When do Saccades begin? Prediction of Saccades as a Time-to-Event Problem

Tim Rolff
tim.rolff@uni-hamburg.de
Universität Hamburg
Hamburg, Germany

Frank Steinicke
frank.steinicke@uni-hamburg.de
Universität Hamburg
Hamburg, Germany

Simone Frintrop
simone.frintrop@uni-hamburg.de
Universität Hamburg
Hamburg, Germany

ABSTRACT

We present a novel view on gaze event classification by redefining it as a time-to-event problem. In contrast to previous models, which consider the classification as discrete events, our redefinition allows for estimating the remaining time until the next saccade event. Therefore, we provide a feature analysis and an initial solution for compensating the latency of wearable eye-trackers build in today's head-mounted displays. Similar to previous classifiers, we utilize oculomotor features such as velocity, acceleration, and event durations. In total, we analyze 104 extracted features of three datasets and apply different regression methods. We identify optimal window sizes for each feature and extract the importance of all extracted windows using recursive feature elimination. Afterwards, we evaluate the performance of all regressors using earlier selected features. We show that our selected regressors can predict the time-to-event better than the baseline, indicating the potential usage of time-to-event prediction of saccades.

CCS CONCEPTS

• **Human-centered computing → HCI theory, concepts and models**; Virtual reality.

KEYWORDS

time-to-event prediction, gaze event classification, virtual reality, feature engineering

ACM Reference Format:

Tim Rolff, Frank Steinicke, and Simone Frintrop. 2022. When do Saccades begin? Prediction of Saccades as a Time-to-Event Problem . In *2022 Symposium on Eye Tracking Research and Applications (ETRA '22), June 8–11, 2022, Seattle, WA, USA.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3517031.3529627

1 INTRODUCTION AND RELATED WORK

Gaze event classification describes the assignment of a label to a captured gaze point for each different type of eye movement, including saccades, fixations, or smooth pursuits. With each eye movement having its unique properties that makes it distinguishable, it allows for the classification of these events from the gaze point data stream of the eye-tracker. This classification has been researched for decades and several algorithms to solve the identification of those events have been proposed [Agtzidis et al. 2016; Andersson et al. 2017; Dar et al. 2021; Komogortsev and Karpov 2013; Salvucci and Goldberg 2000; Startsev et al. 2019; Zemblys et al. 2019]. As it is a well-researched topic, the output of those algorithms is often directly used as event labels for the captured data [Hu et al. 2021; Meghanathan et al. 2015] removing the need for tedious hand labelling of individual gaze events. As directly labelling the data stream is often desired for specific applications, most proposed algorithms are capable of processing the data on the output stream of the eye-tracker containing all gaze points captured so far [Komogortsev et al. 2010; Salvucci and Goldberg 2000; Veneri et al. 2011]. With the rise of deep learning, recent state-of-the-art algorithm utilize features calculated on gaze sequences [Startsev et al. 2019; Zemblys et al. 2019] often outperforming classical gaze label classification. These early deep learning models are typically not designed for online application, which restricts their usage to offline processing.

While most of the listed algorithms can be used on the output stream of the eye-tracker, when deployed in commercial HMD's, their output often has a latency of several milliseconds, either due to latency of the eye-tracker itself [Stein et al. 2021] or due to delay introduced in the filtering process required to smooth the incoming gaze data [Schafer 2011]. However, low latencies are essential to virtual reality (VR) [Stauffert et al. 2020], as eye-trackers in head-mounted displays (HMD) often only provide low sampling frequencies [Stein et al. 2021]. To mitigate those, recent attempts have been made to reduce the latency through the probosal of hardware based solutions [Angelopoulos et al. 2020; Li et al. 2020]. Alternatively, waiting for the gaze events to be classified before rendering the frame, for instance, for foveated rendering [Steinicke 2016; Walton et al. 2021], is usually limited since low rendering latency is required to avoid visual discomfort or VR sickness [Stauffert et al. 2018]. Therefore, the classification of saccades is often not fast enough for further processing by other downstream algorithms, as they lag several milliseconds behind the actual signal [Langbehn et al. 2018; Stein et al. 2021; Sun et al. 2018]. This makes the direct usage of gaze events, especially those of saccades, for VR applications quite challenging. Such applications could be, for example, the estimation of gaze shifts in gaze forecasting [Hu et al. 2021, 2020], blink or saccade detection for redirected walking [Langbehn et al. 2018; Sun et al. 2018], gaze contingent rendering [Arabadzhiyska et al. 2017] or intend based gaze interaction [David-John et al. 2021]. Hence, directly classifying these events for a downstream task might miss the actual timeframe the event has happened or requiring unnatural actions, such as intentional blinking [Langbehn et al. 2018].

ETRA '22, June 8–11, 2022, Seattle, WA, USA

ACM ISBN 978-1-4503-9252-5/22/06…$15.00
https://doi.org/10.1145/3517031.3529627

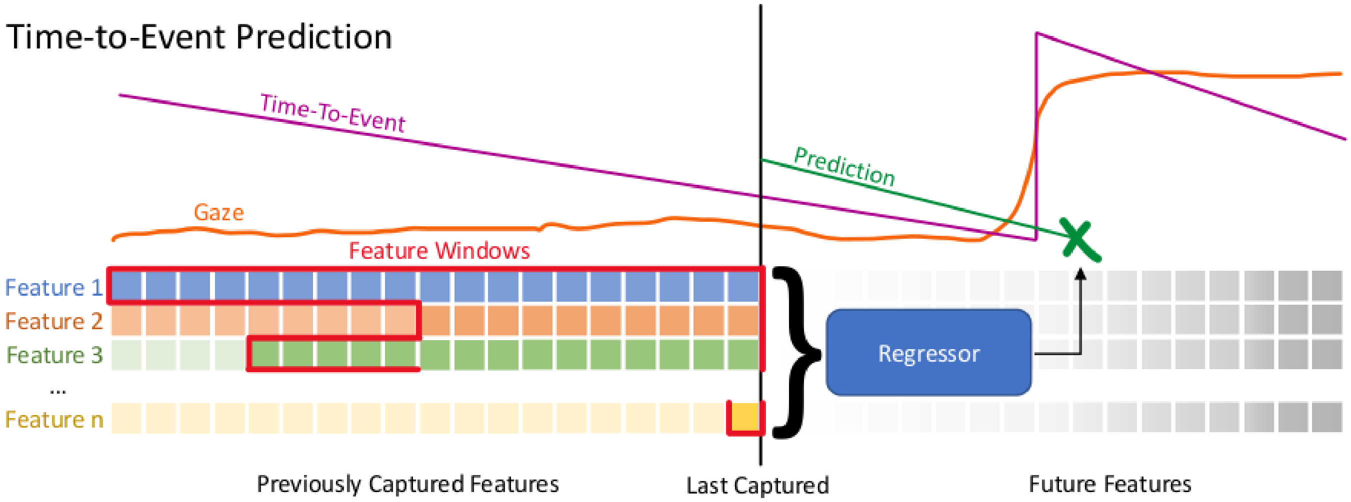

Figure 1: Schematic representation of our approach to the prediction of saccade events given a set of features. The prediction itself is done as an online approach by utilizing previously captured features to predict the remaining time until the start of the next saccade.

Most of the time, however, it is not essential for the above listed use-cases whether the eye-movement for a specific time step is known, but rather when it will change. For example, how many discrete time-steps it will take until a saccade will occur when currently fixating an object. Therefore, predicting the time it will take for a gaze event to change requires to model durations of different eye-movements. Previous work has already shown that these fixation durations are highly variable and depend on multiple factors, for example, on features of the stimulus such as luminance, edge contrast, clutter, or other oculomotor constraints such as saccadic suppression, amplitude of the next saccade, change in saccade direction or viewing time [Dorr et al. 2010; Kowler 2011; Nuthmann 2017; Salthouse and Ellis 1980]. Some of those works already propose statistical models for fixation durations. For example, [Nuthmann 2017; Nuthmann et al. 2010; Walshe and Nuthmann 2021] propose multiple models to predict fixation durations, with [Nuthmann 2017] estimating fixation durations through linear mixture models given local image features in combination with additional oculomotor and spatio-temporal parameters. In this paper, we propose an alternative view on the problem of labelling gaze events by estimating when the event type will change rather than the direct discrete classification of each time step, effectively making it a discrete time-to-event problem [Tutz et al. 2016]. A discrete time-to-event problem models how many time steps it will take until a specific mutually exclusive event is going to happen. These type of problems have long been analyzed in the domains of biostatistics [Bull and Spiegelhalter 1997], social sciences [Yamaguchi 1991], or reliability analysis in engineering [Karim et al. 2019]. Thus, a common synonym for time-to-event problems is *survival analysis*, as the problem is often linked to estimating the time of death. As shown on the right in Fig. 2, we model the time-to-event as the time until the next saccade happens. We purposely choose saccades as our target as most previously mentioned use-cases require a change in state if a saccade has happened. For the estimation of the time-to-event, we utilize different regression techniques using the oculomotor features like velocity, acceleration, or fixation duration captured through the eye-tracker. As depicted in Fig. 1 it would, ideally, allow us to estimate the moment when a saccade will happen given previously captured features. Rephrasing the problem further allows modeling latency as part of the system, as it can be directly accounted for when estimating the timestamp of the change point of the event. We would also like to note that our problem definition differs from the above-mentioned fixation duration studies in that we do not want to predict the total duration of a fixation, but rather the remaining time the current fixation will last only given previously captured features. To our knowledge, no previous research has modeled the classification of saccade events as such a time-to-event problem. Therefore, we will analyze 104 extracted features on their importance to predict the time-to-event of saccades using four different regression techniques. In addition, we will identify the optimal window lengths for each feature and estimate their importance using four different machine learning regressors. For a broader overview of the problem, we will utilize three different real-world and virtual egocentric datasets using four different regression models.

To summarize, our work proposes the following contributions:

- We propose a novel view on the problem of gaze event classification by redefining it from a discrete classification problem of gaze event into a discrete time-to-event problem for the forecasting of saccades.
- We analyze three different datasets and evaluate 104 different features for their optimal window length and importance to predict the time-to-event of saccades.
- We evaluate four different regression methods in their ability to estimate time-to-event of saccades.

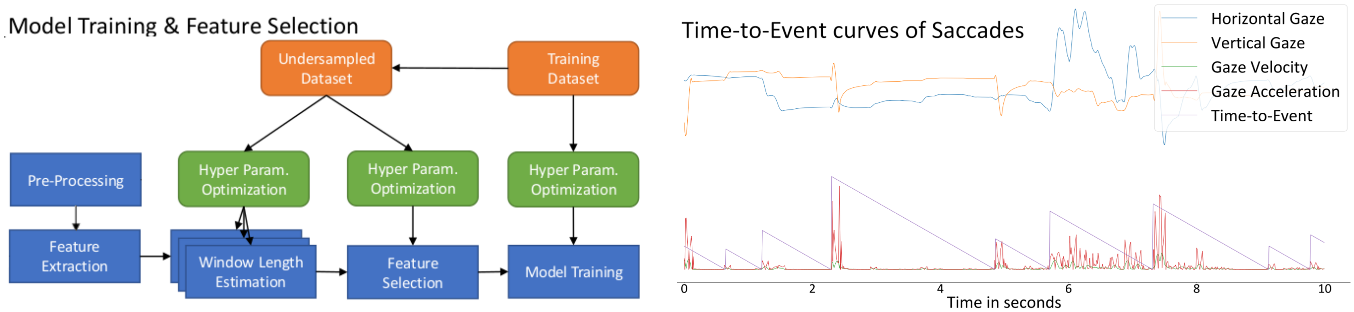

Figure 2: Left: the methodology with which we performed our experiment, further explained in Sec. 2. Right: time-to-event curves (purple) overlaid on top of the gaze acceleration (red) and velocity (green) which are for visualization purposes not to scale. On top, the raw gaze data is shown (blue, orange). Note that a time-to-event curve reaches zero if a saccade event begins, and its local maxima directly after the start of a saccade, as we are not concerned with predicting the duration of a saccade but rather the start of the next saccade event.

2 METHOD

2.1 Overview

In this section, we give an overview of our methodology. In this work, we will explore how well classical regression models can already solve our task: the prediction of the remaining time until a saccade event. We explicitly choose classical regression models for their interpretability but also to provide a well-understood baseline for our specific problem statement. The interpretability of those methods also provides insight into the importance of the selected features, as it is not evident which features should be utilized for the time-to-event prediction due to the novelty of our redefinition. Moreover, this importance can also be used as an indicator for their potential usage in other regression techniques, which do not provide such interpretability. To determine the importance of the features, we will follow the methodology as shown in Fig. 2. First, we pre-process the raw data by applying multiple filters for noise reduction, see Sec. 2.3 for details. Next, we classify the gaze labels for all gaze points if the dataset does not provide them, see Sec. 2.4. Afterwards, we obtain our feature set by extracting 104 different features from the pre-processed data. These features are mainly selected from [David-John et al. 2021; George and Routray 2016] and are computed from the captured pre-processed gaze and inertial measurement unit (IMU) data of the HMD and the eye-tracker (further explained in Sec. 2.5). Given the extracted features, we then perform multiple phases for the feature selection, see Sec. 2.6.

In the feature selection process, we first utilize multiple regression models to identify the optimal window length of each feature, with the window describing the optimal number of the last captured feature points that should be provided to the regressor. The estimation of the window length also helps to identify, which features do not require a sequence, as we also check if a feature has an optimal sequence lengths of one. Using these windows we predict the time-to-event for each new sample. Next, we estimate the importance of each feature window by applying recursive feature elimination, resulting in a rank for each feature window that corresponds to the importance of the particular feature. We would also like to point out that we perform both steps on an undersampled dataset, to avoid adding a bias towards the time-to-event of zero. This is due to the observation that all time-to-event curves will reach that time point eventually, resulting in an imbalance in the dataset. Now, with the window length and the rank, we retrain all regression methods on the full training set to estimate their final performance on our test set.

2.2 Datasets

We analyze the data of three different egocentric virtual and real-world datasets, namely DGaze [Hu et al. 2020], FixationNet [Hu et al. 2021] and EGTEA Gaze+ [Li et al. 2018]. The first provides videos, gaze, and IMU data for head accelerations from 43 participants that were asked to freely explore 2 out of 5 randomly assigned virtual environments. The shown scenes contain different dynamic distractors, which move randomly across the environment. Each participant was instructed to record at least 3 minutes of data without any further specification of a task or explanation of the environment. Further, all participants were provided with a HTC Vive controller, allowing free movement inside the scene. In total, the dataset contains an average sequence length of 20,803 gaze points per session. To capture the data, an HTC Vive in combination with a 7invensun eye-tracker was used.

The FixationNet dataset is similar to the DGaze dataset, such that it can be viewed as an extension of it. It includes the same data modalities, with two dynamic and two static scenes containing 3 variants of the same animals and static objects. The data was captured from 27 participants with the same hardware setup used for the DGaze dataset. However, in contrast to the DGaze dataset, all participants were instructed to solve a specific search task by pointing onto a target while the other two object variants served as distractors. Each participant was tasked to record one static and one dynamic environment with at least 3 trials for each environment, resulting in 162 captured trials. The dataset contains, on average, 12,000 head and gaze points and 7,800 frames per trial with additional information on user ID and task. The gaze data for both datasets was captured with 100Hz.

At last, we evaluate the EGTEA Gaze+ dataset, which captures 29 hours of real-world egocentric videos of cooking activities from 32 participants with a total of 86 sessions [Li et al. 2018]. The participants were tasked to prepare 7 different meal preparation tasks in a kitchen environment. As data points, egocentric videos at 24 Hz and gaze points at 30 Hz were captured for each participant. Afterwards, additional annotations for frequent actions were provided. Unfortunately, the dataset does not contain any additional data points from the IMU of the eye-tracker, such as head acceleration, or orientation.

2.3 Pre-Processing

To pre-process the datasets , we filter the raw gaze data first. As a first step, we transform the raw gaze points from angular view coordinates into normalized screen coordinates ranging from 0 to 1. Further, we set invalid and *nan* values to zero, avoiding the destruction of temporal information and follow a similar filtering process like [Dar et al. 2021]. Here, spikes are removed using a heuristic spike filter [Stampe 1993]. To filter additional noise, two filters are utilized. First, a Savitzky-Golay filter [Savitzky and Golay 1964] with a filter length of 130ms and a polynomial order of 2 is applied to the raw gaze data. Afterwards, we estimate the velocity of each axis independently, using the update rate of the eye-tracker as the time delta. To smooth out fluctuations in velocity, we apply a small median filter of size 3 to each axis and calculate the total velocity using the L^2-norm. Then an analogous approach has been applied for computing the acceleration. As the DGaze dataset contains head accelerations captured with 200Hz, we sub-sampled all input features to match the frequency of the eye-tracker used to capture the dataset, which corresponds to 100Hz for the DGaze and FixationNet. Further, as the EGTEA Gaze+ does not contain IMU data for the computation of head accelerations, we initialized them with zeros with the same frequency as the eye-tracker.

2.4 Eye Movement Classification

After pre-processing, eye-movement classification for the DGaze and FixationNet datasets was performed on the filtered data using the I-VT algorithm [Salvucci and Goldberg 2000], as these datasets do not contain gaze event labels. This algorithm utilizes the velocity of the raw gaze to cluster them into a fixation if they do not exceed the specified threshold of $80°/s$, similar to the one reported by [Hu et al. 2021]. Although, we discarded the clustering step that is usually performed to find the fixation center and only utilized I-VT for the classification into fixations and saccades, assuming that a saccade occurred if the threshold got exceeded. As a result, we obtain a binary label for each time-step of the input, indicating if the participant was fixating at the time or if a saccade was performed. To discard noise, we removed saccades smaller or equal than 10ms and choose 200ms as our minimal fixation duration to comply with typical fixation durations reported by [Holmqvist et al. 2011]. Longer fixation durations also allow to mitigate for the latency added by our method along with the possibility to analyse multiple time-to-event estimates of our model, for example by calucalating uncertainty, mean, or filtering.

2.5 Feature Extraction

For our feature set we extracted 104 features in total, including gaze positions, velocities, accelerations, event labels, head velocities, and durations mostly extracted from gaze and head data. We also include metadata such as user or video IDs. For the selection of features, we mainly based the selection of extracted features on [David-John et al. 2021; George and Routray 2016] and included additional data for head accelerations and event durations. The full list of all extracted features is included in the supplementary. As our gaze data, we utilize the gaze positions that were smoothed through the filtering process described in Sec. 2.4, as well as the head accelerations, and gaze events. We also extracted the gaze velocity, acceleration, normalized screen coordinates from the gaze data using the update rate of the eye-tracker as the time delta. As our target data, we extracted the number of discrete time steps until the desired gaze event happened.

We additionally extracted low-level features over the time-span of each event. However, we still keep the temporal sequence, meaning that we strictly enforce to use feature points that were observed before the currently processed data point. These features include: cluster point of event positions, distance from the cluster center, standard deviation of event, time until the event, time since the last event, duration of the current event and the travelled distance during the current event. Further, we extracted the mean, median, minimum, maximum, standard deviation, skewness and kurtosis for each of the previously extracted features, being the event duration over horizontal, vertical and total velocity, horizontal, vertical and total acceleration and the horizontal and vertical head acceleration. Based on those, we extracted mid-level features, mainly from the event lengths. These include the average event length so far, as well as the standard deviation of those events. Further, we extract the last event length with additional windowed event lengths over all events with window sizes of 3, 5 and 7. As most of the previously described features are only extracted on the same event, we additionally employ window feature extraction regardless of the current gaze event. For the window size, we choose 1 second to capture at least one fixation and saccade. These include the travelled distance, skewness, kurtosis of the horizontal and vertical gaze position and head acceleration.

2.6 Regression Models and Feature Selection

Due to the novelty of the task and to avoid adding a bias by utilizing only one model class, we choose four different regression models for the prediction of the time until the next saccade. We purposefully chose models with different learning strategies, namely k-nearest neighbor (**KNN**) [Altman 1992], AdaBoost [Drucker 1997; Freund and Schapire 1997], linear regression with Nyström kernel approximation [Williams and Seeger 2001] fitted via stochastic gradient descent (**SGD**) [Robbins and Monro 1951], and support vector regression (**SVR**) [Platt et al. 1999]. As our baseline, we used the average time-to-event duration (**Avg.**) on the training set, allowing to estimate the overall performance of our models. We additionally employed Bayesian hyperparameter optimization [Snoek et al. 2012] to find the best set of hyperparameters for KNN's, SGD's and AdaBoost. For the estimation of the optimal window sizes, we initialized the SVR model with constant hyperparameters, due to their computation requirement.

Before feature selection, we split our dataset into training and test sets. We utilize 90% of our data for training and strictly choose different sessions for the training and the test set to avoid data leakage. Further, to reduce data leakage inside the training set, we sub-sampled the generated data and applied random undersampling with a sample size of 50 for each event duration, effectively discarding time-to-event durations with less than 50 samples. This also avoids imbalance caused by the frequent occurrence of shorter time-to-event durations, as each time-to-event sequence will eventually reach a time-to-event duration of zero. To estimate the optimal window length we sampled different window sizes of 1, 10, 20, 40, 60, 80 and 100 samples separately for each feature. With the extracted window sizes, we train a regressor on each and utilize cross validation with a fold size of 7 to find the optimal window length on the training set. To avoid additional bias in the selection of hyperparameters used in each regressor, we employ Bayesian optimization to find a good set of hyperparameters for each feature window separately. However, as a consequence of the size of the datasets we chose 10% of the undersampled training data for the estimation of the hyperparameters with Bayesian optimization resulting in 2624, 2456 and 951 samples for the DGaze, FixationNet and EGTEA Gaze+ dataset. Nonetheless, the evaluation of the window size of each feature was done on the full undersampled training set by retraining a new instance of the best predictor fund in the optimization process.

After estimating optimal window sizes of each feature, we performed another feature selection process to find overall useful features from the extracted windows on all participants. Here, we utilized recursive feature elimination (RFE) [Guyon et al. 2002] with cross validation and discarded the full window instead of singular features for interpretability. During RFE, we utilized linear kernels for the SGD and SVR regressors and retrained all regressors on non-linear kernels afterwards. We also did not apply RFE to KNN's as they do not define feature importance. Instead, we trained all KNN regressors with the full feature set using the optimal window sizes for each feature. Similar to the previous feature selection process, we performed hyperparameter optimization on all regressors. But instead of estimating the best hyperparameter set after the elimination of a feature, we fitted each regressor before the application of RFE on the complete set of features and after the elimination, due to computational reasons.

3 EVALUATION

3.1 Metrics

Commonly used metrics for quantitative measurement of time-to-event data are the concordance index [Harrell Jr et al. 1996], cumulative dynamic AUC [Hung and Chiang 2010] or the brier score [Graf et al. 1999]. The concordance index, measures if two samples were correctly ordered, meaning that a sample with the higher risk score has a shorter time-to-event than a sample with a lower risk. However, we did not utilize a commonly used survival metric, as we are concerned with the exact time point when an event happens rather than the correct order. Instead, we specifically chose the absolute error, as it models the difference between the predicted time-to-event and the exact time point the event happens.

3.2 Results

Table 1 shows the results of all regression methods on all datasets evaluated on the corresponding test sets. Half of the selected methods always outperform our baseline, with the exceptions being KNN and AdaBoost. We found that especially the SGD regressor has shown the best performance on our selected datasets, indicating the potential usage of the selected features for time-to-event prediction of saccades. We also included the 10 most important features of the SGD regressor in Table 2 due to its performance. Surprisingly, we found that among the most important features are the binary indicators of the last fixation and saccade events that are set to one if an event of that type occurred. Further, noticeable are the time since the last saccade as well as the average saccade and fixation duration and their standard deviation. An extensive list of all ranks is provided in the supplementary, along with the window lengths. Note that we cannot compare against any state-of-the-art methods due to the novity of our problem definition, as it is not possible to compare our method directly with the gaze classification models nor with the fixation duration studies mentioned at the begining of the paper due to the reasons given in Sec. 1.

4 CONCLUSION AND DISCUSSION

We redefined the problem of discrete gaze event classification into a time-to-event problem and investigated if classical regression models can be utilized for the prediction of saccade events. We extracted 104 different features, stemming from the raw gaze and the IMU of the HMD and eye-tracker. Using those features, we identified the optimal window length of each feature separately using four different regression techniques. Afterwards, we estimated the optimal feature set using recursive feature elimination, ranking each feature window for their importance to predict the time-to-event of saccade initiations. We found that the most important features are binary event indicators and duration variables of saccades and fixations. This ranking also indicates the potential usage for other regression methods that do not provide the importance of features, such as deep learning or survival analysis methods. As half of the selected regressors outperformed our baseline, we confirm our hypothesis that gaze events can be predicted when redefined as a time-to-event problem. We would also like to emphasize that our model does not aim to replace classical gaze event classifiers but serves as an addition to those. Given that the calculation of features does not depend on future data and the fact that our evaluated methods are fast to compute it allows the real-time application of our models. It, however, remains to be seen if the model is accurate enough for real world applications. Here, more research on decreasing the error is advisable. Possible research directions could be the analysis of additional features like visual stimulus, frontal eye field data features or task data like [Nuthmann 2017] that is already contained in some of the datasets we analyzed. It might also be advisable to research if a calibration phase per user would help to decrease the error. It also requires more research to determine how well the method generalizes to different datasets but given the performance of the SGD regressor shown in Tab. 1 on fully unseen scenes we expect

Dataset	Avg. (Baseline)	KNN	SGD	SVR	AdaBoost
DGaze	420.30	384.89	**301.62**	312.20	353.33
FixationNet	571.70	575.33	**520.25**	521.04	558.83
EGTEA Gaze+	933.15	831.28	**711.96**	748.63	835.49

Table 1: Mean absolute error of all selected regression techniques listed in Sec. 2.6. The error measures the absolute difference between the time-to-event and the prediction on the test sets of the corresponding dataset in milliseconds. The best results are presented as bold text. Note that we do not compare against other methods due to reasons outlined in Sec. 3.

DGaze	FixationNet	EGTEA Gaze+	Mean on all Datasets
Std. of Horiz. Gaze Position	Std. of Vert. Gaze	Time Since Last Fixation	Binary Saccade Indicator
Binary Saccade Indicator	Gaze Event Label	Average Saccade Dur.	Time Since Last Saccade
Total Head Accel.	Std. of Horiz. Velo.	Dur. of current Gaze Event	Binary Fixation Indicator
Gaze Event Label	Binary Saccade Indicator	Time Since Last Saccade	Gaze Event Label
Binary Fixation Indicator	Mean of Horiz. Gaze Accel.	Binary Fixation Indicator	Average Saccade Dur.
Time Since Last Saccade	Horiz. Dist. from Evt. Cluster Pos.	Kurt. of Horiz. Gaze	Std. of Avg. Saccade Dur.
Std. Saccade Dur.	Time Since Last Saccade	Max. Total Gaze Velo.	Dist. from Event Cluster Pos.
Dist. from Event Center	Binary Fixation Indicator	Path Length	Time Since Last Fixation
Dispersion	Ratio	Avg. Dur. of Last 5. Fixations	Avg. Fixation Dur.
Last Fixation Dur.	Max. Vert. Accel.	Binary Saccade Indicator	Kurt. of Horiz. Gaze Pos.

Table 2: The 10 most important features of the SGD regressor trained on the DGaze, FixationNet and EGTEA Gaze+ datasets sorted by their importance in descending order from top to bottom. The *Mean on all Dataset* column shows the overall best features on all datasets, computed as the mean of their importance.

it is a good baseline for the problem at hand. On another note, a comparison against state-of-the-art classification methods when accounting for latency would be a good addition, as this paper did not go further into this topic. Additionally, it might be advisable to evaluate the performance of regressors explicitly designed for time-to-event analysis or deep learning methods, to decrease the measure error. With regard to the error metric, it might also be necessary to investigate additional metrics that take imbalance of time-to-event data into account, especially when also considering other eye-movements such as smooth pursuits.

REFERENCES

Ioannis Agtzidis, Mikhail Startsev, and Michael Dorr. 2016. Smooth Pursuit Detection Based on Multiple Observers. In *Proceedings of the Ninth Biennial ACM Symposium on Eye Tracking Research & Applications* (Charleston, South Carolina) *(ETRA '16)*. Association for Computing Machinery, New York, NY, USA, 303–306. https://doi.org/10.1145/2857491.2857521

Naomi S Altman. 1992. An introduction to kernel and nearest-neighbor nonparametric regression. *The American Statistician* 46, 3 (1992), 175–185.

Richard Andersson, Linnea Larsson, Kenneth Holmqvist, Martin Stridh, and Marcus Nyström. 2017. One algorithm to rule them all? An evaluation and discussion of ten eye movement event-detection algorithms. *Behavior research methods* 49, 2 (2017), 616–637.

Anastasios N Angelopoulos, Julien NP Martel, Amit PS Kohli, Jorg Conradt, and Gordon Wetzstein. 2020. Event based, near eye gaze tracking beyond 10,000 hz. *arXiv preprint arXiv:2004.03577* (2020).

Elena Arabadzhiyska, Okan Tarhan Tursun, Karol Myszkowski, Hans-Peter Seidel, and Piotr Didyk. 2017. Saccade landing position prediction for gaze-contingent rendering. *ACM Transactions on Graphics (TOG)* 36, 4 (2017), 1–12.

Kate Bull and David J Spiegelhalter. 1997. Tutorial in biostatistics survival analysis in observational studies. *Statistics in medicine* 16, 9 (1997), 1041–1074.

Asim H Dar, Adina S Wagner, and Michael Hanke. 2021. REMoDNaV: robust eye-movement classification for dynamic stimulation. *Behavior research methods* 53, 1 (2021), 399–414.

Brendan David-John, Candace Peacock, Ting Zhang, T. Scott Murdison, Hrvoje Benko, and Tanya R. Jonker. 2021. Towards Gaze-Based Prediction of the Intent to Interact in Virtual Reality. In *ACM Symposium on Eye Tracking Research and Applications* (Virtual Event, Germany) *(ETRA '21 Short Papers)*. Association for Computing Machinery, New York, NY, USA, Article 2, 7 pages. https://doi.org/10.1145/3448018.3458008

Michael Dorr, Thomas Martinetz, Karl R Gegenfurtner, and Erhardt Barth. 2010. Variability of eye movements when viewing dynamic natural scenes. *Journal of vision* 10, 10 (2010), 28–28.

Harris Drucker. 1997. Improving Regressors Using Boosting Techniques. In *Proceedings of the Fourteenth International Conference on Machine Learning (ICML '97)*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 107–115.

Yoav Freund and Robert E Schapire. 1997. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of computer and system sciences* 55, 1 (1997), 119–139.

Anjith George and Aurobinda Routray. 2016. A score level fusion method for eye movement biometrics. *Pattern Recognition Letters* 82 (2016), 207–215.

Erika Graf, Claudia Schmoor, Willi Sauerbrei, and Martin Schumacher. 1999. Assessment and comparison of prognostic classification schemes for survival data. *Statistics in medicine* 18, 17-18 (1999), 2529–2545.

Isabelle Guyon, Jason Weston, Stephen Barnhill, and Vladimir Vapnik. 2002. Gene selection for cancer classification using support vector machines. *Machine learning* 46, 1 (2002), 389–422.

Frank E Harrell Jr, Kerry L Lee, and Daniel B Mark. 1996. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Statistics in medicine* 15, 4 (1996), 361–387.

Kenneth Holmqvist, Marcus Nyström, Richard Andersson, Richard Dewhurst, Halszka Jarodzka, and Joost Van de Weijer. 2011. *Eye tracking: A comprehensive guide to methods and measures*. OUP Oxford, Oxford, England.

Zhiming Hu, Andreas Bulling, Sheng Li, and Guoping Wang. 2021. Fixationnet: Forecasting eye fixations in task-oriented virtual environments. *IEEE Transactions on Visualization and Computer Graphics* 27, 5 (2021), 2681–2690.

Zhiming Hu, Sheng Li, Congyi Zhang, Kangrui Yi, Guoping Wang, and Dinesh Manocha. 2020. DGaze: CNN-based gaze prediction in dynamic scenes. *IEEE transactions on visualization and computer graphics* 26, 5 (2020), 1902–1911.

Hung Hung and Chin-Tsang Chiang. 2010. Estimation methods for time-dependent AUC models with survival data. *Canadian Journal of Statistics* 38, 1 (2010), 8–26.

Md Rezaul Karim, M Ataharul Islam, et al. 2019. *Reliability and Survival Analysis*. Springer, Singapore.

Oleg V Komogortsev, Denise V Gobert, Sampath Jayarathna, Sandeep M Gowda, et al. 2010. Standardization of automated analyses of oculomotor fixation and saccadic behaviors. *IEEE Transactions on biomedical engineering* 57, 11 (2010), 2635–2645.

Oleg V Komogortsev and Alex Karpov. 2013. Automated classification and scoring of smooth pursuit eye movements in the presence of fixations and saccades. *Behavior research methods* 45, 1 (2013), 203–215.

Eileen Kowler. 2011. Eye movements: The past 25 years. *Vision research* 51, 13 (2011), 1457–1483.

Eike Langbehn, Frank Steinicke, Markus Lappe, Gregory F Welch, and Gerd Bruder. 2018. In the blink of an eye: leveraging blink-induced suppression for imperceptible position and orientation redirection in virtual reality. *ACM Transactions on Graphics (TOG)* 37, 4 (2018), 1–11.

Richard Li, Eric Whitmire, Michael Stengel, Ben Boudaoud, Jan Kautz, David Luebke, Shwetak Patel, and Kaan Akşit. 2020. Optical gaze tracking with spatially-sparse single-pixel detectors. In *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, 117–126.

Yin Li, Miao Liu, and James M Rehg. 2018. In the eye of beholder: Joint learning of gaze and actions in first person video. In *Proceedings of the European Conference on Computer Vision (ECCV)*. 619–635.

Radha Nila Meghanathan, Cees van Leeuwen, and Andrey R Nikolaev. 2015. Fixation duration surpasses pupil size as a measure of memory load in free viewing. *Frontiers in human neuroscience* 8 (2015), 1063.

Antje Nuthmann. 2017. Fixation durations in scene viewing: Modeling the effects of local image features, oculomotor parameters, and task. *Psychonomic bulletin & review* 24, 2 (2017), 370–392.

Antje Nuthmann, Tim J Smith, Ralf Engbert, and John M Henderson. 2010. CRISP: a computational model of fixation durations in scene viewing. *Psychological review* 117, 2 (2010), 382.

John Platt et al. 1999. Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. *Advances in large margin classifiers* 10, 3 (1999), 61–74.

Herbert Robbins and Sutton Monro. 1951. A stochastic approximation method. *The annals of mathematical statistics* 22, 3 (1951), 400–407.

Timothy A Salthouse and Cecil L Ellis. 1980. Determinants of eye-fixation duration. *The American journal of psychology* 93, 2 (1980), 207–234.

Dario D. Salvucci and Joseph H. Goldberg. 2000. Identifying Fixations and Saccades in Eye-Tracking Protocols. In *Proceedings of the 2000 Symposium on Eye Tracking Research & Applications* (Palm Beach Gardens, Florida, USA) *(ETRA '00)*. Association for Computing Machinery, New York, NY, USA, 71–78. https://doi.org/10.1145/355017.355028

Abraham Savitzky and Marcel JE Golay. 1964. Smoothing and differentiation of data by simplified least squares procedures. *Analytical chemistry* 36, 8 (1964), 1627–1639.

Ronald W Schafer. 2011. What is a Savitzky-Golay filter?[lecture notes]. *IEEE Signal processing magazine* 28, 4 (2011), 111–117.

Jasper Snoek, Hugo Larochelle, and Ryan P Adams. 2012. Practical Bayesian Optimization of Machine Learning Algorithms. In *Advances in Neural Information Processing Systems*, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger (Eds.), Vol. 25. Curran Associates, Inc., Red Hook, USA. https://proceedings.neurips.cc/paper/2012/file/05311655a15b75fab86956663e1819cd-Paper.pdf

Dave M Stampe. 1993. Heuristic filtering and reliable calibration methods for video-based pupil-tracking systems. *Behavior Research Methods, Instruments, & Computers* 25, 2 (1993), 137–142.

Mikhail Startsev, Ioannis Agtzidis, and Michael Dorr. 2019. 1D CNN with BLSTM for automated classification of fixations, saccades, and smooth pursuits. *Behavior Research Methods* 51, 2 (2019), 556–572.

Jan-Philipp Stauffert, Florian Niebling, and Marc Erich Latoschik. 2018. Effects of Latency Jitter on Simulator Sickness in a Search Task. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 121–127. https://doi.org/10.1109/VR.2018.8446195

Jan-Philipp Stauffert, Florian Niebling, and Marc Erich Latoschik. 2020. Latency and Cybersickness: Impact, Causes and Measures. A Review. *Frontiers in Virtual Reality* 1 (2020), 31.

Niklas Stein, Diederick C Niehorster, Tamara Watson, Frank Steinicke, Katharina Rifai, Siegfried Wahl, and Markus Lappe. 2021. A comparison of eye tracking latencies among several commercial head-mounted displays. *i-Perception* 12, 1 (2021), 2041669520983338.

Frank Steinicke. 2016. *Being really virtual.* Springer, Heidelberg, Germany.

Qi Sun, Anjul Patney, Li-Yi Wei, Omer Shapira, Jingwan Lu, Paul Asente, Suwen Zhu, Morgan McGuire, David Luebke, and Arie Kaufman. 2018. Towards virtual reality infinite walking: dynamic saccadic redirection. *ACM Transactions on Graphics (TOG)* 37, 4 (2018), 1–13.

Gerhard Tutz, Matthias Schmid, et al. 2016. *Modeling discrete time-to-event data.* Springer, Heidelberg, Germany.

Giacomo Veneri, Pietro Piu, Francesca Rosini, Pamela Federighi, Antonio Federico, and Alessandra Rufa. 2011. Automatic eye fixations identification based on analysis of variance and covariance. *Pattern Recognition Letters* 32, 13 (2011), 1588–1593.

R Calen Walshe and Antje Nuthmann. 2021. A Computational Dual-Process Model of Fixation-Duration Control in Natural Scene Viewing. *Computational Brain & Behavior* 4, 4 (2021), 463–484.

David R Walton, Rafael Kuffner Dos Anjos, Sebastian Friston, David Swapp, Kaan Akşit, Anthony Steed, and Tobias Ritschel. 2021. Beyond blur: Real-time ventral metamers for foveated rendering. *ACM Transactions on Graphics* 40, 4 (2021), 1–14.

Christopher Williams and Matthias Seeger. 2001. Using the Nyström Method to Speed Up Kernel Machines. In *Advances in Neural Information Processing Systems*, T. Leen, T. Dietterich, and V. Tresp (Eds.), Vol. 13. MIT Press, Boston, USA. https://proceedings.neurips.cc/paper/2000/file/19de10adbaa1b2ee13f77f679fa1483a-Paper.pdf

Kazuo Yamaguchi. 1991. *Event history analysis.* Sage, Newbury Park, California.

Raimondas Zemblys, Diederick C Niehorster, and Kenneth Holmqvist. 2019. gazeNet: End-to-end eye-movement event detection with deep neural networks. *Behavior research methods* 51, 2 (2019), 840–864.

A FEATURE SET

The following features were computed for time-to-event prediction. Note that we take our features mainly from [George and Routray 2016] and [David-John et al. 2021]. We also added durations-based features and features from the inertial measurement unit.

- ***Based on gaze:***
  - ***Computed from singular or adjacent values***:
    Horizontal angular gaze, vertical angular gaze, normalized vertical gaze in screen space, normalized horizontal gaze in screen space, horizontal gaze velocity, vertical gaze velocity, total gaze velocity, horizontal gaze acceleration, vertical gaze acceleration, total gaze acceleration.
  - ***Computed on window:***
    Based on position: clustered horizontal gaze position, clustered vertical gaze position, horizontal distance from clustered gaze position, vertical distance from clustered gaze position, horizontal distance to normalized clustered gaze position, vertical distance to normalized clustered gaze position, euclidean distance from clustered gaze position, standard deviation of horizontal gaze position, standard deviation of vertical gaze position, skewness of horizontal gaze position, skewness of vertical gaze position, kurtosis of horizontal gaze position, kurtosis of vertical gaze position
    Based on velocity: average horizontal gaze velocity, average vertical gaze velocity, median horizontal gaze velocity, median vertical gaze velocity, maximum horizontal gaze velocity, maximum vertical gaze velocity, standard deviation of vertical gaze velocity, standard deviation of horizontal gaze velocity, skewness of horizontal gaze velocity, skewness of vertical gaze velocity, kurtosis of horizontal gaze velocity, kurtosis of vertical gaze velocity, average total gaze velocity, median total gaze velocity, maximum total gaze velocity, standard deviation of total gaze velocity, skewness of total gaze velocity, kurtosis of total gaze velocity
    Based on acceleration: average horizontal gaze acceleration, average vertical gaze acceleration, median horizontal gaze acceleration, median vertical gaze acceleration, maximum horizontal gaze acceleration, maximum vertical gaze acceleration, standard deviation of horizontal gaze acceleration, standard deviation of vertical gaze acceleration, skewness of horizontal gaze acceleration, skewness of vertical gaze acceleration, kurtosis of horizontal gaze acceleration, kurtosis of vertical gaze acceleration, average total gaze acceleration, median total gaze acceleration, maximum total gaze acceleration, standard deviation of total gaze acceleration, skewness of total gaze acceleration, kurtosis of total gaze acceleration, cumulative gaze travel distance
    Other: , Center-Dispersion, dispersion, ratio, gaze travel distance during event (path length)
- ***Based on inertial measurement unit:***
  Horizontal head acceleration, vertical head acceleration, total head acceleration, skewness of horizontal head acceleration over window, skewness of vertical head acceleration over window, kurtosis of horizontal head acceleration over window, kurtosis of vertical head acceleration over window, average horizontal head acceleration, average vertical head acceleration, median horizontal head acceleration, median vertical head acceleration, maximum horizontal head acceleration, maximum vertical head acceleration, standard deviation of horizontal head acceleration, standard deviation of vertical head acceleration, skewness of horizontal head acceleration over all previous samples, skewness of vertical head acceleration over all previous samples, kurtosis of horizontal head acceleration over all previous samples, kurtosis of vertical head acceleration over all previous samples,
- ***Based on event:***
  Gaze event labels, binary indicator of fixation event, binary indicator of saccade event,
- ***Based on duration:***
  Time since the last saccade, time since the last fixation, average fixation duration, average saccade duration, standard deviation of average fixation duration, standard deviation of average saccade duration, last fixation duration, last saccade duration, average of last 3 saccade durations, average of last 3 fixation durations, average of last 5 saccade durations, average of last 5 fixation durations, average of last 7 fixation durations, average of last 7 saccade durations, duration of current event,
- ***Metadata:***
  User ID, frame index, video index

B WINDOW LENGTHS AND FEATURE IMPORTANCE

Here, we also include the estimated optimal window lengths and importance of all features of the SGD regressor, as it reported the best results on all datasets.

Feature	Win. Lenght	Imp.	Feature	Win. Lenght	Imp.
Standard deviation of horizontal gaze position	1	1	Kurtosis of vertical gaze velocity	100	53
Binary indicator of saccade event	10	2	Kurtosis of horizontal gaze velocity	40	54
Total head acceleration	20	3	Vertical gaze acceleration	10	55
Gaze event label	100	4	Maximum vertical gaze acceleration	10	56
Binary indicator of fixation event	10	5	Time since the last fixation	100	57
Time since the last saccade	80	6	Average duration of last 7 fixations	20	58
Standard deviation of saccade duration	60	7	Skewness of vertical gaze position	100	59
Euclidean distance to clustered gaze position	10	8	Average horizontal head acceleration	10	60
Dispersion	20	9	Kurtosis of horizontal head acceleration	80	61
Last fixation duration	20	10	Skewness of horizontal gaze velocity	100	62
Average saccade duration	80	11	Median vertical gaze acceleration	100	63
Maximum horizontal gaze velocity	10	12	Clustered vertical gaze position	80	64
Gaze travel distance during event	10	13	Skewness of horizontal head acceleration	40	65
Standard deviation of horizontal head acceleration	100	14	Average duration of last 5 saccades	20	66
Average duration of last 3 saccades	20	15	Vertical distance to clustered gaze position	100	67
Average fixation duration	40	16	Maximum vertical head acceleration	100	68
Vertical head acceleration	10	17	Average total gaze velocity	20	69
Median horizontal head acceleration	20	18	Kurtosis of vertical head acceleration	100	70
Average duration of last 5 fixations	40	19	Kurtosis of vertical gaze acceleration	100	71
Standard deviation of vertical gaze acceleration	20	20	Average vertical head acceleration	100	72
Cumulative gaze travel distance	40	21	Skewness of vertical head acceleration	100	73
Total gaze velocity	40	22	Kurtosis of vertical head acceleration	40	74
Average duration of last 7 saccades	40	23	Horizontal gaze velocity	20	75
Standard deviation of total gaze acceleration	20	24	Normalized vertical gaze position in screen space	20	76
Kurtosis of horizontal gaze position	20	25	Normalized horizontal gaze position in screen space	10	77
Skewness of total gaze velocity	10	26	Frame index	60	78
Kurtosis of total gaze velocity	100	27	Standard deviation of total gaze velocity	20	79
Horizontal distance to clustered gaze position	20	28	Standard deviation of vertical gaze position	10	80
Standard deviation of vertical gaze acceleration	20	29	Skewness of vertical head acceleration	60	81
Standard deviation of vertical gaze velocity	60	30	Video index	100	82
Dispersion of Euclidean distance to clustered gaze position	10	31	Maximum total gaze acceleration	20	83
Kurtosis of vertical gaze position	80	32	Kurtosis of total gaze acceleration	100	84
Skewness of horizontal gaze position	10	33	User ID	100	85
Total gaze acceleration	40	34	Normalized horizontal distance to from screen center	10	86
Last saccade duration	40	35	Median vertical gaze acceleration	40	87
Average vertical gaze acceleration	100	36	Maximum horizontal head acceleration	10	88
Ratio	100	37	Maximum total gaze velocity	10	89
Standard deviation of fixation duration	80	38	Normalized vertical distance to from screen center	100	90
Horizontal gaze acceleration	20	39	Median vertical head acceleration	100	91
Duration of current gaze event	80	40	Average horizontal gaze velocity	100	92
Average duration of last 3 fixations	60	41	Median vertical gaze velocity	100	93
Horizontal head acceleration	40	42	Maximum vertical gaze velocity	20	94
Standard deviation of vertical head acceleration	100	43	Skewness of vertical gaze acceleration	20	95
Clustered horizontal gaze position	10	44	Vertical gaze velocity	100	96
Median horizontal gaze velocity	100	45	Standard deviation of horizontal gaze velocity	1	97
Skewness of vertical gaze velocity	20	46	Average total gaze acceleration	20	98
Skewness of total gaze acceleration	20	47	Maximum vertical gaze acceleration	20	99
Kurtosis of vertical gaze acceleration	100	48	Skewness of horizontal head acceleration	20	100
Median total gaze acceleration	100	49	Skewness of vertical gaze acceleration	10	101
Kurtosis of horizontal head acceleration	80	50	Average vertical gaze velocity	100	102
Average vertical gaze acceleration	100	51	Vertical angular gaze position	100	103
Median total gaze velocity	20	52	Horizontal angular gaze position	10	104

Table 3: Feature importance (*Imp.*) and estimated optimal window length (*Win. Length*) of the SGD regressor on the DGaze dataset.

Feature	Win. Lenght	Imp.	Feature	Win. Lenght	Imp.
Standard deviation of vertical gaze position	1	1	Skewness of total gaze velocity	10	53
Gaze event label	10	2	Kurtosis of vertical gaze acceleration	80	54
Standard deviation of horizontal gaze velocity	1	3	Average duration of last 3 fixations	100	55
Binary indicator of saccade event	10	4	User ID	10	56
Average vertical gaze acceleration	1	5	Average duration of last 5 saccades	100	57
Normalized horizontal distance to from screen center	1	6	Average horizontal head acceleration	10	58
Time since the last saccade	10	7	Standard deviation of horizontal head acceleration	80	59
Binary indicator of fixation event	10	8	Last saccade duration	80	60
Ratio	1	9	Total gaze velocity	10	61
Maximum vertical gaze acceleration	10	10	Total gaze acceleration	10	62
Standard deviation of saccade duration	10	11	Horizontal distance to clustered gaze position	100	63
Time since the last fixation	10	12	Kurtosis of horizontal head acceleration	60	64
Skewness of horizontal gaze position	1	13	Median vertical head acceleration	10	65
Skewness of total gaze acceleration	10	14	Horizontal gaze velocity	1	66
Kurtosis of total gaze acceleration	10	15	Maximum vertical gaze acceleration	100	67
Median horizontal gaze velocity	1	16	Kurtosis of vertical head acceleration	10	68
Dispersion	20	17	Skewness of horizontal head acceleration	20	69
Average saccade duration	80	18	Clustered vertical gaze position	60	70
Skewness of vertical gaze velocity	10	19	Average duration of last 5 fixations	40	71
Kurtosis of total gaze velocity	10	20	Kurtosis of vertical head acceleration	10	72
Kurtosis of vertical gaze velocity	10	21	Skewness of horizontal head acceleration	20	73
Average vertical gaze acceleration	10	22	Duration of current gaze event	10	74
Standard deviation of horizontal gaze position	20	23	Skewness of vertical gaze acceleration	1	75
Standard deviation of vertical gaze acceleration	20	24	Standard deviation of vertical head acceleration	10	76
Kurtosis of horizontal head acceleration	10	25	Video index	80	77
Kurtosis of vertical gaze position	10	26	Skewness of vertical gaze position	10	78
Average fixation duration	100	27	Normalized vertical distance to from screen center	100	79
Median total gaze acceleration	1	28	Vertical distance to clustered gaze position	10	80
Median vertical gaze acceleration	10	29	Median vertical gaze acceleration	1	81
Euclidean distance to clustered gaze position	10	30	Maximum vertical head acceleration	10	82
Cumulative gaze travel distance	10	31	Skewness of vertical head acceleration	10	83
Frame index	20	32	Standard deviation of total gaze velocity	10	84
Dispersion of Euclidean distance to clustered gaze position	10	33	Median horizontal head acceleration	10	85
Standard deviation of vertical gaze acceleration	10	34	Maximum vertical gaze velocity	1	86
Standard deviation of total gaze acceleration	10	35	Average vertical head acceleration	1	87
Kurtosis of vertical gaze acceleration	10	36	Gaze travel distance during event	20	88
Vertical gaze acceleration	1	37	Vertical gaze velocity	1	89
Vertical angular gaze position	100	38	Kurtosis of horizontal gaze velocity	10	90
Maximum horizontal head acceleration	10	39	Average vertical gaze velocity	1	91
Average duration of last 7 saccades	100	40	Skewness of horizontal gaze velocity	10	92
Kurtosis of horizontal gaze position	10	41	Horizontal angular gaze position	20	93
Standard deviation of fixation duration	80	42	Maximum horizontal gaze velocity	10	94
Median vertical gaze velocity	1	43	Average duration of last 7 fixations	100	95
Total head acceleration	1	44	Clustered horizontal gaze position	1	96
Last fixation duration	100	45	Skewness of vertical head acceleration	10	97
Maximum total gaze velocity	10	46	Median total gaze velocity	1	98
Normalized horizontal gaze position in screen space	40	47	Normalized vertical gaze position in screen space	100	99
Maximum total gaze acceleration	10	48	Horizontal gaze acceleration	1	100
Vertical head acceleration	1	49	Average horizontal gaze velocity	10	101
Horizontal head acceleration	10	50	Average total gaze acceleration	1	102
Average duration of last 3 saccades	80	51	Standard deviation of vertical gaze velocity	10	103
Skewness of vertical gaze acceleration	10	52	Average total gaze velocity	1	104

Table 4: Feature importance (*Imp.*) and estimated optimal window length (*Win. Length*) of the SGD regressor on the FixationNet dataset.

Feature	Win. Length	Imp.	Feature	Win. Length	Imp.
Time since the last fixation	1	1	Dispersion	60	53
Average saccade duration	10	2	Standard deviation of vertical gaze position	60	54
Duration of current gaze event	20	3	Median total gaze acceleration	100	55
Time since the last saccade	20	4	Median vertical gaze acceleration	40	56
Binary indicator of fixation event	60	5	Average vertical gaze acceleration	40	57
Kurtosis of horizontal gaze position	20	6	Standard deviation of vertical gaze velocity	100	58
Maximum total gaze velocity	40	7	Clustered vertical gaze position	1	59
Gaze travel distance during event	40	8	Median total gaze velocity	100	60
Average duration of last 5 fixations	60	9	Kurtosis of horizontal head acceleration	1	61
Binary indicator of saccade event	100	10	Average vertical gaze acceleration	40	62
Total gaze acceleration	20	11	Skewness of vertical gaze acceleration	60	63
Vertical gaze velocity	20	12	Kurtosis of vertical gaze acceleration	20	64
Standard deviation of fixation duration	60	13	Kurtosis of total gaze acceleration	20	65
Normalized horizontal gaze position in screen space	10	14	Median horizontal head acceleration	1	66
Total gaze velocity	100	15	Standard deviation of horizontal head acceleration	1	67
Standard deviation of total gaze acceleration	100	16	Skewness of horizontal head acceleration	1	68
Kurtosis of vertical gaze position	20	17	Standard deviation of vertical gaze acceleration	80	69
Standard deviation of saccade duration	20	18	Average total gaze velocity	100	70
Gaze event label	60	19	Normalized vertical gaze position in screen space	10	71
Last fixation duration	60	20	Dispersion of Euclidean distance to clustered gaze position	60	72
Average duration of last 3 saccades	40	21	Average vertical head acceleration	1	73
Vertical distance to clustered gaze position	10	22	Median vertical head acceleration	1	74
Vertical gaze acceleration	40	23	Maximum vertical head acceleration	1	75
Kurtosis of horizontal gaze velocity	20	24	Standard deviation of vertical head acceleration	1	76
Normalized horizontal distance to from screen center	10	25	Standard deviation of horizontal gaze position	40	77
Average duration of last 7 fixations	80	26	Median vertical gaze acceleration	40	78
Average fixation duration	40	27	Skewness of vertical head acceleration	1	79
Horizontal gaze velocity	40	28	Maximum vertical gaze velocity	40	80
Euclidean distance to clustered gaze position	20	29	Average vertical gaze velocity	40	81
Average duration of last 7 saccades	60	30	Maximum horizontal head acceleration	1	82
Maximum total gaze acceleration	20	31	Standard deviation of vertical gaze acceleration	20	83
Standard deviation of total gaze velocity	100	32	Average horizontal head acceleration	1	84
Average duration of last 5 saccades	20	33	Average total gaze acceleration	100	85
Skewness of vertical gaze position	100	34	Horizontal gaze acceleration	40	86
Cumulative gaze travel distance	100	35	Kurtosis of vertical head acceleration	1	87
Kurtosis of total gaze velocity	20	36	Kurtosis of horizontal head acceleration	1	88
Average duration of last 3 fixations	80	37	Kurtosis of vertical gaze acceleration	10	89
Horizontal distance to clustered gaze position	20	38	Skewness of vertical head acceleration	1	90
Skewness of horizontal gaze position	100	39	Standard deviation of horizontal gaze velocity	20	91
Skewness of horizontal gaze velocity	40	40	Maximum horizontal gaze velocity	60	92
Maximum vertical gaze acceleration	100	41	Skewness of horizontal head acceleration	1	93
Skewness of vertical gaze velocity	80	42	Kurtosis of vertical head acceleration	1	94
Skewness of vertical gaze acceleration	20	43	Median horizontal gaze velocity	40	95
Maximum vertical gaze acceleration	60	44	Average horizontal gaze velocity	40	96
Kurtosis of vertical gaze velocity	20	45	Video index	100	97
Last saccade duration	20	46	Frame index	100	98
Skewness of total gaze acceleration	80	47	User ID	100	99
Clustered horizontal gaze position	40	48	Total head acceleration	1	100
Ratio	40	49	Vertical head acceleration	1	101
Normalized vertical distance to from screen center	20	50	Horizontal head acceleration	1	102
Median vertical gaze velocity	40	51	Vertical angular gaze position	10	103
Skewness of total gaze velocity	20	52	Horizontal angular gaze position	20	104

Table 5: Feature importance (*Imp.*) and estimated optimal window length (*Win. Length*) of the SGD regressor on the EGTEA Gaze+ dataset.