Machine Learning Framework for predicting Empathy using Eye tracking and Pupil Dilation

MSc Research Project
Data Analytics

Akshay Dilip Sayar
Student ID: 20211121

School of Computing
National College of Ireland

Supervisor: Dr. Anu Sahni
Dr. Paul Stynes

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Akshay Dilip Sayar
Student ID:	20211121
Programme:	Data Analytics
Year:	2022
Module:	MSc Research Project
Supervisor:	Dr. Anu Sahni
Submission Due Date:	19/09/2022
Project Title:	Machine Learning Framework for predicting Empathy using Eye tracking and Pupil Dilation
Word Count:	7195
Page Count:	20

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	Akshay Dilip Sayar
Date:	18th September 2022

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	☐
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	☐
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	☐

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Machine Learning Framework for predicting Empathy using Eye tracking and Pupil Dilation

Akshay Dilip Sayar
20211121

Abstract

Empathy is understanding or sensing other persons' emotions. Finding highly empathetic person can help find the right candidate for jobs like nurses, psychology counsellors or team leaders, which require highly empathetic people. Empathy can be measured based on questionnaires. However, this has drawbacks as answers may be incorrect (on-purpose or not) and results can be manipulated. This research proposes a machine learning framework that combines psychological analysis, empathy questionnaire, point-of-gaze using eye-tracking device, and various deep learning frameworks to measure empathy. The point-of-gaze, and pupil of 53 participants were recorded while they watched videos that contained six sad stories which were narrated by actors. Different features like point of gaze based heat map, pupil dilation features, blink percent, saccade percent, blink standard deviation, saccade standard deviation, blink mean, saccade mean and self report parameters like memory questionnaire, age, gender, average distance of point of gaze from left and right eye of actor were extracted from the experiments. Models were trained on parameters from these different experiment individually and combined with target variable generated from the empathy questionnaire. Logistic regression model on all the parameters combined showed the best accuracy of 89%. The result also showed that pupil dilation features contributed the most in prediction followed by eye tracking features.

1 Introduction

In a world where automation is on rise and artificial intelligence and robots replacing human in multiple jobs, there are jobs where human touch or warmth is required, which robots or machine cannot provide. Jobs which are irreplaceable by machine or robots like nurses, doctors, counsellor, or team leaders for example. Nurses, doctors, counsellors treat patient and needs patient to trust them. Good team leaders need to be empathetic towards their team members Simon et al. (2022). All these job roles require people who can understand other's feelings or what they are going through. Therefore, empathetic people are more suitable for these job roles Jongerius et al. (2021), Olson (1995), Redfern et al. (1993).

There are a few ways to find empathy of a person, one of which being questionnaire Davis (1983). Spreng* et al. (2003). Olderbak et al. (2014). The empathy score is based on the responses to a set of questions. However, this method is not very accurate as answers could easily be manipulated or falsely answered. Hence there is a need to find a robust way to detect empathy that is more accurate. In research done by Zhang et al.

(2022), authors analyzed the data of person watching empathetic and non-empathetic videos, to find out whether gaze tracking can help understand if a person is empathetic or not. The analysis that was carried out manually using software to find if a person is empathetic or not which can be automated using machine learning.

The aim of the paper is to create a framework which can detect if a person is empathetic or not using state of the art technologies such as Machine Learning and Deep Learning models. This framework is based on ML models that were trained on point of gaze based heat map, pupil dilation features along with memory questionnaire, age, gender and sadness levels. These features were extracted from deep learning frameworks like yolo, ellseg and dlib. The data for the models were gathered from an experiment which was conducted on 53 participants, where their PoG and pupils were recorded while they watched six sad stories based video. These videos are taken from Cowan et al. (2014), Mackes et al. (2018). Which were created in the research to elicit empathy of a person.

This paper discusses about the related work in section 2 and the work which was used as based to carry out this research. Section 3 discussed about how the experiments were conducted and how to reproduce it in real time to predict if a person is empathetic or not. The design specification and implementation is discussed in section 4 and section 5 respectively. Section 6 presents and discuss the modelling and evaluation of all the model that were trained for this research. Section 7 concludes the research paper and discusses the future work to improve this research.

2 Related Work

There are many techniques to find empathetic person and these different techniques have been used to evaluate empathy over time. These methods were investigated in order to make it ideal and really practical. These techniques include completing questionnaires, checking pupil dilation, tracking eyes, and calculating the empathy score using those results. The subsections that follow present a thorough examination of the literature on different methods used for measuring empathy and literature that were used or helped in this research.

2.1 Measuring Empathy Using Questionnaire

The most common and easy method for measuring empathy is answering questionnaire. There are a lot of questionnaire which can give out score which correlates to empathy of a person. The questionnaire like Interpersonal Reactivity Index (IRI) Davis (1983), Impulsiveness-Venturesomeness-Empathy Questionnaire (IVE-7) Caci et al. (2003) and Toronto Empathy Questionnaire (TEQ) Spreng* et al. (2003) are the most popular and broadly used in real life to determine if a person is empathetic or not. Each of the questionnaire has different set of questions with different scoring formula. These questionnaires are still being used for finding the right person for example Aoki and Katayama (2021) questionnaire are updated for finding right person for nurse's job role and in the same way these questionnaires were modified by researchers based on requirements.

Even though the above questionnaire has high accuracy of predicting the right empathy score of a person, it has some drawbacks which limits from using in this research. While all the above questionnaire are meant to find empathy of a person, the values for empathy for a person was different in all the methods which shows that all the questionnaire are not converging at a single point Stosic et al. (2022). Study conducted Yeo and Kim (2021) shows that the Toronto empathy questionnaire from Spreng* et al. (2003) did not converge properly and gave wrong score in an experiment conducted on medical students.

The problem in the above methods is that all the questionnaire are generic and does not have questions specific to each emotions. Empathy is a function of emotion but emotion alone does not determine empathy Olderbak et al. (2014). A person might be more empathetic to sad emotion but not to happy emotion or vice versa. Because empathy is a function of emotion, empathy should be found out for each emotion and not in generic way. Research in Olderbak et al. (2014) is based on finding empathy based on each emotions namely sad, happy, anger, disgust, fear, surprise. Each emotion has five questions on each affective and cognitive empathy. As this research focuses on sad emotion based empathy, questionnaire of set of 10 question was taken from Olderbak et al. (2014) with sad based emotion focuses on sad emotion based empathy and used for getting the ground-truth values for each experiment conducted to train machine learning model to predict if a person is empathetic or not.

To find the real values of empathy and ground truth value for machine learning model a questionnaire of 10 set of questions was used. Questionnaire contains five questions on sad emotion based cognitive empathy and five questions on sad emotion based affective empathy. The next subsection discuss the literature on visual stimuli used in experiments for determining empathy.

2.2 Visual stimuli for empathetic response

Experiment has been conducted for this research where eyes are tracked and pupils are recorded of participants while showing them video. This visual stimuli is crucial and has to be chosen with care as this will help elicit empathy in participants. Images or videos can be shown to participants as visual stimuli but images will not be of very useful as it will not mimic real scenario Harrison et al. (2007). Many of researchers have conducted experiment for different purpose by showing videos and tracking eyes during the same. The videos used in any of this research could be used as these are peer reviewed. In a similar experiment where participants were made to watch video and empathy was found out Cotler et al. (2020), were the video was taken from a movie "Forrest Gump", the video had multiple emotions in it. Similarly in research Klin et al. (2002), a clip was shown to the participants from a movie to view fixation pattern in autistic people. These movie clips are good and has different emotions in it but this cannot be used in this research as this movie clips are viewed by a lot of person and they might know what happens in the video. Another drawback is that it does not show real time scenario.

Above all limitations are covered in the videos created in research Cowan et al. (2014), Mackes et al. (2018). The videos in the research are not from a movie but actors narrating fictional stories by looking directly into the camera. There are different stories based on different emotion by various actors. Because these actors are narrating the story while looking at camera, it feels like the viewer is having a one on one conversation with the actor which is real life scenario which counsellor, nurse or doctor might face in their day to day job role. As this research focuses on sad emotion based empathy, the videos which are sad emotion based were used. These videos check all the criteria and hence these were used for the experiment.

2.3 Measuring Empathy Using Gaze Tracking

Tracking eyes and its point of gaze can help in predicting if a person is empathetic or not. Gaze will be tracked using eye tracking glasses, which has cameras inside the frame to track pupil movement and a camera on front to find point of gaze and where the person is looking. Eye tracking is tracking of point of gaze and what the eyes and mind are focusing on. Gaze behaviour and gaze pattern can reveal a lot about a person which can help predicting empathy of a person Cowan et al. (2014).

Multiple features from the point of gaze can be extracted which can help in predicting empathy of a person. Features like saccades, blinks, area of interest, fixation duration, smooth eye movements etc as discussed in research Lim et al. (2020). Although in research Warnell et al. (2021) the authors failed to prove correlation of points of gaze pattern to empathy as they found out in some cases the gaze pattern was correlated to empathy and in some cases it was not. Research in Moutinho et al. (2021) tested out the hypothesis of highly anxious people with high empathy has no correlation between high empathy levels and gaze fixation as people with anxiety avoid eye contact but people with low anxiety or no anxiety issue showed high correlation of eye contact and empathy levels. In research conducted by Nebi et al. (2022) it was concluded that in low emotional salience situations compared to high emotional salience situations, gaze behavior corresponds with empathy.

Researchers Cowan et al. (2014) were the first to prove that there is correlation of gaze pattern with empathy. They also proved that the more the person is looking at the other person's eye region the more empathetic that person is. The research paper has proved it by showing heat map of point of gaze of a video which has actors in it narrating a story. The heat map showed area of interest of participant which was dense around eye region for highly empathetic people, less dense for moderately empathetic people and it was scattered around nose, lips and chin region for less empathetic people. This result from this research is very substantial and is leveraged in for the current research. This research paper was used as base for the current research and an open source tool for heat map for area of interest was created. The heat maps as input to CNN model was used in predicting if a person is empathetic or not. The closer the point of gaze on eye the more the empathetic the person is and hence average distance of point of gaze from eye was calculated. These two features were created based on this research paper Cowan et al. (2014). As discussed in paper Klin et al. (2002), average saccades and blink, average and standard deviation of blink and saccades timings were also taken as features.

Study in research Cowan et al. (2014), Zhang et al. (2022) has shown relation between eye tracking and empathy and helped find some features using their analysis. But the analysis was done manually which may be prone to human errors.

2.4 Measuring Pupil Dilation and using it to measure empathy

Pupil is a part of eye which is concentric to iris and present inside it. Pupil takes in the light and displays image in the brain. The pupil size increases if the amount of light is less and size decreases if the amount of light in surrounding is in abundance. The pupil also change its diameter depending on different emotions or feeling Harrison et al. (2007). The research says that the amount of pupil dilation was proportional to the empathy of a person which was found out by answering questionnaire. The research done in Charan et al. (2022) also states that facial expression and pupil dilation was very useful parameter to detect emotion of a person. According to research done by Aktar et al. (2021), pupil dilation starts to happen at an early age, making it a useful indicator of emotional understanding and empathy prediction. The above examples proves that there is corelation between pupil dilation and empathy of a person, and hence it suggest pupil dilation can be considered as a deciding factor for predicting empathy of a person.

This research takes input from infrared camera which is located in the inner side of eye-tracking glasses. Pupil dilation can be measured if pupil is localized in an image and then each frame from the video can be taken and pupil radius or diameter can be predicted from the images to find the pupil dilation. A machine learning or deep learning model can be trained on labelled pupil data to predict pupil diameter in real time. To train a model for pupil detection, three significant datasets are made accessible in the open-source. One of the three being ExCuSe dataset contains 38 000 pragmatically labeled photos of pupils Fuhl et al. (2015). Using the same dataset a highly accurate model for localizing pupil was create Miron et al. (2019). With 55,000 more images than ExCuSe data, Else data contains 93,000 data set and which were labelled using convolutional neural network Fuhl, Santini, Kübler and Kasneci (2016). The latest dataset contains 111,581 pupil images of pupil, the quantity of the images makes this dataset very useful Han et al. (2019). But instead of training a neural network model from scratch which takes a lot of time, capital and resources a pre-trained model could be used to localize pupil and fetch the diameter of pupil in each frame of a video.

In research Kothari et al. (2020), authors has proposed a new neural network model which localizes pupil on real time basis. The model was trained on many open source dataset like NVGaze Kim et al. (2019), OpenEDS Garbin et al. (2019), RITEyes, ElSe Fuhl, Santini, Kübler and Kasneci (2016), ExCuSe Fuhl et al. (2015), PupilNet Fuhl, Santini, Kasneci and Kasneci (2016) and LPW Tonsen et al. (2016), these are almost all the best open source dataset available. The dataset contained a large portion of images gathered from infrared camera and hence it would perform good on similar type of images which are there in my research. The data was augmented using various techniques like flips, rotation, random gamma correction, exposure offset, guassian noise etc which makes the model robust. The author Kothari et al. (2020) considered the pupil to be in ellipse shape and hence there will be two radius to get the shape of pupil which are namely r1 and r1 (radius 1 and radius 2). This resulted in accurately segmenting the pupil and iris. No drawbacks was found in the methodology and in the future work the authors want to test the model on wide variety of people. The code and scripts are provided by the authors on github and is free to use by everyone. Using the code both radius of each eye was found for each frame of the video and then the peak pupil diameter and various percentile of pupil diameter was extracted from the raw data. One of the findings in Zhang et al. (2022) is that the peak pupil dilation is highly correlated to high empathetic people ad hence this feature can be one of the important one for predicting empathy of a person.

2.5 Machine Learning and Deep learning in predicting empathy

Various machine learning models has been used in this research to find empathetic person. Not just for predicting the empathy but in various stages of this research machine learning model was used like extracting features, creating dataset. One of the model was from Bochkovskiy et al. (2020), named yolov5 acronym for you only look once, the model is trained for localizing any shape, or pattern in an image. Two model were trained to localize particularly two object in the frame of video naming point of gaze circle and laptop screen where the video was being played. The the yolov5 return an object coordinates in terms of pixel, one centre of the object and two other points as height and width of the object and last point is the confidence in detecting the object.

The other neural network model that was used is for extracting pupil diameter from each of the frame of the video produced by infrared camera Kothari et al. (2020). The model is based on convolutional neural network model with encoder and decoder based architecture in the neural network which skips connection. The encoder encodes the image to low dimension and then using decoder to decoding the image to high dimension. The model is fast and processes around 50 frames per second in environment with four gigabytes of graphics card.

To conclude yolov5 was transfer learned to localize point of gaze and laptop screen. Ellseg framework was used to extract pupil radius of left eye and right eye. Dlib King (2009) was used to extract 68 landmarks on face of actors narrating the story in the video which are viewed by viewers. From the landmarks, eye coordinates of both eyes of actor was found out which was used to find out average distance of point of gaze from eyes of actors which can be useful parameter to predict empathy of a person Cowan et al. (2014).

3 Methodology

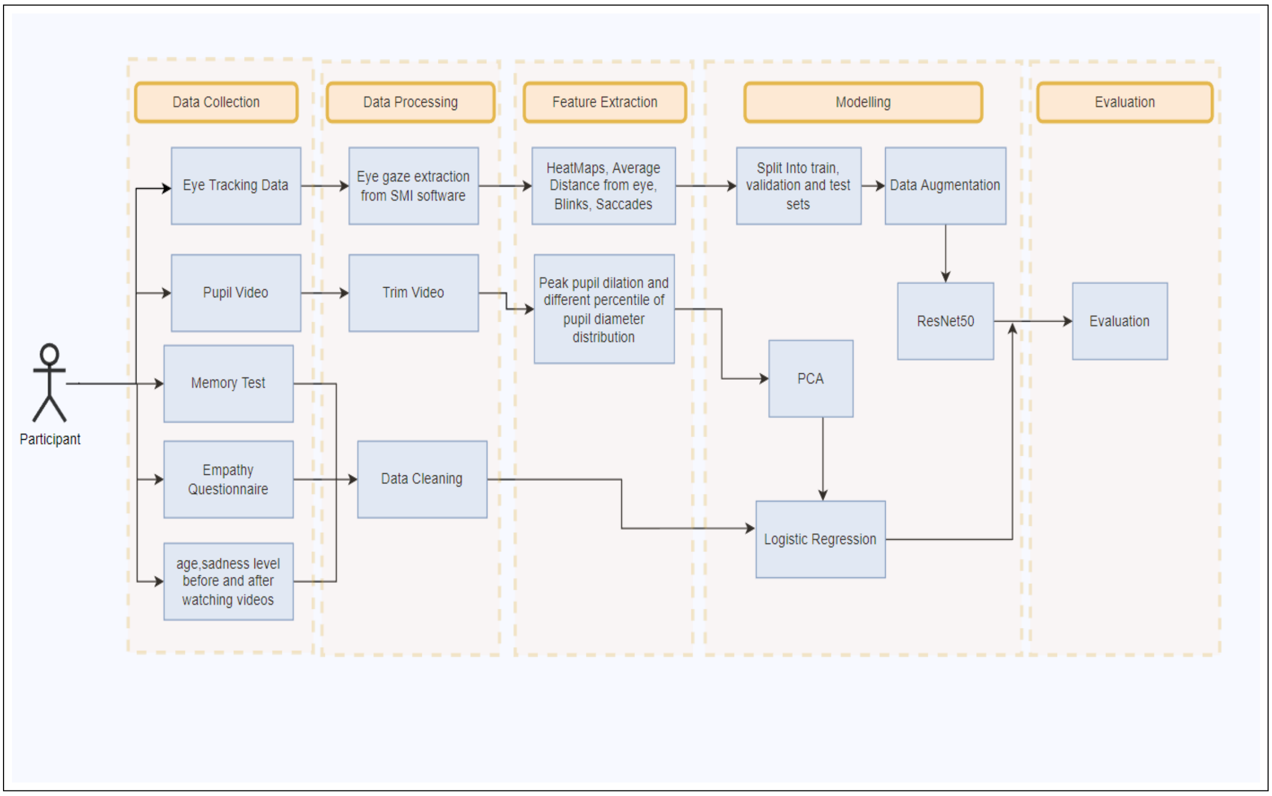

Figure 1: Research Methodology

The whole procedure from conducting experiment, collecting data to evaluation is shown in figure 6. The complete methodology is explained in detail in the following subsections.

3.1 Data Collection

Data collection starts by conducting experiments on participants. The experiment starts with signing the consent form by the participant. After that self reported parameters are noted down like age, gender, sadness level on the scale of 1-10 (1 being least sad and 10 being extremely sad) before the experiment begins. The participants sits on a chair and a laptop is place in front of him. The participants gets a headphone to listen to the audio clearly. Then eye tracking glasses are given to the participant which is then calibrated on three point scale, the three points being top left of the laptop screen, top right and bottom centre of the screen. Then the pupil recording starts on the mobile and video is played on the laptop. The participant is left alone to watch the video. The video is 13 minute long video which is narrated by actors. The stories are sad emotion based stories narrated by four different actors and are used from the research papers Cowan et al. (2014), Mackes et al. (2018). After the video is finished the participants are asked to report the sadness level after watching the video again on the scale of 1-10. Then the participant is asked to memory questionnaire which has ten questions based on the video they just watched. For each right questions is given 1 mark and no negative marks for wrong answers. The final step of the experiment is that the participants has to answer the emotion based empathy questionnaire Olderbak et al. (2014) the score of which will be used as to check if the parson is highly empathetic or not. Data collection for one person ends her and then these same steps are used on 53 participants.

3.2 Data Processing

After the data was collected, it has to be processed to make it useful. In data processing first the video is imported in the SMI software - Begaze and processed which outputs a video with point of gaze shown in the video as orange color dot. Here it was found that data of six participants were corrupted and hence the data points of these person were removed from the dataset. The video from the BeGaze software contains 1000 frames per seconds and this was converted to 25 frame per seconds using opencv module which read the video in 25 frames per second and write the video and this video is further used feature extraction and feature engineering.

After each video is processed in the software, event statistics of each experiment is extracted out which contains event details of for each frame, like category of events, start time and end time of each participant etc. All the data gets extracted in tab separated text file.

The pupil video that was recorded during the experiment contains some extra part before starting the video and after finishing the video which was removed out. After the video was trimmed out, the video is processed using opencv and the video was split into two parts and hence creating two separate videos, each for left and right eye. The process is repeated for the videos of all fifty participants.

3.3 Feature Extraction

The processed data is only useful if right features are extracted from them. Feature extraction was the longest part of the whole process. From the videos that are processed using BeGaze software, heat map was extracted and average distance from left eye and right eye was calculated. Heat map of point of gaze is a great way to find empathy in people Cowan et al. (2014), Mackes et al. (2018). To create heat map of point of gaze from the video, coordinates of point of gaze and coordinates of laptop screen was extracted using yolov5 model Bochkovskiy et al. (2020). Yolov5 model was trained to detect orange circle and laptop screen in the video and two models for two objects were created and the models recognized the object with high confidence as shown in figure 2.

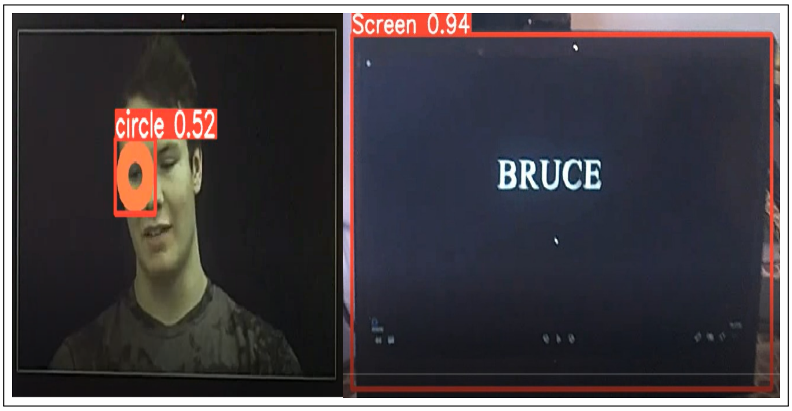

Figure 2: Yolo models localizing point of gaze circle (left) and laptop screen (right)

After the coordinates of point of gaze and laptop is extracted, the point of gaze was again calculated with reference to the frame of laptop so that heat map was only created of points of gaze which were inside the laptop screen frame as shown in figure 3. This would enable to create heat map for all participants on the static frame of reference, even after the distance between the laptop and participants was kept varying (if participants move while watching the video).

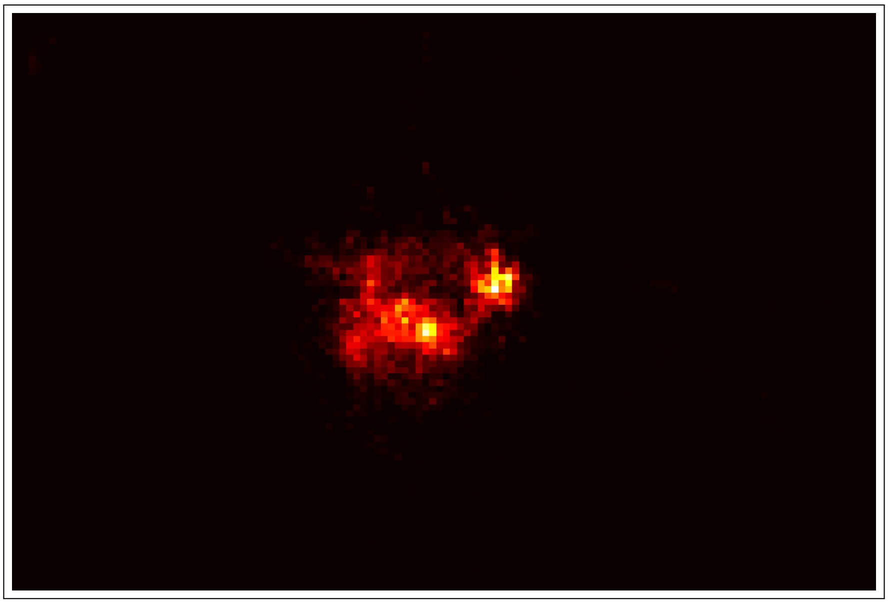

Figure 3: Heat map of point of gaze of a participant

After the heat map was created it was flattened, to one dimension array. It had 12288 columns and hence there was a need to reduce the dimensionality of the images. Principal component analysis was applied on the flattened heat map data. As shown in 4, a total of 34 components could explain 95% of the variance in the data and hence from 12288 columns only 34 were extracted and used in the final model to predict empathetic people.

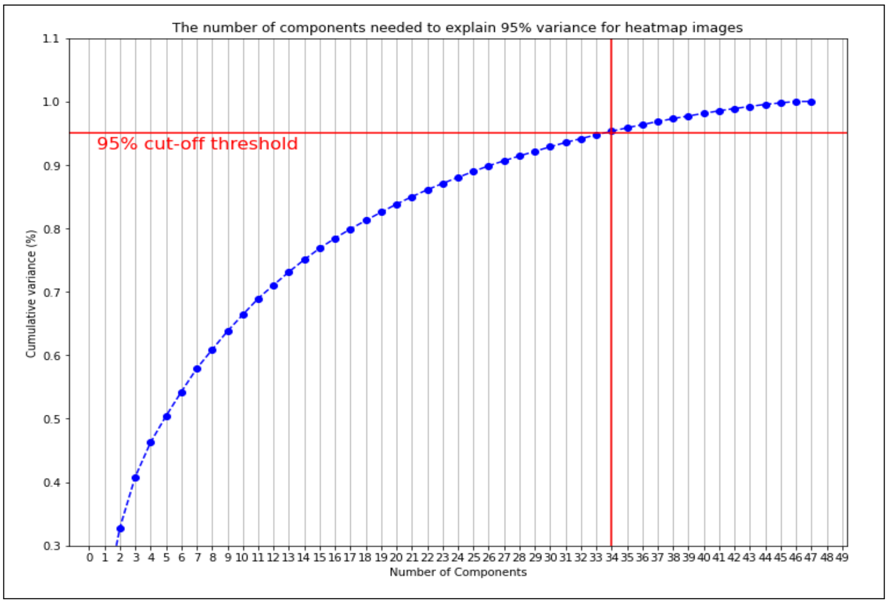

Figure 4: Finding n for PCA on heat map images

The average distance of point of gaze from the actors both eyes are calculated. To locate the actors eyes, dlib library King (2009) was used which locates sixty-eight landmarks on the face of the actors in the videos. With the help of point of gaze coordinates and coordinates of eye, distance from each eye was calculated for each frame of the video. To normalize the data for all the participants, the distance between point of gaze on screen and eyes of actor was divided by distance between two eyes of actor. At the end the sum of distance from all the frame was taken and divided by the total number of frames, this gives normalized average distance of point of gaze from each eye (left and right).

To get peak pupil dilation and other pupil features, ellseg framework Kothari et al. (2020) was used. Some modifications were made to extract just the pupil radii from the script. The pupil video for left and right eyes which were extracted in the last step, will be used as input and as the output two radii for the pupil is extracted for each frame of the video For each participant the script was ran twice, one for left eye and one for right eye which gives data for r1 and r2 (two radii of ellipse) and this was saved in json format. Once all the data was extracted for all the participants and saved to json file format, features from this were extracted. First the average of r1 and r2 was taken for each frame for each eye of a participant. The features include peak pupil dilation, 0,10,25,50,75,90 and 100 percentile of the distribution. All the features were extracted and the data was normalized using both standard scalar and min-max scalar and hence a total of twenty six features were extracted from this pupil dilation data. Model on just pupil dilation data was also applied but because they were correlated to each other, principal component analysis was applied on the data for the final modelling. As seen in figure 5, 99% of variance was explained using just 4 out the 26 features and hence this reduced dimension were used as feature in training machine learning model.

Even statistics from the previous step was processed from the software which had a lot of parameters which were useless and hence only a few parameters were extracted like average blinks and average saccade per experiment. Standard deviation and mean of difference of start and end time of each event of blinks and saccade was also taken out. All these parameters were used and for the final model.

The target variable in the data is sad emotion based empathy questionnaire score.

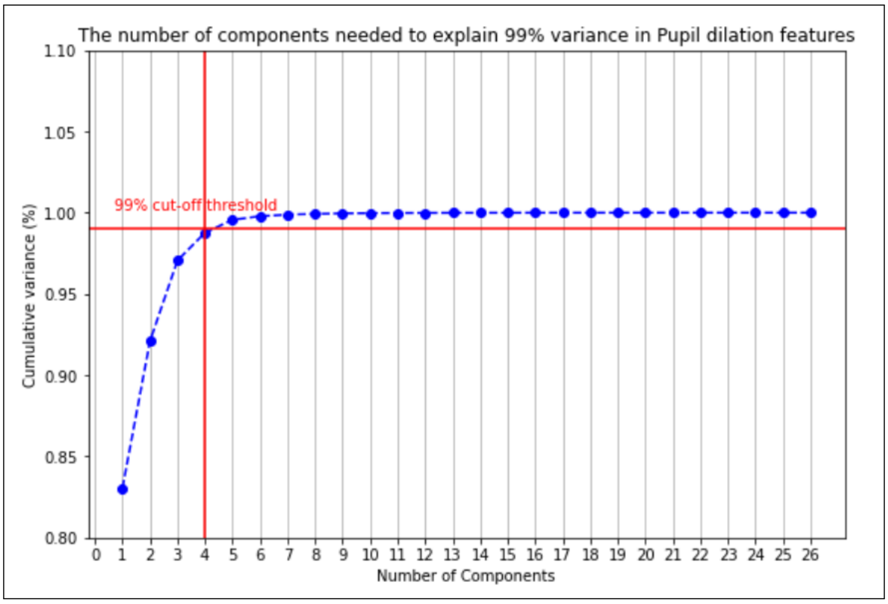

Figure 5: Finding n for PCA on pupil dilation data

The data is a continuous data and because the data points were only 47, it was converted to binary classification problem. The population was of 47 participants and was divided in three groups, non empathetic, moderately empathetic and highly empathetic. Hence in the target variable only highly empathetic are taken as positive and non and moderately empathetic data points are considered as negative. Hence 17 data points were of empathetic people (two person had same score and hence instead of 16, a total of 17 participants were considered as empathetic) and remaining 30 were considered as non empathetic people.

3.4 Modelling

The structured data from pupil data and event statistics along with gender, age, sadness before and after watching the video, average distance from right and left eye was merged into one dataset and was used for modelling. The reduced data of pupil and heat map images was merged to form final data on which models were applied. Test, train split was taken as 18%-8% which resulted in 38 data points for training and 9 data points for testing. Different model like random forest, gradient boosting, extreme gradient boosting, linear regression, naive bayes, knn, svm were applied to the dataset along with hyperparameter tuning and k-fold cross validation for full and subset of features. The models were evaluated and the best performing model was selected for the final prediction. Different models and their performance is discussed in detail in the section 6.

3.5 Evaluation

Predictions from the models were evaluated based on precision, recall, f1 score and accuracy and the best model was chosen. For ResNet50 model with different parameters was evaluated and selected. Evaluation of the models is discussed in detail in section 6.

4 Design Specification

There are two questionnaires in the data collection phase. The first questionnaire was memory based questionnaire which was designed based on the video the participant watched. There were ten questions in the questionnaire, one mark is rewarded for the right answer and no marks are deducted for wrong marks. The questions test the memory and attentiveness of participant and was designed as Microsoft survey based quiz. Each question has four option in it to answer. To give reference to the actors in the video, the actors were given fictional names like Bruce, Emma, Selena and Robert. These names were displayed at the beginning of their part in the video. The questions in the memory questionnaire are as follows:

- Which relative of Bruce was suffering?
- What happened to Bruce's relative (disease)?
- What was stolen from Bruce?
- Why was Bruce sad after the item was stolen from him?
- To which position Selena was applying for?
- What was the actual reason that Selena didn't get the job?
- Which relative of Emma was suffering?
- What disease was Emma's relative suffering from?
- Which pet did Robert have?
- How does Robert describe his friend's sister?

The second questionnaire was empathy questionnaire. These was created as Microsoft survey. There are ten questions in it, out of which five tests affective base empathy and five test cognitive base empathy. Each question has seven options to choose from and the score was given in the range of -3 to 3. The score is not indicator of good or bad, it was just used to find out the score of empathy of a person.

Various architecture and framework has been used to process data and extract features from the processed data. When the video was processed from BeGaze software it was in 1000fps and to convert in 25fps, opencv was used. To create heat map of point of gaze and for finding average distance from eyes, point of gaze and laptop screen has to be extracted from the video. Point of gaze is orange color circle and laptop screen was in quadrilateral shape. To extract these, yolov5 framework Bochkovskiy et al. (2020) was used and was transferred learned to predict point of gaze and laptop screen. To find distance between actor's eyes and point of gaze, actors' eyes coordinate has to be found out and for this dlib King (2009) was used which gives out 68 face landmarks points. To extract pupil radii from the pupil recording, ellseg framework Kothari et al. (2020) was used. This framework is a encoder decoder based convolutional neural network model which was trained on almost all the pupil dataset available which helped in giving the state of the art result for pupil radii and hence pupil dilation.

After the data was ready, principal component analysis was used to reduce the number of features. The right amount of features to be reduced was found out from percent variance. All the features which showed 99% variance was retained. PCA was also applied to heat map image's flattened array which had 12288 features and was reduced to 34 features which had 95% variance. All the data was merge to form a single dataset. Then multiple machine learning models were applied to the data like Linear Regression, Naive Bayes, Random Forest, Gradient Boosting, XGB. The model with highest accuracy which was Logistic regression was used for the deployment.

5 Implementation

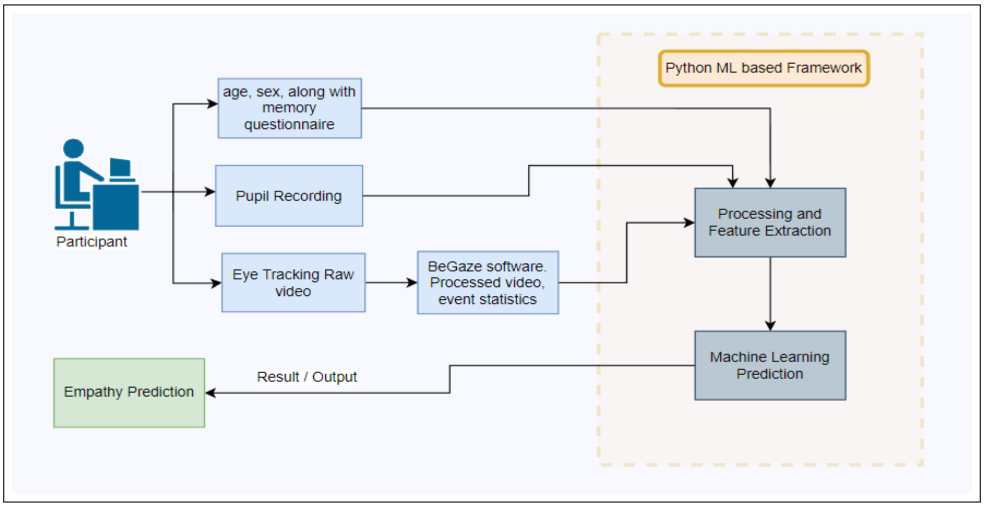

Figure 6: Implementation of ML based framework to find empathetic person in recruitment process

The machine learning based framework was implemented using python and its multiple modules. When deployed to real time it can used to find if a person is empathetic or not. Taking few basic information like gender, sex, before after sadness and memory questionnaire are send directly to the database where it processes the data. The memory questionnaire were implemented by opening the survey link and the mark responses goes to an excel sheet where the score is calculated. One mark is granted for the right answer and no marks are deducted for wrong answer. The empathy questionnaire is also the survey base questionnaire whose response is sent to an excel sheet. Python script takes the score from the excel sheet directly to the framework. The participant is asked to watch the video where actors narrate the stories. While the participant is watching the video, pupil is being recorded and eyes are being tracked. Pupil video is also sent directly to the framework. The only dependency for this framework is to process the raw eye tracking video from the BeGaze software by extracting processed video and event statistics.

Once the participants finishes watching the video all the data is gathered in python based application, where all the data is stored temporarily till the prediction is made. From the raw video, point of gaze, screen coordinates are fetched using the trained yolov5 models. Using this information heat map is made. Using Dlib module face landmarks is found out and average distance from the eyes are extracted. The average distance of point of gaze from actor's left and right eye, processed event statistics along with gender, age, memory question score is combine.

After all the data is merged into one place. Prediction is done which shows if the person is empathetic or not and this information is returned as output of the framework.

6 Evaluation

A total of 47 data points are useful for modelling and evaluation. Results of Empathy questionnaire was analyzed and the distribution is as shown in figure 7. The highest number of participants score between 0.8-1.0. The least score was 0.8 and the highest was 2.9 from the set of participants. Based on the empathy score of participants, the data points were divided into three groups which are low empathetic people, mediocre empathetic people and highly empathetic people. The high empathetic people were classified as '1' and others were classified as '0'. The distribution of memory test score is shown in figure 7. Almost all the participants scored high 8-10)in the memory questionnaire which shows that the participants watched the video with attentiveness. The distribution plot of memory test score shows that the participants were paying attention and were not distracted. This shows that there is credibility in the data points and are valid to use. Analysis showed that there the average difference of before and after sadness level increased by 65%. Further analysis showed that there is 30% between EEQ score and difference in before and after sadness.

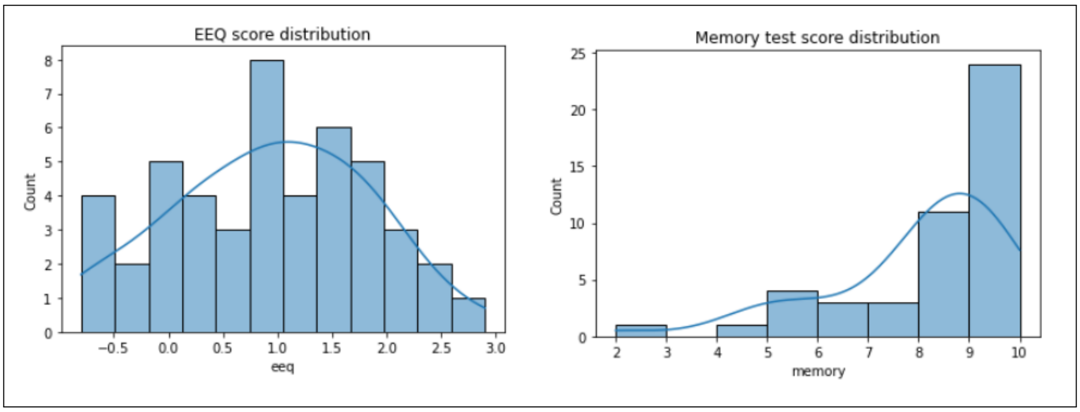

Figure 7: Distribution of EEQ score (left) and memory score distribution (right)

This section has been divided in four parts. The data was initially collected using three experiments: eye tracking, pupil dilation and self reported questionnaires. The model were trained on features from these three experiments and also on data where the features from all the experiments were combined. The framework was specially designed for recruitment process where identification of empathetic people was of utmost importance. If an empathetic candidate is predicted as non empathetic one then the penalty will be more as compared to when a non empathetic candidate is predicted as empathetic, which implies that false positive could be tolerated but not false negative. Therefore, in our framework recall was considered most crucial metric, followed by f1 and accuracy and the model selection was considered based on the same metric priority. Modelling and evaluation for all the four parts are discussed in the following sections.

6.1 Experiment 1 - Modelling with self reported features

While the experiments was conducted some self reported questions were asked like age, sex, sadness before watching the video, sadness after watching the video and a memory questionnaire was asked after watching the video which checks how much the participants can recollect from the video they just watched. Participants were given 1 score for each right question and no negative marks for wrong answers. There were total of 10 questions in the test. Score from this test was taken as feature in this subsection along with the above self reported parameters.

Three different models were trained on this subset of features. The model were Logistic regression, bagging based random forest and gradient boosting with tuned hyperparameters. The model were evaluated based on the four metrics; recall, F1 score, precision and overall accuracy in respective order of importance. The best performing model was logistic regression and random forest with 0.33 recall and 0.5 F1 score.

Models	Recall	F1	Accuracy	Precision
LR	0.33	0.5	0.78	1
Random Forest	0.33	0.5	0.78	1
Gradient Boosting	0	0	0.44	0

Table 1: Accuracy metrics on model trained on self reported features

6.2 Experiment 2 - Modelling with Eye Tracking features

Using the point of gaze, many features were extracted. Features like heat map generated using point of gaze, average distance of point of gaze from left eye and right eye of actors narrating the story. Other features were extracted from the BeGaze software like blink and saccade percent, mean and standard deviation of difference of start and end time of blink and saccade. The heat map image was converted into single dimension array which contained the pixel value from 0-255. A total of 12288 features were there from the heat map itself. To reduce the features PCA was applied and 34 features were selected which showed 95% variance as shown in feature engineering process in 5.

Three different models Logistic regression, random forest and gradient boosting with tuned hyperparameters were applied on the data. These models were evaluated on the basis of four metrics; recall, F1 score, precision and overall accuracy in respective order of importance. The comparison of models are shown in table 2. The table shows that the logistic regression performed the best among the other models and its accuracy and f1 score is better than the results from model LR and RF model from experiment 1. Hence stating the eye tracking features help in detecting empathetic people better than self reported features.

Models	Recall	F1	Accuracy	Precision
LR	0.5	0.6	0.78	1.0
Random Forest	0.25	0.4	0.67	1.0
Gradient Boosting	0.25	0.33	0.55	0.5

Table 2: Accuracy metrics on model trained on eye tracking features

6.3 Experiment 3 - Modelling with Pupil dilation features

Left and right pupils were recorded while participants watched the video. Pupil dilation features were extracted using the pupil radii which was extracted using ellseg framework Kothari et al. (2020). Using the radii data pupil dilation at 0, 10, 25, 50, 75, 90, 100 percentile was fetched out for both left and right eye. The data fetched was normalized based on min-max normalization and standard scalar normalization as discussed in feature engineering section. A total of 26 features were extracted from the data.

Three different models Logistic regression, random forest and gradient boosting tuned hyperparameters were applied on the data. These models were evaluated on the basis of four metrics; recall, F1 score, precision and overall accuracy in respective order of importance. The comparison of models are shown in table 3. Similar to experiment one's result random forest and logistic regression performed exactly similar and showed same accuracy. The accuracy is better than memory based features and eye tracking based features too. This shows that the pupil dilation features are the best features among eye tracking and self reported features. The results based on pupil dilation are not yet good but are better than the above two results.

Models	Recall	F1	Accuracy	Precision
LR	0.67	0.67	0.78	0.67
Random Forest	0.67	0.67	0.78	0.67
Gradient Boosting	0.33	0.4	0.67	0.5

Table 3: Accuracy metrics on model trained on Pupil dilation features

6.4 Experiment 4 - Modelling on all the features

After modelling on the three sets of features individually, models were then trained on all the above features combined. The combined data had all the information from eye tracking, pupil dilation, and self reported parameters. Multiple models were trained on the data.

The data contained structure data and image data as well. Multiple framework were used and applied on image dataset. Frameworks like MobileNet, ResNet, VGG16, VGG19, InceptionResNet, custom tensorflow sequential model with different combinations of hyperparameters and layers. Even after trying neural network models could not learn anything. The probable cause for not learning could be less amount of data to train on (as there were only 47 images). Therefore decision to use machine learning instead of neural networks for image dataset was taken. To be able to use the image dataset in machine learning, the images were flattened and dimension were reduced to thirty-four PCA features.

Different machine learning model were applied like naive bayes, k-nearest neighbours, support vector machines, random forest, gradient boosting, extreme gradient boosting, logistic regression with hyper parameter tuning and applying k-fold cross validation. Out of the above model naive bayes, k-nearest neighbours, support vector machines failed to learn anything from the data and hence were considered useless and hence are not a part of model comparison. The model which showed learning were Random Forest, gradient boosting, extreme gradient boosting (XGB) and logistic regression.

Similar to the previous three experiment's results, logistic regression showed the best accuracy in terms of recall, f1 score and accuracy as shown in table 4. Also in this experiment, logistic regression showed best result as compared to results from the first three experiments. This shows that all the parameters altogether has higher predicting power than individually.

Model Name	Recall	F1	Accuracy	Precision
Logistic regression	1.0	0.89	0.89	0.8
Random Forest	0.67	0.4	0.25	1.0
Gradient Boosting	0.25	0.22	0.22	0.67
XGB	0.25	0.4	0.55	1.0

Table 4: Accuracy Metrics for Models

6.5 Discussion

All the top performing model from all the four experiment was Logistic regression. In experiment four where model was trained on all the features showed highest recall and f1 score as shown in figure 8. This shows that all the features combined has the highest predicting power followed by pupil dilation features, eye tracking features and self reported features. The pupil dilation features performed better than eye tracking and self reported features and were alone good enough to predict empathy of a person as accuracy of experiment three is the higher than experiment one and two.

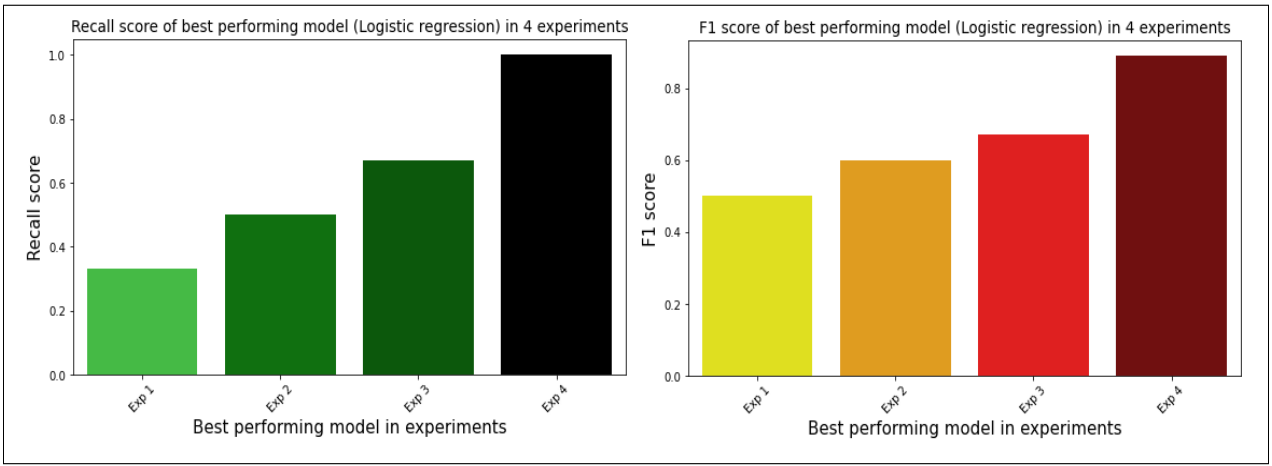

Figure 8: Top performing Model recall and F1 score in different experiments

Because logistic regression performed the best, it was used to extract the most contributing features in the final experiment which had all the features as input. The figure 9 shows the top 17 features extracted from logistic regression. The figure shows that the most contributing feature was feature from pupil dilation followed by heat map features (eye tracking) and self reported features. The figure and the results from all the experiment shows that pupil data and heat map data are actually a good predictor of the empathy of a person.

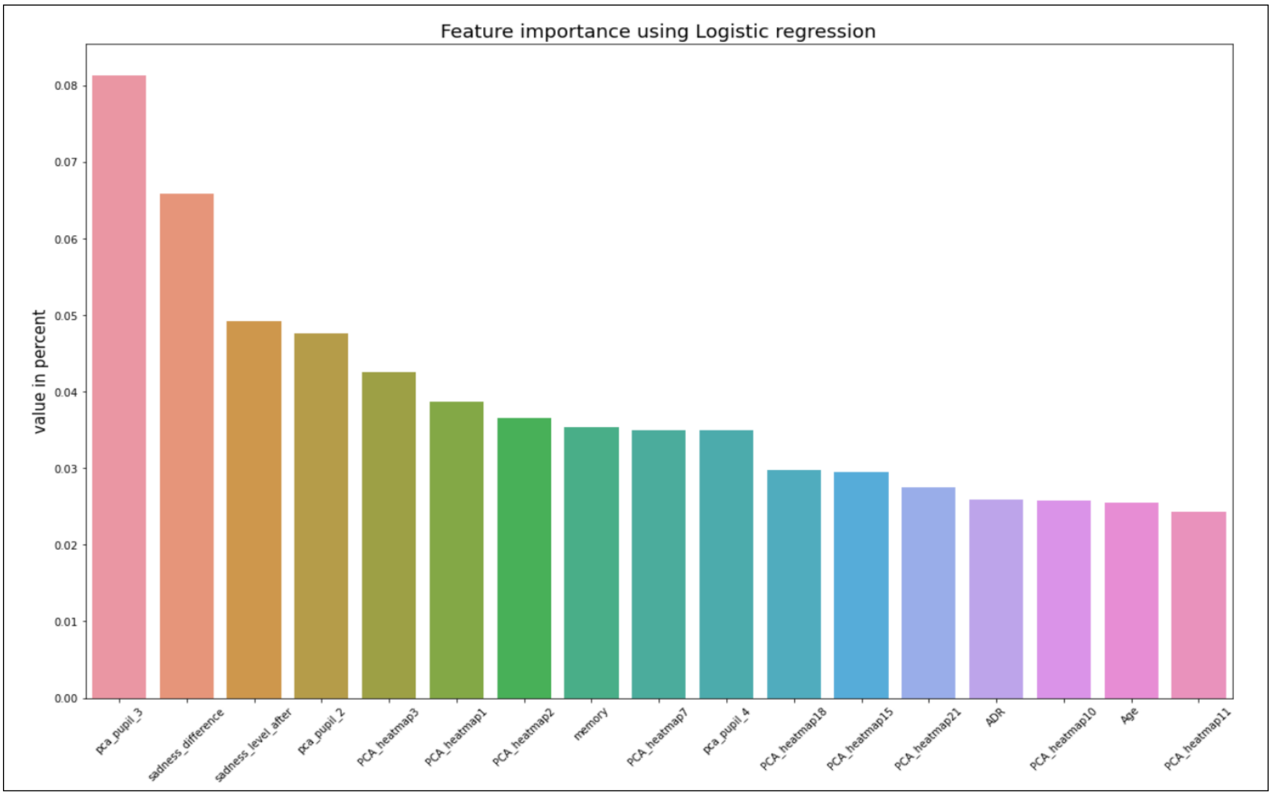

Figure 9: Most important features

7 Conclusion and Future work

The aim of the research was to create machine learning framework to find empathetic people using eye tracking, pupil dilation features and some self reported parameters. Multiple machine learning and deep learning models were applied on whole data and also on subsets of datasets. Best model in all the experiments was logistic regression. While model trained on self reported parameters gave the least accuracy followed by model trained on eye tracking data. Highest accuracy was achieved on on model train on subset of pupil dilation features. Which shows that pupil dilation derived features were the best predictor or empathy followed by eye tracking and self reported parameters. The same was proved when feature importance was found out by logistic regression that was trained on whole dataset. Hence our hypothesis of finding out empathetic person based on pupil dilation and eye tracking was proven right.

This research can potentially be used in recruiting process of nurses, doctors, psychology counsellors or team leaders which needs to be highly empathetic. The framework designed in the research can help find highly empathetic people with 89% confidence (using logistic regression) in the recruitment process. Instead of using empathy questionnaire to find empathy of a person which can be falsely answered or manipulated, use of the framework designed in this research could help find truly high empathetic candidate for the job profile with high accuracy in automated manner (without manual analysis).

Even though the framework predicts with high confidence, there are some limitation. The experiment was only conducted on 53 participants and therefore the data is not enough to train a highly reliable machine learning model. The heat map images were flattened (which looses a lot of information) for training machine learning mode, if more heat maps are provided then the neural network model could be fed with the images and it could learn from the images which would help in increasing the confidence and accuracy of the model. Also in future, instead of taking heat map images of gaze pattern of the whole video if the video itself is used as input to a sequential convolutional neural network, the accuracy of the model could increase drastically as there would be a lot of data to learn from the input. This research is based on sad emotion based empathy, in future models on different emotion based empathy could be made and the empathy of a person can be an predicted by ensemble or average of result of all the emotion based models.

References

Aktar, E., Nimphy, C. A., Kret, M. E., Pérez-Edgar, K., Bögels, S. M. and Raijmakers, M. E. (2021). Pupil responses to dynamic negative facial expressions of emotion in infants and parents, *Developmental psychobiology* **63**(7): e22190.

Aoki, Y. and Katayama, H. (2021). Development of the clinical interpersonal reactivity index to evaluate nurses' empathy, *Nursing & Health Sciences* **23**(4): 862–870.

Bochkovskiy, A., Wang, C.-Y. and Liao, H.-Y. M. (2020). Yolov4: Optimal speed and accuracy of object detection, *arXiv preprint arXiv:2004.10934* .

Caci, H., Nadalet, L., Baylé, F. J., Robert, P. and Boyer, P. (2003). Cross-cultural study of the impulsiveness-venturesomeness-empathy questionnaire (ive-7), *Comprehensive Psychiatry* **44**(5): 381–387.
URL: *https://www.sciencedirect.com/science/article/pii/S0010440X03001056*

Charan, K. S. M., Aswin, A. K., Varshini, K. S. A. and Kirthica, S. (2022). Implementation of pupil dilation in ai-based emotion recognition, *in* R. R. Raje, F. Hussain and R. J. Kannan (eds), *Artificial Intelligence and Technologies*, Springer Singapore, Singapore, pp. 447–454.

Cotler, J. L., Villa, L., Burshteyn, D., Bult, Z., Grant, G., Tanski, M. and Parente, A. (2020). An interdisciplinary approach to detecting empathy through emotional analytics and eye tracking, *J. Comput. Sci. Coll.* **35**(8): 87–95.

Cowan, D. G., Vanman, E. J. and Nielsen, M. (2014). Motivated empathy: The mechanics of the empathic gaze., *Cognition Emotion* **28**(8): 1522 – 1530.
URL: *https://search.ebscohost.com/login.aspx?direct=trueAuthType=ip,cookie,shibdb=bsuAN=9837 livescope=sitecustid=ncirlib*

Davis, M. H. (1983). Measuring individual differences in empathy: evidence for a multidimensional approach., *Journal of personality and social psychology* **44**(1): 113.

Fuhl, W., Kübler, T., Sippel, K., Rosenstiel, W. and Kasneci, E. (2015). Excuse: Robust pupil detection in real-world scenarios, *in* G. Azzopardi and N. Petkov (eds), *Computer Analysis of Images and Patterns*, Springer International Publishing, Cham, pp. 39–51.

Fuhl, W., Santini, T. C., Kübler, T. and Kasneci, E. (2016). Else: Ellipse selection for robust pupil detection in real-world environments, *Proceedings of the Ninth Biennial ACM Symposium on Eye Tracking Research amp; Applications*, ETRA '16, Association

for Computing Machinery, New York, NY, USA, p. 123–130.
URL: *https://doi.org/10.1145/2857491.2857505*

Fuhl, W., Santini, T., Kasneci, G. and Kasneci, E. (2016). Pupilnet: Convolutional neural networks for robust pupil detection, *arXiv preprint arXiv:1601.04902* .

Garbin, S. J., Shen, Y., Schuetz, I., Cavin, R., Hughes, G. and Talathi, S. S. (2019). Openeds: Open eye dataset, *arXiv preprint arXiv:1905.03702* .

Han, S. Y., Kim, Y., Lee, S. H. and Cho, N. I. (2019). Pupil center detection based on the unet for the user interaction in vr and ar environments, *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 958–959.

Harrison, N. A., Wilson, C. E. and Critchley, H. D. (2007). Processing of observed pupil size modulates perception of sadness and predicts empathy., *Emotion* **7**(4): 724.

Jongerius, C., Twisk, J. W., Romijn, J. A., Callemein, T., Goedemé, T., Smets, E. and Hillen, M. A. (2021). The influence of face gaze by physicians on patient trust: an observational study, *Journal of General Internal Medicine* pp. 1–7.

Kim, J., Stengel, M., Majercik, A., De Mello, S., Dunn, D., Laine, S., McGuire, M. and Luebke, D. (2019). Nvgaze: An anatomically-informed dataset for low-latency, near-eye gaze estimation, *Proceedings of the 2019 CHI conference on human factors in computing systems*, pp. 1–12.

King, D. E. (2009). Dlib-ml: A machine learning toolkit, *The Journal of Machine Learning Research* **10**: 1755–1758.

Klin, A., Jones, W., Schultz, R., Volkmar, F. and Cohen, D. (2002). Visual Fixation Patterns During Viewing of Naturalistic Social Situations as Predictors of Social Competence in Individuals With Autism, *Archives of General Psychiatry* **59**(9): 809–816.
URL: *https://doi.org/10.1001/archpsyc.59.9.809*

Kothari, R. S., Chaudhary, A. K., Bailey, R. J., Pelz, J. B. and Diaz, G. J. (2020). Ellseg: An ellipse segmentation framework for robust gaze tracking, *arXiv preprint arXiv:2007.09600* .

Lim, J. Z., Mountstephens, J. and Teo, J. (2020). Emotion recognition using eye-tracking: Taxonomy, review and current challenges, *Sensors* **20**(8).
URL: *https://www.mdpi.com/1424-8220/20/8/2384*

Mackes, N. K., Golm, D., O'Daly, O. G., Sarkar, S., Sonuga-Barke, E. J., Fairchild, G. and Mehta, M. A. (2018). Tracking emotions in the brain – revisiting the empathic accuracy task, *NeuroImage* **178**: 677–686.
URL: *https://www.sciencedirect.com/science/article/pii/S1053811918305068*

Miron, C., Pasarica, A., Bozomitu, R. G., Manta, V., Timofte, R. and Ciucu, R. (2019). Efficient pupil detection with a convolutional neural network, *2019 E-Health and Bioengineering Conference (EHB)*, pp. 1–4.

Moutinho, R., Castro, S. L. and Silva, S. (2021). Look at me: The relation between empathy and fixation on the emotional eye-region in low vs. high social anxiety, *Journal of Behavior Therapy and Experimental Psychiatry* **70**: 101610.
URL: *https://www.sciencedirect.com/science/article/pii/S0005791620300641*

Nebi, E., Altmann, T. and Roth, M. (2022). The influence of emotional salience on gaze behavior in low and high trait empathy: an exploratory eye-tracking study, *The Journal of Social Psychology* **162**(1): 109–127. PMID: 34935601.
URL: *https://doi.org/10.1080/00224545.2021.2001410*

Olderbak, S., Sassenrath, C., Keller, J. and Wilhelm, O. (2014). An emotion-differentiated perspective on empathy with the emotion specific empathy questionnaire, *Frontiers in Psychology* **5**.
URL: *https://www.frontiersin.org/article/10.3389/fpsyg.2014.00653*

Olson, J. K. (1995). Relationships between nurse-expressed empathy, patient-perceived empathy and patient distress, *Image: The Journal of Nursing Scholarship* **27**(4): 317–322.

Redfern, S., Dancey, C. P. and Dryden, W. (1993). Empathy: Its effect on how counsellors are perceived, *British Journal of Guidance & Counselling* **21**(3): 300–309.

Simon, L. S., Rosen, C. C., Gajendran, R. S., Ozgen, S. and Corwin, E. S. (2022). Pain or gain? understanding how trait empathy impacts leader effectiveness following the provision of negative feedback., *Journal of Applied Psychology* **107**(2): 279.

Spreng*, R., McKinnon*, M., Mar, R. and Levine, B. (2003). The toronto empathy questionnaire: Scale development and initial validation of a factor-analytic solution to multiple empathy measures., *Journal of Personality Assessment* **91**(1): 62 – 71.

Stosic, M. D., Fultz, A. A., Brown, J. A. and Bernieri, F. J. (2022). What is your empathy scale not measuring? the convergent, discriminant, and predictive validity of five empathy scales, *The Journal of Social Psychology* **162**(1): 7–25.
URL: *https://doi.org/10.1080/00224545.2021.1985417*

Tonsen, M., Zhang, X., Sugano, Y. and Bulling, A. (2016). Labelled pupils in the wild: a dataset for studying pupil detection in unconstrained environments, *Proceedings of the ninth biennial ACM symposium on eye tracking research & applications*, pp. 139–142.

Warnell, K. R., De La Cerda, C. and Frost, A. (2021). Disentangling relations between attention to the eyes and empathy., *Emotion* .

Yeo, S. and Kim, K.-J. (2021). A validation study of the korean version of the toronto empathy questionnaire for the measurement of medical students' empathy, *BMC Medical Education* **21**(1): 1–8.

Zhang, J., Park, S., Cho, A. and Whang, M. (2022). Significant measures of gaze and pupil movement for evaluating empathy between viewers and digital content, *Sensors* **22**(5).
URL: *https://www.mdpi.com/1424-8220/22/5/1700*