MACHINE LEARNING CLASSIFICATION OF FACIAL AFFECT RECOGNITION DEFICITS AFTER TRAUMATIC BRAIN INJURY FOR INFORMING REHABILITATION NEEDS AND PROGRESS

A Thesis

Submitted to the Faculty

of

Purdue University

by

Syeda Iffat Naz

In Partial Fulfillment of the

Requirements for the Degree

of

Master of Science in Electrical and Computer Engineering

December 2020

Purdue University

Indianapolis, Indiana

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF THESIS APPROVAL

Dr. Lauren Christopher, Chair

Department of Electrical and Computer Engineering

Dr. Brian King

Department of Electrical and Computer Engineering

Dr. Dawn Neumann

Department of Physical Medicine & Rehabilitation

Approved by:

Dr. Brian King

Head of the Graduate Program

I would like to dedicate this work to my parents Syed Ekram Ullah and Shahin Akter. I cannot thank you enough for your love and support.

ACKNOWLEDGMENTS

I want to express my heartiest gratitude to my supervisor Prof. Lauren Christopher, who has mentored and guided me throughout my research work. She has been a very supportive and inspiring mentor and has played an all-important role in making my research experience a fulfilling one. It would have been impossible for me to achieve this so far without her support and trust in me. We all know that research can be very challenging when the result is not up to the expectation. However, her empathetic and positive attitude helped me get through research struggles and be focused. I am glad that I had the opportunity to work under her supervision and have an enlightening experience.

I want to thank Prof. Dawn Neumann for providing us with the TASIT and eye tracking data for our research. I want to thank 'Indiana Traumatic Spinal Cord & Brain Injury Research Grant Program 400-20-113' for sponsoring our research. I appreciate the support.

Also, I would like to thank my project mate Rifat Mueid for being supportive and helpful whenever I needed it. I am also thankful to Ashley Dale and my other lab mates, who always maintained a positive working environment in the lab. It was a delightful experience to work with all of you.

Also, I would like to thank Prof. Brian King and Prof. Dawn Neumann for serving on my thesis committee. It is an honor for me to showcase my work in front of great minds. I would like to show my gratitude towards our ECE department's graduate coordinator Sherrie Tucker. She has always been very patient and prompt to solve all my confusion and answer all my queries with a big smiling face.

I am grateful to my parents who never lose trust in me and have always been supportive. I want to thank my brothers, who are my buddy, my source of joy.

Finally, I would like to thank my Bangladeshi friends here in Indianapolis, who is nothing but a family away from home.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Iffat Naz, Syeda. M.S.E.C.E., Purdue University, December 2020. Machine Learning Classification of Facial Affect Recognition Deficits after Traumatic Brain Injury for Informing Rehabilitation Needs and Progress. Major Professor: Lauren Christopher.

A common impairment after a traumatic brain injury (TBI) is a deficit in emotional recognition, such as inferences of others' intentions. Some researchers have found these impairments in 39% of the TBI population. Our research information needed to make inferences about emotions and mental states comes from visually presented, nonverbal cues (e.g., facial expressions or gestures). Theory of mind (ToM) deficits after TBI are partially explained by impaired visual attention and the processing of these important cues. This research found that patients with deficits in visual processing differ from healthy controls (HCs). Furthermore, we found visual processing problems can be determined by looking at the eye tracking data developed from industry standard eye tracking hardware and software. We predicted that the eye tracking data of the overall population is correlated to the TASIT test. The visual processing of impaired (who got at least one answer wrong from TASIT questions) and unimpaired (who got all answer correctly from TASIT questions) differs significantly. We have divided the eye-tracking data into 3 second time blocks of time series data to detect the most salient individual blocks to the TASIT score. Our preliminary results suggest that we can predict the whole population's impairment using eye-tracking data with an improved f1 score from 0.54 to 0.73. For this, we developed optimized support vector machine (SVM) and random forest (RF) classifier.

1. INTRODUCTION

Traumatic brain injury (TBI) results from a heavy blow or jolt to the head. TBI can be associated with many symptoms, such as executive functioning problems, cognitive problems, and communication problems. All these are rooted in the 'Theory of Mind (ToM).' ToM pertains to the ability to infer others' emotions (affect recognition), intentions, thoughts, beliefs, expectations, and desires. Deficits in emotion recognition and other ToM components are quite common after a traumatic brain injury (TBI); researchers have reported impairments up to 39% of the TBI population [1]. The information needed to make inferences about emotions and mental states comes from visually presented, nonverbal cues (e.g., facial expressions or gestures). ToM deficits after TBI are partially explained by impaired visual attention and the processing of these important cues. Eye-tracking technology allows us to see what participants are looking at in ToM tests.

Eye tracking technology is the raw data to predict ToM tests and give insight into the visual components of ToM. Eye movements are measured to determine where a person is looking, what they are looking at, and how long they look at a particular region. The eye is one of the primary organs that contributes to the perception of the world, and vision is a key component in a person's decision making process. Therefore, eye tracking technology can be very useful in detecting what leads to correct decisions and what leads to bad ones. So in this research, we use eye tracking technology to study patients with TBI and how their visual processing differs from the healthy controls (HC).

Emotional inference deficits are measured by The Awareness of Social Inference Test (TASIT) questions. We have used this collection of videos, which form 59 videos that determine the TASIT score. Patients were asked to watch videos to answer questions about each video. The answers expose the patient's understanding of the

emotions shown in the video (ToM assessment). This collection falls into PART1, PART2, and PART3 videos. Among these, the PART3 videos incorporate the most complex emotions and also have a higher variability of TASIT scores across patients. We want to predict if patients with deficits in visual processing differ from the unimpaired population and, if so, do these visual processing abnormalities contribute to their emotional inference deficits (as measured by the TASIT answers).

The primary contribution of this research is that we use a dynamic analysis of the eye-tracking features data. Among the 59 videos, we found only a specific number of videos play an important role in this research. Only a few videos are highly correlated with the TASIT score. We also found that only some parts of the videos, not the whole, are correlated with the TASIT answers. Eye tracking features in the correlated frames of the videos were used, and the result achieved 73% accuracy and an f1 score of 0.73. Taking only the correlated parts of videos have helped to improve the classification performance. This can help diagnose the TBI impairments and inform rehabilitation treatments.

1.1 Literature Review

Many previous studies have focused on static images for understanding visual attention using eye tracking data. The existing studies used both low and high-level image features. This eye tracking data has been a subject of research for detecting diseases for many years now. Diseases that do not have a clinical biomarker are at risk of being misdiagnosed. This data has been studied as a biomarker of diseases such as autism spectrum disorder (ASD) [2] [3], Alzheimer's disease [4], sports-related concussion [5]. While many studies used eye tracking data to differentiate between people with neurological disorders and control groups, various studies used different approaches to differentiate those.

High functioning autism is a phenomenon when a person has a high level of independence and ability. The eye movement of adult participants with and without

autism was recorded while looking for information within web pages at [3]; the study achieved 74% accuracy in detecting autism. The participants were given a search task, a time-limited browse task, and a synthesis task. The search task and time-limited browse task gave the best performance in discriminating the two groups' data. The study found that increased task complexity did not amplify the discrimination between the groups. Browsing strategies for participants with ASD while viewing a webpage is different. Autistic people are drawn more to images than text. Other findings [2] reported that children with ASD revealed a preference for nonsocial images rather than social stimuli. In the study, two dynamic images were presented side by side. One side features a social stimulus, with children engaging in aerobics and dancing, whereas another side featured a nonsocial stimulus with a series of short sequences of moving geometric shapes. Children who spent more time at geometric shapes show impairment in Autism Diagnostic Observation Schedule (ADOS) [2]. These studies confirm that eye tracking data is key to understanding the perception of impaired individuals vs. healthy controls.

Another study [4] reported a review for studies on Alzheimer patients. The patients had a hard time making saccadic eye movement [see glossary for this kind of definitions 5.1], which is a rapid change of eye position from one fixation point to another compared to healthy people [6]. When people with Alzheimer's disease are directed towards a target, it took a long time to move their attention and showed increased saccadic reaction time; also known as saccade latency.

There has also been research exploring impairments in visual processing after traumatic brain injury (TBI). In [1], difficulties in emotion perception after TBI were presented. The main contribution was to examine the severity of this problem. Static images were presented to detect facial affect recognition in people with TBI (PwTBI). PwTBI showed significant difficulties in recognizing facial effects than controls. While PwTBI has an emotion deficit, it was not clear if the emotion deficit is correlated to eye movement data. There is some research on eye tracking data after TBI. [7] focused on people with different severities of traumatic brain injury (TBI) as well as

asymptomatic controls. Eye tracking tests were performed to measure horizontal and vertical saccades. The research achieved a sensitivity of 0.77 for horizontal saccades while the sensitivity of 0.64 for vertical saccades. The study concluded that eye tracking methods could be a reliable way to quantify the severity of TBI.

While many studies used eye tracking data to differentiate the two groups, various studies used different approaches. Several studies used eye tracking data to detect impairments in TBI patients who are closely related to this study. [8] combined electroencephalogram (EEG) and eye tracking to assess mild traumatic brain injury. They have created tasks, some with high cognitive workloads and some with low workloads. They also generated 'virtual reality driving simulator'. Participants were asked to drive along the coastal driveway, and when a target appears, they are asked to shift their focus to the target while maintaining the lane position. Their preliminary study showed that brain injury does not always lead to observable performance deficits in TBI people. However, they found differences in saccadic performance between TBI and control groups. They also reported increased level effort in the TBI group while performing high cognitive workloads. Another study [9] used eye tracking data of the subjects (military service members) to study differential eye movements (saccades, fixations, smooth pursuits). The paper primarily used standard statistical tools such as mean, variance, and standard deviation to analyze the measures.

In [10], disconjugate eye tracking was used to measure the improved performance of concussion patients overtime during and after medical intervention. The study used an objective, rapid, noninvasive, quantitative algorithm for the assessment of brain-injured subjects. It hypothesized it could prove useful in tracking if a TBI patient is improving or not, especially in cases where the CT scan does not show any significant improvement. In [11], it was hypothesized that there is a deficit in smooth pursuit eye movements (SPEM) in mild TBI patients. The California Verbal Learning Test (CVLT-II) [12] was used to study the performance of the subjects in predictive smooth pursuit and cognitive functioning. This paper demonstrated that TBI patients exhibit deficiency in predictive SPEM, a variability of eye position, and correlation of these

impairments with cognitive impairments. In [13], the experiment was done on eye movement accuracy, quantification of the presence of abnormal eye movements, and reaction time in response to simple environmental stimuli with the help of indices of oculomotor [described in glossary 5.1] performance. I-Portal system and VEST Neuro-Otologic Analysis Software (Neurokinetics, USA) was used to evaluate all the experimental results. The result shows that the excessive amount of saccadic eye movement decreases the fixation point, which leads to impaired recognition in TBI people.

In [14], participants were instructed to interact with the approaching stimulus (soccer ball) while avoiding distractors (pandas heads and cleats). Stimuli traveled a total horizontal distance of 472 pixels. Mean saccadic velocity, mean saccadic amplitude, and the saccadic count were used as salient features. While saccadic velocity is slower and less accurate for people with Parkinson's disease, the eye typically travels farther and faster during a sport like a task for sports-related concussion. There was also a group difference between people having concussion and control for saccadic amplitude.

The study [15] could distinguish severe TBI & moderate TBI using eye tracking data. Participants were asked to track a white dot presented as target stimulus. The dot moved up and down in sinusoidal motion. The vertical smooth pursuit was used to separate the data. Nevertheless, it was not possible to distinguish mild TBI from the control group. ANOVA(analysis of variance) for smooth pursuit variance metrics revealed a significant difference between the groups. Smooth pursuit percentage was calculated as the participant's eyes follow the target within a target's velocity range. The logistic regression model [described in Chapter3] for smooth pursuit variance and smooth pursuit percentage metrics TBI and control groups.

Three research studies [10], [11], [7], included smooth pursuit and/or saccadic eye movements for analysis. These two eye movements have different functional areas and share common brain regions, i.e., brain areas involved in attention and executive functions. [16] found impairment in visual memory following TBI. The impairment

in the group was because of their impaired ability to initiate or utilize a strategy to facilitate their memory. Still images of animals and vehicles were used with fixed and free viewing conditions in this study. Free viewing facilitates the group to freely move their gaze, and fixed viewing mandates them to fixate on one position on the screen. People with TBI performed poorly in the free viewing test compared to the control group.

In [17], participants performed easy and difficult mental arithmetic tasks while fixating a central target. Change in pupil diameter and microsaccade magnitude appeared to discriminate task difficulty adequately. So, the features were used as salient for determining the magnitude of cognitive load on participants. [18] measured the parameters of eye movement while reading in subjects with TBI and found the parameters to be affected by TBI no matter the severity of the injury than controls. [19] showed TBI patients and healthy controls with photographs of male faces, and the result showed that TBI patients paid less attention to the given target and had less dwell time on them.

According to research in published papers, machine learning methods promise a new way of classifying impairment with eye-tracking videos. Machine learning plays a significant role in automated vehicles, medical fields, and many others. Machine learning research using eye tracking data also includes capturing driver's focus and attention while driving. The distracted drivers can be spotted analyzing their eye tracking data with machine learning [20]. While machine learning applications using eye tracking data is common in detecting driver's attention, there has been little research on machine learning to detect emotion recognition deficits. We have expanded our applications on eye tracking data by using the time series segment on video data to detect visual impairments. As an integrated part of this research, a fellow researcher used deep learning to detect facial landmark detection in the videos. Therefore, our research is at the forefront of combining 'Deep Learning' and 'Machine Intelligence' to this impairment classification.

Eye tracking analysis using machine learning is beneficial in detecting diseases and visual impairments. We can process a large number of data using machine learning efficiently. As we are exposed to new data, machine learning can adapt independently without any human involvement. The main property of machine learning is that it produces reliable and repeatable results. It also makes computations easier. So it is a widely preferred tool in many applications as mentioned before. In [21], they used machine learning to classify the subjects. The study selected three categories, including healthy people, brain injury patients, and vertigo patients. Random forest (RF)[3] classifier, a widely used machine learning algorithm, was used for its robustness and better performance than other machine learning algorithms. At first, eye movement images and information such as pupil position and area were extracted as original features. Secondly, those original features were used as training samples for long short-term memory (LSTM) networks to build classifiers, and the classification results of the samples are called evolutionary features. After that, multiple decision trees were built based on evolutionary features. Finally, an RF was constructed with these decision trees, and the results of disease classification were determined by voting. The study showed that advanced machine learning in the pathological analysis of eye movement has apparent advantages and good prospects.

While these studies, as mentioned above, provide promising direction in TBI detection, there has been little work on how people with TBI's eye tracking data changes over time while watching a video without any given direction. Many studies found gaze data, smooth pursuit, saccadic features as differentiation features to distinguish between TBI and HC. While the findings from our research correlated with the current study, this study used time-series data to detect impairment in TBI patients. Saccade, disconjugate eye, and changes in pupil diameter are used as salient features in discriminating people with TBI and control group. Our research applies the ML techniques in a new way, incorporating motion video and eye tracking features important for ToM and predicting the TASIT score that connects the emotion recognition to the eye movement. Other studies used ensemble differences in the groups on static

images or text cues, while this study uses movies where social inference is a significant factor. The previous studies were limited to static or dynamic images with nonverbal clues. There has been little work on directly predicting the TASIT score Table 1.1.

Table 1.1.: Comparison of the Approaches Taken by Existing Literature

Paper	**Approach**	**Result**
[7]	Horizontal and vertical self-paced saccades as a diagnostic marker after TBI, Used ANOVA and logistic regression over mean values of features while watching stimuli	Total 287 Participants; sensitivity 0.77 and specificity 0.78 for horizontal saccades
[15]	Vertical smooth pursuit as a diagnostic marker of traumatic brain injury, used ANOVA and logistic regression over mean values of smooth pursuit while watching stimuli	Total 92 participants; ROC value 0.772 with sensitivity 0.68 and specificity 0.73
[21]	Guided Eye tracking to generate smooth pursuit task (following a dot or scene), used LSTM as evolutionary features and random forest as classifier, a spot of red light moving along a specific trajectory to guide the subjects' eye tracking	60 patients (24 with Brain Injury and 36 with Vertigo) and 36 healthy participants; Accuracy rate for random forest is 0.96
This Study	Eye tracking over videos, used random forest and svm to predict standard test (TASIT) score using time series data to detect ToM impairments	Research found new associations between ToM impairments from video testing and the patient's eye movements; achieved f1 score of 0.73

1.2 Our Contribution

The question we will be answering in this research is if we can predict the TASIT score using eye tracking data, or is there any association between visual impairments and the patient's eye movements. We want to know the root cause of the visual cognition deficit. It is important to know why this impairment occurs in TBI patients

so that further research can be done on TBI patients for removing these deficits. Research studies show that 39% of TBI patients have impairment in facial recognition and other theory of mind (ToM) components (for example, inferences of other expectations or intentions). So in our research, we explore how TBI patients visually process data when trying to recognize someone's facial expression; then, we will be able to find the root cause of the impairments in TBI patients. Those deficits in TBI patients exist because of the way they interpret the scene. Past studies were restricted to processing the static images or dynamic images with no social clues of eye tracking data; we found that dynamic eye tracking data was crucial in predicting visual impairment.

This research's challenges are that we have a relatively small data-set (approximately 100 patients) to perform our research. Some traumatic brain injured patients had recovered from their injury many years ago—the average years since the TBI patients' injury is nine years. We also have a limited number of impaired patients. So the variation in the data is minimal to train a deep learning network. However, breaking up the raw video data into 3-sec chunks and finding a correlation with the target features for those broken up frames helped us realize that not all frames are important to be trained on the classifier. Our main contribution was to use video related eye tracking data and using TASIT scores for emotion recognition. These were instead challenging tasks as the eye movement varies widely from person to person, especially while watching movies, whereas the other studies used directed videos or images to guide patients' eye movement.

In brief, we created highly correlated frames of the dynamic features from eye tracking videos, and we gave these features as input to machine learning (support vector machine and random forest) for classification of impaired and unimpaired population. We successfully built a machine learning model to detect the impairment in visual processing using TASIT scores.

2. DATA PROCESSING STEPS FOR BINARY CLASSIFICATION OF IMPAIRED VS. UNIMPAIRED USING MACHINE LEARNING

Machine learning classification is a supervised learning technique in which the machine learns from the data given to it and then classifies new observations based on the learned parameters. This data set can consist of 2-class or multi-class data.

Classification is used to predict the classes of new data points. Classes are usually referred to as targets, labels, or categories. The result is a mapping function from input variables to output classes. In this chapter, we are classifying impaired and unimpaired patients as a binary classification problem. We have tested our data set on various types of classifiers to predict the impairment in the population. This chapter describes the TASIT test and eye tracking data and then provides the classification result.

2.1 Database Description

The dataset used in this study was created by Indiana University School of Medicine in a project titled "Examining determinants of negative attribution bias in people with traumatic brain injury". For creating this dataset, sample videos are shown to participants (TBI and HC) in a computer equipped with Tobii Studio eye tracking software [22]. At the beginning of the videos, participants are asked to pay attention to a person of interest. When the video is played, the subject's eye gaze data is recorded using the Tobii Studio. Tobii Studio can also detect if there was fixation or saccade [see glossary for descriptions 5.1] when the participants were looking at the videos and their duration in the gaze recording.

The dataset consists of three parts. PART1, PART2, and PART3 contain 28, 15, and 16 videos, respectively. The description of the parts are given below:

- PART1 – Emotion evaluation test: Actors showed one specific emotional state for each video from a total of seven states, which are angry, happy, surprised, revolted, anxious, neutral, and sad. The participants only had to detect which emotional states the actors were expressing for any particular video. Although there is no such thing as a neutral emotion, it is included in this dataset when the person in the scene was not strongly showing any of the other emotions.

- PART2 – Social inference (minimal): Some short scenes were shown to the participants. Each one lasts from 15 to 60 seconds. After a scene had been shown, participants were asked four simple questions.

  - A.What they think someone was doing to the other person
  - B.What they think someone was trying to say to the other person
  - C.What they think someone was thinking
  - D.What they think someone was feeling

  The questions were set in a way so that each time they only needed to answer among these three options: Yes, No, or Do not Know. However, they were encouraged to answer only Yes and No.

- PART3 - Social inference (enriched): Actors simulated relatively complex social interactions. The participants were asked the same set of questions as PART2 for this part.

These videos are shown to both people with TBI and HC. There were, in total, 122 participants. Nevertheless, some of the participants' eye tracking data were missing from the dataset. If missing data percentage was more than 90% for any participant, the data was dropped for that participant. Missing data means eye gaze co-ordinate is missing for more than 90% of the time.

There were 65 male and 57 female participants of different ages ranging from 18yrs to 74yrs. Their education years range from 11yrs to 25yrs. The highest education level is the doctoral-level degree, and the lowest is no diploma with 11yrs of education. TBI patients post amnesia days varies from less than 1hr to greater than 60 days.

The TASIT test is straightforward for people with a standard range of social skills while being difficult for people with TBI or with a social perception deficit. TASIT scores for individual participants are calculated by summing up all the correct answers for all parts (PART1, PART2, PART3). In comparison, people with TBI have difficulty detecting some emotions like sarcasm, lies, angry, revolted. While they have equivalent performance in detecting other emotions such as neutral, happy, and sad. That is why in this study, instead of using ensemble scores of the test, individual video scores are used for classification.

(a) (b) (c)

(d) (e) (f)

Fig. 2.1.: Sample Frames from (a) & (b) PART2 Video13, (c) & (d) PART3 Video1, (e) & (f) PART3 Video11

Among these, the PART3 videos incorporate the most complex emotions and have a higher variability of TASIT scores. The anonymized patient eye-tracking data from the Tobii system was tabulated and provided by a previous project. For our initial binary classification work, we take the answers to the TASIT test questions, and

if all of the answers for that particular video are correct, then the patient is 1=not impaired; otherwise, with an incorrect answer, they are considered 0=impaired. Then, we divided the PART3 videos into 3-second intervals, and for each of the 3 seconds, mostly correlated features were extracted from the eye-tracking data.

2.2 Feature Selection Method

Out of all the PART3 videos, video 13 & video 10 has the most variation in the whole dataset of impaired and unimpaired patients. Therefore we used video 13 & video 10 for our research.

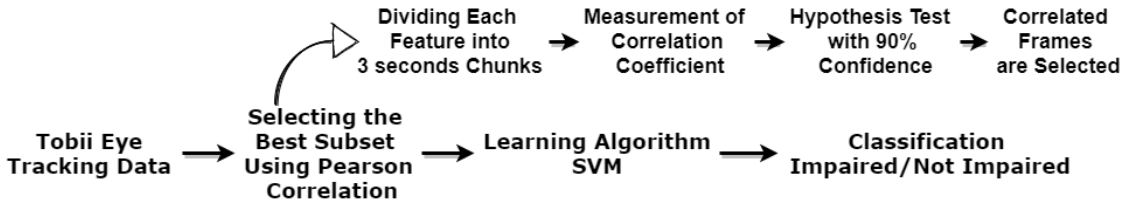

Fig. 2.2.: Flowchart for Feature Selection Method

The data used to train machine learning is extremely important. If the data is noisy, machine learning will perform no better than random guessing. All features need not be used to train machine learning algorithms, as every feature does not correlate to the target variables. Feature selection also makes the training faster and reduces over-fitting.

At first, data are divided into 3-sec chunks for each feature to observe the correlation with the target variables Figure 2.2. If the correlation is strong enough, then a null hypothesis test with 90% confidence will be rejected by observing P-values where the null hypothesis assumes no correlation with the target variables. P-values close to 0 pertains to a significant correlation in correlation coefficients and a low probability of observing the null hypothesis. If each variable has N observations, then the Pearson correlation coefficient is defined as,

$$\rho_{AB} = \frac{1}{N-1}\sum_{i=1}^{n}(\frac{A_i - \mu_A}{\sigma_A})(\frac{B_i - \mu_B}{\sigma_B}) \tag{2.1}$$

μ_A and σ_A are the mean and standard deviation of A, respectively, and μ_B and σ_B are the mean and standard deviation of B.

Only the higher correlated frames containing 3-sec of feature data are combined for the features that showed a significant correlation. After normalizing the features, principal component analysis (PCA) was performed to reduce further dimensions in the data. PCA is the most common method for feature engineering for traditional machine learning methods for reducing high dimensional data to a manageable one. PCA keeps most of the information of the original data. It makes the data analysis more straightforward and more manageable than working with extensive data.

The PCA is a combination of standardization and eigenvalue decomposition of the data. Standardization helps to make sure each variable contributes equally to the analysis. Each variable is converted to a close range to prevent any bias in the results. Mathematically it is done by,

$$X = \frac{X - mean(X)}{StandardDeviation}$$

As we know, principal components retain most of the variation in the data. PCA aims to understand how the variables are correlated with each other and exclude the highly correlated ones so the data dimension can be reduced. Removing the highly correlated variables will remove redundant information from the data. The Co-variance matrix (n x n; where n is the dimension of data) is calculated for the whole dataset to determine the correlation between variables (n-dimensional). If the sign of co-variance value is positive, then the two variables are correlated, and if the sign is negative, they are inversely correlated. Co-variance is calculated by,

$$cov(X,Y) = \frac{1}{N-1}\sum_{i=1}^{N}(X_i - \mu_x)(Y_i - \mu_y) \tag{2.2}$$

In simpler terms, the co-variance matrix helps in summarizing the correlations between all pairs of variables. The maximal amount of variance in data needs to be found to determine the data's principal components. For this purpose, eigenvectors and eigenvalues are calculated from the co-variance matrix. Let A be a square matrix, a vector and a scalar that satisfies $Av = \lambda v$, then λ is eigenvalue corresponding to eigenvector v of A. The eigenvalues are actually the roots of the equation $\det(A-\lambda I) = 0$. The eigenvector corresponding to the largest eigenvalue captures the highest co-variance in the data. The corresponding eigenvector captures the second highest variance to the second largest eigenvalue—that way, all the principal components can be calculated from the data. The eigenvectors are sorted by decreasing eigenvalues, and k eigenvectors are chosen with the largest eigenvalues to form a N x k dimensional matrix where N is the dimension of data. This N x k eigenvector matrix is used to transform the samples onto the new subspace using $y = W'x$ where W' is the transpose of the matrix W.

2.3 Features Used After Down Selection:

Eye tracking features are provided by the Tobii eye tracker, which includes gaze data, saccadic amplitude, relative saccadic direction, fixation co-ordinates, etc. Other features are calculated using the original features such as disconjugate eye, vertical error, horizontal error, distance measure from facial landmarks, etc.

For finding the facial landmarks in videos, Dlib, a cross-platform software library, was used. Sixty-eight facial landmarks detector in Dlib was applied to detect the facial landmarks in the videos. However, the default facial landmarks model does not include the forehead. That is why 68 facial landmark model is modified to 75 facial landmark model to include the forehead Figure 2.3. As the whole idea of emotional expressions can be obtained by seeing a person's whole face, that's why forehead landmarks points are essential in this case Figure 2.4.

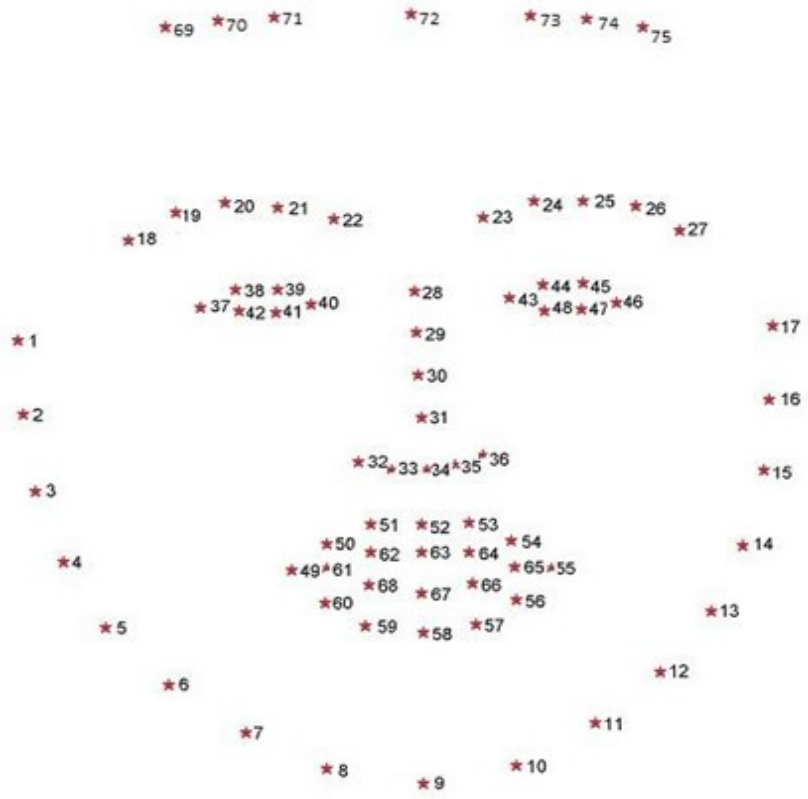

Fig. 2.3.: 75 Facial Landmark Model

The Tobii saccadic measures are calculated based on the fixation locations. It is a visual angle measured in degrees between the previous fixation location and the current fixation location.

- Saccadic Amplitude

  Saccadic amplitude is the distance in degrees (angle) between the previous fixation location and the current fixation location [22]. The Saccadic amplitude is shown in the Figure 2.5.

Fig. 2.4.: Some Frames of Videos after Drawing 75 Facial Landmark Model

- Absolute Saccadic Direction

  The absolute saccadic direction measures the difference in angles between the current fixation location and the horizontal axis. It is calculated based on the fixation locations, as defined by the fixation filter Figure 2.6. [22]

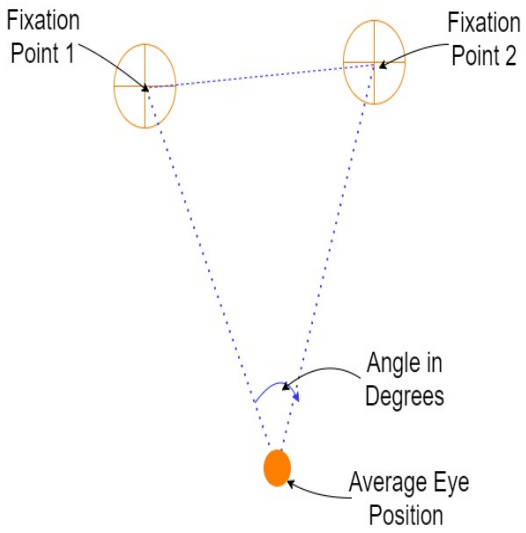

Fig. 2.5.: Saccadic Amplitude

- Relative Saccadic Direction

  The difference in angles between the absolute saccadic direction of the current and the previous saccade is called relative saccadic direction. It is calculated based on the fixation locations, as defined by the fixation filter Figure 2.6. [22]

- Pupil Left and Right Eye

  This feature is the estimated size of the pupil of the left and right eyes. The measure comes from the Tobii eye tracker. If one pupil is more dilated than the other, it can sign acute concussion or brain injury. That is why it can be a differentiating factor for impaired patients.

- Vertical Error

  The vertical error is defined by difference in left and right eye Y coordinates.

$$Verticalerror = Gaze_{yright} - Gaze_{yleft}$$

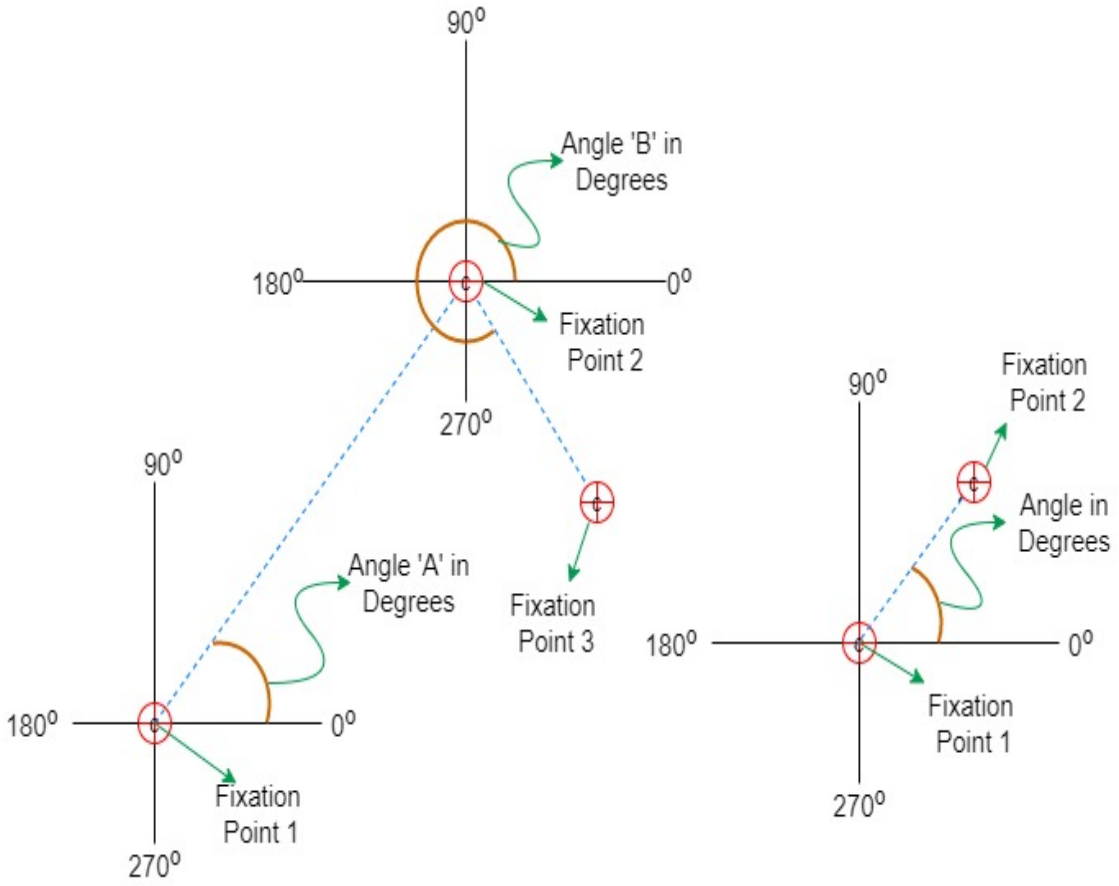

Fig. 2.6.: Relative Saccadic Direction (Left), Absolute Saccadic Direction (Right)

- Horizontal Error

  The horizontal error is defined by the difference in left and right eye X co-ordinates.

$$Verticalerror = Gaze_{xright} - Gaze_{xleft}$$

- Distance Measure The distance between gaze data and each of the facial land-mark points are measured. Only the minimum distance and the corresponding nearest landmark are kept as a feature.

- Disconjugate Eye

  Disconjugate gaze is a measure of the failure of eyes to turn together in the same direction. That means the eyes are not paired when viewing a scene. It is measured by subtracting differences in the right eye's X/Y co-ordinates of consecutive frames from differences in left eye's X/Y co-ordinates of the same consecutive frames.

- Executive Functioning Data

  We use executive functioning data such as animal fluency and letter fluency score as part of the features. The animal fluency test requires the patient to name as many animals as possible within a given 60 second period, whereas the letter fluency test requires them to name as many words as possible for the given letter within a specific time.

The features used in this study are the ones correlated to the target variable, which is the video13 TASIT score. Not all frames are used for these features. Only the correlated frames for which correlation coefficients are significant are used as predictors for the classifier.

2.4 Conclusion

The focus of this chapter is to show the preprocessing steps of eye tracking data. The Tobii eye tracker records eye gaze in the frequency of 500 Hz. The time-series data vary from 4.5k to 18k frames, which is enormous. To make it more manageable and for the extraction of meaningful data, several preprocessing steps are used. The broken-up frames were beneficial to improve the quality of training, as shown in the following chapters, when classifier performance will be discussed. The use of all frames was introducing noise to the data-set, and the classifier has not been able to classify the classes correctly. However, after truncation, the separation is visible, and the classifier's performance improved significantly.

3. MODEL SELECTION AND OPTIMIZATION

Model selection and optimization is a vital part of machine learning. Without an optimized model, the machine learning model will either underfit or overfit the data. We have to ensure the machine learning model we chose for our data gives the best result on the validation data and the unseen data. The fine-tuning of the parameters and observing a model's performance on unseen data is essential in this regard.

3.1 Outlier Detection

Noise and outliers are problematic and affect the performance of machine learning, especially for small data sets. For getting sensible models, cleaning up data is a very crucial step. The usual machine learning methods are not optimized to detect outliers; instead, they are built for detecting normal instances. The isolation Forest algorithm is built to provide an efficient way to detect outliers successfully.

The algorithm's focus is to "isolate" anomalies by creating a forest of random trees using random attributes. The random partitioning produces significantly shorter paths for outliers. Splits happen at random on a random attribute while building a decision tree. The total number of splits determines the level at which the isolation happened. The same process repeats multiple times, and the average number of splits are taken over multiple decision trees. It will provide the anomaly score based on the average number of splits for a given instance. The instances which have higher anomaly scores are labeled as outliers.

The instances are considered outliers if the score is close to 1; they are relatively safe to be regarded as normal instances if the score is significantly lower than 0.5, then, and if all the instances return around 0.5, then the entire sample does not have any distinct anomaly.

3.2 Hyperparameter Optimization

Machine learning models are defined by parameters that automatically are estimated from training data and also by the hyperparameters. Hyperparameters are a part of the model's initial structure and need to be manually tuned. The tuning of machine learning models is one kind of optimization problem. With the right combination of the hyperparameters, the minimum loss or the function's maximum accuracy is successfully found. The optimization is also essential in comparing different machine learning models trained on a dataset.

3.3 Algorithms

There are many states of art machine learning algorithms available for classification problems. However, we use a supervised learning technique to learn the input-output examples' mapping function to predict the output based on new input data. For our case, we are using a Support Vector Machine (SVM) and random forest (RF), which are the widely used binary classifiers and give the best result. This chapter includes other classifiers' results also for comparison.

- Naive Bayes Classifier

  A Naive Bayes classifier is a probabilistic machine learning model based on the Bayes theorem.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \tag{3.1}$$

  The probability of event A (classes) occurring given event B (predictors) has occurred on event A and the probability of event B occurring given event A. The denominator is irrelevant for our purpose as it will always be the same in all conditions, and thus proportionality can be introduced. One assumption is that the predictors are independent of each other, which means the presence of a

particular feature does not correlate to any other feature. Another assumption is that the features have equal weights on the outcome. Though this algorithm works well for a vast data set, the other algorithms outperformed naive Bayes for our dataset.

- k-Means Clustering

  The k-means clustering algorithm is an unsupervised machine learning algorithm. It divides the data points based on a fixed number of clusters. It assumes the centroids (centers) of the clusters and assigns individual data points to a cluster based on the distance from the centroid of the clusters to that data point. After assigning the data points to one of the centroids, the centroids are again recalculated. The whole process iterates until all the data points are assigned to one of the centroids. k-means clustering helps us know the actual organization of the data. Nevertheless, it works more poorly than other supervised learning techniques.

- KNN (K Nearest Neighbor)

  It assumes that similar types of classes stay near to each other. It uses similarity measure or distance. Initialization of K (number of neighbors) is required. According to the distance measure, it sorts the instances, and the first K instances are taken from the sorted collection. Then the labels of these K entries are selected, and the mode of labels is returned for classification problem. Choosing the right K is essential to reduce classification errors.

- Logistic Regression (Predictive Learning Model)

  Logistic regression fits 'S' shaped logistic function. It uses maximum likelihood to select the curve. It is a probabilistic method for classifying a data set in which one or more independent variables determine an outcome. It assigns the probability of data belonging to a class. It provides a quantitative measure that is also suitable for the regression problem.

- Decision Trees

  The decision tree divides the data into a tree-like structure. It breaks the data into smaller parts in an iterative manner using the features associated with the data, and at the same time, the decision tree is gradually developed. To start with building a tree, Gini impurity is first calculated for every feature. Gini impurity is defined by,

$$GiniImpurity = 1 - P^2_{Class1} - P^2_{Class2} \tag{3.2}$$

  Where P_{Class1} , P_{Class2} are class 1 , 2 probabilities, respectively.

  The feature associated with the lowest Gini impurity ends up in the root of the tree. So the root will have a feature with the lowest Gini impurity that means the lowest probability of misclassifying a class. After that, Gini impurity is again calculated on each side of the tree for the rest of the features. The lowest ones take up the nodes and continue like this until it reaches the leaves. In this way, the tree ends up with a tree with decision nodes (features) and leaf nodes (classification). The decision node can have two or more branches depending on the classes.

- Random Forest (RF)

  Random forests are better versions of decision trees for classification, starting by building many decision trees while training and assigning the class that is the classes' mode. Many relatively uncorrelated trees combined improve flexibility resulting in a vast amount of accuracy than any of the individual constituent models. It creates a bootstrapped dataset (sampling randomly with replacement from the original dataset of the same size) for making trees. Instead of selecting all the features, it randomly selects a subset of features and builds the tree in a conventional manner. This process is repeated using another bootstrap dataset to build a different tree. This whole method is iterated 100 or more times to

create a handful of trees to predict the instances. As the prediction does not come from a single tree but a different variety of trees, it reduces over-fitting and results in an overall improvement in the accuracy.

- Support Vector Machine (SVM)

  A support vector machine creates a hyperplane in high dimensional space, which can be used for classification or regression. For linearly separable training data, two parallel hyperplanes separate the two classes of data and, at the same time, try to maximize the distance between the hyperplanes. The kernel functions in SVM only calculate relationships between each data points as if the data are in a high dimensional place. This kernel trick enables algorithms to function in the high dimension without calculating the coordinates in that plane. It helps to separate data in a high dimensional space. We are using a Radial Basis Function (RBF) kernel with free parameter gamma.

  In SVM, the trade-off is between minimizing training error and minimizing model complexity (The parameters of the Kernel function can be chosen from linear to high dimensional feature space, the model complexity increases exponentially from linear to high dimensional feature space). SVM parameters are optimized to minimize both the complexity and the error at their optimum level.

3.4 Voting Scheme

A voting ensemble is an ensemble machine learning model that combines the predictions from other models. It is a scheme used to improve model performance, ideally achieving better performance than any single model used in the ensemble. The voting scheme uses different model structures and gives different weightage to each of the models to get the best out of each model Figure 3.1.

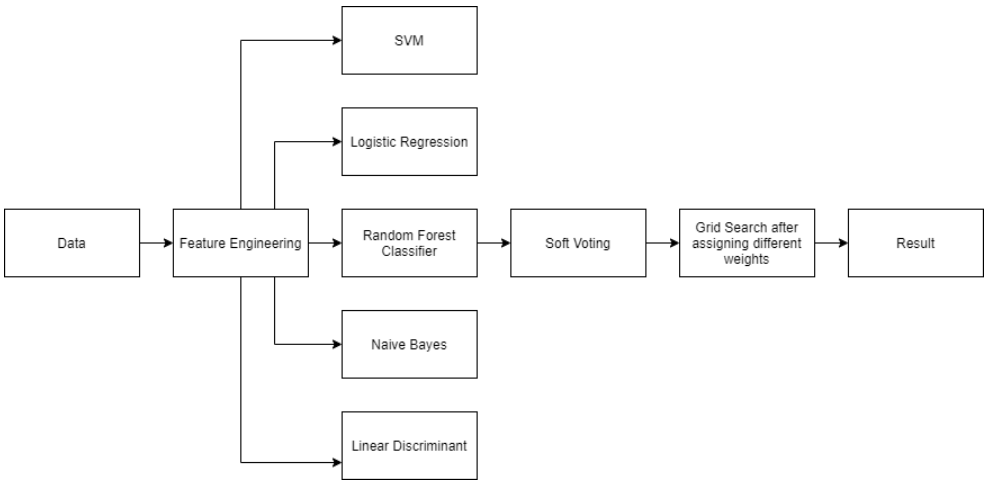

Fig. 3.1.: Flowchart of Voting Classifier

3.5 Description of Classification Metrics

After training the model, the most crucial part is to evaluate the classifier to verify its applicability. The following are the most used and effective measures to know if the trained model gives improved results.

- Cross-validation

  Over-fitting while training is the most common problem in machine learning. The k-fold cross-validation method helps us to determine if the model is over-fitted or not. It divides the dataset into k mutually exclusive subsets, and one set is opt-out for testing during training. This process goes on until all the folds get tested.

- Precision and Recall

  Precision is the ratio of relevant instances among the retrieved instances, whereas recall is the ratio of relevant instances retrieved among the total relevant instances. These measures help in quantifying how well a classifier is on minority class in case of imbalanced data.

$$Precision = \frac{TruePositives}{TruePositives+FalsePositives}$$

$$Recall = \frac{TruePositives}{TruePositives+FalseNegatives}$$

- F1-Score:

  The F1-score is the harmonic mean of precision and recall. The range of F1-score can be between 0 to 1, where 1 means perfect precision and recall and 0 being the worst score.

$$f1 = 2 * \frac{Precision.Recall}{Precision+Recall}$$

- ROC curve (Receiver Operating Characteristics)

  The ROC curve is another measure of validation of a model. ROC curve shows the trade-off between the true positive rate (tpp) and the false positive rate (fpr). The model with correctly classified data will have ROC value 1. We can tune our model to have the best combination to maximize tpp and minimize fpr by selecting an optimum threshold from the ROC curve. The point which gives the minimum distance from the ROC value of 1 is our optimum threshold.

4. BINARY TASIT SCORE PREDICTION USING MACHINE LEARNING

In this chapter, we are predicting the TASIT Score instead of classifying TBI and HC. TASIT Score for all three parts is different for each patient. There is a score for every video/episode for each participant, calculated from how many correct answers they gave for that particular video. Not every participant does well in all episodes. There is a variation of scores within even HCs in answering the questions. Depending on the test's difficulty level, some participants do well on the test, and some do not. In PART1, participants had a hard time differentiating revolted and angry. They got confused if the actor/actress was showing revolted or angry emotions. Sometimes they also mix up in differentiating happy and surprised. So detecting emotional cues is not always easy for even HCs. We set ground truth data based on the TASIT score. We are using individual video scores to detect impairment in participants for that particular video as described in Chapter 2, page 12.

As we know, people with TBI will not necessarily have visual defects. Our focus is on classifying the impaired in visual from the unimpaired population. That directly correlates to the TASIT Score. We want to predict the TASIT Score using eye tracking data. As we are trying to detect visual impairment in the population, so we will be focusing on reducing the false-negatives as part of an improvement in our research. The rest of the chapter shows how we progressed and improved our impairment detection over each of the experiments.

4.1 Classification Using All Parts

Videos from all parts are used in this section. The features from Table 4.1 used for this experiment. We normalized the features before feeding the data to a classifier.

Also we applied a simple imputer to impute the missing values in the data. The recursive feature elimination method is used for selecting the best subset of features. 5-fold cross-validation (CV) is used to select the best hyperparameters. The train test ratio for this experiment is 0.75/0.25. The classes are unimpaired (got all answers to TASIT questions correctly) vs impaired (got at least one answer wrong).

4.1.1 Features Used

In this experiment, we have used engineered features from eye tracking data. All of the parts including PART1, PART2 & PART3 data was used to classify impairment based on the TASIT Score.

Table 4.1.: Engineered Features for All Parts

Features
Number of Saccade
Average Saccadic Duration
Percentage of Saccade
Average Horizontal Error
Average Vertical Error
Number of Fixation
Average Fixation Duration
Percentage of Fixation in Face
Percentage of Fixation in Eye
Percentage of Fixation in Mouth
Percentage of Fixation in Forehead
Average Distance from Intersection of Eye and Nose
Average Saccadic Amplitude
Average Saccadic Direction
Average Relative Saccadic Direction
Average High Frequency Data of Gaze Point
Average High Frequency X Coordinate Data of Right Eye

4.1.2 Result

For all parts, splitting of train and test data was done according to a 50/50 ratio. 5-fold cross-validation (CV) is used to test the model. CV is used to prevent any over-fitting and selection bias during training. The goal of a CV is to predict how a model will perform on totally unseen data. CV uses the data to tune model hyperparameters and returns model performance. CV gives us insight into how a model will perform in a generalized independent set.

Cross-validation accuracy for all parts is 0.69, but f1 score for impaired class is poor with f1 score of 0.54, as we can see in the Figure 4.1 and Table 4.2. That is why we trained our model to predict for the three parts separately to improve the result on the minority class as we will see in Section 4.2.

Table 4.2.: Classification Results only on Testing Dataset for All Parts

	precision	**recall**	**f1-Score**	**support**
Impaired	**0.52**	**0.57**	**0.54**	497
Unimpaired	0.78	0.74	0.76	1022
Weighted Average	0.70	0.69	0.69	1519

4.2 Classification Separately for PART1, PART2 and PART3 Videos

Videos from different parts are used separately to predict the impairment. This experiment shows that videos from PART3 results are more correlated to TASIT Score than other parts. The features used in the experiment are the same as the Table 4.1. 5-fold CV is used to prevent any over-fitting and selection bias during training.

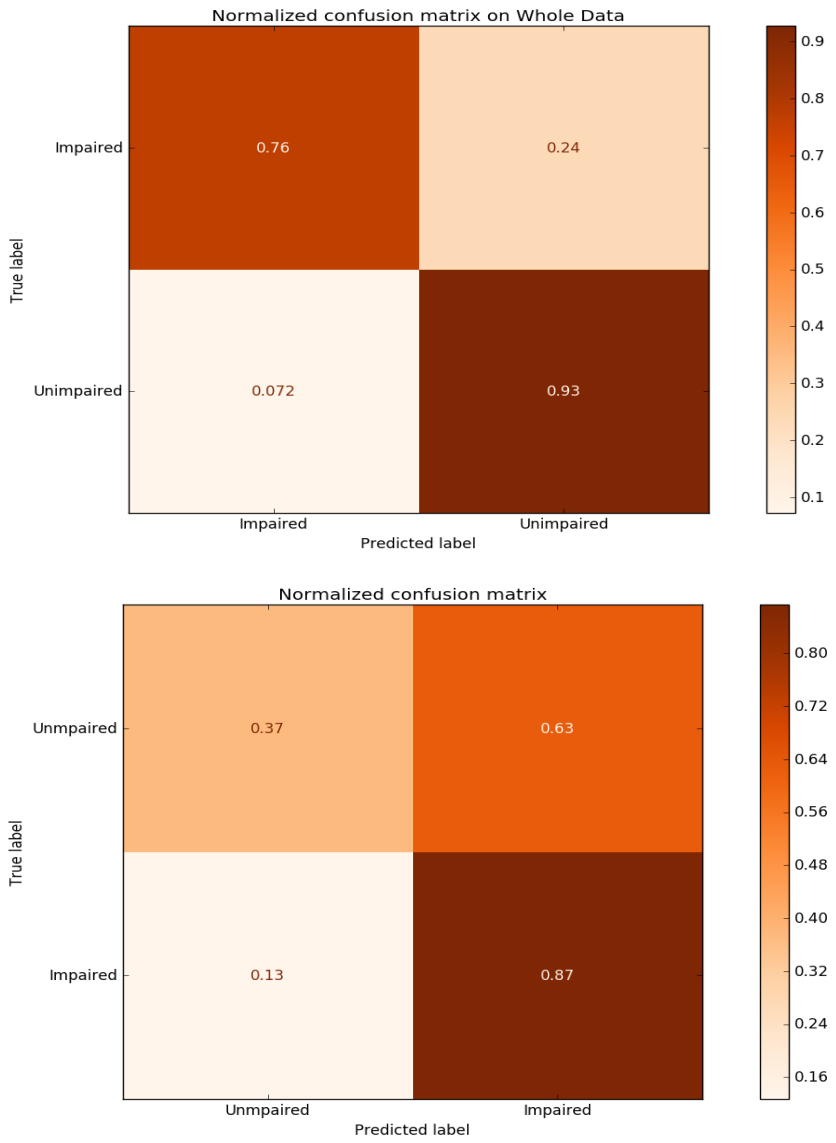

Fig. 4.1.: a) Confusion Matrix All Parts Whole Data b) Confusion Matrix All Parts Test Data

4.2.1 PART1 result

For PART1, splitting of train and test data was done according to the 80/20 ratio. A voting classifier is implemented for this experiment. Soft voting is used for detecting

the classes. A weighted voting scheme is implemented in this regard Chapter 3. The f1 score for impaired population for PART1 on test data is 0.40 Figure 4.2.

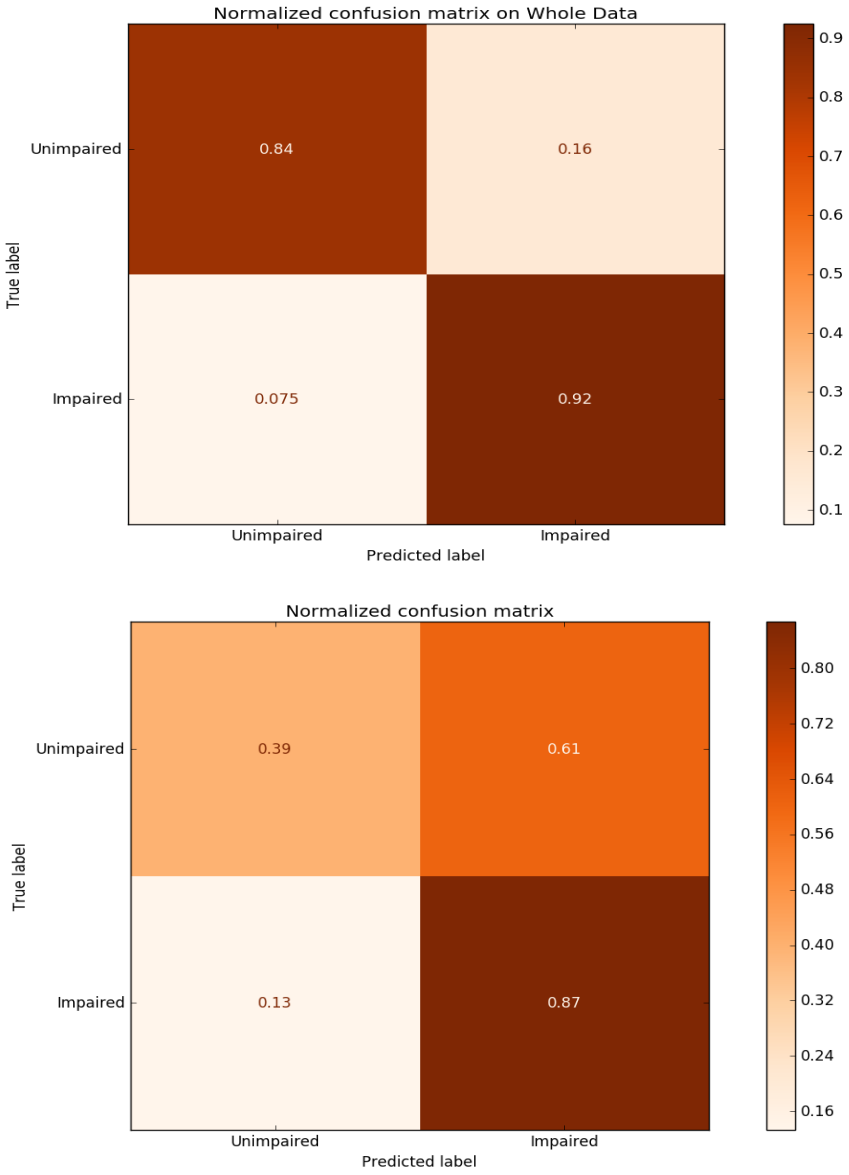

Fig. 4.2.: a) Confusion Matrix PART1 Whole Data b) Confusion Matrix PART1 Test Data

Table 4.3.: Classification Results only on Testing Dataset for PART1

	precision	**recall**	**f1-Score**	**support**
Impaired	**0.40**	**0.39**	**0.40**	104
Unimpaired	0.87	0.87	0.87	466
Weighted Average	0.78	0.78	0.78	570

4.2.2 PART2 result

For PART2, splitting of train and test data was done according to the 80/20 ratio. A voting classifier is implemented for this experiment. Soft voting is used for detecting the classes. A weighted voting scheme is implemented in this regard. Cross-validation score of a voting classifier for PART2 with test data is 0.65. The f1 score for impaired population for PART2 on test data is 0.32 Figure 4.3.

Table 4.4.: Classification Results only on Testing Dataset for PART2

	precision	**recall**	**f1-Score**	**support**
Impaired	**0.29**	**0.36**	**0.32**	67
Unimpaired	0.81	0.76	0.79	247
Weighted Average	0.70	0.68	0.69	314

4.2.3 PART3 result

For PART3, splitting of train and test data was done according to the 80/20 ratio. A voting classifier is implemented for this experiment. Soft voting is used for detecting the classes. A weighted voting scheme is implemented in this regard. Cross-validation score of a voting classifier for PART3 with test data is 0.70. The f1 score for impaired population for PART3 on test data is 0.58 Figure 4.4.

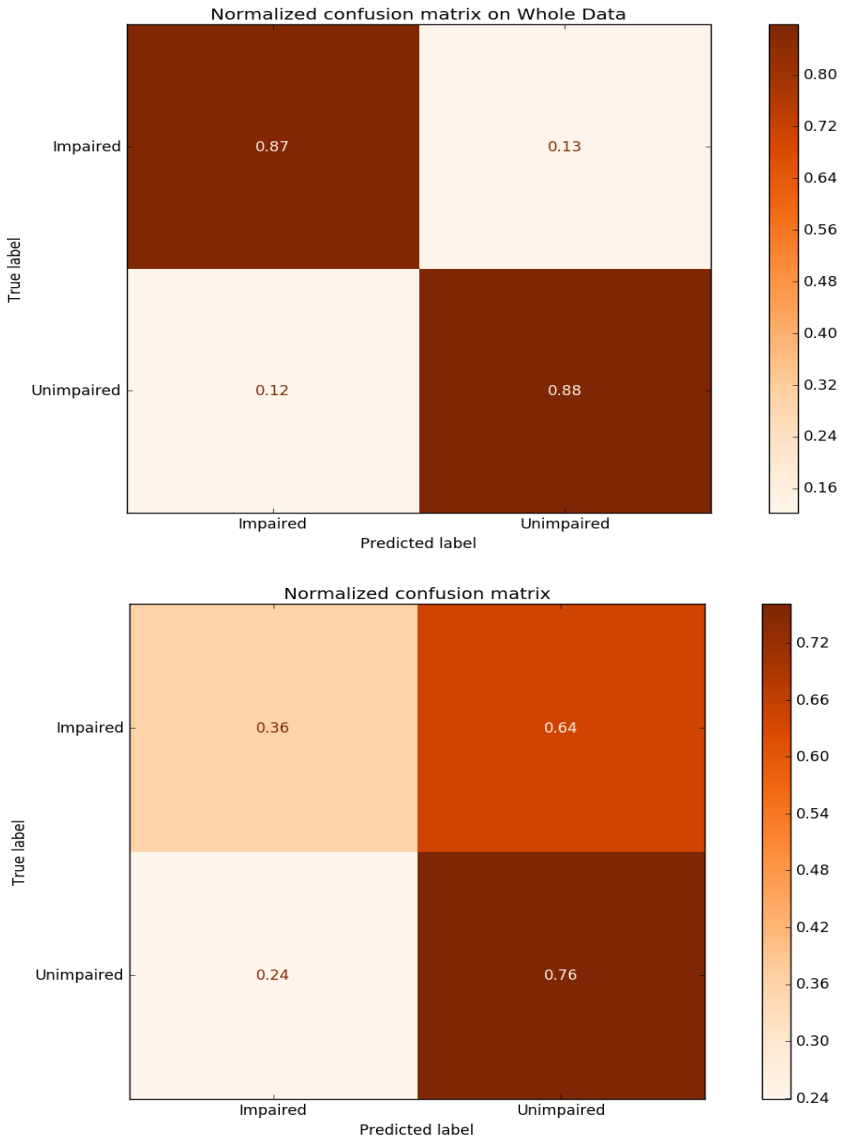

Fig. 4.3.: a) Confusion Matrix PART2 Whole Data b) Confusion Matrix PART2 Test Data

Comparing the precision, recall and f1-score Table 4.3, 4.4, 4.5, we can say PART1 and PART3 are more correlated to TASIT score than PART2. Among them, PART3 gives the best result for impaired population.

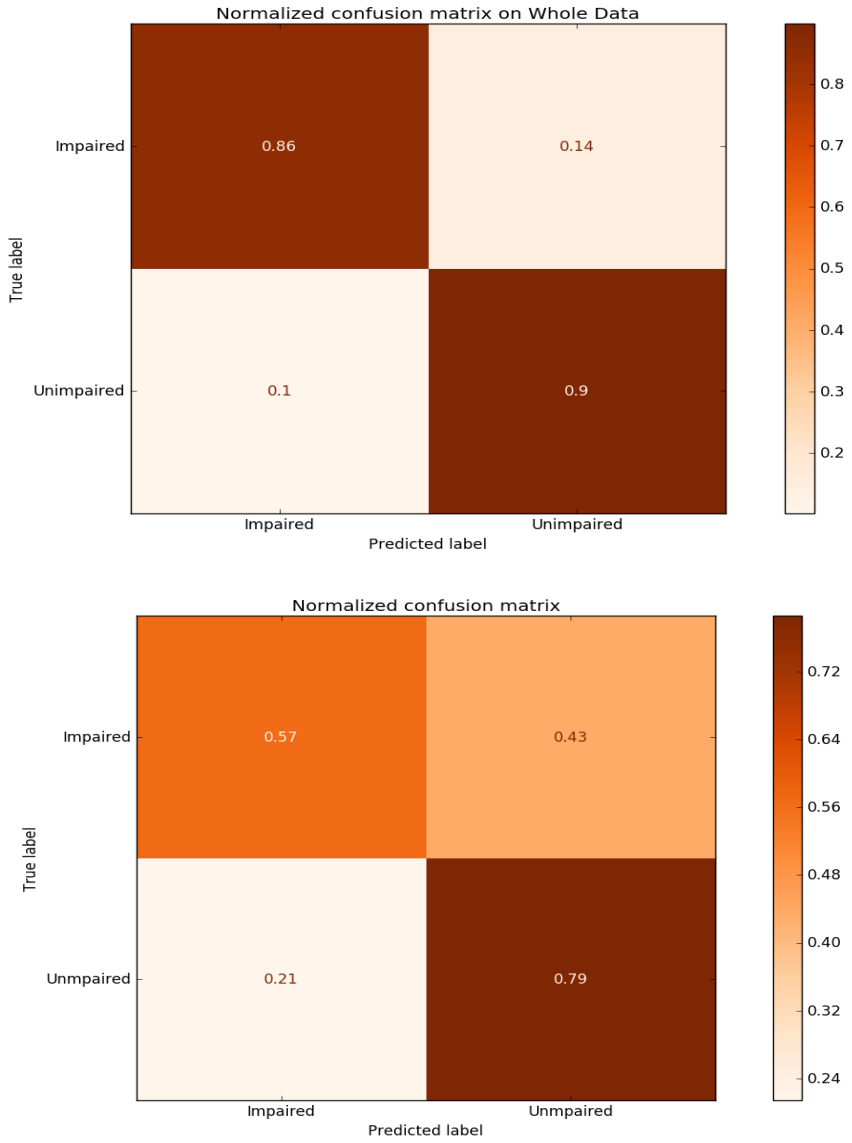

Fig. 4.4.: a) Confusion Matrix PART3 Whole Data b) Confusion Matrix PART3 Test Data

4.3 Video Wise Prediction

We took each video data separately to train our model. It helped us to determine which videos play an essential part in the TASIT test. We only took the video,

Table 4.5.: Classification Results only on Testing Dataset for PART3

	precision	**recall**	**f1-Score**	**support**
Impaired	**0.61**	**0.57**	**0.58**	122
Unimpaired	0.76	0.79	0.77	210
Weighted Average	0.70	0.70	0.70	332

which had a significant amount of impaired population compared to the unimpaired population. Otherwise, machine learning will not have the data to train on.

We found that videos that show surprise, or sarcasm; those video TASIT scores have a high correlation with eye tracking data. So we took Video10 (Sarcasm), Video13 (lie) from PART3, for our experiment, which also had the highest impairments among the population.

4.3.1 Result for PART3 Video13

The whole video of Video13 from PART3 is used separately to predict the impairment. The features used in the experiment are the same as the Table 4.1. 5-fold CV is used as same before to prevent any over-fitting and selection bias during training Table 4.6.

Table 4.6.: Classification Results only on Testing Dataset for PART3 Video13 Whole Video

	precision	**recall**	**f1-Score**	**support**
Impaired	**0.64**	**0.64**	**0.64**	11
Unimpaired	0.80	0.80	0.80	20
Weighted Average	0.74	0.74	0.74	31

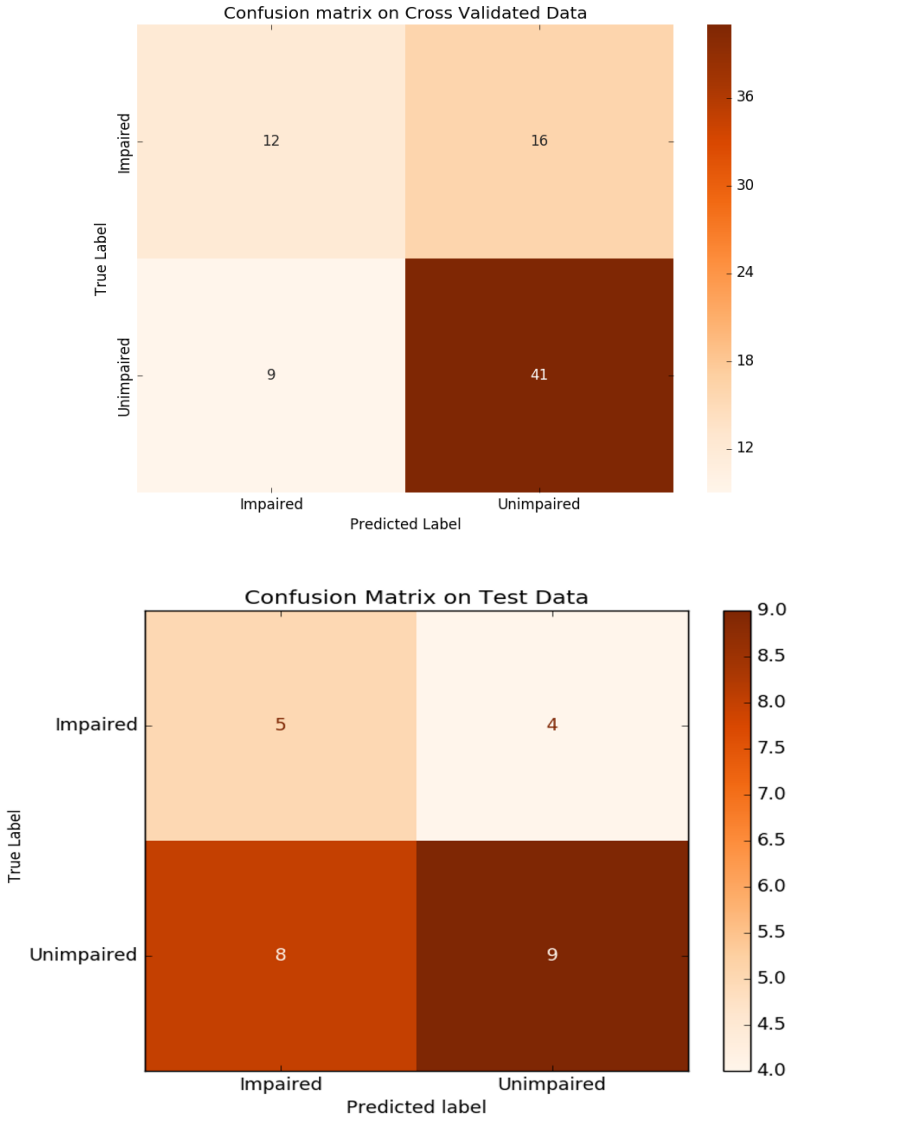

Fig. 4.5.: a) Confusion Matrix PART3 Video13 Cross Validation Data b) Confusion Matrix PART3 Video13 Test Data

4.3.2 Result for Part3 Video10

The whole video of Video10 from PART3 is used separately to predict the impairment. The features used in the experiment are the same as the Table 4.1. 5-fold CV

is used as same before to prevent any over-fitting and selection bias during training Table 4.7.

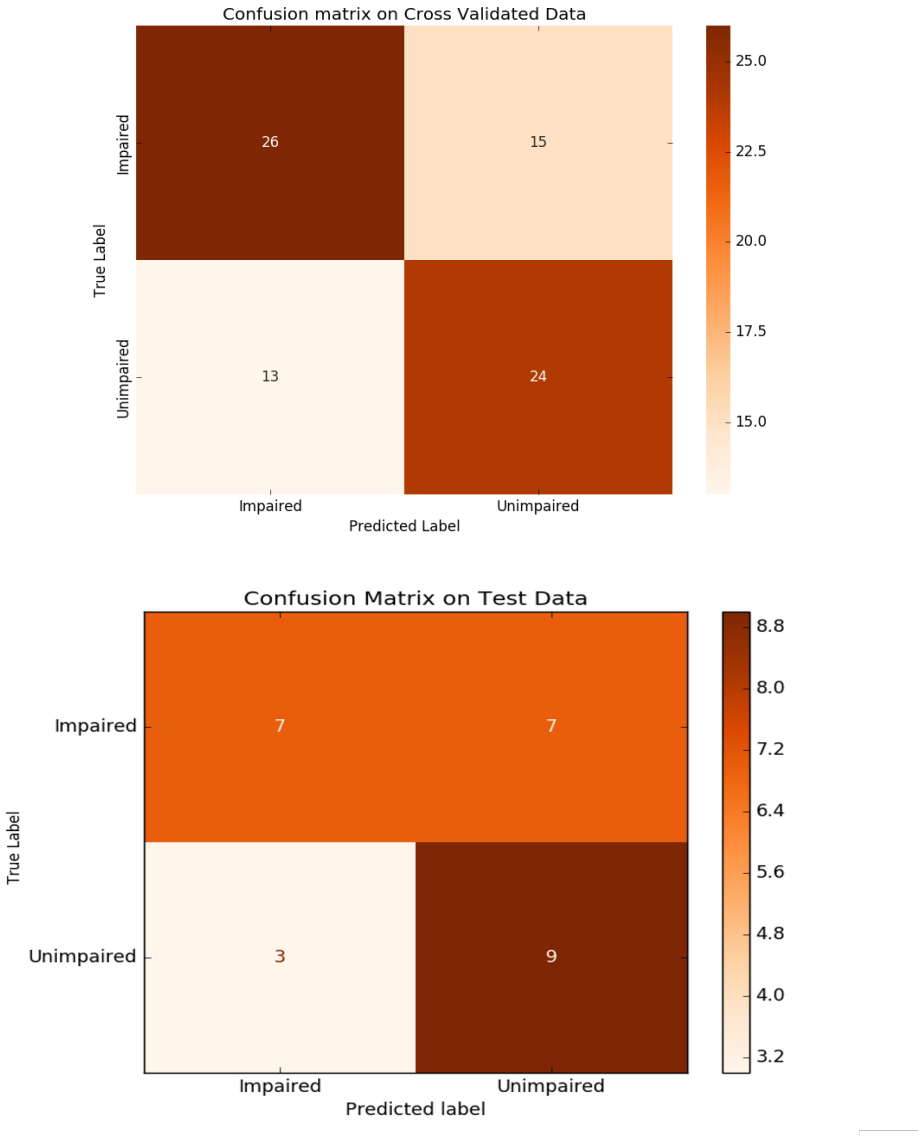

Fig. 4.6.: a) Confusion Matrix PART3 Video10 Cross Validation Data b) Confusion Matrix PART3 Video10 Test Data

Table 4.7.: Classification Results only on Testing Dataset for PART3 Video10 Whole Video

	precision	**recall**	**f1-Score**	**support**
Impaired	**0.70**	**0.50**	**0.58**	14
Unimpaired	0.56	0.75	0.64	12
Weighted Average	0.64	0.62	0.61	26

4.4 Dividing Video Data into 3-sec Chunks

The next experiment divided the video into 3-second blocks. In addition, the feature data was no longer averaged, now the eye tracking features were presented to the machine learning as vectors of time data instead of taking the whole video data. The saliency is determined based on model performance. We included the vectors that improved the model performance and excluded those which results in decreased performance.

4.4.1 Result for Salient Part of Part3 Video13

We took PART3 Video13 as part of our experiment as it had a significant amount of impaired population-based on TASIT Score (less than four answers correct - impaired, four answers correct - unimpaired). The salient part of the video13 is given in the Table 4.8 and the classification result is shown in the Figure 4.9, Table 4.9.

4.4.2 Result for Salient Part of Part3 Video10

We took PART3 Video10 as part of our experiment as it also had a significant amount of impaired population-based on TASIT Score (less than four answers correct - impaired, four answers correct - unimpaired). The salient part of the video10 is given in the Table 4.10. The classification result is shown in the Figure 4.10, Table 4.11.

Table 4.8.: Salient Part of The Engineered Features for PART3 Video13

Features	**Video Segment**
Horizontal Error	6.5th sec - 16.5th sec
Relative Saccadic Direction	30th sec - 33th sec
Distance from Fixation to Centroid of Whole Face	33th sec - 36th sec
Vertical Error	20th sec - 30th sec
Disconjugate Eye	6.5th sec - 13th sec
Distance from Fixation Point to Nearest Landmark	13th sec - 16.5th sec
Fixation Coordinate Y	6.5th sec - 16.5th sec

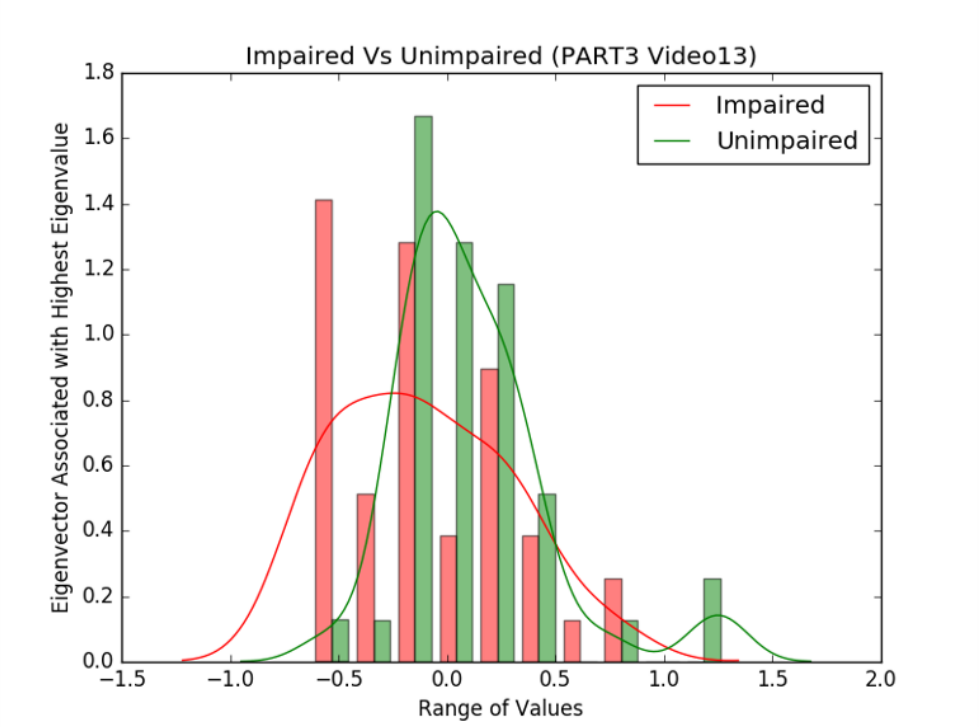

Fig. 4.7.: Plot for Eigenvector after PCA for PART3 Video13 Using Salient Part

We can see in Figure 4.12, the rf model gives relatively good result than other models. Of course, the result slightly changes after each run for cross-validation. For that, we used SVM on top of rf to get the best and consistent result on testing data.

Table 4.9.: Classification Results only on Testing Dataset for PART3 Video13 Using Salient Data

	precision	**recall**	**f1-Score**	**support**
Impaired	**0.60**	**0.75**	**0.67**	8
Unimpaired	0.83	0.71	0.77	14
Weighted Average	0.75	0.73	0.73	22

Table 4.10.: Salient Part of The Engineered Features for PART3 Video10

Features	**Video Segment**
Fixation Duration	13th sec - 16.5th sec
Saccadic Amplitude	6.5th sec - 16.5th sec
Absolute Saccadic Direction	3rd sec - 13th sec sec
Relative Saccadic Direction	16.5th sec - 20th sec
Vertical Error	23rd sec - 33th sec
Disconjugate Eye	3rd sec - 6.5th sec
Distance from Fixation Point to Nearest Landmark	20th sec - 23rd sec
Saccadic Velocity(Saccadic Amplitude/Duration)	Whole Video

Table 4.11.: Classification Results only on Testing Dataset for PART3 Video10 Taking Salient Data

	precision	**recall**	**f1-Score**	**support**
Impaired	**0.75**	**0.71**	**0.73**	17
Unimpaired	0.71	0.75	0.73	16
Weighted Average	0.73	0.73	0.73	33

4.5 Comparison

Using some videos, not all, made a difference in model performance and helped us predicting TASIT score and impairment for a particular video. We realized not all videos help us in detecting impairment. While some videos are comfortable and not discriminatory, some videos are salient. The f1 score improvement is 0.54 to 0.73

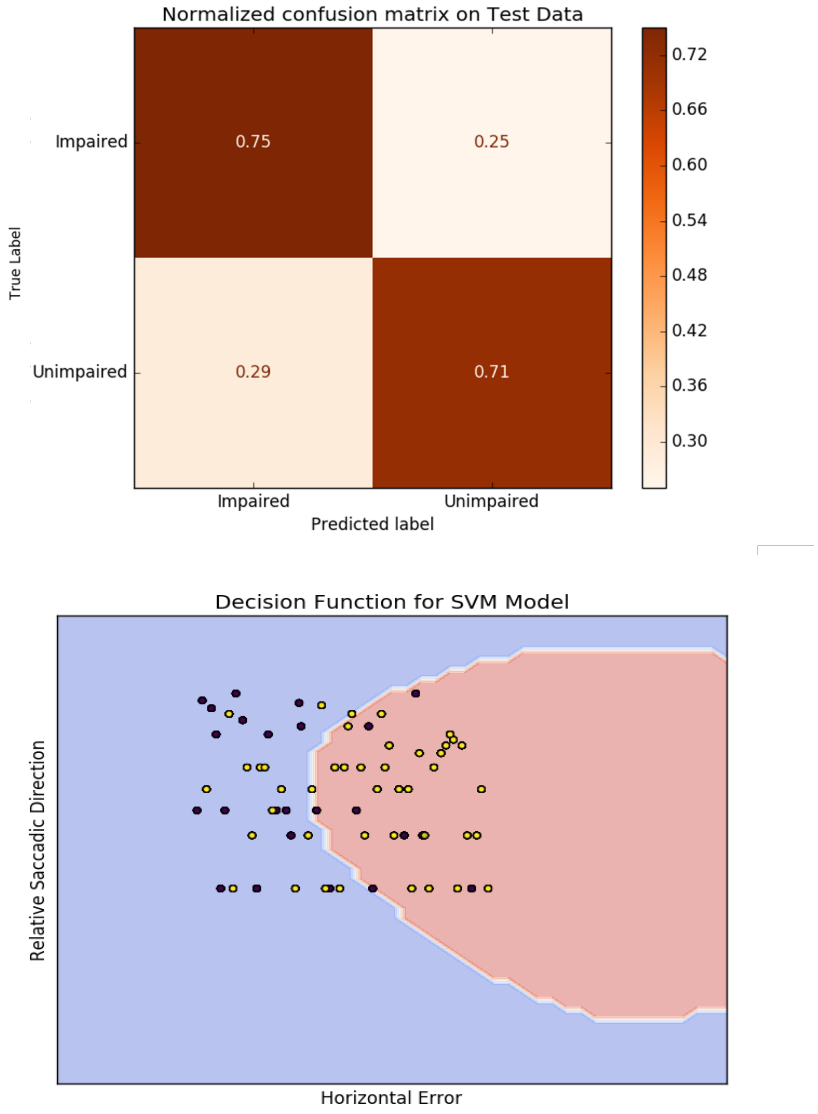

Fig. 4.8.: a) Confusion Matrix for PART3 Video13 Taking Salient Part of Data b) Decision Boundary for SVM Model

Table 4.12, which was challenging given the data is noisy, and the population size is small. We can see taking only the salient parts of the videos; we can significantly improve the result and minimize the missed or false detection of impaired populations.

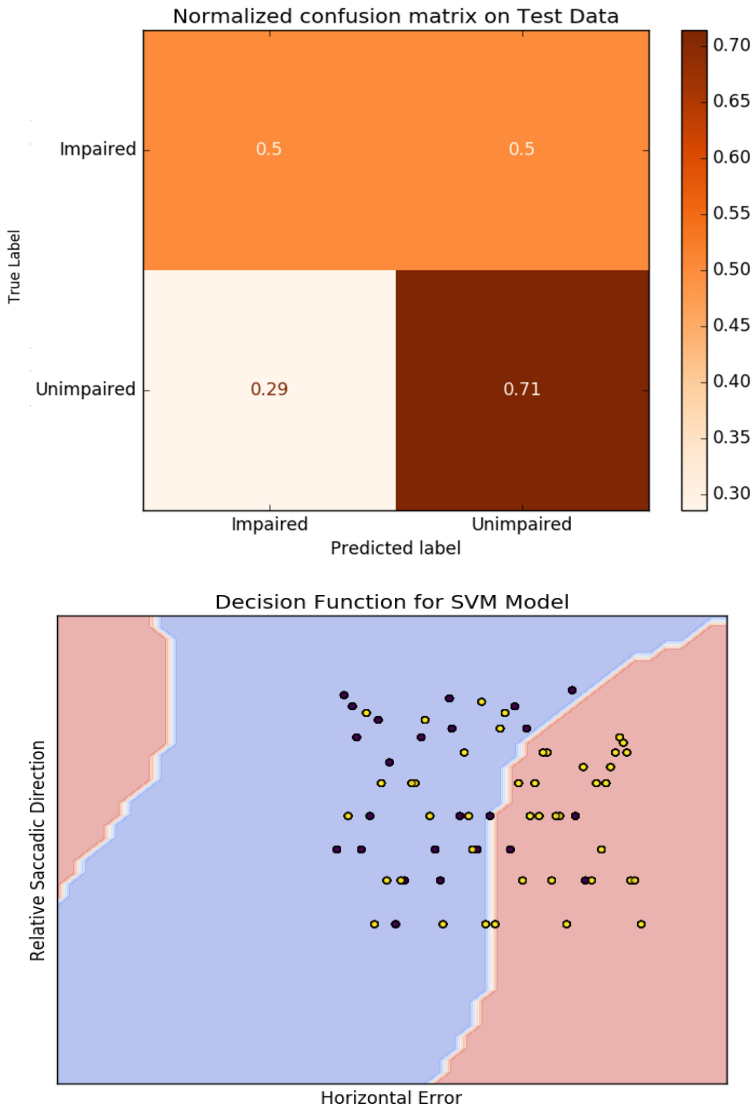

Fig. 4.9.: a) Confusion Matrix for PART3 Video13 Taking Salient Part of Data b) Decision Boundary for SVM Model

As we can see in Table 4.12, we significantly improved the detection of impaired population and successfully reduced the false negatives over the experiments.

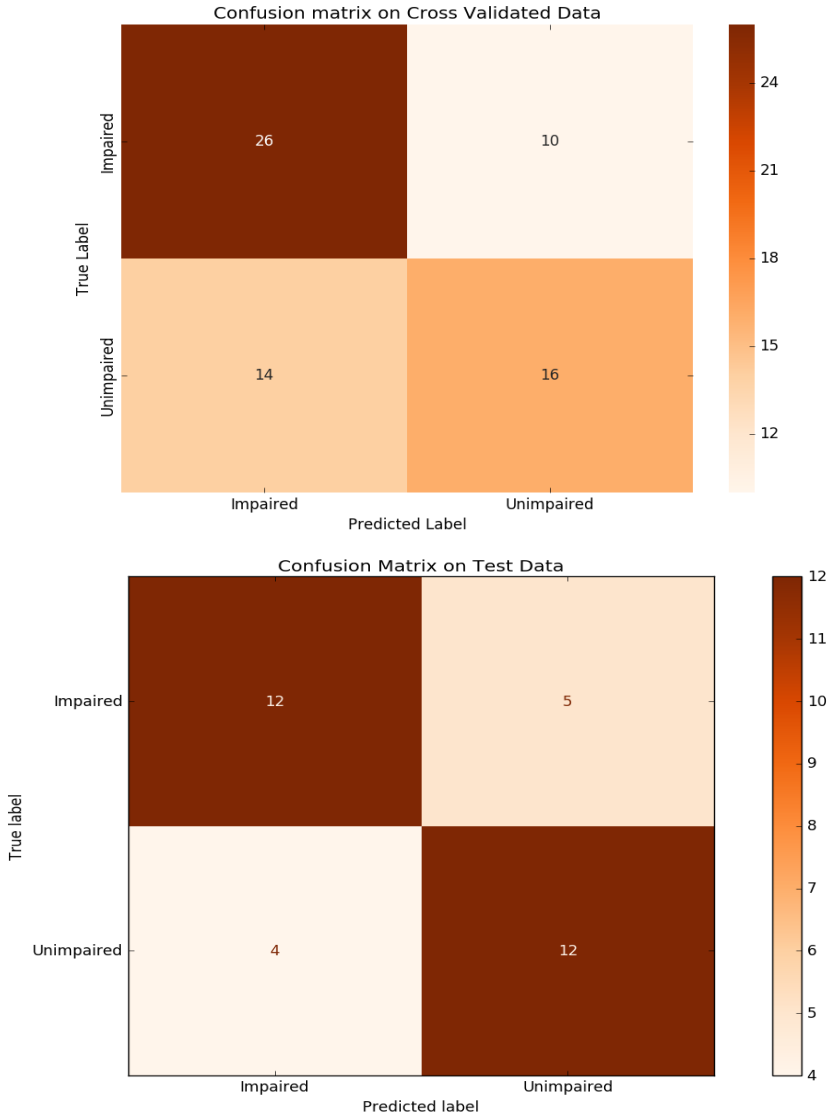

Fig. 4.10.: a) Cross Validation Confusion Matrix for Salient Part of PART3 Video10 on Train Data b) Confusion Matrix for PART3 Video10 on Test Data

4.6 Conclusion

Our contribution is finding the salient videos and determining the salient part of those videos correlated to the TASIT score, which, in turn, correlates to eye tracking data. Dividing the video data into a 3-second vector is a significant finding of this

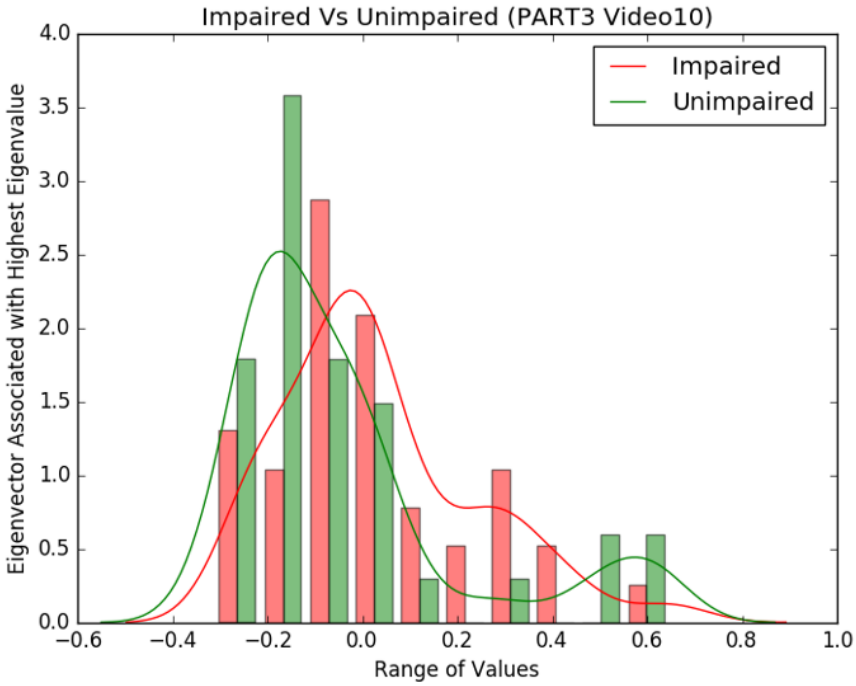

Fig. 4.11.: Plot for Eigenvector after PCA for PART3 Video10 Using Salient Part

Table 4.12.: Comparison of the Different Methods for Impaired Population

Methods	Precision	Recall	f1-Score
Whole Parts	0.52	0.57	0.54
PART1 Separate	0.4	0.39	0.4
PART2 Separate	0.29	0.36	0.32
PART3 Separate	0.61	0.57	0.58
PART3 Video10 Separate	**0.70**	0.50	0.58
PART3 Video13 Separate	0.64	0.64	0.64
PART3 Video10 Salient	**0.75**	**0.71**	**0.73**
PART3 Video13 Salient	0.60	**0.75**	**0.67**

research. The participants do not have to take the test for all videos, which can be taxing. The TASIT test can be modified and the participants will only need to take the test on the particular videos, which is salient.

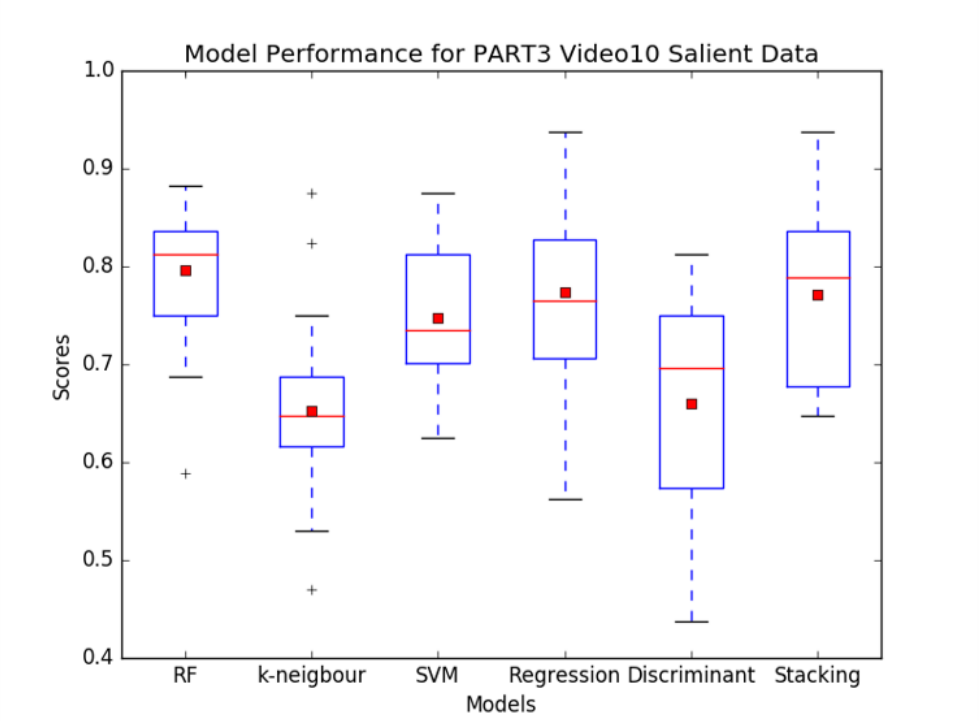

Fig. 4.12.: Model Selection for PART3 Video10

5. CONCLUSION

Traumatic Brain Injury research is an essential field in the public health issue, as there are millions of TBI patients in the United States and around the world. This research can be useful in keeping track of the impairment caused after TBI. Much research has been done on TBI patients to detect their visual impairment. While in other studies, eye tracking data on still images or videos with relatively less complex structures were used to differentiate the groups. However, in this research, videos with complex emotional features are shown to participants for collecting eye tracking data. Then eye tracking feature vectors were extracted containing dynamic information such as saccadic movement, saccadic direction, fixation distance from different facial regions of interests, saccadic velocity.

The gaze data on dynamic images can detect facial recognition impairments in TBI patients and the control group. We have found that the eye position error (horizontal and vertical error), saccade data shows better accuracy than fixation data. Especially saccadic measures produce the most crucial classification features, confirmed by other research that we have reviewed earlier. We also used distances from major facial regions (center of the whole face) as the features, which further helped improve the classifier's accuracy. We expected that percentage of fixation in the region of interests would paint a better picture to classify impaired from unimpaired. Nevertheless, they do not show improvement in the classifier's performance compared to other features. However, our result is consistent with the other studies, which showed that a significant difference is found in eye position errors and saccadic movement in impaired population.

We had a collection of 59 videos for the TASIT test. The PART3 videos incorporate the most complex emotions and have a higher variability of TASIT scores. We found only a few videos among 59 videos are essential for this research. We found the

videos with sarcasm shows more promising result in predicting TASIT scores. Visual processing abnormalities, while watching those particular videos, are evident for the impaired population. We can say visual processing abnormalities contribute to their emotional inference deficits (measured by the TASIT answers).

While ANOVA or t-test used in most of the studies are useful for finding the primary difference in eye tracking measures between the classes, machine learning is an excellent tool to detect impairment in a person's visual processing. Random forest classifier, support vector machine have proved to be a vital tool in our research to classify impaired and unimpaired people. As our sample population is small (≤ 100), we kept our model simple to prevent any overfitting. For a small dataset like ours, the model's high complexity is better on training data but worse on testing data. Keeping the problem simple helped us to find a model that is better in performance and prediction.

Our main contribution to this research is that we used a vector of eye tracking features, not only the features averaged over the entire video. Other research did not use dynamic features. We also successfully found a correlation between eye tracking data and TASIT scores. Our research is also confirming the significant finding which has been reported in other research. We have improved the facial landmarks model to include the forehead to have full facial information. We incorporated the facial feature information in the video along with the eye tracking data. The improved result is an f1 score of 0.73, whereas the baseline for the f1 score was 0.54 in the impaired population. Furthermore, we achieved a precision score of 0.75, improved from 0.52 in the impaired population using support vector machine (SVM) and random forest (RF). Taking only the correlated parts of videos (related to sarcasm) have improved the classification performance.

We can conclude that the videos associated with complex emotions reflect visual impairments to eye tracking data. Later, this can help make the TASIT test short and be more focused on the simulation of relatively complex social interactions than the simple ones.

5.1 Future Works

More complex time series machine learning structures like RNN can predict the TASIT score in future studies. Now the prediction is at the binary level. In the future, it would be more helpful if we could predict multi-level impairments in the population. Emotion-related complex features or salient images can also be incorporated to differentiate those groups along with gaze data. Also, the accuracy of our model can be improved by collecting more samples of data. Although the emotional deficit is a complex problem, we made significant progress in detecting the impairment. In the medical field, achieving what we have was challenging, especially in this small population. Our population had much bigger variation in years since injury compared to other studies, which affected machine learning's performance. However, we are hopeful that the performance can be further improved if more data is available. Also, in this research, we have not used the audio data. Maybe incorporating audio data can also prove to be useful for the prediction.

REFERENCES

REFERENCES

[1] D. R. Babbage, J. Yim, B. Zupan, D. Neumann, M. R. Tomita, and B. Willer, "Meta-analysis of facial affect recognition difficulties after traumatic brain injury." *Neuropsychology*, vol. 25, no. 3, p. 277–285, 2011.

[2] E. C. Bacon, A. Moore, Q. Lee, C. Carter Barnes, E. Courchesne, and K. Pierce, "Identifying prognostic markers in autism spectrum disorder using eye tracking," *Autism*, vol. 24, no. 3, pp. 658–669, 2020.

[3] V. Yaneva, S. Eraslan, Y. Yesilada, R. Mitkov *et al.*, "Detecting high-functioning autism in adults using eye tracking and machine learning," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 2020.

[4] J. Beltrán, M. S. García-Vázquez, J. Benois-Pineau, L. M. Gutierrez-Robledo, and J.-F. Dartigues, "Computational techniques for eye movements analysis towards supporting early diagnosis of alzheimer's disease: a review," *Computational and mathematical methods in medicine*, vol. 2018, 2018.

[5] N. Snegireva, W. Derman, J. Patricios, and K. Welman, "Eye tracking technology in sports-related concussion: a systematic review and meta-analysis," *Physiological measurement*, vol. 39, no. 12, p. 12TR01, 2018.

[6] K. Holmqvist, M. Nyström, R. Andersson, R. Dewhurst, H. Jarodzka, and J. Van de Weijer, *Eye tracking: A comprehensive guide to methods and measures*. OUP Oxford, 2011.

[7] M. Hunfalvay, C.-M. Roberts, N. Murray, A. Tyagi, H. Kelly, and T. Bolte, "Horizontal and vertical self-paced saccades as a diagnostic marker of traumatic brain injury," *Concussion*, vol. 4, no. 1, p. CNC60, 2019.

[8] M. L. Ettenhofer, "Integrated eye tracking and neural monitoring for enhanced assessment of mild tbi," The Henry M. Jackson Foundation Bethesda United States, Tech. Rep., 2018.

[9] B. Caplan, J. Bogner, L. Brenner, D. X. Cifu, J. R. Wares, K. W. Hoke, P. A. Wetzel, G. Gitchel, and W. Carne, "Differential eye movements in mild traumatic brain injury versus normal controls," *Journal of Head Trauma Rehabilitation*, vol. 30, no. 1, pp. 21–28, 2015.

[10] U. Samadani, R. Ritlop, M. Reyes, E. Nehrbass, M. Li, E. Lamm, J. Schneider, D. Shimunov, M. Sava, R. Kolecki *et al.*, "Eye tracking detects disconjugate eye movements associated with structural traumatic brain injury and concussion," *Journal of neurotrauma*, vol. 32, no. 8, pp. 548–556, 2015.

[11] M. Suh, R. Kolster, R. Sarkar, B. McCandliss, J. Ghajar, Cognitive, N. R. Consortium *et al.*, "Deficits in predictive smooth pursuit after mild traumatic brain injury," *Neuroscience letters*, vol. 401, no. 1-2, pp. 108–113, 2006.

[12] D. C. Delis, "California verbal learning test," *Adult version. Manual. Psychological Corporation*, 2000.

[13] A. Danna-Dos-Santos, S. Mohapatra, M. Santos, and A. M. Degani, "Long-term effects of mild traumatic brain injuries to oculomotor tracking performances and reaction times to simple environmental stimuli," *Scientific reports*, vol. 8, no. 1, pp. 1–11, 2018.

[14] N. G. Murray, B. Szekely, A. Islas, B. Munkasy, R. Gore, M. Berryhill, and R. J. Reed-Jones, "Smooth pursuit and saccades after sport-related concussion," *Journal of Neurotrauma*, vol. 37, no. 2, pp. 340–346, 2020.

[15] M. Hunfalvay, C.-M. Roberts, N. P. Murray, A. Tyagi, K. W. Barclay, T. Bolte, H. Kelly, and F. R. Carrick, "Vertical smooth pursuit as a diagnostic marker of traumatic brain injury," *Concussion*, vol. 5, no. 1, p. CNC69, 2020.

[16] Y. Deitcher, Y. Sachar, and E. Vakil, "Effect of eye movement reactivation on visual memory among individuals with moderate-to-severe traumatic brain injury (tbi)," *Journal of clinical and experimental neuropsychology*, vol. 42, no. 2, pp. 208–221, 2020.

[17] K. Krejtz, A. T. Duchowski, A. Niedzielska, C. Biele, and I. Krejtz, "Eye tracking cognitive load using pupil diameter and microsaccades with fixed gaze," *PloS one*, vol. 13, no. 9, p. e0203629, 2018.

[18] A. V. Reddy, R. Mani, A. Selvakumar, and J. R. Hussaindeen, "Reading eye movements in traumatic brain injury," *Journal of optometry*, vol. 13, no. 3, pp. 155–162, 2020.

[19] E. Vakil, O. Aviv, M. Mishael, S. Schwizer Ashkenazi, and Y. Sacher, "Direct and indirect measures of context in patients with mild-to-severe traumatic brain injury (tbi): The additive contribution of eye tracking," *Journal of clinical and experimental neuropsychology*, vol. 41, no. 6, pp. 644–652, 2019.

[20] Y. Liang and J. Lee, *Driver Cognitive Distraction Detection Using Eye Movements*, 12 2007, pp. 285–300.

[21] Y. Mao, Y. He, L. Liu, and X. Chen, "Disease classification based on eye movement features with decision tree and random forest," *Frontiers in Neuroscience*, vol. 14, p. 798, 2020.

[22] T. P. AB, *Tobii Pro Lab User's Manual*, Computer software, Danderyd, Stockholm, 2014, Last accessed: 11/20/2020. [Online]. Available: http://www.tobiipro.com/

[23] K. A. Dalrymple, M. Jiang, Q. Zhao, and J. T. Elison, "Machine learning accurately classifies age of toddlers based on eye tracking," *Scientific reports*, vol. 9, no. 1, pp. 1–10, 2019.

APPENDIX

ACRONYMS

TBI

Traumatic Brain Injury

TASIT

The Awareness of Social Inference Test

ToM

Theory of Mind

ASD

Autism Spectrum Disorder

SVM

Support Vector Machine

RF

Random Forest

LSTM

Long Short Term Memory

EET

Emotion Evaluation Test

PCA

Principal Component Analysis

KNN

K Nearest Neighbour

RNN

Recurrent Neural Network

GLOSSARY

Fixation

A fixation occurs when eyes are focused on a particular spot for an extended period, usually ranges from 150 to 300 milliseconds.

Saccadic Eye Movement

A saccadic eye movement is a rapid change in both eyes' movement between two fixation points in the same direction.

Saccadic Amplitude

Saccadic amplitude is the distance in degrees (angle) between the previous fixation location and the current fixation location.

Saccade Latency

The delay to initiate a saccade is called saccade latency. rnnsaccadel

Horizontal and Vertical Saccade

The horizontal and vertical saccade are referred respectively as the horizontal and vertical saccadic eye movement.

Smooth Pursuit

Smooth pursuit is the voluntary movement in both eyes when closely following a moving object.

Discojugate Eye Tracking

Disconjugate gaze is a measure of both eyes' failure to turn together in the same direction.

Saccadic Velocity

The velocity at which eyes change position from one fixation to another is called saccadic velocity.

Saccadic Count

Total number of saccades is called saccadic count.

Microsaccade Magnitude

The very small saccades are referred to as microsaccades. The microsaccade amplitude/microsaccade magnitude is typically less than 0.1.

Oculomotor Response

Oculomotor nerve is the third of 12 pairs of cranial nerves in the brain. This nerve is responsible for the eyeball and eyelid movement. The response initiated by the oculomotor nerve is referred to as oculomotor response.

Null Hypothesis Test

Null hypothesis test based on the idea that there is no relationship in the population and that the relationship the sample reflects is occurred by chance.

P-values

In statistical testing, the p-value is the probability of obtaining test results at least as extreme as the results observed, assuming that the null hypothesis is correct.

T-test

A t-test is a type of inferential statistic used to determine if there is a significant difference between the means of two classes.

ANOVA

Analysis of variance (ANOVA) is the same as a t-test; the only difference between them is that ANOVA is applicable for more than two classes while the t-test determines the difference between two groups.

TASIT

Social inference deficits are measured by The Awareness of Social Inference Test (TASIT) questions.

ToM (Theory of Mind)

Theory of mind pertains to the ability to infer others' emotions (affect recognition), intentions, thoughts, beliefs, expectations, and desires.

Absolute Saccadic Direction

The absolute saccadic direction measures the difference in angles between the current fixation location and the horizontal axis.

Relative Saccadic Direction

The difference in angles between the absolute saccadic direction of the current and the previous saccade is called relative saccadic direction.

Vertical Error

The vertical error is defined by difference in left and right eye Y coordinates.

Horizontal Error

The horizontal error is defined by the difference in left and right eye X coordinates.

Executive Functioning Data

We use executive functioning data such as animal fluency and letter fluency score as part of the features. The animal fluency test requires the patient to name as many animals as possible within a given 60 second period

Cross-validation

The k-fold cross-validation method divides the dataset into k mutually exclusive subsets

Precision and Recall

Precision is the ratio of relevant instances among the retrieved instances

F1-Score

The F1-score is the harmonic mean of precision and recall. The range of F1-score can be between 0 to 1

ROC curve

ROC curve shows the trade-off between the true positive rate (tpp) and the false positive rate (fpr). The model with correctly classified data will have ROC value 1.