Eye Swiping: Choosing People With Pupillary Response

Mohd Saif Khan
120803, HCI
mohd.saif.khan@uni-weimar.de

Abstract

We live in a world where Social media has given rise to a powerful sense of interconnection. This constant connection and communication have affected our social habits and sense of community in a way that could not have been predicted a few years ago. It has given rise to several dating applications in which people swipe left or right to choose or reject the people they find attractive. There are physiological responses like increased heart rate, pupillary changes in the eyes, skin conductance changes in the body when people find someone attractive. In this paper, we are examining the diameter of the pupil in the eyes to detect attractiveness. We have built and trained machine learning classifiers to predict if a person is feeling attracted to an image of a person, by observing the changes in pupil diameter for a given interval of time.

1 Introduction

Social media is a powerful force that is influencing billions of lives all over the globe. People have become more connected and can access a host of information at an enormous pace. This boom of social media has resulted in a change in people's psychology and behavioral patterns. People are getting happier with the number of likes and upvotes on their social media profiles. This feeling of connectedness has given rise to many dating applications where people can find their potential partners. These dating applications follow a simple interface and provide lots of profiles to choose their companions. People can add their photos, personal information, likes or dislikes, etc in their profiles. Now other people can see their profile and can swipe left to reject and swipe right to select. It has made dating a whole lot easier than it used to be.

According to (Szwoch, 2015), emotion is the mental state in humans that arises spontaneously and lasts for a few seconds or minutes and is different from the mood that can last a few hours or even a few days. These emotions can be classified and can be used to understand the human psyche and feelings. Emotion recognition based on physiological signals has gained lots of traction in recent years. (Kołakowska et al., 2014) has shown its application in scenarios like safe driving cars, voice assistants, etc. With correct information about human emotions based on physiological signals, we can create systems that don't require manual human intervention.

When people find someone attractive, the body produces different physiological signals as explained by (Szwoch, 2015). If we were to measure these responses, we would be able to tell if a person is feeling attracted to another person or not. Pupil diameter is one such physiological signal which could reveal the feeling of attraction. The pupil diameter changes whenever we encounter positive or negative stimuli in our surroundings. These changes in pupil diameter can tell us if a person is feeling attracted to a another person's profile on a dating application.

In this paper, we have tried to differentiate the feeling of attractiveness experienced by a person by observing their pupil diameter. The pupil diameter is measured for a little time using a device called Tobii pro.(Tobii Pro AB, 2014) Then we cleaned these readings by filling up some missing values by interpolation and smooth the signal, so it shouldn't contain sudden peaks and valleys. We added each reading, to a dataset of readings of pupil diameter. We also used a dataset of pupil diameters that was used to classify negative, positive, and neutral feelings by viewing an image.(Jami, 2020) We fed these datasets to Time Series classification algorithms that generated the classifiers to predict if a person is feeling attracted or not. The classifiers were giving out good accuracies on the test datasets.

2 Related Work

The pupillary studies have been done and studied several times in the past. (Alshehri and Alghowinem, 2013) studied various eye movement features like pupil size, time of the first fixation, first fixation duration, fixation duration, and fixation count to detect emotional states in humans. Their data did not show significance in the average pupil size differences between pleasant and unpleasant clips stimulation. (Lanatà et al., 2011) investigated the relationship between emotions and information coming from the eyes, i.e. pupil size variation and eye gaze tracking. They achieved an accuracy of 90% for neutral images and 80% for images of high arousal. In a similar work, (Aracena et al., 2015) generated a mapping between pupil size and gaze behavior and emotional states (positive, negative, and neutral) provoked by visual stimuli (IAPS images[(Lang et al., 1999)]). Pupil size and gaze position were recorded using an eye tracker. The pupil tends to constrict in size when people see unpleasant images and, the pupil becomes dilated when the emotion results due to pleasant pictures.

The information given by eyes has also been used to classify a person as attractive or not attractive. (Cornelissen et al., 2009) uses fixation patterns and pupil diameter to analyze a person's attractiveness and found the difference in fixation patterns in males and females. Eye fixations were used by (Valuch et al., 2015) as the basis for the experiment. They find that attractive faces hold or capture attention more effectively than less attractive faces based on saccades. (Getov et al., 2016) discusses the different gaze patterns of individuals towards attractive and unattractive faces. People will display distinctive gaze patterns when looking at attractive faces and will spend longer viewing times on facial features. (Kościński, 2007) discusses the general patterns of facial preferences and finds that Averageness and symmetry are preferred by both males and females.

The first concrete step to classify human emotions based on pupil diameter was presented by (Babiker et al., 2013). The pupil is controlled by the autonomic nervous system(ANS) that acts unconsciously and regulates bodily functions, such as the heart rate, digestion, respiratory rate, pupillary response, urination, and sexual arousal.[(Wikipedia, 2021a)] In this paper, authors were able to differentiate between negative and positive emotions using pupil diameter and observed that negative emotions have a higher effect on pupil diameter. In the follow-up paper, (Babiker et al., 2015) introduced a k-nearest neighbor algorithm to classify emotions. They achieved high accuracies on the test datasets.

3 Experimental design

The purpose of this study is to classify an image of a person as attractive or not attractive. We designed an experiment that facilitated the collection of pupil diameter data for different participants. The experiment was conducted in a controlled environment to minimize the effect of the external environment, peer pressure, and any other stimuli. We used the Tobii Pro eye-tracking device to record the size of the pupil diameter. It is a screen-based eye-tracker that is capable of capturing the information of pupil diameter at 60 Hz. The experiment was designed in Psychopy software. The Tobii pro provides an interface to interact with the device directly using python and, all the gaze parameters like pupil diameter, gaze position, etc. could be stored and used.

The experiment followed a within-subject design method and consists of 30 different baseline and stimuli images each, that appear randomly for a participant. The images mainly contain the face stimuli with less background information. Since the pupil is hypersensitive to the change in color of the stimuli, each stimuli image is preceded by a baseline image. The baseline image is a scrambled version of the original stimuli image to maintain the color contrast of the image.

The experiment has five different stages seen in Figure 1 and, each state serves a definitive purpose. The experiment starts with a welcome screen, then a loop of baseline image, stimuli image, and feedback slider screen is presented to get the response from a participant about the attractiveness of the image. The participant can choose a value between 1 and 4, where one is the most attractive and four is the least attractive. The loop continues until all the images are presented to the participant. After the loop of images is completed, the user is presented with a Thank You screen. In the following section,

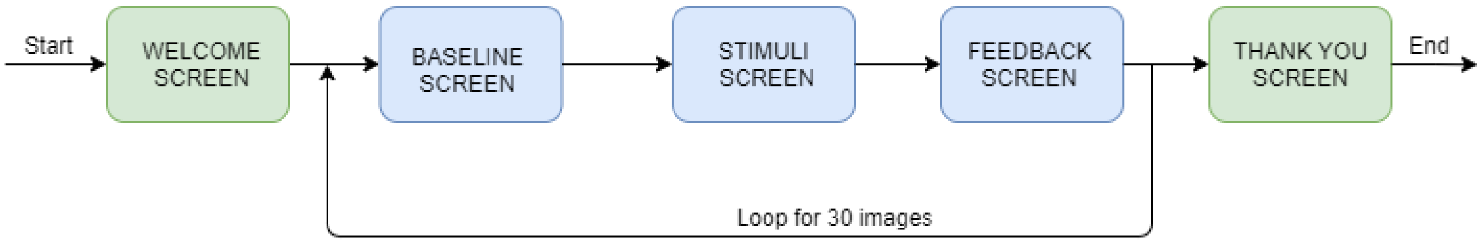

Figure 1: Experimental setup

we will explain each stage of the experiment in detail.

1. Welcome Screen- This screen serves as the gateway to the experiment. The participants can enter their details and choose the gender of images according to their preference. It also serves as a short introduction screen that provides information about the experiment to the participant. The participant must click next to start the experiment.

2. Baseline Screen- The participant is shown a baseline image that is a scrambled version of the original stimuli image. It will minimize the effect of a sudden change in color contrast and lighting and makes the eyes ready for the main stimuli image. This screen is visible for 2 seconds and, the baseline diameter is averaged over the period to give the average baseline pupil diameter value.

3. Stimuli Screen- The participant is shown a high-quality image of a person. The image is visible for 5 seconds and, Tobii pro captures the pupil diameter reading at a frequency of 60 Hz in that period.

4. Feedback screen- The Feedback screen contains a slider that contains values from 1 to 4. The screen waits for the input given by the participant. The participant can select the value on the slider based on the attractiveness of the image. Value 4 corresponds to the most attractive and value 1 corresponds to the least attractive. After the user selects the value on the feedback screen, the loop continues.

5. Thank You Screen- This screen tells the participant that the experiment is being completed and gives them instructions to end the experiment.

The readings taken by the experiments are noisy and incomplete. When a participant blinks during the experiment, the Tobii pro doesn't take the reading. This results in gaps in the readings that need to be corrected. We employed the spline interpolation technique to fill the gaps between readings. The data were then smoothed to get rid of sudden peaks or valleys that might skew the learning of Machine learning models. Finally, the data contains a time series of pupil diameter and, the attractiveness value given by the participant during the experiment.

4 Data

We have used two separate datasets in our experiments. The first dataset was compiled by (Jami, 2020) for his master's thesis and will be referred to as *TanveerDS*. We have prepared the second dataset by following a controlled experiment and will be referred to as *ControlledDS*. Both datasets use the Tobii eye tracker to take readings of pupil diameter.

In the *TanveerDS*, there were 10 participants, and for each participant, ninety different images were used for different classes of emotions. There are three different classes of emotions in the dataset which are positive, negative, and neutral. For each image, there are 300 readings of pupil diameter. Let's call this dataset the *original raw* dataset, but the total number of readings were not enough to train a classifier. To increase the number of data points for the classifier, he applied augmenting techniques to the *original raw* data. Data Augmentation introduced a few synthetic data points to the original data set. (Le Guennec et al., 2016) explained several techniques for augmenting time series data in their research. Window Warping was used to augment additional data using the *original raw* dataset. Window Warping is a technique that involves either increasing or decreasing the speed of an interval in a time series. A small part of the data was selected randomly, and then with a prob-

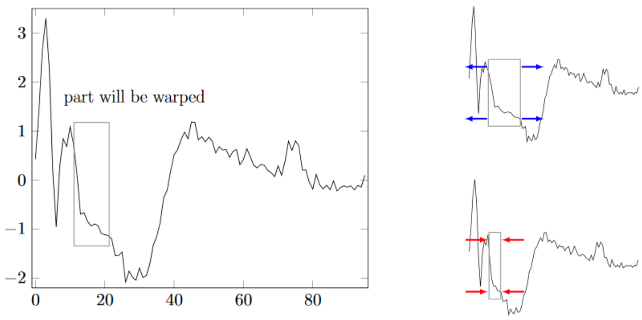

Figure 2: Data Augmentation using Window Warping

ability, this region was stretched or squeezed to obtain a new data point. This can be seen in Figure 2.

Therefore the dataset becomes bigger, and more data points of similar kind were introduced in the dataset. Lets call this dataset, the *augmented* dataset. He then split the *augmented* dataset into test and train datasets for training and testing of the classifier. Using this *augmented* dataset, he created another dataset which we will call *augmented_train* dataset. The training dataset of *augmented_train* dataset was made with the training dataset of *augmented* dataset. The test dataset was chosen randomly from the *original raw* dataset. Finally, we will run our evaluations on both *augmented* and *augmented_train* datasets.

The *ControlledDS* was recorded to classify attractiveness by seeing an image of a person. It contains two separate classes to classify an image as attractive or unattractive. There are 30 images on which a participant has to decide the corresponding attractiveness. For each image, there are 300 readings of pupil diameter at a frequency of 60 readings per second. In 5 seconds, the participants can blink their eyes, and the data is not recorded when the eyes are closed. This is being shown in Figure 3. This results in a gap in the data points obtained. To mitigate this problem, spline interpolation was applied to fill the gaps between the data points. The corrected data point is shown in Figure 4.

Due to Covid-19, there was a restriction in place prohibiting us from taking readings with more than two people in a room, therefore we were only able to take a few readings within our group. A classifier needs a large amount of data for training, and as a result, we were not able to build a classifier for this *ControlledDS* dataset. In the future, someone can

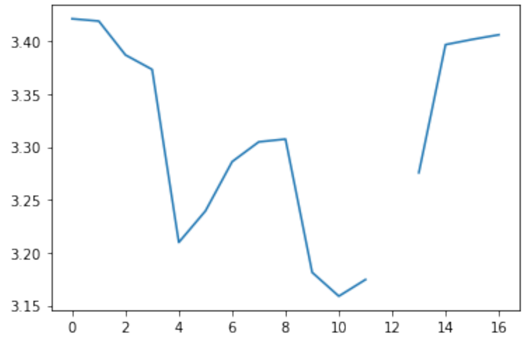

Figure 3: Gaps in data due to blinking

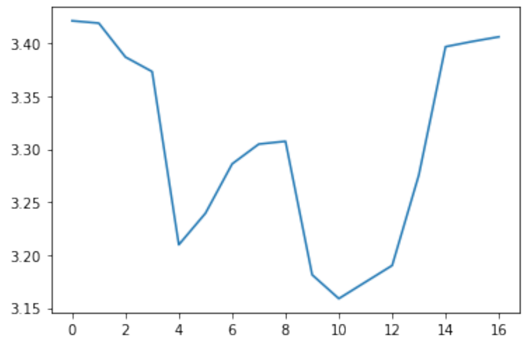

Figure 4: Spline Interpolation to fill the gap

take the appropriate number of readings and use our implemented algorithms to build the classifiers.

5 Dynamic Time Warping(DTW)

The main aim of the experiment was to differentiate between positive and negative emotions. In both cases, we get data that varies over time. Hence we can say that to classify the emotions as negative or positive, we have to classify the time series data of the emotions. As discussed by (Jeremy Zhang, 2020), in a conventional way, the time points can be converted to vectors and a Euclidean distance can be calculated but this technique has a lot of disadvantages when comparing time series data as seen in Figure 5. It shows that this technique is very limiting and does not give proper information about the similarity between the series.

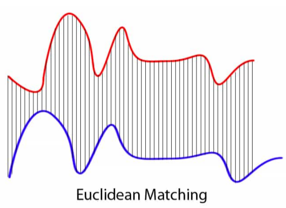

Figure 5: Euclidean Matching

DTW, on the other hand, is a distance-based technique to measure similarities between time series of different lengths i.e. if the sequences have different speeds. This technique was developed to identify the same words even if they were said with different intonation and speed. For example, if a person greets someone in the morning before his cup of coffee. He will say 'Gooodd moornning', in a long and dull intonation but in the office, he will say a crisp 'Good Morning'. So to identify both utterances of 'Good Morning' as similar, (Brown and Rabiner, 1982) proposed the method that serves as a proper similarity measure. (Jeremy Zhang, 2020) mentions the stark differences between comparing time series with Euclidean distance as a measure of comparison which we can see in Figure 6. The sequences are 'Warped' Non Linearly in time. DTW provides warping invariance which is a peak-to peak and valley-to-valley alignment of 2 different time series.

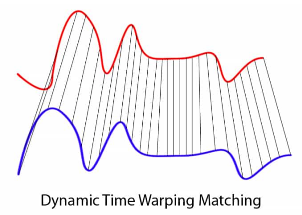

Figure 6: Dynamic Time Warping matching

This technique can be applied to any time-varying data like Speech Recognition, Pattern Recognition, Power Series Analysis, etc.(Wikipedia, 2021b) Therefore we can apply this technique in our application and learn about the distance measures between positive and negative emotions. The series of positive, negative, and neutral emotions is represented below as an example in Figure 7. When exposed to stimuli the pupil diameter first decreases in size and then it increases. The pupil diameter is represented as a function of time. The total length of the time frame is 5 seconds and the value of pupil diameter is recorded at a frequency of 60.

The local peaks and valleys of the data in the series are also plotted as red points on the graph. The main idea is to study the data from its deepest valley to its highest peak and run a DTW measure on this slice of data, as this is the main response of pupil diameter when it is exposed to visual stimuli. The slice of data from Figure 7 ie, from the lowest

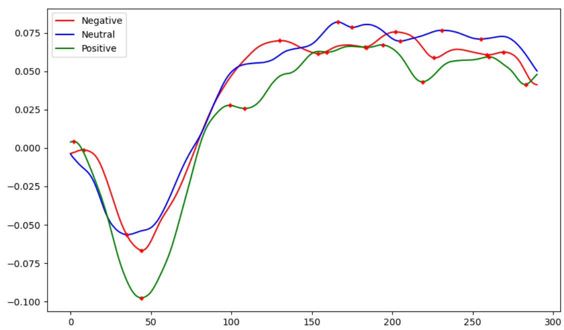

Figure 7: Series of positive, negative, and neutral emotions

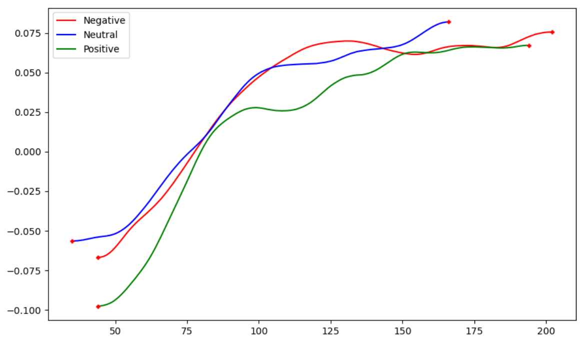

Figure 8: Lowest valley to the highest peak of the series

valley to the highest peak is represented in Figure 8.

DTW on the slice of the data: The DTW algorithm is run on all the responses of the experiment. We got the similarity measure of series with DTW that can be seen in the Figure 9. We get the following insights from our experiment.

1. Positive and neutral emotions are more similar to each other.

2. The negative emotions are the most dissimilar.

6 Time Series classification

The readings obtained during the experiments contain a time series of pupil diameter over an interval. The goal of the study was to classify the emotions of the participants based on the response of pupil diameter. Therefore, we applied time series classification algorithms to differentiate the feeling of attractiveness based on the images.

During the last decade, there have been many advances in the number and the quality of algo-

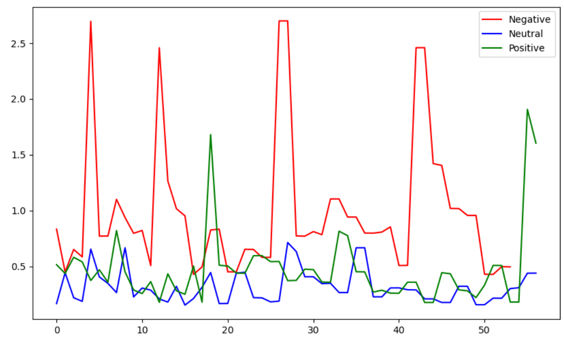

Figure 9: DTW similarity measure on the series

rithms for time series classification. Since time series inherently have natural temporal ordering, a lot of deep learning models like CNN can also be applied. In this study, we have tried different algorithms, and they all perform reasonably well on both emphTanveerDS datasets. The various classes of time series classification algorithms are based on the types of features the algorithm uses. The features can be global(whole series) or local(bins, random intervals, sliding window, etc.). The algorithms can be also be classified as Dictionary-based, Interval-based, Frequency-based, Shapelet-based, Distance-based, Deep Learning based, and Ensemble/Hybrid. In the next section, we will discuss the algorithms used in the study.

6.1 KNN with DTW

K-Nearest Neighbour algorithm is a distance-based algorithm where the Euclidean distance measure has been replaced by Dynamic time warping. As discussed in the previous sections that DTW measures the similarity between two sequences that are not aligned in speed. It is a simple algorithm and does not require much hyperparameter training. (Switonski et al., 2019) applied the technique to motion data and obtained considerable results.

6.2 Time Series Forest Classification

In a time series, every value can be used as a regular feature, and an algorithm like Nearest Neighbor is sensitive to distortions and may result in unsatisfactory accuracy performances. In a random forest algorithm, features are calculated over an interval of time and are used to build a decision tree over which we can train a classifier. In a decision tree classifier, the better the split criteria, the better is the result. (Deng et al., 2013) employs a new measure called Entrance gain to identify the high-quality splits that help to learn more from less amount of data. It randomly picks features from each tree node and is comprehensively less computationally expensive than the Nearest Neighbor algorithm.

6.3 WEASEL

(Schäfer and Leser, 2017) presented a dictionary-based classification algorithm in which a sliding window is run across the time series to extract discrete features that are used to build a classifier. This algorithm uses SFA(Symbolic Fourier approximation), that was developed by (Schäfer and Högqvist, 2012) to approximate high dimensional features in time series. It uses varying lengths of windows that also considers the order of windows, to better capture the characteristics of each class. It produces a small and discriminative feature set that reduces computation without affecting the accuracy.

6.4 Random interval spectral ensemble (RISE)

It is a frequency-based algorithm presented by (Lines et al., 2018) that extracts spectral features from a time series. It uses the idea of tree-based ensembles used by (Deng et al., 2013) and build trees on random intervals to construct a Random Forest. This algorithm is also used in the current state of the art for Time Series classification: HIVE-COTE algorithm.

6.5 Mr-SEQL

It is a shapelet-based univariate time series classifier that trains linear classification algorithms like linear and logistic regression presented by (Nguyen et al., 2020). The features are extracted from multiple symbolic representations of time series, such as SAX(Symbolic aggregate approximation) in the time domain and SFA(Symbolic Fourier approximation) in the frequency domain. They have extended a symbolic sequence classifier (SEQL) proposed by (Ifrim and Wiuf, 2010) to work with SAX and SFA. It delivers an accurate and interpretable time series classifier that takes less time to train. This algorithm can also be applied to a variable-length time series.

6.6 ROCKET

ROCKET(RandOm Convolutional KErnel Transform) is an ensemble method in which (Dempster et al., 2019) have built a Linear Classifier using random convolutional kernels. It is specifically built to have less computational complexity, require less training, and can scale to large datasets without any difficulty. It transforms time series using many random convolutional kernels, i.e., kernels with random lengths, weights, bias, dilation, and padding. The transformed features are used to train a logistic regression classifier.

We have applied all the above algorithms to both *TanveerDS* datasets. The results with small datasets are encouraging and can further improve if the number of data points increases. The downside to a large dataset is that it will require more time to train the classifiers, but we will get better classifier performances. Apart from these algorithms, we looked at some other algorithms like Inception time((Fawaz et al., 2019)), TS-chief((Shifaz et al., 2019)), HIVE-COTE((Lines et al., 2018)), BOSS((Schäfer, 2015)), Shapelet Transform((Hills et al., 2013)) and Resnet((He et al., 2015)). These algorithms were taking a significant amount of time to train the classifier on the machines at our disposal. They required more computational resources than we had available on our systems. Therefore, we were not able to generate any results for them. We would have achieved a better outcome by using the above mentioned algorithms.

7 Results

We have applied multiple algorithms to *TanveerDS*(*augmented* and *augmented_train*) dataset and have obtained the following results as shown in Table 1. We will discuss the results of both the datasets separately.

7.1 *augmented_train* dataset

Since the *augmented_train* dataset uses the test set from the *original raw* dataset, there is a greater chance that this dataset contains fewer repeated examples from the augmentation stage. So we can rely more on this dataset for accurate results because it has fewer augmented readings. But because it is a small dataset, the final results were not encouraging. Nonetheless, we can achieve significant observations from the results.

We can see that K- Nearest neighbor has performed the worst as expected with an accuracy of 0.2976. Forest classifier with an accuracy of 0.3184 was a bit more accurate than KNN, but it was expected because the distance and tree-based algorithms are not known to perform well. We know that a pupillary response can provide us with parameters like the amplitude, the Latency, the Rise Time, and the Recovery Time. Since distance-based algorithms don't consider time points contiguously in an interval, they are less affected by the parameters, and therefore they perform poorly.

The more sophisticated algorithms implement the classifier using an interval of time or frequency, and therefore they perform better. In some way, the classifiers mentioned below create their features using the parameters mentioned above. The ROCKET and WEASEL got an accuracy of around 0.34. The shapelet-based Mr-SEQL achieved an accuracy of 0.363, which is also way ahead of the KNN. But the best algorithm came out to be the RISE algorithm that achieved an accuracy of 0.407. We can conclude from our readings that frequency-based algorithms are more suited to classify pupillary data.

7.2 *augmented* dataset

We can quickly observe that the algorithms are overfitting in this dataset. It is the result of the data augmentation that (Jami, 2020) did for the thesis. The test dataset out of this dataset was less diverse, so the classifiers were overfitted. Even the KNN algorithm has achieved the state of art accuracy of 0.875, which is even better than WEASEL. It is surprising since WEASEL is a state of the art algorithm in Time Series classification. Therefore we will conclude that classifiers were overfitting over the *augmented* dataset and the results obtained are less reliable.

8 Conclusion

The results support earlier findings that it is easier to differentiate between Neutral and Affective States. Using DTW, we found out that a negative state is much more pronounced than a positive or a neutral state. The pupillary reaction is quite an unstable signal that changes very frequently. As the signals contained many invalid regions due to eye blinks and edge artifacts, it is a great challenge to generate smooth and continuous signals from them. Every individual responds differently to the same stimuli and, this can also result in inaccuracies.

We used six different algorithms and build classifiers for each algorithm using *TanveerDS* datasets.

Algorithm/Classifier	*augmented_train* Accuracy	*augmented* Accuracy
KNN with DTW	0.297	0.875
Time Series Forest	0.318	0.872
WEASEL	0.348	0.866
RISE	0.407	0.706
Mr-SEQL	0.363	0.882
ROCKET	0.342	0.882

Table 1: Evaluation results on the *augmented_train* and *augmented* test sets with different algorithms.

We have shown with our experiments that we can identify the feeling of attractiveness in a person with the help of machine learning analysis of pupil diameter. The classifiers can differentiate between the neutral and affective states. As mentioned in the results, if a classifier takes 'rise time' as a feature, it can have better efficiency. Therefore we should only choose the algorithms which use an interval of the signal, in the future.

RISE algorithm comes out to be the best algorithm on the *original raw* test dataset, followed by ROCKET and WEASEL algorithms. It is sufficient to say that with a large dataset, the performances of the classifiers will increase. We have added these classifiers in the code of our experiments discussed in Section 3 and, can directly predict if the participant is feeling attracted towards an image or not.

9 Future Work

There are a lot of possibilities for future work using this project as a base. We can implement better machine learning algorithms by running the algorithms on machines that have considerable computational power. The state-of-the-art algorithms are computation-intensive and require a lot of GPU power. We can also increase the number of reading from participants that will result in a bigger dataset. A bigger dataset will train the Machine learning algorithms more effectively. The classifiers will not overtrain and will result in overall high-performance accuracies. We can also develop a user-friendly application that will be easy to use and, it will convert this experiment into an efficient use case.

10 Code Implementation

The code implementation for creating the experiment with tobii can be found in the following Github repo .

`https://github.com/saifkhan-m/TobbiSwiping`

The code implementation of the designing the classifiers can be found in the following Github repo .

`https://github.com/saifkhan-m/EyeSwiping`

References

M. Alshehri and S. Alghowinem. 2013. An exploratory study of detecting emotion states using eye-tracking technology. In *2013 Science and Information Conference*, pages 428–433.

C. Aracena, S. Basterrech, V. Snáel, and J. Velásquez. 2015. Neural networks for emotion recognition based on eye tracking data. In *2015 IEEE International Conference on Systems, Man, and Cybernetics*, pages 2632–2637.

A. Babiker, I. Faye, and A. Malik. 2013. Pupillary behavior in positive and negative emotions. In *2013 IEEE International Conference on Signal and Image Processing Applications*, pages 379–383.

Areej Babiker, Ibrahima Faye, Kristin Prehn, and Aamir Malik. 2015. Machine learning to differentiate between positive and negative emotions using pupil diameter. *Frontiers in Psychology*, 6.

M. Brown and L. Rabiner. 1982. Dynamic time warping for isolated word recognition based on ordered graph searching techniques. In *ICASSP '82. IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 7, pages 1255–1258.

Piers Cornelissen, Peter Hancock, Vesa Kiviniemi, Hannah George, and Martin Tovée. 2009. Patterns of eye-movements when male and female observers judge female attractiveness, body fat and waist-to-hip ratio. *Evolution and Human Behavior*, 30.

Angus Dempster, François Petitjean, and Geoffrey I. Webb. 2019. Rocket: Exceptionally fast and accurate time series classification using random convolutional kernels.

Houtao Deng, George C. Runger, Eugene Tuv, and Vladimir Martyanov. 2013. A time series forest for classification and feature extraction. *CoRR*, abs/1302.2277.

Hassan Ismail Fawaz, Benjamin Lucas, Germain Forestier, Charlotte Pelletier, Daniel F. Schmidt, Jonathan Weber, Geoffrey I. Webb, Lhassane Idoumghar, Pierre-Alain Muller, and François Petitjean. 2019. Inceptiontime: Finding alexnet for time series classification. *CoRR*, abs/1909.04939.

Ivan Getov, Greg Gettings, A.J. Villanueva, and Chris Belcher. 2016. Differential viewing strategies towards attractive and unattractive human faces.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2015. Deep residual learning for image recognition. *CoRR*, abs/1512.03385.

Jon Hills, Jason Lines, Edgaras Baranauskas, James Mapp, and Anthony Bagnall. 2013. Classification of time series by shapelet transformation. *Data Mining and Knowledge Discovery*, 28.

Georgiana Ifrim and Carsten Wiuf. 2010. Bounded coordinate-descent for biological sequence classification in high dimensional predictor space. *CoRR*, abs/1008.0528.

Tanveer Al Jami. 2020. The affective pupil: Automized emotion recognition for human-machine interaction. `https://github.com/skeletor069/affective_pupil`.

Jeremy Zhang. 2020. Dynamic time warping. Https://towardsdatascience.com.

Agata Kołakowska, Agnieszka Landowska, Mariusz Szwoch, Wioleta Szwoch, and Michał Wróbel. 2014. Emotion recognition and its applications. *Advances in Intelligent Systems and Computing*, 300:51–62.

Krzysztof Kościński. 2007. Facial attractiveness: General patterns of facial preferences. *Anthropological Review*, 70:45–79.

A. Lanatà, A. Armato, G. Valenza, and E. P. Scilingo. 2011. Eye tracking and pupil size variation as response to affective stimuli: A preliminary study. In *2011 5th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth) and Workshops*, pages 78–84.

P. J. Lang, M. M. Bradley, and B. N. Cuthbert. 1999. International affective picture system (iaps): Technical manual and affective ratings. Technical report, University of Florida.

Arthur Le Guennec, Simon Malinowski, and Romain Tavenard. 2016. Data Augmentation for Time Series Classification using Convolutional Neural Networks. In *ECML/PKDD Workshop on Advanced Analytics and Learning on Temporal Data*, Riva Del Garda, Italy.

Jason Lines, Sarah Taylor, and Anthony Bagnall. 2018. Time series classification with hive-cote: The hierarchical vote collective of transformation-based ensembles. *ACM Trans. Knowl. Discov. Data*, 12(5).

Thach Le Nguyen, Severin Gsponer, Iulia Ilie, Martin O'Reilly, and Georgiana Ifrim. 2020. Interpretable time series classification using linear models and multi-resolution multi-domain symbolic representations.

Patrick Schäfer and Ulf Leser. 2017. Fast and accurate time series classification with weasel. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, CIKM '17, page 637–646, New York, NY, USA. Association for Computing Machinery.

Patrick Schäfer. 2015. The boss is concerned with time series classification in the presence of noise. *Data Mining and Knowledge Discovery*, 29.

Patrick Schäfer and Mikael Högqvist. 2012. Sfa: A symbolic fourier approximation and index for similarity search in high dimensional datasets. pages 516 – 527.

Ahmed Shifaz, Charlotte Pelletier, François Petitjean, and Geoffrey I. Webb. 2019. TS-CHIEF: A scalable and accurate forest algorithm for time series classification. *CoRR*, abs/1906.10329.

Adam Switonski, Henryk Josinski, and Konrad Wojciechowski. 2019. Dynamic time warping in classification and selection of motion capture data. *Multidimensional Systems and Signal Processing*, 30(3):1437–1468.

Wioleta Szwoch. 2015. Emotion recognition using physiological signals. In *Proceedings of the Mulitimedia, Interaction, Design and Innnovation*, MIDI '15, New York, NY, USA. Association for Computing Machinery.

Tobii Pro AB. 2014. Tobii pro lab. Computer software.

Christian Valuch, Lena Pflüger, Bernard Wallner, Bruno Laeng, and Ulrich Ansorge. 2015. Using eye tracking to test for individual differences in attention to attractive faces. *Frontiers in Psychology*, 6.

Wikipedia. 2021a. Autonomic nervous system — Wikipedia, the free encyclopedia. [Online; accessed 1-January-2021].

Wikipedia. 2021b. Dynamic time warping — Wikipedia, the free encyclopedia. [Online; accessed 9-January-2021].