


# Will You Take This Turn? Gaze-Based Turning Activity Recognition During Navigation

Negar Alinaghi ✉ 

Geoinformation, TU Wien, Austria

Markus Kattenbeck ✉ 

Geoinformation, TU Wien, Austria

Antonia Golab ✉

Geoinformation, TU Wien, Austria

Ioannis Giannopoulos ✉ 

Geoinformation, TU Wien, Austria

Institute of Advanced Research in Artificial Intelligence (IARAI), Vienna, Austria

---

## Abstract

Decision making is an integral part of wayfinding and people progressively use navigation systems to facilitate this task. The primary decision, which is also the main source of navigation error, is about the turning activity, i.e., to decide either to turn left or right or continue straight forward. The fundamental step to deal with this error, before applying any preventive approaches, e.g., providing more information, or any compensatory solutions, e.g., pre-calculating alternative routes, could be to predict and recognize the potential turning activity. This paper aims to address this step by predicting the turning decision of pedestrian wayfinders, before the actual action takes place, using primarily gaze-based features. Applying Machine Learning methods, the results of the presented experiment demonstrate an overall accuracy of 91% within three seconds before arriving at a decision point. Beyond the application perspective, our findings also shed light on the cognitive processes of decision making as reflected by the wayfinder's gaze behaviour: incorporating environmental and user-related factors to the model, results in a noticeable change with respect to the importance of visual search features in turn activity recognition.

**2012 ACM Subject Classification** Computing methodologies → Activity recognition and understanding; Computing methodologies → Supervised learning by classification

**Keywords and phrases** Activity Recognition, Wayfinding, Eye Tracking, Machine Learning

**Digital Object Identifier** 10.4230/LIPIcs.GIScience.2021.II.5

## 1 Introduction

While decision making has seen considerable research interest in various fields beyond GIScience and LBS (see e.g., [45, 38, 17]), in our domain, up until now, very few attempts have been made to understand the behavioural correlates of decision making in human wayfinding (see Section 2). This is in contrast to the fact that understanding these decision making processes, which in wayfinding are primarily about turning activity, shows scientific and application perspectives: It can deepen the understanding of the cognitive aspects related to spatial decision-making and provide the grounds for engineering cognitively adequate wayfinding assistance systems, e.g., by monitoring user behaviour and detecting potentially wrong activities before they are actually performed.

The current work takes a step into this direction: we provide evidence that the decision making processes in terms of turning activity are reflected by gaze behaviour and the processes are, moreover, also highly affected by user-related and environmental factors. To this end, we report on an in-situ pedestrian wayfinding study ( $N = 52$ ) involving high precision GNSS and mobile eye tracking, and perform a turn-activity behaviour classification based on a series



© Negar Alinaghi, Markus Kattenbeck, Antonia Golab, and Ioannis Giannopoulos;  
licensed under Creative Commons License CC-BY 4.0

11th International Conference on Geographic Information Science (GIScience 2021) – Part II.

Editors: Krzysztof Janowicz and Judith A. Verstegen; Article No. 5; pp. 5:1–5:16

Leibniz International Proceedings in Informatics



LIPICs Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

of machine learning (ML) experiments, before the actual action of taking a turn. As there exists a vast body of literature which studies how basic eye-movement events reflect cognitive processes, we deliberately trained the classifiers based on a commonly-used set of gaze features. In addition to that, we also selected familiarity (measured as a self-report binary feature indicating whether the wayfinder is familiar or unfamiliar with the environment) as a simple example of a user feature, and number and spatial arrangement of street branches at each junction (see Section 2 for a reason) as indicators of environmental complexity. We provide evidence that wayfinders' gaze behavior (and, hence, their cognitive processes) are highly influenced by these factors while reaching a decision situation.

## 2 Related Work

Gaze has raised particular interest in research on cognitive load and wayfinding assistance alike. Given the aforementioned research goals, we review apt prior evidence relating to, first, gaze behaviour, cognitive load and (spatial) decision making. Second, we touch on wayfinding studies which look into gaze behaviour during this spatial activity. Third, prior work on using machine learning methods for activity classification based on eye tracking data is reviewed in order to gain an insight which classification techniques have shown promising results.

### 2.1 Gaze Analysis in Decision Making and Beyond

Since Yarbus' seminal work [55] in the 1960s, studying gaze behaviour and understanding how it reflects cognitive processes has drawn significant attention by researchers (see e.g., [22] for an early example, revealing that different cognitive tasks yield specific fixation patterns). In particular, the examination of eye movements in decision making and its underlying models has seen a long-lasting research interest (see [35] for an overview). So far, however, it remains undecided whether gaze behaviour triggers or reflects an actual choice which has been made before, and choices according to decision theory are frequently not well reflected in gaze behaviour (see [46, p. 117] for both claims). Having said this, existing literature points towards the fact that gaze bias is a good indicator for the decision to be made: A bias in gaze behavior can be observed approximately the last five seconds before we make a choice [17] and this effect is commonly referred to as the *Gaze Cascade Effect* [45]. The impact of this effect has also been studied for spatial decision making: Wiener and colleagues [53], e.g., used an eye-tracking experiment investigating the relation of gaze behavior, spatial decision making and route learning strategies, and demonstrated the importance of visual decision making in wayfinding performance. In another study, Wiener and colleagues [54] report on four studies, in which eye tracking was used to understand spatial decision making during wayfinding. Their results suggest that gaze and fixations depend on both, the task (either route learning or spatial exploration) and the environmental features.

Beyond the exploitation of fixations and saccades, however, scholars have also used pupil dilation and spontaneous eye blink rate (see [10]) as higher frequency eye tracking devices have become increasingly widespread. Numerous studies have, thereby, linked pupillometry to cognitive effort since Hess and Polt (see [20] according to [33]) found the link between higher-order cognitive processes and pupil features. Since we use a mobile eye-tracking device in our in-situ wayfinding study, we will not use any pupil-based features in our turn-activity classification task due to the frequency of 200Hz. Indeed, we will focus on a subset of fixation- and saccade-related features suggested in the literature in order to gain an insight whether real-time classification may come into range. Similarly, we have deliberately chosen not to exploit the gaze bias due to the sensory complexity this would have yielded.

## 2.2 Studying Gaze in Research on Wayfinding

Gaze has seen increased interest in recent years (see [26, p. 2–9] for an overview) with respect to both, research on wayfinding and gaze-based wayfinding assistance systems (see [14] for a dissertation spanning both aspects). Scholars have, thereby, studied gaze in wayfinding with respect to both, real-world (including indoor) and VR environments. Schnitzler and colleagues [43], for example, studied three different wayfinding aids (signage, paper or digital map) for indoor environments. By segmenting the route into 28 segments, they found, e.g., evidence for an increased visual attention at crucial decision points along a route, i.e., those which are important for floor changes. Comparing gaze behaviour during a real-world and VR wayfinding studies, Dong and colleagues [8] recently provide evidence for both, differences and commonalities of gaze behaviour between both conditions during a map reading task. For example, while the real-world condition yielded higher fixation times on the map, more fixation time was spent on self-localization in the first-person view in VR; both conditions, however, show similar fixation frequencies. In [52] differences in gaze behaviour based on a real-world wayfinding study are reported. Participants learned part of a route by incidental learning, whereas the remaining part was learned intentionally. Regardless of the learning mode, the visual salience of landmarks was a good predictor for fixation time. Intentional and incidental learning, however, show difference with respect to the structural salience of landmarks (longer fixations in case of intentional learning). [50] studies how gaze behaviour reflects subjective risk perception in bicycle riders by considering fixation duration, sight vector length and gaze angle. The results revealed a positive correlation of familiarity with the environment and cycling experience and fixation duration, gaze angle and sight vectors. Very recently, Liu and colleagues [30] study the impact of an important environmental feature, namely the regularity of road patterns, on the duration of fixations in general, mean duration, the duration of the first fixation and the fixation count. Depending on the task (shortest route selection or relative direction of destination), participants' gaze was more impacted by the (ir)regularity of the road pattern across tasks than by signs or other visual stimuli attached to buildings, a finding which is in line with the notion of an increase in wayfinding decision situations (see [16]).

## 2.3 Machine Learning for Gaze Data

Having a look at which machine learning techniques have been used to conduct classification tasks based on eye tracking, the use of Support Vector Machines (SVMs) is prevalent across different eye tracking technologies and usage scenarios. Using electrooculography to monitor eye movements, Bulling and colleagues [4], e.g., trained an SVM classifier based on a very small sample of participants ( $N = 8$ ) and achieved reasonable results regarding both, precision and recall, for an activity classification task including five everyday tasks. Similar findings for the classification of everyday tasks were reported by [44] combining gaze and motion features. Using 229 saccadic and fixation-related features, Kiefer and colleagues [25] provide evidence that six different activities which are commonly performed on maps (e.g., comparing the size of polygons when comparing e.g., lakes, or following a line representing a road) can be distinguished using a SVM classifier (achieved accuracy 78%). Liao and colleagues [29] use a Random Forest classifier and compare its performance to six different classifiers for classifying tasks according to Downs and Stea [9, pp. 125–135], based on a dataset collected during a real-world study. They use saccadic- and fixation-related features and find an accuracy of 67% when using a time window of 17 seconds.

Based on this evidence we conclude, first, to use features which are hand-crafted from eye movements like all of the aforementioned studies do and deliberately choose not to use deep learning methods. Second, regarding the reports on SVM and Random Forest algorithms in earlier studies, as well as our own pilot-test findings (see Section 4.2), we base our machine learning experiments on tree-based techniques. Using these approaches is also in line with other machine learning literature which suggests that algorithms which split the feature space are considered among the best models when it comes to small-to-medium sized structured or tabular data [48].

### 3 Data Collection, Pre-processing and Feature Extraction

Based on our goal of turning activity recognition, this section provides information on data collection and all data pre-processing steps, including the extraction of gaze features. We start with a short description of the human-subject in-situ study as we use a dataset which was collected in order to address several research problems, next to the turn activity recognition. We, then, move on to explain the synchronization of the different sensory data used. Finally, we provide details on the segmentation of the gaze data and the extraction of saccade- and fixation-related features.

#### 3.1 Data Collection

The data used in this paper is part of a larger data collection effort<sup>1</sup> addressing human spatial behavior in real-world wayfinding scenarios<sup>2</sup>. This goal renders it also a valuable source for answering other questions regarding decision-making in navigation. The data collection, which took place in 2020, required participants to walk two routes (length ranging from 0.9km to 1.3km), one of which was located in an area they were familiar<sup>3</sup> with, whereas they were unfamiliar with the other. They had to find their way by means of auditory, landmark-based<sup>4</sup>, turn-by-turn route instructions, which were provided to them on demand and as many times as they requested. We tracked participant's behaviour using a mobile eye-tracking device (PupilLabs Invisible) recording gaze positions at 200Hz and a high precision GNSS receiver (PPM 10-xx38) tracking their position in time. In total,  $N = 52$  people (27 female and 25 males,  $M(age) = 26$  years,  $SD(age) = 8.3$ ) participated in the outdoor experiments resulting in  $N = 104$  trials out of which  $N = 86$  were contained for further processing<sup>5</sup>.

#### 3.2 Data Pre-processing

To obtain the gaze-related data at each decision-point, several steps had to be taken. First, junctions had to be matched to the GPS tracks obtained in order to be able to align eye-tracking and GPS data. Subsequently, the point in time when a junction can be perceived as junction had to be approximated.

---

<sup>1</sup> This experiment has been described for the first time in [19] and here we only present the parts we used in our analysis.

<sup>2</sup> The data used in this paper, will be made available at <https://geo.geoinfo.tuwien.ac.at/resources/> (DOI: 10.5281/zenodo.4298703).

<sup>3</sup> During an online study participants indicated areas and places therein, with which they are familiar and provided a route connecting two of these places.

<sup>4</sup> Points of Interest were used as landmarks and chosen according to the algorithm described in [40].

<sup>5</sup> Trials had to be excluded, for example due to equipment malfunction or non-cooperative participant behaviour.

### 3.2.1 Matching Junctions to GPS Tracks

In order to match junctions to GPS trajectories, we make extensive use of VGI data (OpenStreetMap) in a multistep procedure:

**Step 1: Retrieve intersections** We retrieve all intersections of type *car and pedestrian* or *pedestrian only* within a distance of  $30m$  from the GPS track using the *Intersections Framework* [11].

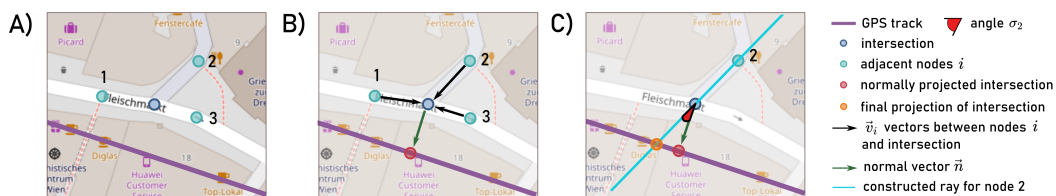
**Step 2: Filter intersections** To avoid having unreasonably short segments, we exclude any pedestrian intersections in proximity of  $50m$  to road intersections. Furthermore, due to the fact that many auxiliary path segments which are not meaningful for real-world wayfinding task (e.g., ways in parks would be obstructed by barriers) exist in OSM data, two raters individually checked each of the junctions using a web map. Subsequently, they agreed on a set of rules for exclusion and checked each of the candidate-intersections during a pair review session, ultimately eliminating 695 intersections.

**Step 3: Labeling** Based on the route instruction, we assign to each of the intersections one of the classes *turn left (TL)*, *turn right (TR)* or *non-turn (NT)*.

**Step 4: GPS track recovery** Low quality GPS tracks, which were inevitable due to the urban environment, were retraced using a polyline onto which the GPS trajectory was normally projected. This projection is applied to refine the GPS points coordinates and, at the same time, preserve the timestamps of the original points.

**Step 5: Match intersections to GPS track** We matched each of the intersections along a route to their corresponding GPS point (see Figure 1) by drawing a ray oriented according to the ways meeting in the given intersection, and intersecting it with the GPS trace.

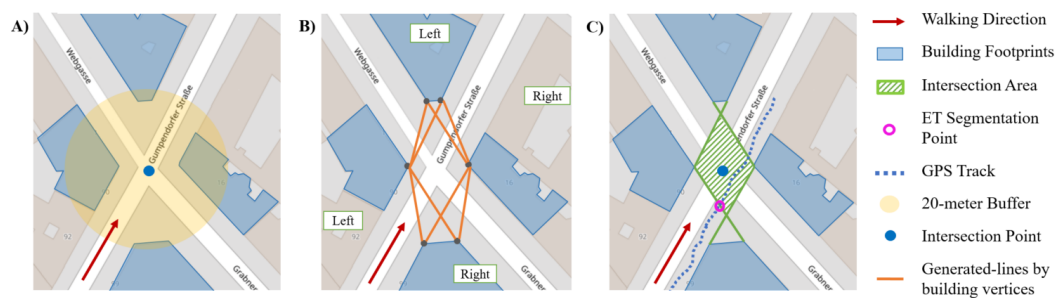
Taken together, the procedure described above enabled us to identify the GPS point of our trajectory representing the junction and, consequently, also the corresponding timestamp at which this junction point was reached. We exploited this timestamp for segmenting the eye tracking data for our turning activity recognition experiment.



**Figure 1** Identification of GPS points corresponding to an intersection **A**: Extract adjacent nodes to which the intersection node is connected by way entities **B**: Determine the normal vector  $\vec{n}$  (green) and construct vectors  $\vec{v}_i$ , which represent the direction between each adjacent node 1,2,3 and the intersection. **C**: Construct rays passing through the adjacent nodes and the intersection. We choose the ray with the smallest angle  $\sigma_i$  between vectors  $\vec{n}$  and  $\vec{v}_i$  (in this example ray from node 2) and intersect it with the GPS track to determine the GPS position.

### 3.2.2 Eye Tracking Data Segmentation

Since we wanted to monitor gaze behaviour during decision making in real-world wayfinding, we need to identify the point in time (again based on the GPS track) at which participants were able to perceive a decision situation as such. This can only be the case where participants have a line of sight to further wayfinding options, i.e., a junction is not perceived as a point by wayfinders but as a larger area where different elements of the street network meet.



■ **Figure 2** Extracting intersection boundary in order to have the closest point to decision-situation in reality, which is essential for eye tracking data segmentation.

Figure 2, illustrates the process for those cases in which buildings were located at a junction<sup>6</sup>: We select building footprints within a distance of  $20m$  to the intersection point for both sides of the current route segment (Figure 2 A). We, then, select the three vertices of each building polygon which are closest to the intersection point and connect each of these with the selected vertices on the opposite side of the street (the set of orange lines in Figure 2 B). Finally, we find the intersection boundary by choosing those lines from the set of orange lines which are oriented as perpendicular as possible to the given route segment (Figure 2 C). Having extracted the intersection boundaries at each junction, the closest GPS point was selected and its corresponding timestamp was derived. This point in time was found in the raw gaze-position data and flagged as the beginning of the intersection area, i.e., the point in time at which a decision would have been made already.

### 3.3 Feature Extraction

Inspired by prior evidence (see Section 2), we extracted 28 fixation-/saccade-related features (see Table 1). Since a  $200Hz$  gaze-recording frequency does not allow for velocity-based feature extractions, we used the IDT [42] as a commonly-used dispersion-based algorithm to detect fixations (gaze-dispersion threshold:  $0.02\ deg$ ; time threshold:  $100\ ms$ ) [15, 14]. All saccades were calculated based on this set of fixations. We extracted all of these features for two to ten seconds (referred to as *window* from now on) before a participant would reach the intersection area (see above). We, thereby, aim to cover the decision process as Brunye and colleagues provide evidence in [3] that ‘[w]hile intersections may prompt a decision, and elicit overt behavior that reflects a decision, the process of arriving at a spatial decision often occurs before arriving in [sic] the intersection.’

Inspired by the work on wayfinding decision situations [16], we decided to include two environmental features: namely the number of ways an intersection comprises and the skewness of these street segments. Both of these features were obtained using the *Intersection Framework* [11] and can, hence, be easily obtained on a global scale. In addition to that, we included familiarity with the environment traversed as a binary variable because prior evidence strongly suggests that familiarity has an impact on both spatial behaviour (see e.g., [50, 18, 37, 13]) and visual search behaviour (see e.g., [23, 51]). Having taken all steps for data preparation, we ended up having nine tabular datasets of dimension  $804 * 32$ .

<sup>6</sup> If no buildings were located at a junction, we found the boundary of the intersection by using a threshold of  $3.75seconds$  from the projected junction point. This resembles  $5m$  based on the assumption of an average walking speed of  $4.5km/h$  [27].



■ **Table 1** Gaze-based features extracted for turn-activity classification. For each row, the rightmost column indicates the final number of features we gained by applying the statistical measures, presented in the first column, on the gaze features represented in the second column. For instance, the first row of saccade-based features, encodes eight features, namely, mean/min/max/var of both, saccade amplitude and saccade duration.

Fixation-based Features		
mean, min, max, var frequency	duration, dispersion, dispersion X, dispersion Y -	16 1
Saccade-based Features		
mean, min, max, var skewness frequency	amplitude, duration amplitude -	8 1 1
g-l ratio (the ratio between long and short saccades)	amplitude	1

## 4 Machine Learning Experiments

### 4.1 Data Preprocessing

Prior to model training, all categorical features should be converted to numerical values, and the sample distribution across all turn-activity classes should be balanced. In order to resolve the imbalanced class issue in our dataset (“NT” class containing 636 samples, had almost four times more samples than both “TL”, containing 91 samples, and “TR”, containing 77 samples), we used Synthetic Minority Over-sampling Technique (SMOTE) for oversampling [5]. By default, SMOTE oversamples all classes to achieve equal sample sizes based on the largest sample size. Therefore, we ended up having 636 samples per class which sums up to 1908 samples in total.

### 4.2 Machine Learning Experiments

As mentioned above (see Section 3.1), we set our sights on classifying wayfinders’ turning activity, mainly by using their gaze behavior, and to do so we extracted some basic gaze-based features and prepared the data for the classification task.

During pilot testing three different models were utilized, namely, SVM-RBF, CART and Random Forest. As described in Section 2, Support Vector Machine (SVM) classifiers have been particularly popular for gaze data classification tasks. Therefore, this algorithm was implemented first for the classification task (test accuracy:  $0.58 \pm 0.05$ ). The Decision Tree (test accuracy:  $.61 \pm 0.06$ ) creates a system of rules to roughly split the feature space into several regions. Following the advantage of tree structures, the Random Forest algorithm (test accuracy:  $.77 \pm 0.06$ ) was next deployed to capture the insight of several Decision Trees in order to divide the feature space more precisely. The Random Forest algorithm yielded better test accuracy which supports the assumption that tree-based structures are more suitable for modeling the data at hand (see e.g., [48, 34, 28]). Additionally, as tree-based ensemble algorithms are reported to have multiple advantages over normal tree-based structures, including the handling of multi-collinearity in features (see e.g., [7]), we next deployed a Gradient Boosted Trees algorithm. All these classifiers were compared in the pilot testing based on achieved accuracies and their potential power to model the provided input data: Gradient Boosted Trees showed the most promising results. We, therefore, chose this class of ML algorithm to base our results on and will provide further details about this algorithm, only.

### 4.3 XGBoost Classifier

Gradient boosting is a very powerful technique for building predictive models [6]. The main idea behind this algorithm stems from the so-called “Hypothesis Boosting Problem” which states that several poor hypotheses (so called weak learners) can be converted into a very good hypothesis [24]. Compared to Random Forest classification, Gradient Boosted Trees have a lot of model capacity so they can model very complex relationships and decision boundaries. While there are various implementations of Gradient Boosted Trees available, we used XGBoost due to its scalability and high efficiency [6]. The Boosting method in XGBoost combines weak learners (i.e., trees) sequentially so that each new tree corrects the error of the previous one. We used Python 3.8 for our implementation and made use of different packages including xgboost 1.3.3 and scikit-learn 0.24.1.

Splitting the data into 70% and 30% (train, validation) and using 20-fold validation, we performed a randomized search to tune hyper parameters. The hyper parameters performing best were: *colsample-bylevel* = 0.7, *colsample-bytree*: 0.8, *learning-rate*: 0.2, *max-depth*: 15, *min-child-weight*: 0.5, *n-estimators*: 800, *reg-lambda*: 1.0 and *subsample*: 0.5. Using the trained model, we classified the data and plotted the *mlogloss* and *merror* to check for signs of overfitting (cross entropy was used as the loss function). In addition to that, in order to avoid training and validation sessions with samples from the same user we deployed a “leave one out” method, in which all samples from one participant were left out and only used for validation. The results of our XGBoost classifier are presented in Section 5.

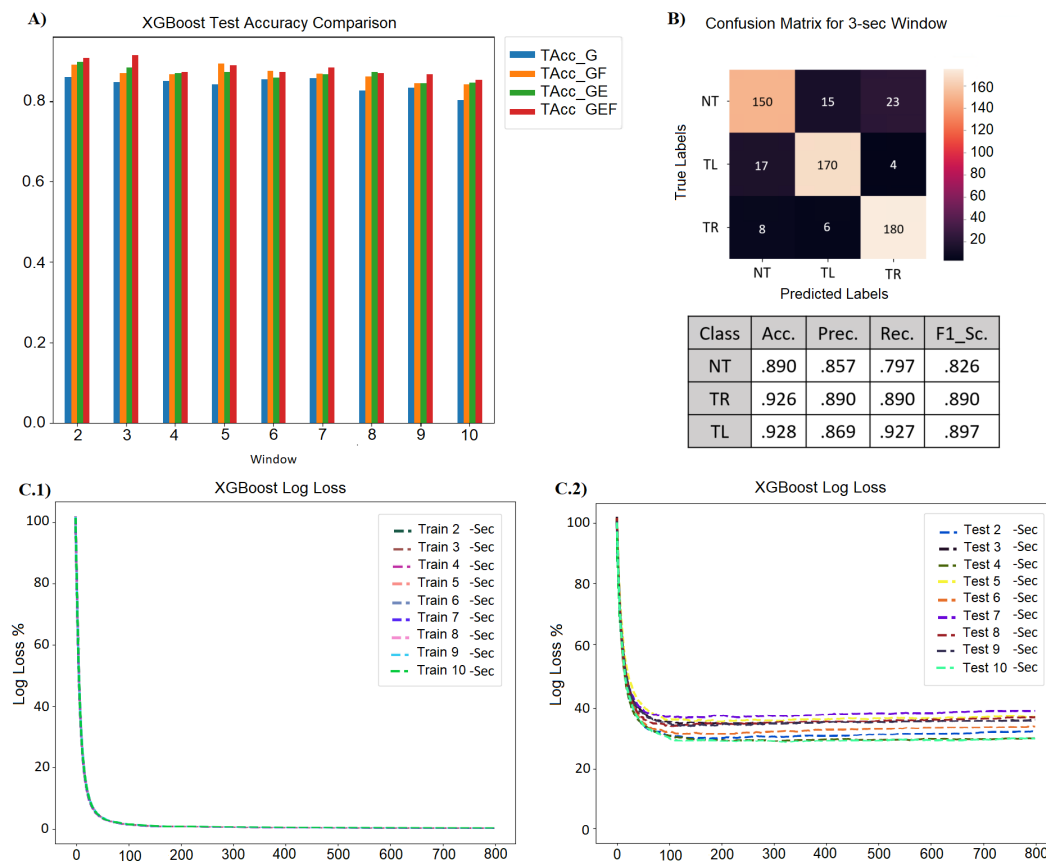
As mentioned above, the basic idea behind boosting algorithms is to build a weak model, making conclusions about the various feature importance and parameters, and then using those conclusions to build a new, stronger model trying to reduce the misclassification error. The most important part which allows us to come closer to an interpretation of the results is the so-called feature importance. In order to preserve both the *global* impact of features on the model and the *individualized* impact of features on a single prediction, we used the Tree SHapley Additive exPlanations (SHAP) method to calculate feature importance as it provides consistent results through handling collinearity of features automatically [31]. The interpretation of results using these features importance, is presented and discussed in Section 6.

## 5 Results

This section provides the results of our machine learning experiments for turning activity classification. We distinguish three classes, namely non-turn (NT), turn left (TL) and turn right (TR). We run our experiments based on four different sets of features: *Gaze-based (G)* features, *Gaze-based + Familiarity (GF)* features, *Gaze-based + Environmental (GE)* features, and *Gaze-based + Environmental + Familiarity (GEF)* features. As explained in Section 3.2.2, we considered these sets of features in different time windows (i.e., from two to ten second) before the actual action of turning.

The overall results in terms of *Test Accuracy* for two- to ten-seconds windows are presented in Figure 3A. Additionally, Figure 3B, presents the confusion matrix for the three-seconds window and its corresponding evaluation metrics including accuracy, precision, recall, and F1-score per class. The false positive/negative rate within turn classes (turn-left/-right) are lower than between non-turn and each of the turn classes. This observation indicates that the model has a better performance distinguishing turn left and right, than differentiating non-turn from turn classes. Figures 3C.1 and C.2, represent, as an example, the logloss error for training and test based on the GEF-dataset for all time intervals. The results show two





**Figure 3** This figure contains, **A:** Test accuracy of XGBoost classifier for four categories of features: Gaze features (TAcc-G), Gaze and Familiarity features (TAcc-GF), Gaze and Environmental features (TAcc-GE), and Gaze and Familiarity and Environmental features (TAcc-GEF); as shown in the figure adding more factors to incorporate user and environmental aspects boosts the model performance; **B:** Confusion matrix calculated for three-seconds window, and evaluation metrics including *Accuracy*, *Precision*, *Recall* and *F1-Score* per class; this indicates that the model has a better performance distinguishing left and right turns compared to non-turns from turns. **C.1** and **C.2:** Log loss plots for XGBoost classifier, respectively for *Train* and *Test* subsets, illustrating the model behavior during the training and test sessions.

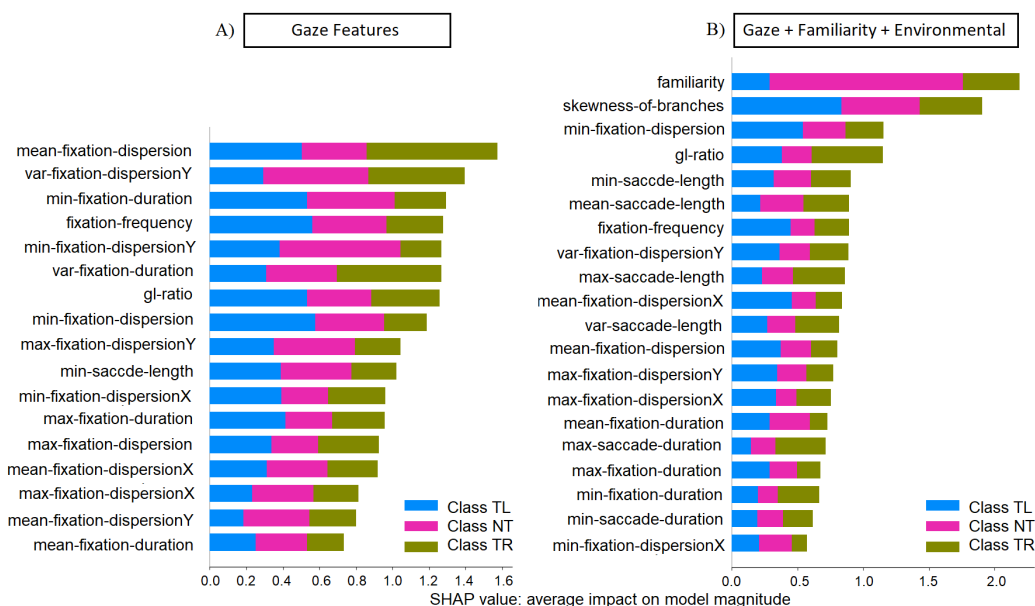
important patterns: 1) the accuracy increases as the distance to the decision point decreases; 2) adding environmental and familiarity features considerably enhances model accuracy. In order to gain a deeper understanding of the results, Table 2 provides details regarding Train (TrAcc) and Test accuracies (TAcc) and the kappa coefficient for all time intervals and all combination of features. The results indicate that using XGBoost, the highest accuracy of 91.4% in turn-activity classification was achieved for the three-seconds window. Therefore, our analysis of feature importance focuses on this timeframe. Figure 4 presents the SHAP importance plots for the features in A showing the gaze features only, and B showing all three categories of features. These plots indicate that binary familiarity of participants and the skewness of the street segments at an intersection are very important for the accuracy of our classification results. A deeper understanding of the differences in importance for different feature classes can be gained from Figure 5, which shows the SHAP importance ranks (see Figure 4) for features per group of features. To visually distinguish features

## 5:10 Will You Take This Turn?

from each other, we use a combination of background colors and outline patterns: Saccadic features are depicted in yellow, fixation-related features are given in green, environmental features are shown in pink and blue denotes the familiarity feature. Count-based features are shown using a dashed outline, whereas a dotted outline marks all duration-based features and a solid outline highlights all dispersion/length-based features. This figure shows a clear pattern: Adding user-related and environmental features increases the overall number of saccadic features.

■ **Table 2** Turn activity classification results for two- to ten-seconds window (W), and the four feature sets: “Gaze” features (G), “Gaze + Familiarity” features (G+F), “Gaze + Environmental” features (G+E), “Gaze + Familiarity + Environmental” features (G+F+E).

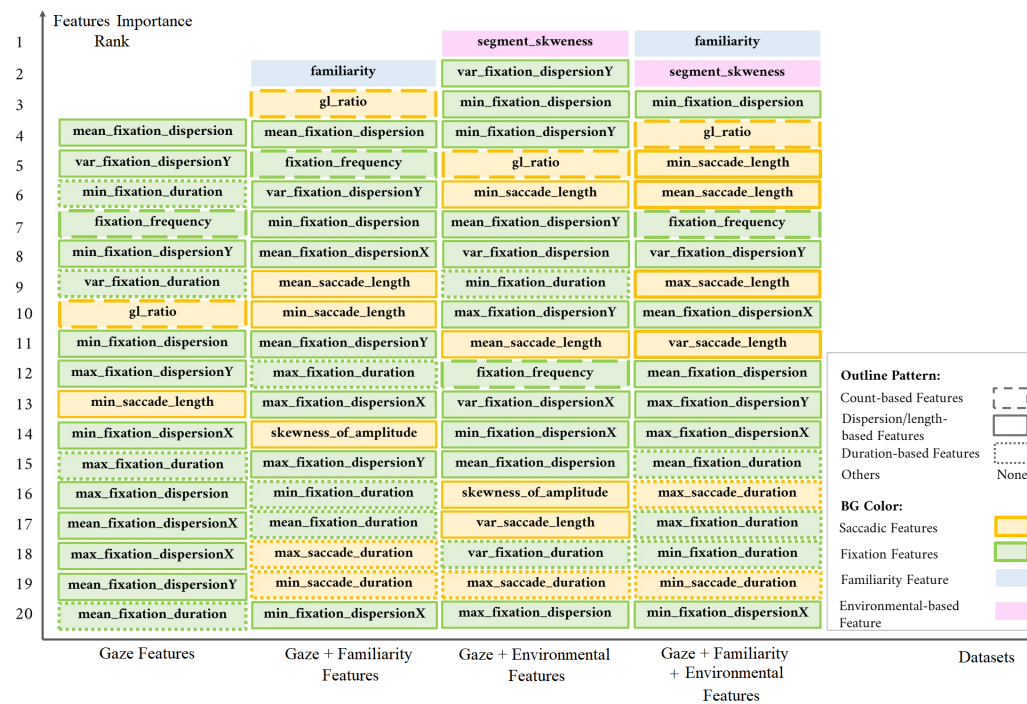
W	G			G+F			G+E			G+F+E		
	TrAcc	TAcc	Kappa	TrAcc	TAcc	Kappa	TrAcc	TAcc	Kappa	TrAcc	TAcc	Kappa
2	.842	.860	.748	.868	.891	.825	.881	.898	.772	.883	<b>.907</b>	.835
3	.872	.848	.812	.857	.869	.783	.890	.884	.741	.891	<b>.914</b>	.896
4	.853	.851	.744	.857	.867	.786	.872	.870	.788	.883	<b>.873</b>	.777
5	.838	.842	.772	.861	<b>.893</b>	.824	.876	.873	.784	.868	.890	.786
6	.827	.853	.789	.876	<b>.876</b>	.822	.898	.858	.756	.883	.873	.799
7	.823	.857	.802	.895	.868	.803	.851	.866	.830	.919	<b>.883</b>	.775
8	.815	.826	.774	.866	.861	.814	.838	<b>.872</b>	.782	.869	.870	.817
9	.850	.833	.735	.895	.845	.760	.842	.845	.705	.879	<b>.867</b>	.785
10	.847	.802	.730	.847	.841	.769	.847	.846	.716	.859	<b>.852</b>	.814



■ **Figure 4** SHAP importance plot for three-seconds window, derived from model, **A:** using only gaze-based features, and **B:** using all features. (Due to space restrictions, only the plot for these two feature sets are presented.)

## 6 Discussion

The results presented in Section 5, primarily, indicate that turning activity in real-world navigation scenarios using gaze-movement data and a combination of user/environmental-related features can be predicted reasonably well. It is outlined in Figure 3B, that adding



■ **Figure 5** Introducing familiarity and environmental features has a major impact on the feature importance patterns: Saccadic features which, according to the literature, reflect visual search behavior (see e.g., [36, 47]), become more dominant by encoding more information about the environment and user experience into the model.

familiarity feature increases the model accuracy, on average across all window sizes, from 83% (obtained by applying only gaze features) to 86%. Likewise, by adding environmental features to gaze data, we can see the same 3% rise in model performance from 83% (on average for all time intervals) to 86%. More importantly, adding both, user and environmental features, on top of gaze-based features boost the model performance to almost 87% on average and best performance of 91%. Taken together, these results, in general are in line with the decision-situation model proposed by Giannopoulos and colleagues [16], which state that the complexity of a wayfinding decision situation depends on user-related and environmental features, alike. The results also reiterate the findings by Brunye and colleagues [3] in two respects: Our results also suggest that the decision pattern for turning at intersections is deeply influenced by environmental elements and experience with the environment. Second, we provide evidence that the decision making in wayfinding occurs before wayfinders arrive at an intersection as, across feature sets, the highest accuracy was achieved for two to three seconds before this point in time. These findings provide the ground for discussion from both, theoretical and application aspects, and the following subsections are devoted to present them in more detail.

### 6.1 Findings on Application Level

Turning activity is one of the most important decisions made during wayfinding as almost all navigation errors occur at decision situations [1]. As a consequence, a large body of literature exists, which tries to propose solutions to either minimize the number of navigation

errors a wayfinder makes (e.g. [41]), or to minimize the impact of these errors by, for example, precalculating all possible routes close to the shortest path from the upcoming junctions [1]. A system which is capable to detect potential navigation errors would be able to pro-actively draw the wayfinders attention to this potential error and, thereby, help to avoid a wrong decision. Moreover, it has been long on debate that one of the potential drawbacks of wayfinding assistance systems, is their negative impact on humans spatial cognition and abilities (in long term) by constantly distracting users attention from the environment (see e.g. [2],[29], [39]). This detrimental effect is caused by either the need of constant visual interaction with the system, or by providing a more than necessary amount of information to the user. It has been suggested that gaze-based interactions in general can help limiting this distraction [15]. While researchers have sought for solutions to reduce the number of wayfinding instructions for years (e.g., [12]), our results allow for a personalized wayfinding assistance by monitoring gaze behavior for turning activity recognition and providing corrective instructions only if needed [32]. As an example, in case a wayfinder is supposed to continue straight ahead but shows a tendency to take a turn, the system can provide her/him with a corrective instruction; otherwise there is no need to distract the wayfinder by repeating the same instruction).

## 6.2 Findings on Theoretical Level

The findings of this paper in terms of *the role of each feature in visual search behavior during wayfinding decision situations*, can deepen our understanding of decision making in real-world wayfinding. Considering only gaze-based features (see left column of Figure 5), we observe two patterns: First, in total 15 out of 17 most important features are fixation-related. One potential explanation, which is in line with the fact that fixation-related features convey signs of attention and information processing [22], is that this dominance reflects the importance of *attention* at decision point. Prior evidence (see [3, 53]), suggests already that deciding to take a turn is a dynamic decision making process consisting of stimulus processing and action preparation which requires an attention shift. The second pattern we observe from Figure 5, relates to the equal number and importance of duration-based and dispersion-based features (note that the number of dispersion-related features are three times more than duration-based features and the same ratio holds here): Two fixation duration based features are among the 50% most important features, which is in line with the known positive correlation of fixation duration with cognitive processing and scene perception [21]. Inline with current evidence (see e.g., [22, 49]), these two patterns suggest that using only gaze information, the turn-activity decision is more reflected by features which are representative for cognitive processing and scene perception. Adding familiarity with the environment as a user-related feature (second left column of Figure 5), boosts the performance of the model and, at the same time, increases the number of saccadic features to six. Simultaneously, the importance of duration-based features (both fixation and saccade-related) slightly decreases. Saccadic features in general represent visual search activity rather than attention or concentration, and more specifically saccade amplitude, is known to reflect search task difficulty [36] and presence of high-frequency visual information [47]. This result can, hence, be interpreted as changing the importance of visual search in predicting a turn-activity decision. This would be in line with the fact that fixation behavior is particularly biased by familiarity or level of expertise [50]. Adding familiarity renders itself most important; moreover, it also redirects the model's attention towards saccadic features. This finding, though, leaves much room for further research: Arguably, familiarity somehow encapsulates most of the fixation behavior in itself, and lets the model seek for meaningful patterns in saccadic features. This

observation requires further assessment of familiarity in order to unfold the differences in gaze behavior of familiar vs unfamiliar wayfinders. Similar to adding user-related features, an increase in presence of visual search related features (i.e., saccadic features), occurs also when adding environmental factors (third left column of Figure 5). We used skewness of street segments and the number of branches an intersection has, as proxy for the difficulty of the decision situation in terms of urban configuration. Due to the collinearity of both features, only skewness is retained by the SHAP procedure. This may be an indicator for the fact that skewness is a better (i.e., more precise than the number of branches) proxy for the difficulty of the decision situation, with respect to environmental configuration. The most important pattern was revealed when we added both familiarity and environmental features (right column of Figure 5). As is shown in Figure 5, familiarity remains the most important feature, which can be explained by the fact that gaze-behavior is heavily biased by user-related factors, and particularly the level of expertise of which familiarity with the environment is an instance of. In addition to that, a slight increase in number *and* importance of saccadic features, and moreover, a sharp drop in importance for duration-based features arises. This observation may indicate that environmental and user-related features encode important aspects of the decision situation and, thereby, redirect the model's focus more to visual search behavior.

## 7 Conclusion and Future Work

In this paper we provide evidence that it is possible to classify the turning activity of wayfinders based on a combination of gaze-data, wayfinder familiarity (measured as a binary variable) and two environmental factors regarding the morphology of an intersection. To this end, we used data collected during an in-situ wayfinding experiment and compared different time windows and different classification algorithms. The highest accuracy was found for an XGBoost classifier, which achieved 91% overall accuracy in turn-activity classification for a window size of three seconds. We discussed our findings from two main perspectives: First, we shed light on the applicational prospects with respect to wayfinding assistance systems capable to detect potential navigation errors; second, we explored which gaze features are most important and provided explanations based on the existing literature. Taken together, these results raise further research questions. While we have deliberately used a basic set of gaze features combined with a single user-related and two basic environmental features, it would be interesting to see whether more sophisticated features (e.g., incorporating different levels of familiarity) convey even more information regarding the decision making process. Besides, according to our findings, adding user-related and environmental factors to gaze features, alters the pattern of fixation-related and saccadic features, in terms of the number and importance of each gaze-event category. It is, therefore, worthwhile to disentangle the relationship between these gaze features and the familiarity/environmental factors. Another interesting pattern which needs deeper exploration, relates to count-based fixation and saccadic features. So far, these have been used mostly as an indication of semantic importance and search-task difficulty [21]. Therefore, a future study can introduce semantic information of areas of interest (AOI) to the model, and observe why count-based features retain their importance in the model. Eventually, the differences in false positive/negative rates between different classes observed in the confusion matrix, may suggest missing environmental or user characteristics in the model. Therefore, a future study may for instance, incorporate head/body movements to the model to see if this behavioral source of information can lead to

better classification results. In addition to that, our findings open up several other questions regarding the general research direction of this paper:

1. When is the actual onset of the decision process for turn-activity, does this point in time vary for different phases of wayfinding (as proposed by [9]), and is this point in time stable as travel time/distance covered increases?
2. Why are some of the features more effective with respect to predicting particular classes of turn-activity? For instance, why is familiarity more effective on predicting non-turn activity as suggested by Figure 4?
3. Does the so-called cascade effect also hold for turn-activity recognition based on gaze-behaviour in real-world navigation?

---

## References

- 1 D. Amores, E. Tanin, and M. Vasardani. A proactive route planning approach to navigation errors. *Int J of Geographical Information Science*, pages 1–37, 2020.
- 2 A. Brügger, K. Richter, and S. Fabrikant. How does navigation system behavior influence human behavior? *Cognitive Research: Principles and Implications*, 4(1):5, December 2019.
- 3 T. T. Brunyé, A. L. Gardony, A. Holmes, and H. A. Taylor. Spatial decision dynamics during wayfinding: intersections prompt the decision-making process. *Cognitive Research: Principles and Implications*, 3(1), 2018.
- 4 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*, 33(4):741–753, 2011.
- 5 N. V Chawla, K. W Bowyer, L. O Hall, and W P. Kegelmeyer. Smote: synthetic minority over-sampling technique. *J of artificial intelligence research*, 16:321–357, 2002.
- 6 T. Chen, T. He, M. Benesty, V. Khotilovich, Y. Tang, H. Cho, et al. Xgboost: extreme gradient boosting. *R package version 0.4-2*, 1(4), 2015.
- 7 T. G Dietterich. Ensemble methods in machine learning. In *International workshop on multiple classifier systems*, pages 1–15. Springer, 2000.
- 8 W. Dong, H. Liao, B. Liu, Z. Zhan, H. Liu, L. Meng, and Y. Liu. Comparing pedestrians’ gaze behavior in desktop and in real Envs. *Cartography and Geographic Information Science*, 47(5):432–451, 2020.
- 9 R. Downs and D. Stea. *The world in the head Maps in minds: Reflections on cognitive mapping*. Harper & Row, 1977.
- 10 M. K. Eckstein, B. Guerra-Carrillo, A. T. Miller Singley, and S. A. Bunge. Beyond eye gaze: What else can eyetracking reveal about cognition and cognitive development? *Developmental Cognitive Neuroscience*, 25:69–91, 2017.
- 11 P. Fogliaroni, D. Bucher, Nikola J., and I. Giannopoulos. Intersections of our world. In *10th Int Conf on Geographic Information Science*, 2018.
- 12 A. U. Frank. Pragmatic information content—how to measure the information in a route. *Foundations of geographic information science*, page 47, 2003.
- 13 N. Gale, R. G. Golledge, W. C. Halperin, and H. Couclelis. Exploring spatial familiarity. *Professional Geographer*, 42(3):299–313, 1990.
- 14 I. Giannopoulos. *Supporting Wayfinding Through Mobile Gaze-Based Interaction*. PhD thesis, ETH Zurich, 2016. URL: <https://www.research-collection.ethz.ch/handle/20.500.11850/116375>.
- 15 I. Giannopoulos, P. Kiefer, and M. Raubal. Gazenav: gaze-based pedestrian navigation. In *Proc of MobileHCI 2015*, pages 337–346, 2015.
- 16 I. Giannopoulos, P. Kiefer, M. Raubal, K. Richter, and T. Thrash. Wayfinding Decision Situations: A Conceptual Model and Evaluation. In *Proc of GIScience 2014*, 2014.



- 17 K. Gidlöf, A. Wallin, R. Dewhurst, and K. Holmqvist. Using eye tracking to trace a cognitive process: Gaze behaviour during decision making in a natural Env. *J of Eye Movement Research*, 6(1):1–14, 2013.
- 18 L. Gokl, M. Mc Cutchan, B. Mazurkiewicz, P. Fogliaroni, and I. Giannopoulos. Towards Urban Env Familiarity Prediction. *Advances in Cartography and GIScience of the ICA*, 2(November):1–8, 2019.
- 19 A. Golab, M. Kattenbeck, G. Sarlas, and G. Giannopoulos. It’s also about timing! when do pedestrians want to receive navigation instructions. *SPAT COGN COMPUT*, 2021. accepted for publication.
- 20 E. H Hess and J. M. Polt. Pupil size as related to interest value of visual stimuli. *Science*, 132(3423):349–350, 1960.
- 21 K. Holmqvist, M. Nyström, R. Andersson, R. Dewhurst, H. Jarodzka, and J. Van de Weijer. *Eye tracking: A comprehensive guide to methods and measures*. OUP Oxford, 2011.
- 22 M. Adam Just and Patricia A. C. Eye fixations and cognitive processes. *Cognitive psychology*, 8(4):441–480, 1976.
- 23 A. Kafkas and D. Montaldi. Familiarity and recollection produce distinct eye movement, pupil and medial temporal lobe responses when memory strength is matched. *Neuropsychologia*, 50(13):3080–3093, 2012.
- 24 M. Kearns. Thoughts on hypothesis boosting. *Unpublished manuscript*, 45:105, 1988.
- 25 P. Kiefer, I. Giannopoulos, and M. Raubal. Using eye movements to recognize activities on cartographic maps. *Proc of SIGSPATIAL 2013*, pages 478–481, 2013.
- 26 P. Kiefer, I. Giannopoulos, M. Raubal, and A. Duchowski. Eye tracking for spatial research: Cognition, computation, challenges. *SPAT COGN COMPUT*, 17(1-2):1–19, 2017.
- 27 R. V Levine and A. Norenzayan. The pace of life in 31 countries. *J of cross-cultural psychology*, 30(2):178–205, 1999.
- 28 B. Li, L. Peng, and B. Ramadass. Accurate and efficient processor performance prediction via regression tree based modeling. *Journal of Systems Architecture*, 55(10-12):457–467, 2009.
- 29 H. Liao, W. Dong, H. Huang, G. Gartner, and H. Liu. Inferring user tasks in pedestrian navigation from eye movement data in real-world Envs. *IJGIS*, 33(4):739–763, April 2019.
- 30 B. Liu, W. Dong, Z. Zhan, S. Wang, and L. Meng. Differences in the gaze behaviours of pedestrians navigating between regular and irregular road patterns. *ISPRS Int J of Geo-Information*, 9(1), 2020.
- 31 S. Lundberg and S. Lee. A unified approach to interpreting model predictions. *Proc. of NIPS 2017*, page 4768–4777, 2017.
- 32 B Mazurkiewicz, M. Kattenbeck, and I Giannopoulos. Navigating your way! increasing the freedom of choice during wayfinding. In *11th International Conference on Geographic Information Science (GIScience 2021) - Part II*. Schloss Dagstuhl–Leibniz-Zentrum für Informatik, 2021.
- 33 N. Venkata Kartheek Medathati, R. Desai, and J. Hillis. Towards inferring cognitive state changes from pupil size variations in real world conditions. In *Proc of ETRA 2020*, 2020.
- 34 I Nitze, U Schulthess, and H Asche. Comparison of machine learning algorithms random forest, artificial neural network and support vector machine to maximum likelihood for supervised crop type classification. *Proc. of the 4th GEOBIA*, 35, 2012.
- 35 J. L. Orquin and S. Mueller Loose. Attention and choice: A review on eye movements in decision making. *Acta Psychologica*, 144(1):190–206, 2013.
- 36 M. H Phillips and J. A Edelman. The dependence of visual scanning performance on saccade, fixation, and perceptual metrics. *Vision research*, 48(7):926–936, 2008.
- 37 L Piccardi, M Riseti, and R Nori. Familiarity and Enval Representations of a City: A Self-Report Study. *Psychological Reports*, 109(1):309–326, August 2011.
- 38 R. Pieters and L. Warlop. Visual attention during brand choice: The impact of time pressure and task motivation. *Int J of Research in Marketing*, 16(1):1–16, 1999.

- 39 C. A. Rothkopf, D. H. Ballard, and M. M. Hayhoe. Task and context determine where you look. *J of Vision*, 7(14):1–20, 2007.
- 40 A. Rousell and A. Zipf. Towards a landmark-based pedestrian navigation service using OSM data. *ISPRS Int J of Geo-Information*, 6(3), 2017.
- 41 R. A. Ruddle, E. Volkova, B. Mohler, and H. H. Bülthoff. The effect of landmark and body-based sensory information on route knowledge. *Memory & cognition*, 39(4):686–699, 2011.
- 42 D. D. Salvucci and J. H. Goldberg. Identifying fixations and saccades in eye-tracking protocols. In *Proc of ETRA 2000*, page 71–78, 2000.
- 43 V. Schnitzler, I. Giannopoulos, C. Hölscher, and I. Barisic. The interplay of pedestrian navigation, wayfinding devices, and Enval features in indoor settings. In *Proc of ETRA 2016*, pages 85–93, 2016.
- 44 Y. Shiga, T. Toyama, Y. Utsumi, K. Kise, and A. Dengel. Daily activity recognition combining gaze motion and visual features. *Adjunct Proc of UbiComp 2014*, pages 1103–1111, 2014.
- 45 S. Shimojo, C. Simion, E. Shimojo, and C. Scheier. Gaze bias both reflects and influences preference. *Nature Neuroscience*, 6(12):1317–1322, 2003.
- 46 N. Stewart, F. Hermens, and W. J. Matthews. Eye Movements in Risky Choice. *J of Behavioral Decision Making*, 29(2-3):116–136, 2016.
- 47 B. W. Tatler, R. J. Baddeley, and B. T. Vincent. The long and the short of it: Spatial statistics at fixation vary with saccade amplitude and task. *Vision research*, 46(12):1857–1862, 2006.
- 48 J. Treboux, D. Genoud, and R. Ingold. Decision tree ensemble vs. nn deep learning: efficiency comparison for a small image dataset. In *Proc. of the IWBI 2018*, pages 25–30, 2018.
- 49 P. Unema. Differences in eye movements and mental work-load between experienced and inexperienced motor vehicle drivers. *Visual search*, pages 193–202, 1990.
- 50 R. von Stülpnagel. Gaze behavior during urban cycling: Effects of subjective risk perception and vista space properties. *Transportation Research Part F: Traffic Psychology and Behaviour*, 75:222–238, 2020.
- 51 Q. Wang, P. Cavanagh, and M. Green. Familiarity and pop-out in visual search. *Perception & Psychophysics*, 56(5):495–500, September 1994.
- 52 F. Wenzel, L. Hepperle, and R. von Stülpnagel. Gaze behavior during incidental and intentional navigation in an outdoor Env. *SPAT COGN COMPUT*, 17(1-2):121–142, 2017.
- 53 J. M. Wiener, O. de Condappa, and C. Hölscher. Do you have to look where you go? Gaze behaviour during spatial decision making. *Proc of CogSci 2011*, pages 1583–1588, 2011.
- 54 J. M. Wiener, C. Hölscher, S. Büchner, and L. Konieczny. Gaze behaviour during space perception and spatial decision making. *Psychological Research*, 76(6):713–729, 2012.
- 55 A. L. Yarbus. *Eye movements and vision*. New York: Plenum Press, 1967.