Automatic Detection of Mind Wandering Based on Eye Movement from the Mementos Data Set

Max van Dijk
Supervisors: Xucong Zhang, Bernd Dudzik, Hayley Hung
EEMCS, Delft University of Technology, The Netherlands

June 19, 2022

A Dissertation Submitted to EEMCS faculty Delft University of Technology,
In Partial Fulfilment of the Requirements
For the Bachelor of Computer Science and Engineering

Abstract

Mind wandering is a phenomenon that is used to describe moments where a person's attention appears to shift away to something that is not related to the primary task, which can have a negative influence on the task performance. In this research, the aim is to create a viable algorithm that can automatically detect mind wandering based on eye movement from the Mementos data set. A method to automatically detect mind wandering could be used in online education in order to help students study more effectively, for example. The Mementos data set contains, unlike previous research, videos captured in an uncontrolled environment using inexpensive equipment. Features based on fixations and saccades were used to create two algorithms which were able to perform better than chance, having an average AOC-ROC of 0.63 and 0.59, as well as having an average F1 score of 0.046 and 0.041 compared to the chance-based model with an F1 score of 0.029.

1 Introduction

Mind wandering is a term that is often used to describe things such as day dreaming or moments where the attention shifts away. There is, however, not a universal definition of mind wandering. One definition that has frequently been used, as well as the definition that is used during this research says that the executive components of attention appear to shift away from the primary task during mind wandering (Smallwood & Schooler, 2006). This definition is often joined with the constraint that the attention is not shifted due to external factors or the person interacting with the external environment. These periods of mind wandering can result in a decreased task performance and a superficial representation of the external environment. Research has shown that mind wandering occurs frequently and during various different activities (Killingsworth & Gilbert, 2010). An activity in which mind wandering has extensively been researched is that of reading, where it has been found that an increase of mind wandering has a negative influence on reading comprehension (Reichle et al., 2010). Mind wandering has, however, also been shown to have positive effects. Evidence suggests that mind wandering can help people in their daily lives by enabling cognitive operations, which in turn help us remember to execute delayed intentions in the future (Baird et al., 2011).

An area that has also been researched is that of automatic detection of mind wandering. One field where such a detection method could prove useful is in online education. With the increase of online education in the recent years, the number of students who receive online education has increased drastically. There are also many platforms such as Coursera and edX, where students are able to study various topics by participating in MOOCs[1], which are often in the form of pre-recorded lectures that can be viewed online. Farley et al. (2013) conducted an experiment where participants were shown an online lecture with a duration of one hour. The participants were, during this period, asked multiple times whether they were experiencing mind wandering at that moment. The responses of the participants indicated that mind wandering occurred in 43% of the cases. If mind wandering could be detected in online education, actions could be taken the moment mind wandering is detected which could help students to study more effectively.

Eye movement has already been used in various studies to detect mind wandering. Bixler and D'Mello (2015) have tested automated gaze-based detection of mind wandering during reading. In this study it was found that certain types of fixations were longer, reading times

[1] Massive Open Online Courses

were longer than expected, and more words were skipped when mind wandering occurred. With their detection algorithm, they were able to detect mind wandering 18% better than chance. Another similar study, which also used eye movements to detect mind wandering during reading, found that the eye movement was generally less complex during mind wandering compared to reading normally (Uzzaman & Joordens, 2011). Less research has been done in the context of reactions to video content. One study used eye movement to detect mind wandering during video lectures and found that the number of fixations on the instructor increased while mind wandering (Zhang et al., 2020). Additionally, the average duration of fixations on the slides increased, while the dispersion decreased.

During this research, the context in which mind wandering will be detected differs from previous research. Firstly, the data set that will be used for this research, called Mementos, differs from data sets that have been used in previous research. The Mementos data set is a data set containing 1995 annotated individual responses to various segments of music videos. (Dudzik et al., 2021) The videos in this data set have been recorded using the participant's webcam and were recorded in their personal, uncontrolled environment. This differs greatly from most studies, which have been done in controlled environments and used a single camera for all participants, thus making sure that the conditions for each recording are the same. Furthermore, this research focuses on perceived mind wandering, meaning mind wandering that was detected from watching the recordings. Previous research often either probed the participant whether they were experiencing mind wandering, or would have the participant report moments where they realise they are mind wandering. Lastly, many studies that have focused on detecting mind wandering from eye movement have used expensive eye tracking equipment to measure several different eye features. The costs of this type of equipment reduces the accessibility and could hinder wide spread adoption.

The main question this research will aim to answer is whether it is possible to create an algorithm that automatically detects mind wandering, based on eye movement from the Mementos data set, and performs better than chance. In addition to the main research question, there are two sub-questions that will be researched. The first question will focus on finding indicators for mind wandering in the Mementos data set, whereas in the second question the differences between the gaze features will be compared during mind wandering and outside of mind wandering. These questions were answered by analyzing the videos in Mementos in order to find periods of mind wandering, after which descriptive statistics for fixations and saccades were used as features to train two different classifiers. The performance metrics of these classifiers were then compared to a chance-based model in order to analyse the performance.

2 Methodology

This section will focus on presenting the methods and materials that were used during this research in order to answer the research questions. Specifically, it will first showcase how the indicators for mind wandering in Mementos were decided and the methods that were used for annotating the Mementos data set. Following this, the ways in which the gaze data were extracted and the features that were used are discussed. Lastly, the machine learning algorithms that are used are shown.

2.1 Annotating the Data Set

The Mementos data set that is used for this study is a data set with the purpose of modelling affect and memory processing in response to music videos (Dudzik et al., 2021). As the data set was not specifically made for the purpose of detecting mind wandering, there is no information that shows when mind wandering has occurred in the data set. Since the aim of this research is to automatically detect mind wandering, periods of mind wandering that occur in Mementos must first be annotated. In order to find these periods, indicators of mind wandering needed to be decided on.

Mind Wandering Indicators

As this research is part of a wider study which focuses on automatic detection of mind wandering using audiovisual data, deciding on the indicators for mind wandering, as well as the annotating itself were done together with the peers part of this larger study. Two things were done to decide on the indicators that were used. Firstly, individual research was done with the aim of finding signs that could indicate mind wandering. Research that specifically focused on mind wandering, together with more general research regarding an individual's attention were used for this. Secondly, using the research that was gathered regarding the indicators, numerous videos in Mementos were watched together with the peer group in order to find clear instances of mind wandering. When these instances were found, the period of mind wandering would be analysed to find additional indicators or to make the indicators more specific. The indicators that were found, together with their description can be seen in Table 1.

Indicator	Description
Smiling	Sometimes a smile can be an indication of reminiscing or thinking of a good memory
Looking up / Rolling eyes	Looking up or rolling eyes for a certain amount of time could indicate that the individual is trying to recollect something
Squinting eyes	Can indicate the person is focusing on something that is likely unrelated to the music video
Frowning	A frown could be an indication of the individual thinking of negative memories
Sound of a person	When a person makes sounds that are not related to music, such as speaking to themselves, that could be interpreted as mind wandering

Table 1: Indicators for mind wandering in Mementos that were used for annotating periods of mind wandering

Annotation of the Videos

The indicators that were found using the previously mentioned methods could then be used as a guide while annotating periods of mind wandering. As the peer group contained a total of five people, groups of two and three were used for the annotating itself. There are multiple reasons for annotating the data set in groups. Since the indicators that were decided on were not used as concrete rules, it was not always clear whether a person was mind wandering. Being in a group, and being able to discuss with peers could clear this doubt. Additionally, some of the indicators can occur in subtle ways that could easily be missed when watching individually. In a group, these subtle expressions could more easily be noticed. The VGG Image Annotator[2] software was used to annotate periods of mind wandering that were found in each video.

Using this method, a total of 549 videos of Mementos were annotated. Of these videos, 54 videos were deemed unusable for this research due to various reasons such as bad video quality, or participants not being in frame for a large part of the video. Of the usable videos, 52 instances of mind wandering were found, meaning that approximately 10% of the annotated videos from Mementos contain mind wandering.

2.2 Extracting the Features

After the annotation of the data set, the gaze data was extracted from the annotated videos. The software that was chosen for the gaze extraction was OpenFace[3]. OpenFace is a toolkit for analysis of facial behaviour, including facial landmark detection, head pose estimation and eye gaze estimation. This software was chosen as it has been shown to have a state-of-the-art performance for gaze estimation, and is able to analyse videos in real-time (Baltrusaitis et al., 2018). For each video that is analysed by OpenFace, a CSV file is created which contains a gaze direction vector for each eye, as well as the gaze direction in radians, averaged for both eyes.

Once all the gaze data had been extracted from videos, the raw gaze data was converted into eye movements. Fixations and saccades have been chosen to represent eye movement during this research. Fixation are moments where the gaze is fixed on a single location and saccades are moments between fixations where there is quick eye movement. Fixations and saccades have been used in many studies regarding eye movement during mind wandering and these studies have shown that there is a correlation between the fixation duration and mind wandering (Frank et al., 2015; Foulsham et al., 2013; Reichle et al., 2010). Another reason for choosing fixations and saccades is the uncontrolled environment. Because of the uncontrolled environment, the participants are in many different positions and angles from the camera which make it difficult to use other features that change depending on the positions. Fixations and saccades, on the other hand, remain the same regardless the position or angle.

Fixation Estimation

In order to estimate the fixations and saccades from the estimated gaze date, a fixation identification algorithm is necessary. The algorithm that was chosen for this is the Velocity-Threshold Identification (I-VT) fixation classification algorithm (Salvucci & Goldberg, 2000). The I-VT algorithm uses velocity of the directional shifts in order to identify fixations, which is often given in degrees per second (°/s). The I-VT algorithm was chosen over other existing methods, as many other algorithms rely on locations on the screen in order to identify fixations, which can not easily be done in an uncontrolled environment. The velocity was calculated per frame for each video using the gaze direction that was estimated using OpenFace. The angle between two consecutive gaze directions was calculated and was then divided by the time between the two samples to produce the velocity.

The next step in the I-VT algorithm is to set a threshold that is used to identity saccades by marking each sample with a larger velocity than the threshold as a saccade. Consecutive

[2] https://www.robots.ox.ac.uk/vgg/software/via/
[3] https://github.com/TadasBaltrusaitis/OpenFace

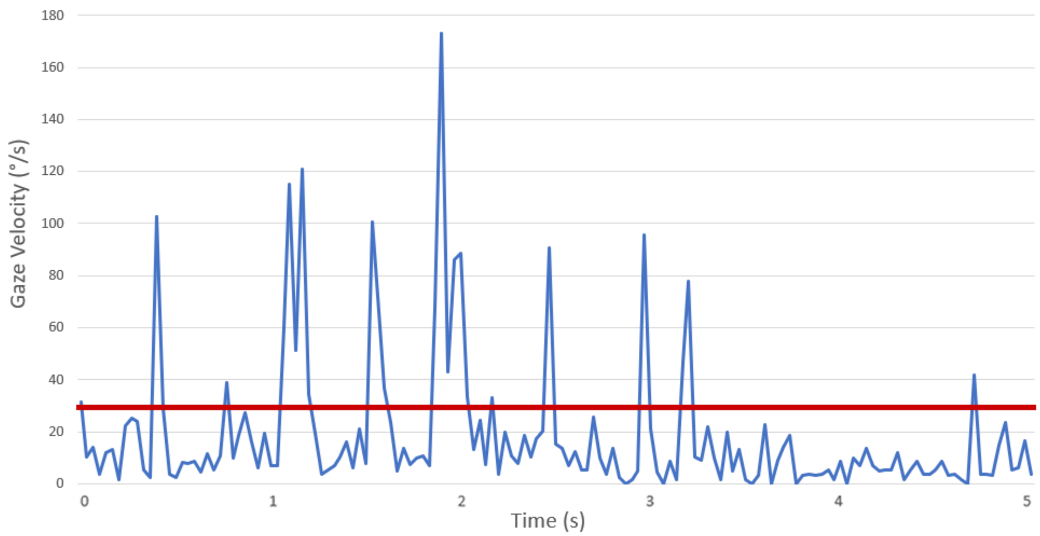

Figure 1: Example of the I-VT algorithm for identifying fixations using a threshold of 30 °/s, where samples above threshold correspond to saccades and samples below threshold correspond to fixations

samples that have been marked as saccades are collapsed into a single saccade with the combined length and the periods between these saccades are identified as fixations. Previous studies have used velocity thresholds between 20 and 30 °/s (Sen & Megaw, 1984; Olsen, 2012). Olsen (2012) suggests that in noisy data, a higher threshold may be preferable to a low threshold, as many values might be wrongly classified as saccades. Therefore, the threshold that was chosen for this research was 30 °/s. In Figure 1, an example can be seen of the I-VT algorithm on a 5 seconds section from a video of mementos, where the red line corresponds to the velocity threshold.

Fixations can have various different lengths and there is no universal definition for the average fixation and saccade duration. Most studies, however, agree that fixations typically have a duration between 150 to 300 msec, with some fixations lasting up to 1000 msec (Galley et al., 2015; Tullis & Albert, 2013). Some videos in Mementos for which the fixations and saccades had been estimated had exceptionally long average fixation durations such as 5 seconds. Because of these outliers, videos with an average fixation duration exceeding 1 second were removed from the data set. These values often indicated that OpenFace could not accurately extract the gaze data, likely due to factors such as video quality or lighting conditions. A total of 60 videos have been removed in this matter, with most of the removed videos belonging to 8 individuals. This resulted in a total of 435 videos and 51 instances of mind wandering.

Lastly, as it would be difficult to detect small periods of mind wandering in a 60 seconds video, with the shortest period of mind wandering being 1 second, the data from each video was split into segments of 5 seconds. This duration is somewhat arbitrary, but is consistent with similar previous studies (Krasich et al., 2018; Smilek et al., 2010; Uzzaman & Joordens, 2011). For each segment, descriptive statistics were calculated for both the fixations and saccades, resulting in a total of 11 features: *fixation count, total fixation duration, maximum fixation duration, minimum fixation duration, mean fixation duration, median fixation duration, total saccade duration, maximum saccade duration, minimum saccade duration, mean saccade duration,* and *median saccade duration.* Segments were labeled positive instances of mind wandering if it contained at least 1 second of mind wandering, resulting in 5020 negative instances of mind wandering and 73 positive instances.

2.3 Machine Learning Algorithms

Once all the features had been extracted and the data set containing the features was completed, the machine learning algorithm had to be created. The algorithm that was chosen for this is the Support Vector Machine (SVM). The SVM was chosen for the algorithm as they have been shown to be robust, memory efficient and can be used with imbalanced data. Additionally SVM's have been shown to be successful in similar research (Hutt et al., 2019; Stewart et al., 2017). The Support Vector Classifier (SVC) implementation from the Scikit-learn library was used for this research (Pedregosa et al., 2011).

Since the data that are used for the algorithm are highly imbalanced, with approximately 99% of the samples belonging to one class, measures were taken to prevent the algorithm from classifying everything to the majority class. Two different methods were used to solve this problem. The first method made use of the Syntetic Minority Oversampling Technique (SMOTE), which is an oversampling technique that generates synthetic samples for the minority class (Chawla et al., 2002). This allows the SVM algorithm to train on balanced data. The second method uses the class-weighted SVM, which works by assigning higher misclassification penalties to instances of the minority class.

3 Experimental Setup and Results

Algorithm Set-Up

Support Vector Machine algorithms are not scale invariant, and it is therefore recommended to scale the data before training the algorithm. The data for this research was first standardized to have a mean of 0 and a variance of 1 using the standard scaler from Scikit-learn. Next, the features were scaled to a range between 0 and 1 using the min-max scaler from the Scikit-learn library.

Using the SMOTE method, the scaled data was first split into a training set and a test set with 70% of the data being in the training set, and 30% in the test set, making sure that individuals who appear in the training set do not appear in the test set. Following this, the Imbalanced-learn library implementation of SMOTE was applied to the data belonging to the training set, resulting in a balanced training set (Lemaître et al., 2017). Next, a technique known as grid-search was performed on the training data. The Scikit-learn grid-search algorithm uses 5-fold stratified cross validation to exhaustively search over specified parameter values to find the SVM with the best performance. Table 2 shows the parameter values that were used during grid-search. The estimator that was found by grid-search was used to classify the samples in the test set, after which several different performance metrics were calculated. This process was repeated 10 times to calculate the mean performance metrics.

With the class-weighted SVM Method, the data was also split into a training set and a test set, again with a distribution of 70% and 30% and without overlap of individuals.

Parameter	Values
Kernel	linear, rbf, poly
C	0.1, 0.5 ,1, 5, 10, 100
Gamma	10, 1, 0.1, 0.001, 0.0001

Table 2: Parameter values used during grid-search to find optimal estimator

In contrast to the SMOTE method, no oversampling or undersampling was done and the training set was directly used in the grid-search algorithm. The grid-search algorithm was done in the same way as was done during the SMOTE method, using the same parameter values. This time, however, the class-weighted SVM algorithm was used instead of the normal SVM algorithm, automatically adjusting weights inversely proportional to the class frequencies. The estimator that was found was then used to classify the samples in the test set to calculate the performance metrics. This process was repeated 10 times as well to calculate the mean performance metrics.

Lastly, a third chance-based classifier was used as a baseline to which the other classifiers can be compared. From the SVM using SMOTE and the class-weighted SVM, the average percentage of samples classified as mind wandering were calculated. It was found that on average 40% of the samples in the test set were classified as mind wandering by the two classifiers. The chance-based classifier works by randomly classifying 40% of the samples in the test set as mind wandering. This was also repeated 10 times to find average performance metrics. The performance metrics that are used are the AOC-ROC, F1 score, recall, and precision. These metrics were chosen as they give a good representation of the performance in data sets with imbalanced data, unlike metrics such as the accuracy. The F1 score of the chance-based model can be compared to the trained models to see whether they perform better than chance. Additionally, the AOC-ROC will also give an indication whether the performance is better than chance, since an AOC-ROC score of 0.5 represents a model performing at chance level.

Feature Values

Feature	MW Mean (SD)	MW Median	NMW Mean (SD)	NMW Median
Fixation count	20.6 (8.5)	20	16.7 (9.1)	15
Total fixation dur	4.0 (0.7) s	4.1 s	4.2 (0.7) s	4.4 s
Max fixation dur	1.0 (0.7) s	0.8 s	1.3 (0.9) s	1.1 s
Min fixation dur	47 (70) ms	33 ms	114 (500) ms	33 ms
Mean fixation dur	0.27 (0.24) s	0.21 s	0.43 (0.58) s	0.28 s
Median fixation dur	0.16 (0.22) s	0.10 s	0.30 (0.57) s	0.13 s
Total saccade dur	1.0 (0.65) s	0.90 s	0.84 (0.71) s	0.63 s
Max saccade dur	130 (95) ms	100 ms	133 (187) ms	67 ms
Min saccade dur	33 (0.11) ms	33 ms	35 (80) ms	33 ms
Mean saccade dur	50 (20) ms	43 ms	53 (100) ms	41 ms
Median saccade dur	38 (15) ms	34 ms	41 (98) ms	34 ms

Table 3: Comparison of the mean, standard deviation and median values for all features, for both mind wandering (MW) and not mind wandering (NMW). Calculated using the segments discussed in section 2.2.

In Table 3, the mean, standard deviation and median values are shown for all features that were calculated for the segments. It is shown for both during mind wandering (MW) and outside of mind wandering (NMW). From this table it can be seen that there is a significant difference for some features between MW and NMW and that the largest differences are in features using fixations, while features using saccades are fairly similar for MW and NMW. Furthermore, many features have a large standard deviation and a significant difference between the mean and median value, which indicate that there are outliers as well as a large variance in the data.

More observations can be made about the mean fixation duration feature. One thing that can be seen is that the mean fixation duration during mind wandering is much lower than outside of mind wandering. This contradicts previous research which have shown that the mean fixation duration increases during mind wandering (Frank et al., 2015; Foulsham et al., 2013; Reichle et al., 2010). Additionally, the mean fixation duration outside of mind wandering is much larger than what previous research has shown, as the average fixation duration is typically between 150 to 300 msec (Galley et al., 2015; Tullis & Albert, 2013). All these factors may indicate that the way in which the eye movement was extracted, as well as the way in which the fixations were estimated, were not suitable for the Mementos data set, resulting in inaccurate results.

Classification

Method	AOC-ROC (SD)	F1 (SD)	Recall (SD)	Precision (SD)
SMOTE	0.63 (0.035)	0.046 (0.0078)	0.65 (0.10)	0.024 (0.0041)
Class-Weighted	0.59 (0.048)	0.041 (0.010)	0.57 (0.11)	0.021 (0.0053)
Chance	0.51 (0.049)	0.029 (0.0090)	0.42 (0.097)	0.015 (0.0047)

Table 4: Comparison of the average performance metrics for the SVM's classifying mind wandering using either SMOTE to account for the imbalanced data or using class-weighted SVM. The average performance metrics of a chance-based classifier is also shown.

Table 4 shows the average performance metrics of the classifier using SMOTE and the class-weighted classifier. From the table can be seen that there is not a large difference between these two classifiers, although SMOTE slightly outperforms the other in all performance metrics. Both methods resulted in a very low F1 score, caused by the low precision score of the classifiers. The chance-based classifier is also shown in Table 4 and when the two SVM classifiers are compared to the chance-based classifier, it can be seen that both classifiers significantly outperform the chance-based classifier. The chance-based classifier has an AOC-ROC of 0.51 which is close to the expected AOC-ROC value of 0.5 for a chance classifier, while the other two classifiers have an AOC-ROC that is significantly larger than that. The F1 score of the two SVM classifiers are also significantly larger than that of the chance-based classifier, due to having both higher recall and precision values. From these results, it can be concluded that both the SVM using SMOTE, as well as the class-weighted SVM perform better than chance.

Figure 2 and 3 also give additional metrics in the form of confusion matrices and ROC-Curves for the two SVM classifiers. The figures were taken from the best performing classifiers out of the 10 times grid-search was performed. From the matrices it can be seen that that there are many false positives, samples that have been classified as mind wandering while actually having the label non mind wandering. This is also the cause of the low F1 scores for the classifiers, since the many false positives lead to a low precision value, which subsequently leads to a low F1 score.

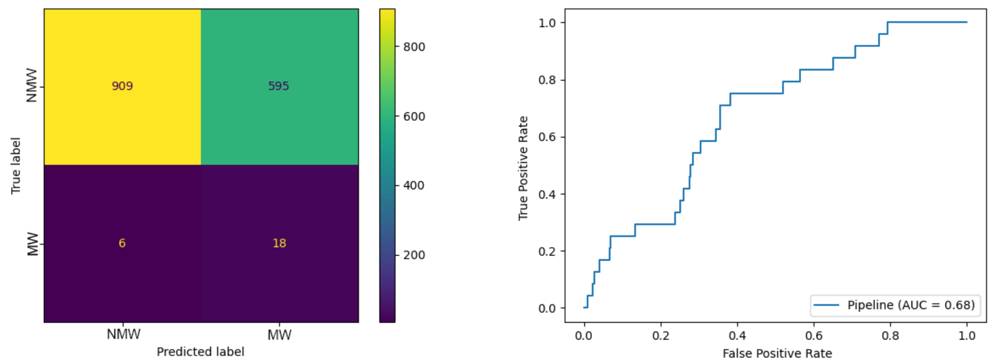

Figure 2: Confusion matrix and ROC-Curve for the SVM using the SMOTE method. The best performing case out of the 10 times the grid-search process was repeated is shown.

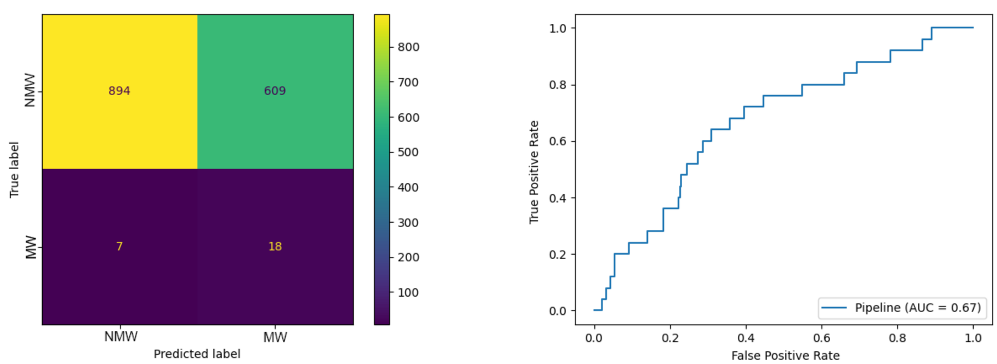

Figure 3: Confusion matrix and ROC-Curve for the SVM using the class-weighted method. The best performing case out of the 10 times the grid-search process was repeated is shown.

4 Responsible Research

This research has significant ethical implications, most importantly the Mementos data set. The Mementos data set contains videos of many different individuals captured in their personal environment and great care must be taken to make sure that the privacy of these individual remains safe. In order to guarantee the individuals' privacy, an End User License Agreement (EULA) was signed in the beginning of the research by all members involved in the study. The EULA stated several agreements and limitations which had to be agreed upon before access to the Mementos data set was granted. This made sure that the data could not be shared to third parties and that the data was only used for academic purposes, among other conditions. For example, no images of the individuals in the data set could be shown in the paper and it was made sure that all software that was used, was done locally and that the data was not shared to third parties. For example, the VGG Image Annotator and the OpenFace software that were used during this research were done completely offline. Besides the Mementos data set, there are no further elements with a significant ethical implication.

As for the reproducibility of this research, most parts of this research are completely reproducible as almost all steps that have been done, have been clearly explained and should lead to the same results as have been shown. The ways in which the gaze was extracted and how the fixations were calculated are all clearly explained and should produce identical results. Furthermore, the ways in which the features were created and how the machine learning algorithms were used should also be fully reproducible. The largest factor that influences the reproducibility in this research is the annotating of the Mementos data set itself. Even though indicators for mind wandering were created that were kept in mind while annotating the data set, actually finding instances of mind wandering is still highly subjective and there were many instances where it was not clear whether the person was mind wandering. If the annotating were to be reproduced, this would likely result in a significantly different data set and could consequently produce very different results.

5 Discussion

Although both classifiers perform better than chance, with the classifiers having an F1 score improvement of 59% and 41% over the chance-based model, the results that were found in this research are still worse compared to what previous studies have found. This was somewhat expected as the focus during this research was on detecting mind wandering in uncontrolled environments. The research of Stewart et al. (2017) for example resulted in an F1 score of 0.39, and the research of Hutt et al. (2019) found an F1 score of 0.59. These studies had, however, many more positive samples of mind wandering in their data. Both studies had between 20% and 25% of the data consisting of positive mind wandering samples, while this is only 1.5% in this research.

There could be several reasons as to why the results of this research differs from previous research. The largest influence on these results are most likely due to the gaze extraction and the annotation, as most other techniques that have been used during this research have already successfully been used in previous research. Using SMOTE together with SVM for example, has shown good performances in previous research, which have also used similar features. Consequently, it is likely that the difference in results is partly because of the extraction of eye movement. Due to the various different camera qualities, lighting conditions and positions, OpenFace was likely not able to accurately perform gaze estimation, resulting in noisy gaze data for many of the videos in the data set.

Inaccurate gaze data would also lead to inaccurate fixations and saccades estimates. Due to the inaccurate gaze data, choosing a single threshold for the I-VT algorithm to estimate fixations can result in inaccurate fixations and saccades estimates. For example, low noise data generally requires a lower threshold for detecting saccades, while videos with high noise require a higher threshold (Olsen, 2012). As there was a lot of noisy data extracted from the videos in Mementos, a higher threshold was chosen for this research. However, doing this might have resulted in some videos with low noise having inaccurate fixations and saccades estimates. Moreover, the data in Table 3 shows that the features that were calculated from the fixations and saccades have a large variance and that there is a large difference between the mean and median in some cases, even though the largest outliers have been removed. Furthermore, the data shown does not correlate with existing research, which shows that the fixation duration increases during mind wandering, whereas the results of this research show the opposite (Frank et al., 2015; Foulsham et al., 2013; Reichle et al., 2010).

Lastly, the way in which mind wandering was detected is also an important factor. There were likely many more instances of mind wandering that could not be detected just from watching videos, such as instances were the person does not make any different expressions while mind wandering. Other research in this field have used probe-caught and self-caught instances of mind wandering, where either the participant is asked whether they are mind wandering at that moment, or the participant reports when they are mind wandering themselves. This would likely lead to many more instances of mind wandering, which could have resulted in a better algorithm than is currently presented.

6 Conclusions and Future Work

The main purpose of this study was to research whether it is possible to create a machine learning algorithm that could automatically detect mind wandering based on gaze data from the Mementos data set and perform better than chance. Previous research has already shown that this is possible for other data sets, but many of these studies used data with videos captured in controlled environments, while the focus of this research was to see whether it was possible to detect mind wandering in an uncontrolled environment using inexpensive equipment.

In order to answer this question, videos in the mementos data set were annotated by a group to mark sections where mind wandering was thought to have occurred. For the annotated videos, several descriptive statistics were calculated for the fixations and saccades that were used to represent the eye movement. Next, two classifiers were trained, one using SMOTE to oversample the training data of the minority class which was subsequently used on a SVM classifier. The other classifier did not oversample or undersample the training data, but used a class-weighted SVM to account for the class imbalance. The results have shown that the two classifiers that were used perform better than chance, having an average AOC-ROC of 0.63 and 0.59. Furthermore, average F1 scores of 0.046 and 0.041 were found for the two classifiers, with the chance-based classifier having an average F1 score of 0.029. Although the classifiers perform better than chance, it is still not a reliable way to to detect mind wandering in the Mementos data set.

Since one important limitation of this research is likely the accuracy of the gaze estimation, future research could focus on making the extracted gaze data more accurate. OpenFace was used to perform gaze estimation during this research, but there are many more tools which are also designed to do this. Some other tool might be able to more accurately extract eye movement from the videos in Mementos, which could lead to improved features and a better classifier. Another method would be to only use videos which are of a certain quality, making it more probable that the gaze estimation is done accurately.

Further research could also be done to research the performance of classifiers when other features that are not related to fixations and saccades are used. For example, one of the indicators that was used to annotate mind wandering in Mementos was that the individual would look upwards or downwards during mind wandering. These kind of eye movement are not taken into account when only using fixations and saccades and it might therefore be better to use some other features which do take this into account, although it will likely be difficult due to the uncontrolled environments. Lastly, as this research is focused on perceived mind wandering, it might be beneficial to have the videos in mementos be annotated by a larger group. This could lead to more instances of mind wandering being found, as well as being more certain that the individual is actually mind wandering, as this can be quite subjective.

References

Baird, B., Smallwood, J., & Schooler, J. W. (2011, December). Back to the future: Autobiographical planning and the functionality of mind-wandering. *Consciousness and Cognition*, *20*(4), 1604–1611. Retrieved from `https://doi.org/10.1016/j.concog.2011.08.007`

Baltrusaitis, T., Zadeh, A., Lim, Y. C., & Morency, L.-P. (2018, May). OpenFace 2.0: Facial behavior analysis toolkit. In *2018 13th IEEE international conference on automatic face &amp gesture recognition (FG 2018).* IEEE. Retrieved from `https://doi.org/10.1109/fg.2018.00019`

Bixler, R., & D'Mello, S. (2015, September). Automatic gaze-based user-independent detection of mind wandering during computerized reading. *User Modeling and User-Adapted Interaction*, *26*(1), 33–68. Retrieved from `https://doi.org/10.1007/s11257-015-9167-1`

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002, June). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, *16*, 321–357. Retrieved from `https://doi.org/10.1613/jair.953`

Dudzik, B., Hung, H., Neerincx, M. A., & Broekens, J. (2021). Collecting mementos: A multimodal dataset for context-sensitive modeling of affect and memory processing in responses to videos. *IEEE Transactions on Affective Computing*, 1–1. Retrieved from `https://doi.org/10.1109/taffc.2021.3089584`

Farley, J., Risko, E. F., & Kingstone, A. (2013). Everyday attention and lecture retention: the effects of time, fidgeting, and mind wandering. *Frontiers in Psychology*, *4*. Retrieved from `https://doi.org/10.3389/fpsyg.2013.00619`

Foulsham, T., Farley, J., & Kingstone, A. (2013, March). Mind wandering in sentence reading: Decoupling the link between mind and eye. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, *67*(1), 51–59. Retrieved from `https://doi.org/10.1037/a0030217`

Frank, D. J., Nara, B., Zavagnin, M., Touron, D. R., & Kane, M. J. (2015, June). Validating older adults' reports of less mind-wandering: An examination of eye movements and dispositional influences. *Psychology and Aging*, *30*(2), 266–278. Retrieved from `https://doi.org/10.1037/pag0000031`

Galley, N., Betz, D., & Biniossek, C. (2015, 01). Fixation durations - why are they so highly variable? In (p. 83-106).

Hutt, S., Krasich, K., Mills, C., Bosch, N., White, S., Brockmole, J. R., & D'Mello, S. K. (2019, June). Automated gaze-based mind wandering detection during computerized learning in classrooms. *User Modeling and User-Adapted Interaction*, *29*(4), 821–867. Retrieved from `https://doi.org/10.1007/s11257-019-09228-5`

Killingsworth, M. A., & Gilbert, D. T. (2010, November). A wandering mind is an unhappy mind. *Science*, *330*(6006), 932–932. Retrieved from `https://doi.org/10.1126/science.1192439`

Krasich, K., McManus, R., Hutt, S., Faber, M., D'Mello, S. K., & Brockmole, J. R. (2018, August). Gaze-based signatures of mind wandering during real-world scene processing. *Journal of Experimental Psychology: General*, *147*(8), 1111–1124. Retrieved from `https://doi.org/10.1037/xge0000411`

Lemaître, G., Nogueira, F., & Aridas, C. K. (2017). Imbalanced-learn: A python toolbox to tackle the curse of imbalanced datasets in machine learning. *Journal of Machine Learning Research*, *18*(17), 1-5. Retrieved from http://jmlr.org/papers/v18/16-365.html

Olsen, A. (2012). The tobii i-vt fixation filter algorithm description. In *Proceedings of the symposium on eye tracking research.* Tobii Technology.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Reichle, E. D., Reineberg, A. E., & Schooler, J. W. (2010, August). Eye movements during mindless reading. *Psychological Science*, *21*(9), 1300–1310. Retrieved from https://doi.org/10.1177/0956797610378686

Salvucci, D. D., & Goldberg, J. H. (2000). Identifying fixations and saccades in eye-tracking protocols. In *Proceedings of the symposium on eye tracking research &amp applications - ETRA '00.* ACM Press. Retrieved from https://doi.org/10.1145/355017.355028

Sen, T., & Megaw, T. (1984). The effects of task variables and prolonged performance on saccadic eye movement parameters. In *Theoretical and applied aspects of eye movement research, selected/edited proceedings of the second european conference on eye movements* (pp. 103–111). Elsevier. Retrieved from https://doi.org/10.1016/s0166-4115(08)61824-5

Smallwood, J., & Schooler, J. W. (2006, November). The restless mind. *Psychological Bulletin*, *132*(6), 946–958. Retrieved from https://doi.org/10.1037/0033-2909.132.6.946

Smilek, D., Carriere, J. S., & Cheyne, J. A. (2010). Out of mind, out of sight: Eye blinking as indicator and embodiment of mind wandering. *Psychological Science*, *21*(6), 786–789. Retrieved 2022-06-18, from http://www.jstor.org/stable/41062288

Stewart, A., Bosch, N., Chen, H., Donnelly, P., & D'Mello, S. (2017). Face forward: Detecting mind wandering from video during narrative film comprehension. In *Lecture notes in computer science* (pp. 359–370). Springer International Publishing. Retrieved from https://doi.org/10.1007/978-3-319-61425-0_30

Tullis, T., & Albert, B. (2013). Behavioral and physiological metrics. In *Measuring the user experience* (pp. 163–186). Elsevier. Retrieved from https://doi.org/10.1016/b978-0-12-415781-1.00007-8

Uzzaman, S., & Joordens, S. (2011, December). The eyes know what you are thinking: Eye movements as an objective measure of mind wandering. *Consciousness and Cognition*, *20*(4), 1882–1886. Retrieved from https://doi.org/10.1016/j.concog.2011.09.010

Zhang, H., Miller, K. F., Sun, X., & Cortina, K. S. (2020, January). Wandering eyes: Eye movements during mind wandering in video lectures. *Applied Cognitive Psychology*, *34*(2), 449–464. Retrieved from https://doi.org/10.1002/acp.3632