in the United States*. Oxford University Press, USA, 2006.

[11] V. G. Coronado, L. C. McGuire, M. Faul, D. E. Sugerman, and W. S. Pearson, "Traumatic brain injury epidemiology and public health issues," *Brain injury medicine: Principles and practice*, vol. 84, 2012.

[12] A. J. Schneier, B. J. Shields, S. G. Hostetler, H. Xiang, and G. A. Smith, "Incidence of pediatric traumatic brain injury and associated hospital resource utilization in the united states," *Pediatrics*, vol. 118, no. 2, pp. 483–492, 2006.

[13] V. Rajajee, "Traumatic brain injury: Epidemiology, classification, and pathophysiology," in *UpToDate*, UpToDate Waltham, MA, 2019.

[14] D. M. Bernstein, "Subject review: Recovery from mild head injury," *Brain injury*, vol. 13, no. 3, pp. 151–172, 1999.

[15] G. Teasdale and B. Jennett, "Assessment of coma and impaired consciousness: a practical scale," *The Lancet*, vol. 304, no. 7872, pp. 81–84, 1974.

[16] NIH, "How do healthcare providers diagnose traumatic brain injury?." https://www.nichd.nih.gov/health/topics/tbi/conditioninfo/diagnose. Accessed: 2021-02-10.

[17] N. R. Council *et al.*, "Synopsis of brain injury detection methods," in *Review of Department of Defense Test Protocols for Combat Helmets*, National Academies Press (US), 2014.

[18] CDC, "How do healthcare providers diagnose traumatic brain injury?." https://www.cdc.gov/traumaticbraininjury/get_the_facts.html. Accessed: 2021-01-29.

[19] R. L. Kane, T. Roebuck-Spencer, P. Short, M. Kabat, and J. Wilken, "Identifying and monitoring cognitive deficits in clinical populations using automated neuropsychological assessment metrics (anam) tests," *Archives of Clinical Neuropsychology*, vol. 22, no. Suppl_1, pp. S115–S126, 2007.

[20] C. Calvert, "Tbi research review: Unidentified brain injury,"

[21] CDC, "Report to congress: Traumatic brain injury in the united states." https://www.cdc.gov/traumaticbraininjury/pubs/tbi_report_to_congress.html. Accessed: 2021-01-13.

[22] eMarketer, "Statista digital market outlook: Number of smartphone users in the united states from 2018 to 2025 (in millions)*." https://www.statista.com/statistics/201182/forecast-of-smartphone-users-in-the-us/. Accessed: 2021-01-22.

[23] C. A. Opperman and G. P. Hancke, "Smartphones as a platform for advanced measurement and processing," in *2012 IEEE International Instrumentation and Measurement Technology Conference Proceedings*, pp. 703–706, IEEE, 2012.

[24] G. M. Harari, S. R. Müller, M. S. Aung, and P. J. Rentfrow, "Smartphone sensing methods for studying behavior in everyday life," *Current opinion in behavioral sciences*, vol. 18, pp. 83–90, 2017.

[25] M. A. Azam, A. Shahzadi, A. Khalid, S. M. Anwar, and U. Naeem, "Smartphone based human breath analysis from respiratory sounds," in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 445–448, IEEE, 2018.

[26] S. N. Murthy, F. Asani, S. Srikanthan, and E. Agu, "Deepseas: Smartphone-based early ailment sensing using coupled lstm autoencoders," in *2020 IEEE International Conference on Big Data (Big Data)*, pp. 4911–4918, IEEE, 2020.

[27] I. García-Magariño, C. Medrano, I. Plaza, and B. Oliván, "A smartphone-based system for detecting hand tremors in unconstrained environments," *Personal and Ubiquitous Computing*, vol. 20, no. 6, pp. 959–971, 2016.

[28] A. A. Farhan, J. Lu, J. Bi, A. Russell, B. Wang, and A. Bamis, "Multi-view bi-clustering to identify smartphone sensing features indicative of depression," in *2016 IEEE first international conference on connected health: applications, systems and engineering technologies (CHASE)*, pp. 264–273, IEEE, 2016.

[29] M. Ashley, G. O'Shanick, and L. Kreber, "Early vs. late treatment of traumatic brain injury," *Vienna, VA: Brain Injury Association of America*, 2009.

[30] N. Ahmed, J. I. Rafiq, and M. R. Islam, "Enhanced human activity recognition based on smartphone sensor data using hybrid feature selection model," *Sensors*, vol. 20, no. 1, p. 317, 2020.

[31] A. Madan, M. Cebrian, S. Moturu, K. Farrahi, *et al.*, "Sensing the" health state" of a community," *IEEE Pervasive Computing*, vol. 11, no. 4, pp. 36–45, 2011.

[32] A. Priye, S. W. Bird, Y. K. Light, C. S. Ball, O. A. Negrete, and R. J. Meagher, "A smartphone-based diagnostic platform for rapid detection of zika, chikungunya, and dengue viruses," *Scientific reports*, vol. 7, no. 1, pp. 1–11, 2017.

[33] F. Lipsmeier, K. I. Taylor, T. Kilchenmann, D. Wolf, A. Scotland, J. Schjodt-Eriksen, W.-Y. Cheng, I. Fernandez-Garcia, J. Siebourg-Polster, L. Jin, *et al.*, "Evaluation of smartphone-based testing to generate exploratory outcome measures in a phase 1 parkinson's disease clinical trial," *Movement Disorders*, vol. 33, no. 8, pp. 1287–1297, 2018.

[34] A. Mehrotra and M. Musolesi, "Using autoencoders to automatically extract mobility features for predicting depressive states," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 2, no. 3, pp. 1–20, 2018.

[35] R. LiKamWa, Y. Liu, N. D. Lane, and L. Zhong, "Moodscope: Building a mood sensor from smartphone usage patterns," in *Proceeding of the 11th annual international conference on Mobile systems, applications, and services*, pp. 389–402, 2013.

[36] A. Bogomolov, B. Lepri, M. Ferron, F. Pianesi, and A. Pentland, "Daily stress recognition from mobile phone data, weather conditions and individual traits," in *Proceedings of the 22nd ACM international conference on Multimedia*, pp. 477–486, 2014.

[37] K. Fan, C. Li, and K. Heller, "A unifying variational inference framework for hierarchical graph-coupled hmm with an application to influenza infection," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 30, 2016.

[38] P. Admin, "Pcci's new mypci app informs individuals of covid-19 exposure risk." https://pccinnovation.org/pccis-new-mypci-app-informs-individuals-of-covid-19-exposure-risk/. Accessed: 2021-02-12.

[39] B. Suffoletto, A. K. Wagner, P. M. Arenth, J. Calabria, E. Kingsley, J. Kristan, and C. W. Callaway, "Mobile phone text messaging to assess symptoms after mild traumatic brain injury and provide self-care support: a pilot study," *The Journal of head trauma rehabilitation*, vol. 28, no. 4, pp. 302–312, 2013.

[40] M. E. Narad, G. Bedell, J. A. King, J. Johnson, L. S. Turkstra, J. Haarbauer-Krupa, and S. L. Wade, "Social participation and navigation (span): description and usability of app-based coaching intervention for adolescents with tbi," *Developmental neurorehabilitation*, vol. 21, no. 7, pp. 439–448, 2018.

[41] S. B. Juengst, K. M. Graham, I. W. Pulantara, M. McCue, E. M. Whyte, B. E. Dicianno, B. Parmanto, P. M. Arenth, E. R. Skidmore, and A. K. Wagner, "Pilot feasibility of an mhealth system for conducting ecological momentary assessment of mood-related symptoms following traumatic brain injury," *Brain injury*, vol. 29, no. 11, pp. 1351–1361, 2015.

[42] J. R. Lindsay, "The significance of a positional nystagmus in otoneurological diagnosis," *The Laryngoscope*, vol. 55, no. 10, pp. 527–551, 1945.

[43] J. Maruta, S. W. Lee, E. F. Jacobs, and J. Ghajar, "A unified science of concussion," *Annals of the New York Academy of Sciences*, vol. 1208, no. 1, p. 58, 2010.

[44] J. Maruta, J. Tong, S. W. Lee, Z. Iqbal, A. Schonberger, and J. Ghajar, "Eye-trac: monitoring attention and utility for mtbi," in *Sensing Technologies for Global Health, Military Medicine, Disaster Response, and Environmental Monitoring II; and Biometric Technology for Human Identification IX*, vol. 8371, p. 83710L, International Society for Optics and Photonics, 2012.

[45] J. Xie, R. Girshick, and A. Farhadi, "Unsupervised deep embedding for clustering analysis," in *International conference on machine learning*, pp. 478–487, PMLR, 2016.

[46] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pp. 785–794, 2016.