Eye tracking-based Reading Activity Recognition with Conventional Machine Learning Algorithms

JULIAN MEIJERINK*
Supervisors: Guohao Lan, Lingyu Du†
EEMCS, Delft University of Technology, The Netherlands

June 21, 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

*J.A.M.Meijerink@student.tudelft.nl
†G.Lan@tudelft.nl, Lingyu.Du@tudelft.nl

Abstract

The use of eye-tracking as a tool to provide cognitive context is rising in real-world systems. Though extensive research has been done on using machine learning and deep learning to classify sedentary activities using data captured by eye-trackers, there is a gap in analyzing the impact of the usage of different sedentary activities and feature extraction methods on the performance. In this paper, conventional machine learning algorithms are used to classify reading activities, captured by eye tracking devices. Multiple data pre processing methods and filters are used to extract fixations out of the raw data captured by the eyetrackers. Out of these fixations, a total of 16 features are used to classify activities. Using the optimal configurations found, a 0.99 user dependent and a 0.76 user independent score is obtained. All obtained results are compared to results obtained by peers performing similar research, who use either a different data set or deep learning instead of conventional machine learning. The papers using deep learning had vastly strong performance for all user dependent evaluations, but performed poorer in the user independent evaluation. Overall, conventional machine learning performed better on user independent evaluation, where this paper obtained the best results for user independent evaluation, most likely due to the fact that Japanese reading material was used, which has very distinctive reading directions.

1 Introduction

The use of computers and applications are ever rising, and alongside this growth, more creative and intricate tools are being used to make applications more usable and diverse. The use of human visual behavior is one such relevant emergence in providing cognitive context. Research has shown that human visual behavior is strongly related to taking in information [1], attention [2], human emotion [3] and memory access [4].

Eye trackers use the merit human visual behavior provide by tracking eye movements. These measured signals can be processed and analyzed, which can provide the aforementioned cognitive context. Such eye trackers are already finding real-world use, as various eye-tracking applications have been found useful in improving human performance, safety and productivity. For instance, eye tracking can allow us to see into workers' visual attention and identify why accidents in the workplace occur [5]. In the gaming industry, eye tracking enables game developers to get better customer insights into the gaming experience [6]. Furthermore, the usage of eye tracking in cars have increased our knowledge about what drivers' behavior and attention, more specifically how they distribute their glances when interacting with non-driving related tasks, which can be used to improve safety [7].

Though extensive research has been done on classifying sedentary activities using both conventional machine learning, as well as deep learning, there is a gap in analyzing the impact of using different data sets and feature extraction methods. A complete analysis of related work is done in section 2. To fill this gap in research, this paper focuses on classifying reading activities using conventional machine learning. Following, the obtained results will be compared to the results obtained by peers, who either use a different data set, or use deep learning instead of conventional machine learning. To do this, the following research questions have been formulated:

1. To achieve good recognition accuracy, what are the best features that need to be extracted and used for training conventional machine learning algorithms, e.g., k-Nearest Neighbors (K-NN), Suport Vector Machine (SVM), and Random Forest Tree?

2. What is the impact of different subjects on the recognition performance?

3. How do the three above algorithms compare on different datasets in terms of accuracy, memory usage, inference latency, and robustness against heterogeneity among subjects?

This paper begins by discussing related work, and how it impacts this paper's research in section 2. In section 3 the methodology is explained, which includes all algorithms and methods used to obtain data needed for the results. Subsequently, in section 4, a look is taken at what evaluation methods were used to obtain the results, all obtained results for all different configurations and algorithms are given, and the results are compared to those obtained by peers. Next, in section 5 all ethical aspects of this research are reflected on. In section 6, we discuss the possibilities for future work, as well as limitations encountered during this research. Finally, in section 7 we conclude this paper by summarizing the goal, the findings and provide the answers to the aforementioned research questions.

2 Related Work

Extensive research has been done on classifying different activities using deep learning and other feature extraction methods [8-13]. For instance, Bulling et al. has done research in classifying different sedentary activities using an SVM in both a stationary [9] and mobile setting [10]. Furthermore, Kunze et al. investigated classifying different reading activities using machine learning, obtaining a 99% and 74% accuracy score for user dependent and user independent classification respectively [11]. Kunze et al. also explored the usage of Electroencephalography to track eye movements, and recognize the activities and habits of users while they are reading, obtaining a 97% accuracy score for user-dependent classification [12]. Lan et al. presented GazeGraph, a system using human gazes as the sensing modality for cognitive context sensing and activity recognition, in which their system outperforms existing solutions by 45% on average over three different data sets, when a large training data set was available [8].

Though classifying different sedentary activities using different feature extraction methods has been researched extensively, there is a gap in comparing performance of classifying these different activities using different feature extraction methods. This research focuses on classifying reading activities specifically, using conventional machine learning and aims to compare the obtained results to papers performing similar research on different data sets and using deep learning instead of conventional machine learning.

3 Methodology

In this section, the methodology followed in this research will be explained. This includes an explanation of the data used, data preprocessing methods, filtering methods, feature extraction methods and evaluation procedures.

3.1 Data

In this project, a public data set [11] was used. It contains recordings from eight Japanese subjects, of which exactly half were male and exactly half were female, aged between 21 and 32. Though in the beginning a slightly different data set with an additional 9th subject was used, the data of this 9th subject was later removed, which makes the used data set identical to the one used in [11]. The eye movements of the subjects performing different reading activities was captured using over the span of fifteen minutes, at 30Hz. This amounts to 27.000 raw data points, where each point has an x- and

y-coordinate. To capture these eye movements, the head mounted SMI Eye Tracking Glasses [15] were used. The subjects were all tasked with performing six different types of reading activities: reading a novel, a manga, a fashion magazine, a newspaper, a scientific paper and a textbook, all in Japanese. Notably, the Japanese novels were in a distinct top to bottom writing style, rather than a left to right writing style.

3.2 Fixations and Saccades

When a subject is performing a task for which visual attention is required, they will need to take in information. Research has shown that when eyes are taking in information, their movements consists of two main actions: fixations and saccades. When we read, look at a scene, or search for an object, we continually make saccades, while between the saccades, our eyes remain relatively still during fixations for about 200-300 ms [16]. These fixations and saccades form certain patterns, out which the performed activity can be predicted. For example, when reading, most saccades in reading English are made from left to right. However, readers do not relentlessly go forward: About 10-15% of the saccades are regressions (right-to-left movements along the line or movements back to previously read lines) [16]. Using machine learning, we can predict reading activities by analyzing different gaze patterns for each performed activity.

3.3 Data Preprocessing

During the eye tracking process, signals are prone to noise. There are two reasons this noise occurs: noise inherent in the measurement device, and small motions of the eye during a fixation [17]. These small motions are called microsaccades, and extensive research has been done on this particular topic [18]. The noise and microsaccades make it more difficult to distinguish saccades from fixations, and the noisy data would make predicting less accurate. To deal with the noise in the data, multiple processing methods have been used to remove noise in the signal.

The first step is to normalize all the data, that is, to map all values to have values between a minimum of 0 and maximum of 1. Though mathematically this should not change the outcomes of any calculations, it is easier to work with and evaluate results that are between this range.

The next step is to remove outliers. During the eye tracking process, subjects may look away from the reading activity, or the eye tracker may give completely faulty measurements, in which case the outlying data points are not a part of the reading eye movements. All data points with a distance over a certain threshold away from the mean of all points are removed.

3.3.1 Peak Filter

The peak filter is based on a method used in the master thesis of P. Olsson [14]. The peak filter is the start of distinguishing the fixations from the saccades in the data set. A peak is a pair of data points where the distance between those two points is larger than the distance between the prior and subsequent pair of data points. Figure 1 shows a few data points in which peaks are highlighted. As one can imagine, the idea is to separate the fixations by the longer distances.

After obtaining all the peaks, we also filter out all peaks with a distance less than a certain threshold. This is to remove peaks caused by noise of the Eye-tracker and peaks not caused by a saccade.

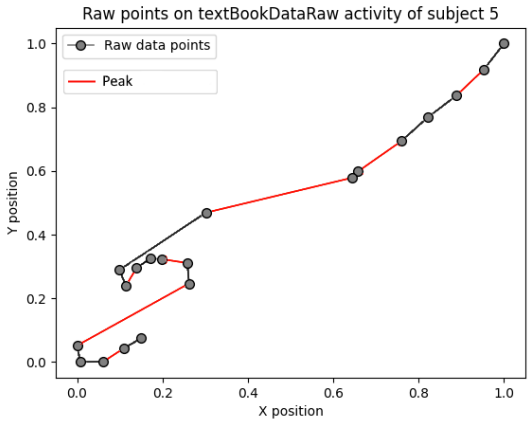

Figure 1: Gaze estimation sample segment of subject 5 reading a textbook. Here the red lines represent peaks, as their length is longer than the prior and subsequent pair.

3.3.2 Median Filter

The next step is crucial to separating the fixations from one another. Median filtering is a nonlinear signal processing technique that is useful for noise suppression in images. In this case, it is used to remove noise in eye tracking data, rather than an image. The idea behind median filtering is that all points are traversed and replaced by the median of some amount of points before and after the current point considered. As the median is constantly taken of a small set of points, outliers are filtered out. The set of points considered in each traversal is called the window, and the amount of points in this window is the window size. The window size has a large impact on how much filtering is done, and must be carefully chosen. A window size too small will result not filter out enough noise, whereas a window size too large would filter out too many points and thus valuable information. In this case, we would like to have a point for each fixation. A window size much smaller than the average number of points a fixation contains would erroneously filter noise within a fixation. A window size spanning over lots of fixations would also wrongly make one fixation out of multiple. Thus, a window size spanning a little over one fixation on average is chosen.

3.4 Feature Extraction

The features are the things we use to classify a set of points into a certain activity. The features used are thus essential to the performance and accuracy of this classification. The features come in two different forms: fixation-based and saccade-based features. These are mainly based on the paper *Combining Low and Mid-Level Gaze Features for Desktop Activity Recognition* [13]. A general overview of the features is given in table 1. Below, they are described in detail.

3.4.1 Fixation-based Features

When the eyes are focusing on a specific point, they are fixating, and this is called a fixation. Research has shown that during fixations, new information is extracted from the text [22]. The duration, frequency and location of the fixations can give us information on the activity performed. On average, the duration of a fixation is 200-250 ms, however, these durations have a high variance [16].

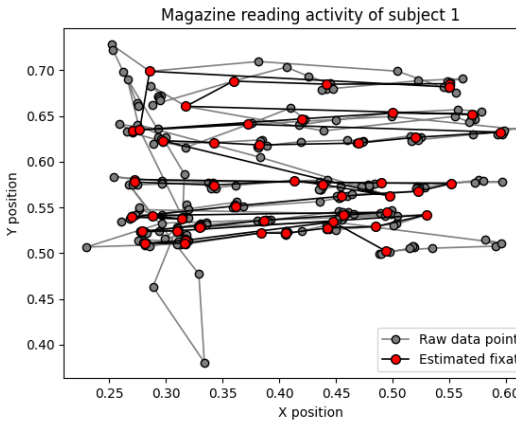

Figure 2: A small sample of subject 1 reading a magazine is shown. Clutters of raw data are grouped into a fixation.

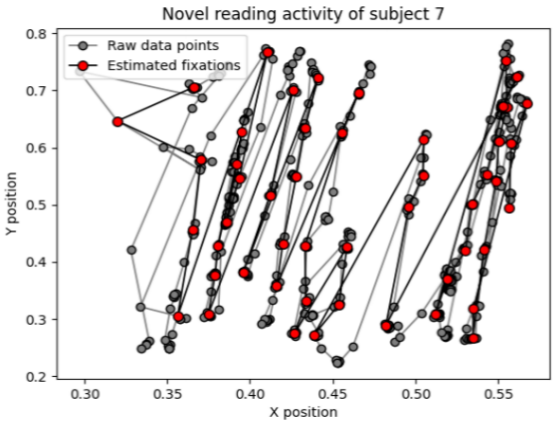

Figure 3: A small sample of subject 7 reading a novel is shown. The top to bottom reading style is apparent.

The total range of a fixation duration spans from 100 ms to over 500 ms, even for fairly simple text [22]. Since different reading activities contain different amounts of information and differ in intensity, the fixation duration can inform us about the performed activity. The mean fixation duration as well as the variance of fixation durations are features. The fixation count represents the amount of fixations estimated in the considered window. Finally, the dispersion area tells us something about the spread of the fixations. It is the average distance from every point to the mean of all points.

3.4.2 Saccade-based Features

Alongside the fixation-based features, 11 saccade based-features are constructed. The 8 direction-based features segregate all saccades into one of the eight directions, where every saccade is mapped to a direction when it falls in the 45 degree window of that direction. Another direction-based feature is the average angle of every saccade in the window. Finally, the mean saccade length and the variance of the saccade length complete the 16 features in total.

Category	Sub-Category	Features
Fixation based	duration	duration-mean, duration-variance
	dispersion area	dispersion-mean, dispersion-variance
	count	fixation-count
Saccade based	saccade direction	right,right-up, up, left-up, left, right-down, down, left-down, average angle
	length	saccade-length, saccade-variance

Table 1: An overview of features extracted from the data. In total 16 features were extracted, of which 5 were fixation-based and 11 were saccade-based.

3.5 Evaluation

To classify, the k-nearest neighbour, random forest tree and Support vector machine classifiers were used. They are well known machine learning classifiers and open source. In the evaluating process, both k-fold cross validation, which is user dependent, as well as Leave-One-Out Cross-Validation (LOOCV), which is user independent was used. Compared to k-fold cross validation, LOOCV is approximately an unbiased estimate of the true prediction error, but it has a higher variance [19]. Since the new subject's bias and personal behavior are unaccounted for in LOOCV, it will not perform as well compared to using k-fold, but it will give a score more representative of one to be expected in the real-world.

4 Experimental Setup and Results

In this section, the setup and configurations of the algorithms used will be discussed. Following, the results obtained are analyzed. Finally, these the results obtained will be compared with the results obtained by peers doing similar research on different data and using different feature extraction methods.

4.1 Experimental Setup

The accuracy of classifying a set of features into the correct activity depends heavily on the configurations of algorithm used. All three algorithms were fine-tuned to optimize classification performance. The K-nearest neighbor classifier was optimized with parameters leaf size = 1 and k = 8. The Random forest classifier was optimized with max depth = 420 and the number of estimators = 60. The SVM classifier was tuned using C = 1 and gamma = 0.001. To assess the performance, the accuracy score, calculated as a ratio of correct over total predictions, was used.

4.2 Results

Below in table 2, all different tested window sizes with their overlap are also shown. In nearly every case, the windows with the highest overlap yielded the best scores, for both 10-fold CV and LOOCV. The more data was extracted, the better the algorithms were able to classify. The optimal window size was 150s for the 10-fold CV and 125s for LOOCV, however, results were very similar for both window sizes, and the overlap was a much more important factor for the accuracy score. In both the 10-fold cross validation and in the LOOCV, the K-NN had the median score. In 10-fold cross validation, the Random Forest Tree scores best. Research has shown that the Random Forest Tree classifier achieves better classification results when the data that is used multi-dimensional data such as hyper spectral or multi-source [20]. In our case, we are indeed dealing with multi-dimensional features, as the data used contained a total of 16 features. The SVM performed best in LOOCV, and worst in the 10-fold cross validation. Futhermore, in nearly every case, the windows with the highest overlap yielded the best scores, for both 10-fold CV and LOOCV. The more data was extracted, the better the algorithms were able to classify. The optimal window size was 150s for the 10-fold CV and 125s for LOOCV, however, results were very similar for both window sizes, and the overlap was a much more important factor for the accuracy score.

In general, the SVM classifier is claimed to have good generalization properties compared to conventional classifiers [23]. However, here the SVM fails to learn the complexity of the features extracted, and yields only mediocre scores.

Window size		SVM		K-NN		Random Forest	
Window size	Overlap	10-fold CV	LOOCV	10-fold CV	LOOCV	10-fold CV	LOOCV
100s	75%	0.83	0.71	0.89	0.70	0.80	0.75
	80%	0.87	0.72	0.92	0.68	0.82	0.76
	85%	0.89	0.71	0.95	0.69	0.83	0.75
	90%	0.92	0.70	0.98	0.69	0.83	0.75
125s	75%	0.82	0.71	0.90	0.67	0.83	**0.76**
	80%	0.85	0.71	0.94	0.69	0.84	0.75
	85%	0.90	0.69	0.97	0.69	0.86	0.75
	90%	0.93	0.70	0.98	0.69	0.87	0.74
150s	75%	0.87	0.71	0.92	0.70	0.85	0.74
	80%	0.89	0.71	0.94	0.70	0.85	0.75
	85%	0.89	0.69	0.96	0.68	0.85	0.75
	90%	0.93	0.70	**0.99**	0.69	0.88	0.74
175s	75%	0.85	0.71	0.91	0.69	0.84	0.75
	80%	0.85	0.71	0.95	0.69	0.85	0.74
	85%	0.90	0.72	0.96	0.69	0.86	0.74
	90%	0.93	0.71	0.98	0.70	0.87	0.74

Table 2: Overview of all obtained accuracy scores for all three algorithms for both the 10-fold cross validation and LOOCV. Best obtained results in the 10-fold cross was by the Random Forest Tree with a window size of 150s and a 90% overlap.

4.3 Peer Results Comparison

This research has been conducted alongside 4 peers performing similar research, either using a different data set of eye tracking data, or using deep learning instead of conventional machine learning to classify the activity. In 3 an overview with all the obtained scores by peers and this research is given. Both papers using deep learning obtained near perfect scores for the user independent (K-fold CV) classifying sessions.

	Paper 1		Paper 2			Paper 3		Paper 4		Paper 5	
Algorithm	CNN		LSTM			Conventional Machine Learning					
						Sedentary		Desktop		Reading	
	dep	ind	dep	ind	Classifier	dep	ind	dep	ind	dep	ind
Sedentary	0.99	0.69	0.98	0.67	K-NN	0.77	0.48	0.84	0.54	0.91	0.71
Desktop	0.99	0.39	0.95	0.32	SVM	0.86	0.52	0.95	0.60	0.85	0.75
Reading	0.99	0.67	0.98	0.31	Forest	0.94	0.65	0.82	0.58	0.99	0.67

Table 3: An overview of all scores obtained by peers performing similar research. The resreach done with deep learning looked at all three data sets, whereas research performed with conventional machine learning used only one data set. The user independent classifying is k-fold CV and the user dependent classifying is LOOCV.

This is not surprising, as deep learning is known to be powerful when dealing with complex data. More specifically, conventional machine-learning techniques are limited in their ability to process natural data in their raw form, whereas deep learning is very good at discovering intricate structures in high-dimensional data [24].

The other peers also performing research using conventional machine learning on a different data set obtained good results as well, with the paper *Eye Tracking-Based Desktop Activity Recognition with Conventional Machine Learning* using data of desktop activities obtaining a 0.94 score and the paper *Eye Tracking-Based Sedentary Activity Recognition with Conventional Machine Learning* using data of sedentary activities obtaining a 0.95 score.

More interesting are the user dependent (LOOCV) classification sessions. The deep learning research papers obtained less impressive scores, with decent percentages for the sedentary activity and the reading data by the paper using a Convolutional Neural Network (CNN), and worse results for the Long Short Term Memory (LSTM) algorithm. However, it seems overall the deep learning algorithms struggle significantly with classifying user dependent data. The conventional machine learning papers performed quite well on the user dependent data. With *Eye Tracking-Based Sedentary Activity Recognition with Conventional Machine Learning* obtaining the best score of 0.65, *Eye Tracking-Based Desktop Activity Recognition with Conventional Machine Learning* obtaining the best score of 0.60 and this paper obtaining the best score 0.75. It is to be expected for user dependent scores to be lower than user independent scores, as more data is unknown, and user bias and personal features are unaccounted for.

4.4 Feature Importances

When classifying using machine learning, it is important to choose features carefully for optimal accuracy. In figure 4, an overview of all features used with their importance to the classifying decision is given. The directionbased features make up most of the importances, with the right, up, left and down directions, as well as the average saccade angle being the most important direction-based features.

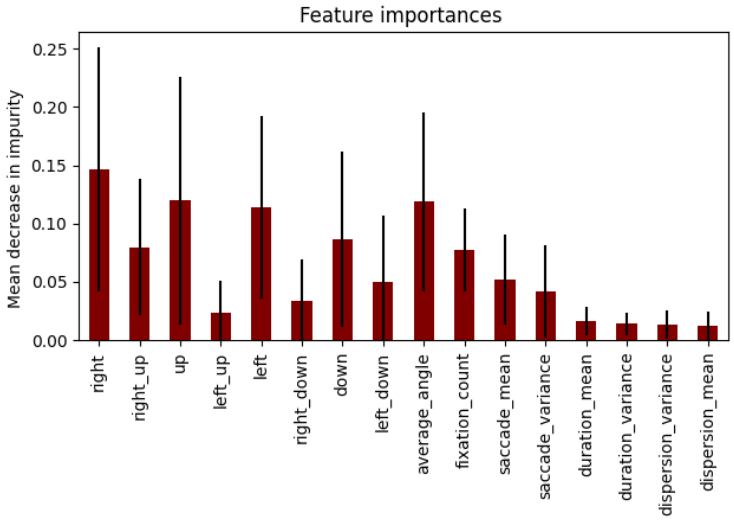

Figure 4: All features with their average importance when a classifying decision is made. The first 8 features are the direction-based features, followed by the rest of the features ordered on importance.

This is not surprising, considering the Japanese reading materials used differ strongly in direction, as is depicted in figures 2 and 3. The duration and dispersionbased features were the least

important. While durations have been shown to differ per activity [16], their variance is too high to extract useful information from during reading activities.

5 Responsible Research

This research has used real-world data from 8 participants. Moreover, the use of machine learning and Artificial Intelligence are involved in numerous ethical issues [21]. It is therefore important to consider the ethical aspects of this research.

5.1 Ethical Aspects

This research used recorded gaze data of eight Japanese participants. The data is publicly available, and originally recorded by [11]. During the recording process, only the gaze data was recorded and stored. Furthermore, the participants have remained unknown after the research, ensuring their privacy. This research solely used machine learning for analytical purposes, and no real life decisions have been made based on any machine learning algorithms. However, as mentioned in 1, eye tracking is being implemented in real world products more and more. It is therefore of the essence that when important decisions are made based on the classification or any other determination by the machine learning algorithm, it is done so with sufficient ethical consideration.

5.2 Reproducibility

To make sure the research done in this paper is reproducible, *The Machine Learning Reproducibility Checklist* [25] has been followed. An overview of the machine learning algorithms used is provided in section 3.5 and are open source libraries. Futhermore, all hyper parameters chosen are given in section 4.1. The dataset used was obtained by *I know what you are reading Recognition of Document Types Using Mobile Eye Tracking* [11] and is publicly available. In section 3.3, it is also described how one subject is left out and all data was preprocessed. All details on training and test splits have been explained in section 3.5. An overview of different window sizes and overlaps considered is given table 2, and methods used to assess performance are described in 4.1. Finally, all source code used is available upon request. All steps and measures have been taken to make this research reproducible.

6 Discussion and Future Work

In this section, the results will first be discussed and compared to results obtained by relevant research. Following, future work is considered, and limitations in this paper are discussed.

The research performed in this paper are similar to *I know what you are reading Recognition of Document Types Using Mobile Eye Tracking* [11]. It used the same data set and also used conventional machine learning to classify activities. In that paper, a score of 0.99 and 0.74 was obtained for the user independent and user dependent evaluation sessions respectively. Those are similar to the results obtained in this paper, with the user independent score obtained by *I know what you are reading Recognition of Document Types Using Mobile Eye Tracking* being slightly more accurate than obtained in this paper, whereas this research obtained a slight improvement on the user dependent data. An explanation for this could be that in this paper, all algorithms were configured to make the user dependent score optimal, from which the user independent scores followed. Moreover, window

size and overlap may have allowed the algorithms to not over fit and classify more general, and thus perform better when classifying user dependent data.

Though this research achieved promising results in classifying reading activities, one of the limitations was data used. As can be seen in figure 3 Japanese writing style is both vertical and horizontal, which is different from non East Asian languages. Future work could analyze reading activities in different, and perhaps multiple languages.

7 Conclusion

This research aimed to find out *How to design and implement different feature extraction methods for eye movement signal? To achieve good recognition accuracy, what are the best features that need to extracted and used for training conventional machine learning algorithms, e.g., kNN, SVM, and decision tree?* The research was carried out by using a K-NN, SVM and Random Forest Tree classifier to classify reading activities performed by 8 subjects. In this paper, as described in section 3, numerous methods have been used to pre process data, and using two filters group the processed data into fixations. In table 2, an overview of all different window sizes, overlap percentages and their accuracy per classifier is given. The window size of 150 seconds with an overlap of 90% was found optimal, and an accuracy score of 0.99 and 0.76 was obtained for the 10-fold cross validation and Leave One Out cross validation respectively. In section 4.1, the optimal filter thresholds and hyper parameters which best extracted features are given. In section 4.4, the used features and their importance have been analyzed, where it became clear the directionbased features were the most important features, and it was found that the duration and dispersionbased features did not improve accuracy, as their variance was too high.

The second research question formulated was *What is the impact of different subjects on the recognition performance?* In table 2, the difference in accuracy on user dependent and user independent evaluation is shown. The user dependent scores were higher, than those obtained for user independent scores, which was not surprising considering during the user independent evaluation, the classifiers could not account for user bias and personal traits.

Lastly, the paper sought to find out *How do the three above algorithms compare on different datasets in terms of accuracy and robustness against heterogeneity among subjects?* In table 3, an overview of the obtained scores by peers performing similar research on different data sets or using deep learning instead of conventional machine learning to classify is shown. The deep learning algorithms performed incredibly well on the user dependent evaluation, where the CNN obtained a 0.99 accuracy score, whereas the LSTM obtained the best score of 0.98. Regarding user independent evaluation however, all scores were not as strong. The papers using deep learning sometimes obtained some poorer results, with some accuracy scores being lower than 0.40. The papers using conventional machine learning obtained somewhat better results, with accuracy scores ranging from 0.48 to 0.75. This paper, classifying reading activities obtained the strongest user independent evaluation accuracy scores, most likely due to the fact that Japanese reading material was provided, which has very distinctive reading directions.

References

[1] Land, M., 2006. Eye movements and the control of actions in everyday life. Progress in Retinal and Eye Research, 25(3), pp.296-324. https://doi.org/10.1016/j.preteyeres.2006.01.002

[2] A. Bulling, D. Roggen, and G. Tröster, "What's in the Eyes for Context-Awareness?," IEEE Pervasive Computing, vol. 10, no. 2, pp. 48-57, Apr. 2011, doi: 10.1109/mprv.2010.49.

[3] Roman Bednarik, Hana Vrzakova, and Michal Hradis. 2012. What Do You Want to Do Next: A Novel Approach for Intent Prediction in Gaze-based Interaction. In Proceedings of the Symposium on Eye Tracking Research and Applications (ETRA '12). ACM, New York, NY, USA, 8390. https://doi.org/10.1145/2168556.2168569

[4] D. Melcher and E. Kowler, "Visual Scene Memory and the Guidance of Saccadic Eye Movements," Vision Research, vol. 41, nos. 25-26, 2001, pp. 3597-3611.

[5] How to improve human performance with eye tracking Tobii Pro. (2018). Tobiipro. https://www.tobiipro.com/applications/industry-human-performance/

[6] A. N. Singh, "Practical applications of Eye Tracking Technology," Medium, 13-May-2019. [Online]. Available: https://ankitnsingh.medium.com/practical-applications-of-eye-tracking-technology-d30bfe0c131e. [Accessed: 30-Apr-2022].

[7] Ahlström, C., Kircher, K., Nyström, M., Wolfe, B. (2021). Eye Tracking in Driver Attention ResearchHow Gaze Data Interpretations Influence What We Learn. Frontiers in Neuroergonomics, 2. https://doi.org/10.3389/fnrgo.2021.778043

[8] G. Lan, B. Heit, T. Scargill, and M. Gorlatova, "GazeGraph," Proceedings of the 18th Conference on Embedded Networked Sensor Systems, Nov. 2020, doi: 10.1145/3384419.3430774.

[9] A. Bulling, J. A. Ward, H. Gellersen, and G. Tröster, "Eye Movement Analysis for Activity Recognition Using Electrooculography," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 33, no. 4, pp. 741-753, Apr. 2011, doi: 10.1109/tpami.2010.86.

[10] A. Bulling, J. A. Ward, and H. Gellersen, "Multimodal recognition of reading activity in transit using body-worn sensors," ACM Transactions on Applied Perception, vol. 9, no. 1, pp. 1â21, Mar. 2012, doi: 10.1145/2134203.2134205.

[11] K. Kunze, Y. Utsumi, Y. Shiga, K. Kise, and A. Bulling, "I know what you are reading - Recognition of Document Types Using Mobile Eye Tracking" Proceedings of the 17th annual international symposium on International symposium on wearable computers ISWC '13, 2013, doi: 10.1145/2493988.2494354.

[12] K. Kunze, Y. Shiga, S. Ishimaru, and K. Kise, "Reading Activity Recognition Using an Off-the-Shelf EEG Detecting Reading Activities and Distinguishing Genres of Documents," 2013 12th International Conference on Document Analysis and Recognition, Aug. 2013, doi: 10.1109/icdar.2013.27.

[13] N. Srivastava, J. Newn, and E. Velloso, "Combining low and mid-level gaze features for desktop activity recognition," Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, vol. 2, no. 4, p. 189, 2018.

[14] P. Olsson, "Real-time and offline filters for eye tracking," 2007, Master Thesis, KTH Royal Institute of Technology.

[15] "SMI Eye Tracking Glasses," Imotions. https://imotions.com/hardware/smi-eye-tracking-glasses/ (accessed Jun. 02, 2022).

[16] K. Rayner, "Eye movements in reading and information processing: 20 years of research." Psychological Bulletin, vol. 124, no. 3, pp. 372-422, 1998, doi: 10.1037/0033-2909.124.3.372.

[17] F. Ratliff and L. A. Riggs, "Involuntary motions of the eye during monocular fixation." Journal of Experimental Psychology, vol. 40, no. 6, pp. 687-701, 1950, doi: 10.1037/h0057754.

[18] H. Collewijn and E. Kowler, "The significance of microsaccades for vision and oculomotor control," Journal of Vision, vol. 8, no. 14, pp. 20-20, Dec. 2008, doi: 10.1167/8.14.20.

[19] D. Berrar, "Cross-Validation," Encyclopedia of Bioinformatics and Computational Biology, pp. 542-545, 2019, doi: 10.1016/b978-0-12-809633-8.20349-x.

[20] M. Belgiu and L. Drăguţ, "Random forest in remote sensing: A review of applications and future directions," ISPRS Journal of Photogrammetry and Remote Sensing, vol. 114, pp. 24-31, Apr. 2016, doi: 10.1016/j.isprsjprs.2016.01.011.

[21] C.-H. Chao, "Ethics Issues in Artificial Intelligence," 2019 International Conference on Technologies and Applications of Artificial Intelligence (TAAI), Nov. 2019, doi: 10.1109/taai48200.2019.8959925.

[22] K. Rayner and S. A. Duffy, "Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity," Memory Cognition, vol. 14, no. 3, pp. 191-201, May 1986, doi: 10.3758/bf03197692.

[23] A. Widodo and B.-S. Yang, "Support vector machine in machine condition monitoring and fault diagnosis," Mechanical Systems and Signal Processing, vol. 21, no. 6, pp. 2560-2574, Aug. 2007, doi: 10.1016/j.ymssp.2006.12.007.

[24] Y. LeCun, Y. Bengio, and G. Hinton, "Deep Learning," Nature, vol. 521, no. 7553, pp. 436-444, May 2015, doi: 10.1038/nature14539.

[25] J. Pineau, "The Machine Learning Reproducibility Checklist (v2.0," 2020. [Online]. Available: https://www.cs.mcgill.ca/ jpineau/ReproducibilityChecklist.pdf.