Machine Learning Analysis of Pedestrians' Hazard Anticipation from Eye Tracking Data

Andreas Gregoriades[1], Loukas Dimitriou[2], Maria Pampaka[3], Harris Michail[1], and Michael Georgiades [4]

[1] *Cyprus University of Technology, Limassol, Cyprus*
[2] *University of Cyprus, Nicosia, Cyprus*
[3] *The University of Manchester, Manchester, UK*
[4] *Neapolis University of Pafos, Pafos, Cyprus*

Abstract

Pedestrian tourists are considered the most vulnerable road users of urban mobility environments. Tourists are a special category of pedestrians, exhibiting different visual behaviour to residents due to their enthusiasm and unfamiliarity with the environment. These characteristics of pedestrian tourists influence their hazard perception. Eye tracking technology became popular in investigating pedestrian safety problems after findings that eye-gaze behaviour is linked with human attention and hazard anticipation. The majority of eye-tracking studies to date use stationary technology that may miss out important properties relating to environmental dynamics that cannot be accurately simulated. This study employs a novel method utilising mobile eye-tracking technology in naturalistic settings to investigate the application of machine learning in identifying differences between tourist and resident pedestrians' visual behaviour. Eye tracking metrics are used to train an Extreme Gradient Boost (XGBoost) model to examine whether tourists have less hazard perception than residents when visiting destinations with opposite driving conventions to their own. Preliminary results with a small group of tourist and resident pedestrians demonstrate how such machine learning models could be used in real-time by agent-based systems that utilise wearable augmented reality displays to support hazard perception of tourist pedestrians.

Keywords [1]

Pedestrian safety, Mobile eye tracking, XGBoost classification, Wearable Augmented Reality Displays.

1. Introduction

Pedestrians constitute 22% of all road traffic fatalities worldwide [1]. Tourists represent a vulnerable category of road users due to their unfamiliarity with the environment and traffic rules at destinations they visit [2][3]. These, in combination with their curiosity and enthusiasm for exploration, reduce their hazard perception making tourists more vulnerable to accidents.

Hazard perception refers to the anticipation of traffic hazards, is a critical component of road safety and is directly linked to pedestrians' and drivers' visual behaviour [4]. Novice road users in comparison with experienced road users are less effective in anticipating safety-relevant traffic events [5][6] thus a need to assist novice users based on experts' knowledge is required [4]. Tourists visiting countries with

different road conventions compared to those in their origin country are considered as novice road users since they are not familiar with neither the environment, nor the driving rules. Research indicates that 54% of tourists experience problems crossing the road as pedestrians when they are unfamiliar with a country's road convention [7]. A recent driver safety study attributed many accidents involving tourist drivers to attentional factors, such as increased cognitive workload and reduced hazard perception while adapting to new traffic environments (i.e. finding out where to look at and from where to expect incoming cars at intersections) [8]. Like drivers, to maintain their safety when crossing roads, pedestrians have to process different information from the environment such as the road infrastructure characteristics, traffic density, incoming vehicles direction, sounds, other pedestrian movements and other visual distractions such as illuminated advertisements [9]. Despite this variety of factors that influence pedestrian safety most studies use either surveys [10], or unrealistic synthetic environments through simulated experiments to draw conclusions. These approaches, however, might miss important information [11] and lead to inaccurate conclusions. Recently, eye tracking technology has emerged as a promising method for analyzing safety by examining humans' visual behaviour and head movements. This technology enables capturing richer information relevant to pedestrians' safety after evidence linking eye movements with attention. Therefore, what we are looking at often corresponds to what we are attending cognitively. Mental attention is a cognitive resource that is limited and is consumed when processing visual information. Changes in cognitive effort are manifested through changes in the physiology of the human eye such as pupil dilations that regulate the amount of light that enters the eye. Pupils' dilations have also been linked to cognitive factors such as workload, surprise, attention, emotional arousal [12] and hazard anticipation[13]. Therefore, eye tracking data can be used to reveal information about mental processes, which are not easily accessible through behavioural performance measures alone. Most eye tracking studies, however, use stationary eye tracking in a lab with a computer screen as the visual scene to be analysed. Such artificial settings suffer from low realism.

In this study, we use mobile eye tracking equipment (i.e., Tobii glasses) to investigate differences between two groups of participants (i.e., residents and tourists) during a road crossing scenario and use eye tracking metrics to train a machine learning (ML) model to analyse hazard perception. This is a continuation of our previous work in naturalistic eye tracking [14]. The main aim of the study is to investigate how the visual behaviour of resident and tourist pedestrians differs when it comes to hazards and how tourists hazard anticipation can be enhanced in real time using wearable technology and tacit knowledge (visual behaviour for improved hazard perception) from expert road users. Due to the large volume of data generated by eye tracking equipment the use of ML is utilised in this study to assist in automating the hazard perception support of pedestrians. A popular classification technique is used, namely, XGBoost, to train a model and evaluate its patterns to identify differences between tourists and residents. Such models can be utilised by autonomous agents in transportation networks to evaluate the state of pedestrians' hazard anticipation in real time.

The study proposes the use of Wearable Augmented Reality Displays (WARD) equipped with eye tracking capability to assist in pedestrians' hazards anticipation. WARD can overlay the physical world with digital content and thus can provide users with important safety information. They have been recently used to enhance the situation awareness of highway workers[15] and in other applications such as healthcare [16]. The analysis highlights ways to enhance tourist safety by analyzing eye gaze, eye pupil and head movements data through ML techniques and the use of such models in prospective multi-agent systems that would utilise WARD and Fog/Edge computing to process and store data [17], and make inferences about hazard anticipation of pedestrians.

The paper is organized as follows. The next section describes the literature relating to workload, hazard anticipation and use of eye tracking techniques for their assessment. This is followed by a section describing the methodology followed by the data preprocessing and training of an XGBoost model. The paper concludes with preliminary results, a discussion and conclusions.

2. Literature Review

Pedestrian safety countermeasures usually include infrastructural changes or policies [2]. These, however, are expensive and time consuming. Alternatively, technological developments such as Intelligent Transportation Systems that integrate information and communication technologies within

the transportation infrastructure can provide real time information to road users through software agents running on the Fog/Edge of a computer network, offering time-sensitive and location-based services for autonomous vehicles or WARD agents as in our case [31].

Currently, in-vehicle information systems are becoming popular means for improving driver hazard perception, and the same principle could be applied for pedestrians. Such systems are designed based on knowledge from safety literature highlighting that visual and auditory clutter overloads drivers and pedestrians, and in combination with distractions (i.e., advertisements, cell phones, in-vehicle conversations, billboards, etc.) can interfere with visual search strategies [19] such as, where to focus attention to infer information necessary to maintain safety and anticipate risk. These, in combination with traffic rules and environment unfamiliarity, can increase the risk of accidents [20]. This risk becomes critical when pedestrians are crossing roads or engage with new road infrastructure where hazard perception is key [21]. Tourists are more vulnerable to overloading due to unfamiliarity with the environment and the driving conventions which decreases their hazard perception [3].

Cognitive workload and hazard perception of pedestrians can be measured in different ways. Methods are usually categorized into subjective, performance-based, and physiological. Subjective techniques, including surveys, are commonly used in safety research, with NASA task load index being a popular option. Performance-based measures are usually classified in terms of primary and secondary task performance where users are engaged in a secondary task to the main task and through its assessment the level of workload can be inferred, since it shows user's spared cognitive capacity; examples of such measures include vehicle lane departures, lateral deviations, task completion time, reaction time, accuracy, and error rates, with poor performance indicating that a driver is overloaded. Physiological measures encompass audiology, cardiovascular, respiratory, neurophysiology, and ophthalmic physiology [22]. The latter refers to metrics of pupil dilations, and is the method used in this study. Physiological methods are advantageous, because they can assess workload and hazard perception in real-time since increased workload or anticipation of hazard evoke small and involuntary fluctuations in pupil dilation due to the attentional demands imposed by a cognitive task [23].

Pupil dilation is considered a reliable and valid psychophysiological measure of the amount of cognitive effort devoted to a given visual stimuli and a suitable indicator for hazard perception [24]. Techniques for assessing workload using pupillometry include the task-evoked pupillary response [25] and the index of cognitive activity (ICA) [12]. Both refer to variations in dilation as reactions to cognitive processing. Hazard perception can be assessed in addition to pupillometry using metrics such as fixation count on certain areas of the visual scene (i.e., Areas of interest), fixation duration, time to first fixation, fixation heat map, scan-paths, etc.

When it comes to designing solutions to tackle safety, eye tracking insights can be used for specifying and validating requirements in of prospective systems, with methods such as the one presented in[18] demonstrating the optimization of an information system's user interface based on knowledge extracted from eye tracking experiments.

2.1.   Eye tracking and Visual behaviour

Visual attention refers to the cognitive processes that guide the selection of relevant information from visual scenes and the filtering out of irrelevant information. Numerous studies agree that where we direct our eyes, and for how long often correlate with attentional selection and information processing [26]. Eye-tracking technology has gained popularity in different safety and consumer-related disciplines, due to its ability to provide accurate information on visual attention of participants in experiments. Attention is expressed in the form of fixations (i.e., periods when eyes are relatively still and the visual system absorbs information about what is being looked at) and visualization patterns represented by saccades, which refer to fast eye movements between stimuli.

Most eye-tracking studies assume that humans process the visual information they focus on [27], also known as the eye-mind hypothesis [28] with Hyönä [29] reporting this is true when the visual environment in front of our eyes is relevant to the task we are about to perform.

Eye tracking technology can be used alone or in combination with other physiological measures such as electrooculography. Alone they can measure participants' attention in a visual environment through 1) eye fixations in Areas of Interest (AOI) that denote important areas in a visual scene (e.g.,

hazard perception indicators in this study), 2) saccadic behaviour that indicate visual exploration or confusion, 3) goal seeking behaviour through scan-path analysis of transitions of fixations between AOIs, and 4) head movements and acceleration in three axes as used in this study.

Eye tracking data is obtained on a millisecond basis therefore millions of observations can be recorded which makes the analysis of such raw data difficult. Eye tracking software can provide support for this process, but this is limited to generic features that might not fit the needs of the researcher/user.. Therefore, researchers use ML to analyse raw eye tracking data to generate new features from the data. Example ML applications on eye tracking data include classification problems from different domains including safety, psychology and systems design. No work however has been reported on using mobile eye tracking with ML on pedestrian hazard perception.

On the contrary the applications of eye tracking in safety related studies mainly used stationary apparatus with static imagery [30]. Mobile eye tracking studies are limited due the complexity of analyzing the data and difficulty in controlling confounding variables but can collect richer data and can be used to explain pedestrian safety issues in a more holistic manner. Alternative methods to eye tracking include pedestrian observation from cameras or pedestrian simulators with limited pedestrian movement [11]. The latter though suffers from a limited level of realism that could lead to biased conclusions.

Recent efforts to improve pedestrian safety focus on increasing pedestrians' situation awareness/hazard perception through assistive wearable technologies and sensors. Such systems identify potential dangers and provide warnings. Eye tracking results from this study can be used to infer the information needs of pedestrians and thus assist when designing new pedestrian safety technologies such as wearable augmented reality glasses by minimizing the design space exploration of potential technological solutions. However, more in depth understanding of such designs can be achieved when both simulations and naturalistic visual behaviour analyses are combined, as in this study. The latter can explore the problem in its natural settings to identify main requirements which can be then refined and evaluated through simulations.

3. Methodology

This study proposes a novel analytical framework for pedestrian safety analysis based on explainable ML models with mobile eye tracking data extracted from a road crossing experiment with six participants in naturalistic settings. The technique used to assess hazard perception is based on pupillary variations when fixating in AOI linked to hazardous areas in the visual scene. The method consists of six steps: (1) designing the experiment, selection of road section and specifying the hazard AOIs; (2) participant selection, familiarization with equipment and procedure, conducting the experiment and collecting the data; (3) selecting the set of eye movement variables from the raw data that are relevant to hazard perception and cognitive workload; (4) pre-processing data associated with hazard AOIs; (5) Train an XGBoost classifier to identify patterns that explain the link between the selected variables and the target variable (tourists or resident pedestrians); and (6) interpreting the learned XGBoost model using an explainable ML method to highlight differences among the two groups.

3.1. Participants and Procedure

The experiment used a targeted sample of participants. Six healthy participants (3 male and 3 female) with an average age of 35 years (standard deviation of 7.19 years) participated in this study. Three participants were tourists from a country with opposite driving rules to Cyprus and the other three were residents of Cyprus familiar with the rules and area where the experiment was conducted. The purpose and procedure of the experiment were explained to participants in advance, and it was made clear that they could abort the experiment at any time. Participants were given enough time to familiarize with the Tobii glasses prior to undertaking the experiment. The eye-tracker was calibrated for each participant prior to the experiment. The selection of the experiment's location was based on tourist visitation data obtained from local authorities. The road section (Figure 1) includes a one-way road enclosed by buildings (occluded) with no visibility of incoming vehicles from side roads until the pedestrians reached the intersection points (kerb), to avoid providing any advance information to

participants regarding imminent vehicles and hence test their vigilance level and visual behaviour at kerb (K). The road section had no clear indication of incoming vehicle direction, hence helped to convey visual scanning behaviours from habitual knowledge of participants from different driving conventions, that might expect incoming vehicles from specific direction. To minimise the effect of confounding variables such as different traffic flow for each participant, the experiment was conducted when the road section under study was closed for planned road works (2 days) but not at the points of interest and not visible by the participants. Neither the residents nor the tourists were aware of this closure so as not to influence their visual behaviour. Participants engaged the pre-specified route of Figure 1, with a free walk scenario (observe stimuli as they do normally), crossing the road at the point shown in the figure. The flow of vehicles in this road section is indicated with the blue arrows.
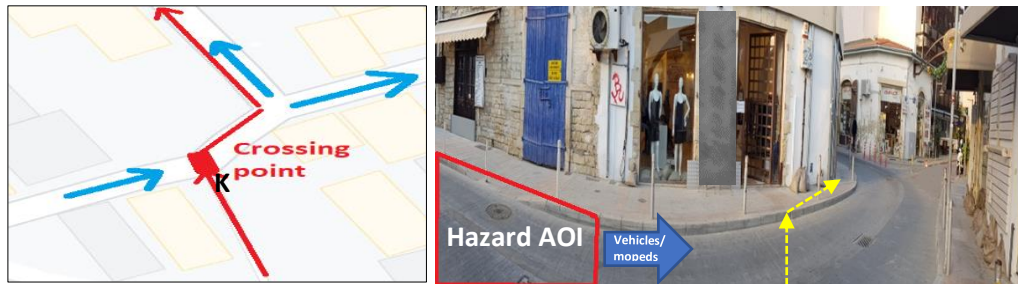

Figure 1. Left: Road section and path that participants had to follow (red arrows, allowed vehicles flow in blue arrows, kerb indicated by the letter K). Right: Wide angle view of the infrastructure on the kerb with the hazard AOI as an overlay polygon and the path participants followed in dotted arrows. The name of the shop has been blurred for anonymity purposes(gray rectangle).

3.2.   Extracting patterns from XGBoost models

Data from the experiment were used to train a ML model that allowed the automated identification of patterns from the large amount of eye tracking data. ML techniques are classified into supervised and unsupervised techniques, the former requiring labelled data while the latter not, and they are applied into two broad types of problems, namely classification (e.g., predicting the probability that an input set of data is associated with a categorical output variable) and regression (e.g., predicting the value rather than a category) problems. ML techniques have been applied on eye movement data from cognitive science and other domains to classify human performance on cognitive tasks. In ML, features correspond to measurable properties that best characterize the problem under study.

The classification technique employed in this study is used to classify participants based on eye tracking data into tourist or residents and in this way identify patterns that characterize each group. For this task a binary Extreme Gradient Boost (XGBoost) [32] model was developed in Python. XGBoost is a newer version of the gradient boosting decision tree model and has been extensively used in various problems including human performance due to its excellent performance by preventing overfitting with regularization and computational efficiency. XGBoost is an ensemble method since it combines multiple classification and regression trees, each composed of several nodes that represent variables in the dataset. During XGBoost, multiple decision trees are trained, with each tree built based on the result of the previous developed tree.

The binary XGBoost was trained to predict hazard anticipation using as features the visual behaviour data of participants in the road crossing experiment. In this work, for each participant several observations were made by the eye tracker regarding participants' eyes and head movements every millisecond. These were set as initial features of the classifier and were refined afterwards through feature selection while developing the model.

Although ML algorithms have been proven effective in prediction problems, when it comes to the interpretation of their results they are classified into white and black box techniques [33]. Black box techniques such as the XGBoost ensemble method used in this work and deep neural networks, can produce better results in terms of performance but provide little insights into how they come up with the outcome. Thus, they suffer from low interpretability. This is important in situations such as the one

presented in this study where the reason that cause low hazard anticipation is key to specifying the requirements for future solutions to this problem.

A popular method for black box explanation used in this study is the SHAP (SHapley Additive explanation) technique that is based on cooperative game theory [34], and allows black box ML models to explain the impact of input variables on the model's outcome. SHAP is used to extract the patterns of the trained XGBoost model.

3.3. Outcome variable and data labelling

An XGBoost classifier is trained in this study with outcome class variable being the tourist/resident property of participants and predictors their eye tracking observations. The class variable was annotated using the Tobii Pro Lab software based on the tourist/resident status of participants.

Independent variables or features used to train the model included eye movement parameters relevant to cognitive effort and hazard anticipation based on the literature. The initial features were developed using statistical properties of raw eye tracking variables such as mean and standard deviation of fixation coordinates (on 2 dimension space), pupillary and head movement data. Pupillary data was selected for fixations data points only, to account for points in the visual scene that participants attended cognitively. Additional features were the average duration values for fixations, and their normalized metrics by dividing the total duration of each parameter by the total time participants took engaging the intersection. Other pupillary features were the standardised pupillary score and the pupillary moving standard deviations. Head movement data was collected from the glasses' gyroscope readings (degrees/s) with regards to lateral and vertical head movements (yaw, pitch).

As a hazard perception feature we considered the successful fixations of participants in an area of the road from where potential hazards could occur [35], and these have been prespecified by the researchers based on incoming vehicles, bicycles and mopeds direction. Fixating on the area where a hazard may occur does not always indicate that the hazard has been anticipated, however, studies [36] have shown that most glances in that area occur because the road user is anticipating the hazard. Thus, the hazard anticipation variable was determined from pedestrians' fixations on specific intersection areas and was coded from the recorded eye tracker videos and the raw data that were collected by the device. Specifically, to code the hazard AOIs a predetermined hazard zone was defined on the visual scene coordinates, that refer to the position of incoming vehicles (Figure 1). Anticipatory glances were labelled accordingly based on participants' fixations in these hazard AOIs. Tobii Pro Lab software allows the mapping of eye gaze/fixation data onto still images (2D) such as snapshots of the environment under study (road crossing in our case) and thus hazard hits could be assessed using AOI coordinates on this image.

3.4. Data pre-processing

Raw data from the eye tracker consisted of multidimensional time series data of all variables of interest (262K rows of data tuples for all participants during their interaction with the road crossing scenario) collected approximately every 20 milliseconds (50Hz sampling rate) and containing information about, horizontal, vertical head movements, head accelerometer data, fixation coordinates on x and y axes, fixation duration, eye gaze direction on x, y axes, eye gaze coordinates, saccades, AOI (Hazard areas) hits, pupil diameter, and pupil assessment confidence (indicator of the eye tracker's level of confidence it is correctly measuring the pupil). For the road crossing scenario only tuples that fall within the time interval in which pedestrians engaged with the road crossing scenario were selected, resulting in 62K observations from the initial dataset, thus eliminating irrelevant data prior to participants entering the scenario zone. The time intervals were specified in the Tobii pro lab software prior to extracting the raw data. Tuples' timestamps were used for filtering the eye tracking data.

Additional filtering activities were performed to ensure sufficient eye tracking quality. Thus, all data points with low pupil confidence (estimated by the eye tracker) were removed. To normalize the data, the z score, mean and median values were extracted and used as additional features during the ML training, to find the best fitting for the problem. Raw data pre-processing was performed in Python.

To identify differences in hazard perception and workload level among the two groups, the number of variations in pupils' diameter in the timeseries was used similarly to [12]. The points of interest in the pupils' time-series data refer to temporal sections with increases/decreases of pupil's diameter which indicate participants' cognitive processing due to hazard anticipation. We hypothesize that, since the pupillary data selected for processing are the ones associated with hazard AOIs, the pupillary variations are a response to hazard perception for those AOIs. To achieve this the moving standard deviation approach is employed with a sliding window on the pupillary timeseries (Figure 2). This was essential to minimise noise and highlight pupilar variations relevant to hazard perception. Thus, the higher the pupillary variations the higher the hazard anticipation. The size of the sliding window was identified after trying a range of window sizes until the points of high variability became apparent and the trained ML model produced best predictive performance.
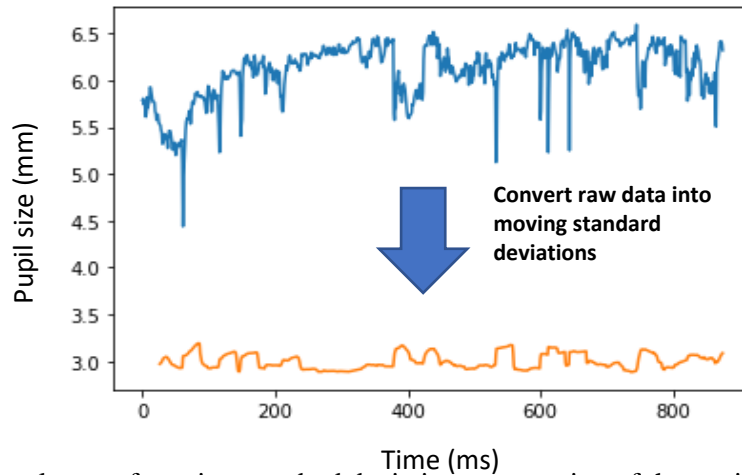

Figure 2. Example use of moving standard deviation on a portion of the pupil dilation data of one participant. Top series(blue) refer to raw data and bottom to moving standard deviations used later as one of the features in the XGBoost model training.

Head movement data was also collected corresponding to gyroscope measures, yaw-movement (turning of the head sideways), pitch-movement (nodding the head up/down), roll-movement (tilting the head to the side). Positive yaw values denote high lateral head variability, when participants were turning their heads at the kurb possibly to scan for hazards. A positive slope in the yaw data denotes turning the head to the right and a negative slope turning to the left. The accelerometer measures the acceleration of the head unit and denotes the speed with which the head turns due to surprise. Acceleration is measured along three axes in the head unit coordinate system in m/s^2. Head movement data was also essential in assessing participants hazard perception.

3.5. Training the XGBoost model

The XGBoost model was trained using the data described in the previous section. During XGBoost training, firstly, we specified the classifier's performance metric to optimize. In this case, since the output variable is binary (tourist/resident) and we wanted to maximise the performance of the model in predicting both states of the class variable while also addressing class imbalances, the metric we used was the Area Under the Curve (AUC). During model learning the data was split into training and testing sets. A 70/30 stratified training-test set split, meaning that 70% of data was used to train the model and 30% was used to validate the model's accuracy after training. This split was stratified, which means that both the train and test set maintained a roughly equal proportion of data of both classes (tourists, residents). The trained XGBoost model achieved an AUC of 0.89 which demonstrates the predictive performance of the model.

During model optimization, hyperparameter tuning was performed using an exhaustive search approach (GridSearch). The best performing model is selected based on the hyperparameters and features that maximise AUC. Feature selection is an important step in training a model in ML that evaluates the importance of each feature (eye tracking variable) on classifiers performance. During

feature selection several variables were eliminated using the ANOVA F-test and the Scikit-learn SelectKBest function.

4. Results

Since XGBoost is not considered an interpretable model, it was imperative to use the SHAP technique [34] that visualizes the contribution of each input feature. A SHAP value is assigned to each of the model's features based on its marginal contribution to the model's output.

The SHAP summary plot depicted in Figure 3 shows the features based on their importance in classifying tourists/residents. The red color presents a larger value of the feature, while the blue indicates a smaller value. The horizontal axis represents the SHAP value of each data point. A positive SHAP value on X-axis demonstrates the increase in probability (log odds) that the data belong to a tourist rather than a resident participant, while negative values imply that the subject is a resident. This diagram enables us to visualize the relationships between feature values (i.e., red/blue/purple dots on the horizontal lines next to each feature representing the intensity of the feature value) and their associations with the output variable(resident/tourist). The impact of each feature on the output variable can be examined further using the dependency plots depicted in Figure 4 where feature's values is plotted on the x-axis and the SHAP value of the feature on the y-axis. Dependency plots enable analysts to drill down into each feature to examine how the feature interacts with the output variable and with other model variables.
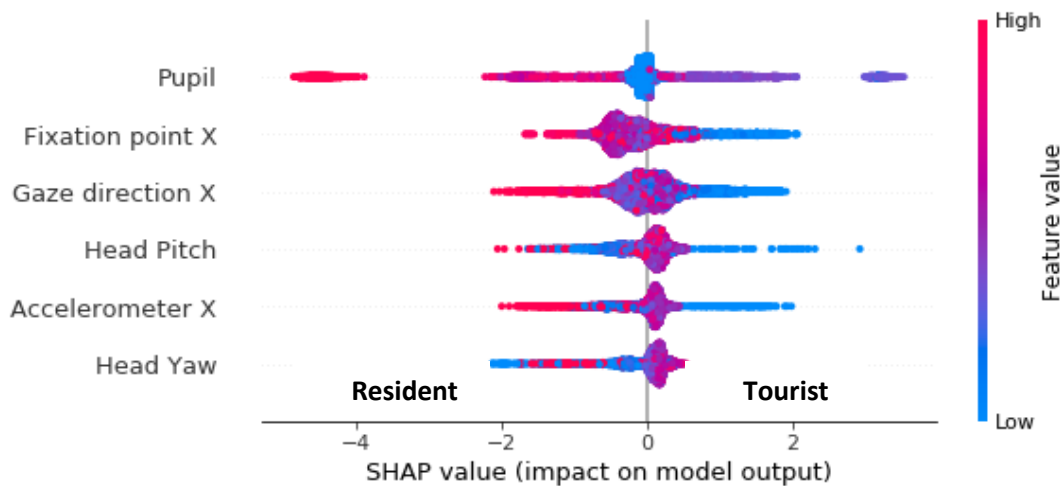

Figure 3. SHAP summary plot with features on Y axis and probability of belonging to either state of the target variable on x axis

From the summary plot it can be observed that pupil variability (Pupil feature in figure 3) is the most influential feature in the model for the data associated with the hazards' AOI under investigation. Increased pupil variability (indicated with red colour observations) is associated more with residents. This indicates that the residents are more aware of the road hazards at the intersection, and they are anticipating them. On the contrary, tourists show low variability on the hazards AOI. Cyprus follows the right-hand driving convention while all tourist participants were familiar with the left-hand drive convention. This difference in prior knowledge resulted in tourists failing to attend the safety relevant areas when crossing the road which highlighted the need for information support in the form of warning when incoming vehicles or mopeds are electric and thus silent.

With regards to fixations on the x-axis (with [0,0] being the upper leftmost point on the coordinate system, the values of x refer to x after subtracting the mean value of x), since high values of x are associated more with residents, this denotes that residents were fixating more on the right side of the intersection that was associated with the path that they had to follow according to Figure 3. High lateral head movement (Yaw) and high head acceleration (Accelerometer X) were associated more with residents than tourists which indicate that residents were more vigilant and were scanning the area for hazards with rabid lateral head movements, in contrast to tourists that scanned the area less. Similarly

high head pitch movements were associated more with residents than tourists that pitch with less intensity but more often possibly to regularly check the surface where they were walking.

A drill down analysis of the summary plot is performed using the dependency plots of Figure 4. This show the link between pupil variability and class variable (resident/tourist on Y axis), in the bottom chart. Thus, when the pupil variability increases, the odds of being a resident pedestrian also increase, which denotes that those resident participants have better awareness of the hazard since they attend and process the hazard AOI more. The dependency plot also show the interactions between Yaw movement and pupil variability that confirms that residents are more vigilant with more lateral head movements. This top chart in Figure 4 show that high pupil variability(bright red dots as opposed to blue/purple) is associated more with lateral head moves for resident pedestrians rather than tourists.
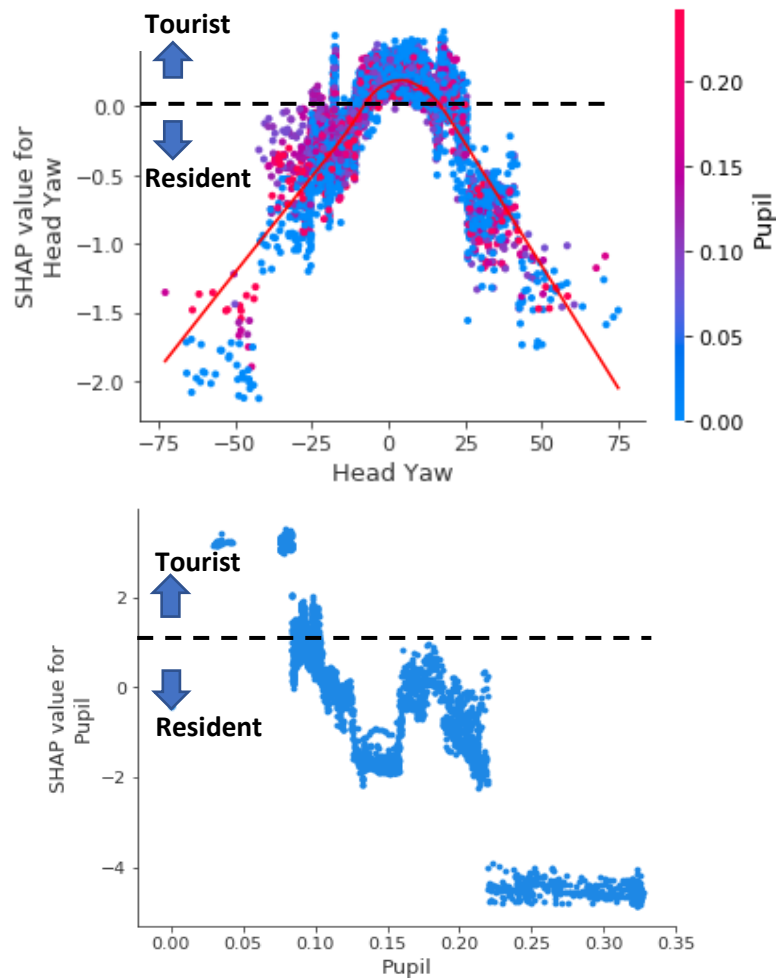

Figure 4. Dependency plots of pupil variability(bottom) with class variable's odds on y axis (resident/tourist), and lateral head movements(top) with pupil variability and class variable (top). The dotted line distinguish residents from tourists scores.

5. Discussion and Implications

These findings provide novel insights to understand the heterogeneous impacts of pupil and head movement on hazard perception using the SHAP method. The results point to the conclusion that tourists' hazard anticipation needs to be supported. Different techniques to alert pedestrians of imminent threats include the use of auditory, vibration and visual cues. The former two fail in accurately orienting the attention of pedestrians on the direction of the risk, thus the use of visual cues in WARD is recommended.

Traditional approaches to pedestrian safety challenges, such as infrastructural changes, are expensive and sometimes less desirable. For instance, the "look left/right" road markings in pedestrian

crossings can warn the pedestrians but such signs do not exist in all road sections, plus they warn only if the pedestrians attend this information. WARD can resolve these issues through dynamic virtual signs that attract attention and could be part of a multi-agent system. Such intelligent multi-agent transportation systems can integrate information and communication technologies within the transportation infrastructure to provide real-time information to road users through software agents running on the Fog/Edge of a computer network, offering time-sensitive and location-based services for autonomous vehicles or WARD agents as in our case [31]. A conceptual model of a multi-agent system that could be utilised in the scenario presented in this study is depicted in Figure 5. WARD agents communicate with intersection agents on fog/edge servers that can store and process real time data from the WARD agents and make predictions (using trained ML models) of the likelihood that pedestrians will miss a hazardous event and accordingly provide them with warnings. WARD agents will have data collection and visualization functionalities. The training of the ML models could be performed on the edge/fog using as training data the eye tracking/head movement observations of all pedestrians, their geolocation and road infrastructure scene properties (from WARD) using object detection capabilities as proposed by [16]. Data can be labeled based on pedestrians' status (expert residents or novice/tourists).
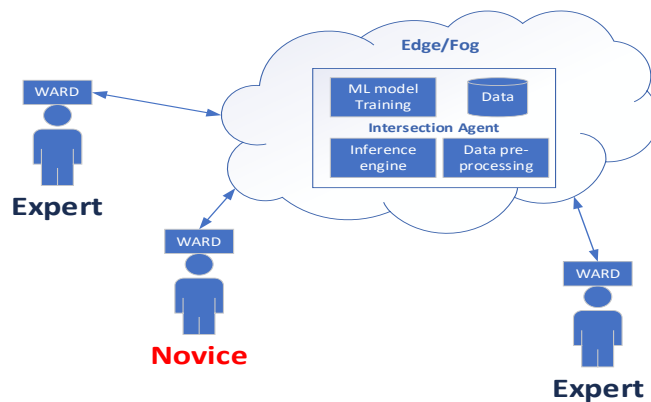

Figure 5. Conceptual model of a multi-agent system for pedestrian hazard perception support

6. Conclusions

This work proposes an analytical method for evaluating the hazard perception of pedestrians using mobile eye tracking in naturalistic settings. This work constitutes one of the few studies that use mobile eye tracking in naturalistic settings for pedestrian hazard perception. The study verifies the hypothesis that resident pedestrians attend more safety-relevant information than tourists and highlights important information-needs of tourists compared to residents in order to maintain adequate hazard perception.

The results of this study provide evidence of several issues with regards to tourist pedestrian safety, that points out the need for hazard anticipation support using WARD and trained ML models that utilise eye gaze and head movement observations. The study proposes the assessment of pedestrians' hazard perception in naturalistic settings through the utilization of a multi-agent system and fog/edge computing. Future directions, include the simulation of a prospective muti-agent WARD system architecture using a pedestrian scenario to verify its utility and response time performance.

References

[1]   WHO, *Pedestrian safety: a road safety manual for decision-makers and practitioners*. 2013.

[2]   J. Wilks, B. Watson, and I. J. Faulks, "International tourists and road safety in Australia: Developing a national research and management programme," *Tour. Manag.*, vol. 20, no. 5, pp. 645–654, 1999.

[3]   L. (Don) A. N. Dioko and R. Harrill, "Killed while traveling – Trends in tourism-related

mortality, injuries, and leading causes of tourist deaths from published English news reports, 2000–2017 (1H)," *Tour. Manag.*, 2019.

[4]   K. Shimazaki, T. Ito, A. Fujii, and T. Ishida, "Improving drivers' eye fixation using accident scenes of the HazardTouch driver-training tool," *Transp. Res. Part F Traffic Psychol. Behav.*, vol. 51, pp. 81–87, 2017.

[5]   M. J. Lazaro, M. H. Yun, and S. Kim, "Stress-level and attentional functions of experienced and novice young adult drivers in intersection-related hazard situations," *Int. J. Ind. Ergon.*, vol. 90, p. 103315, 2022.

[6]   A. Meir and T. Oron-Gilad, "Understanding complex traffic road scenes: The case of child-pedestrians' hazard perception," *J. Safety Res.*, 2020.

[7]   FCO, "UK citizens and road accidents abroad," 2008.

[8]   C. Thompson and M. Sabik, "Allocation of attention in familiar and unfamiliar traffic scenarios," *Transp. Res. Part F Traffic Psychol. Behav.*, 2018.

[9]   H. Tapiro, T. Oron-Gilad, and Y. Parmet, "The effect of environmental distractions on child pedestrian's crossing behavior," *Saf. Sci.*, 2018.

[10] E. Papadimitriou, S. Lassarre, and G. Yannis, "Human factors of pedestrian walking and crossing behaviour," *Transp. Res. Procedia*, vol. 25, pp. 2002–2015, 2017.

[11] S. Kalantarov, R. Riemer, and T. Oron-Gilad, "Pedestrians' road crossing decisions and body parts' movements," *Transp. Res. Part F Traffic Psychol. Behav.*, vol. 53, pp. 155–171, 2018.

[12] R. Buettner, S. Sauer, C. Maier, and A. Eckhardt, "Real-time Prediction of User Performance based on Pupillary Assessment via Eye-Tracking," *AIS Trans. Human-Computer Interact.*, vol. 10, no. 2, pp. 26–56, 2018.

[13] N. Kim, J. Kim, and C. R. Ahn, "Predicting workers' inattentiveness to struck-by hazards by monitoring biosignals during a construction task: A virtual reality experiment," *Adv. Eng. Informatics*, vol. 49, p. 101359, 2021.

[14] A. Gregoriades and L. Dimitriou, "Naturalistic analysis of tourist pedestrians' spatial cognition," in *Smart Innovation, Systems and Technologies*, 2020, pp. 3–13.

[15] S. Sabeti, O. Shoghli, M. Baharani, and H. Tabkhi, "Toward AI-enabled augmented reality to enhance the safety of highway work zones: Feasibility, requirements, and challenges," *Adv. Eng. Informatics*, vol. 50, p. 101429, 2021.

[16] Y. Ghasemi, H. Jeong, S. H. Choi, K.-B. Park, and J. Y. Lee, "Deep learning-based object detection in augmented reality: A systematic review," *Comput. Ind.*, vol. 139, p. 103661, 2022.

[17] T. S. J. Darwish and K. Abu Bakar, "Fog Based Intelligent Transportation Big Data Analytics in The Internet of Vehicles Environment: Motivations, Architecture, Challenges, and Critical Issues," *IEEE Access*, vol. 6, pp. 15679–15701, 2018.

[18] J. A. Diego-Mas, D. Garzon-Leal, R. Poveda-Bautista, and J. A. Alcaide-Marzala, "User-interfaces layout optimization using eye-tracking, mouse movements and genetic algorithms," *Appl. Ergon.*, vol. 78, pp. 197–209, 2019.

[19] A. Gregoriades, C. Florides, V. P. Lesta, and M. Pampaka, "Driver Behaviour Analysis through Simulation," in *2013 IEEE International Conference on Systems, Man, and Cybernetics*, 2013, pp. 3681–3686.

[20] B. Wallace, "Driver distraction by advertising: genuine risk or urban myth?," *Munic. Eng.*, vol. 156, no. 3, pp. 185–190, 2003.

[21] T. Rosenbloom, R. Mandel, Y. Rosner, and E. Eldror, "Hazard perception test for pedestrians," *Accid. Anal. Prev.*, vol. 79, pp. 160–169, 2015.

[22] R. L. Charles and J. Nixon, "Measuring mental workload using physiological measures: A systematic review," *Appl. Ergon.*, vol. 74, pp. 221–232, 2019.

[23] S. Goldinger and M. Papesh, "Pupil dilation reflects the creation and retrieval of memories," *Curr. Dir. Psychol. Sci.*, vol. 21, pp. 90–95, 2012.

[24] K. Kitazawa and T. Fujiyama, "Pedestrian vision and collision avoidance behaviour: Investigation of the Information Process Space of pedestrians using an eye tracker," in *Pedestrian and Evacuation Dynamics*, 2009, pp. 95–108.

[25] G. Marquart, C. Cabrall, and J. de Winter, "Review of Eye-related Measures of Drivers' Mental Workload," *Procedia Manuf.*, vol. 3, pp. 2854–2861, 2015.

[26] J. Theeuwes, A. Belopolsky, and C. N. L. Olivers, "Interactions between working memory,

attention and eye movements," *Acta Psychol. (Amst).*, 2009.

[27] B. Strobel, M. A. Lindner, S. . Saß, and O. Köller, "Task-irrelevant data impair processing of graph reading tasks: An eye tracking study," *Learn. Instr.*, vol. 55, pp. 139–147, 2018.

[28] M. A. Just and P. A. Carpenter, "A Theory of Reading: From Eye Fixations to Comprehension.," *Psychol. Rev.*, vol. 87, pp. 329–354, 1980.

[29] J. Hyönä, "The use of eye movements in the study of multimedia learning," *Learn. Instr.*, vol. 20, no. 2, pp. 172–176, Apr. 2010.

[30] F. Muñoz-Leiva, J. Hernández-Méndez, and D. Gómez-Carmona, "Measuring advertising effectiveness in Travel 2.0 websites through eye-tracking technology," *Physiol. Behav.*, 2019.

[31] R. Mahmud, K. Ramamohanarao, and R. Buyya, "Application Management in Fog Computing Environments: A Taxonomy, Review and Future Directions," *ACM Comput. Surv.*, vol. 53, no. 4, Jul. 2020.

[32] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016, pp. 785–794.

[33] A. Barredo Arrieta *et al.*, "Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI," *Inf. Fusion*, vol. 58, pp. 82–115, 2020.

[34] S. M. Lundberg *et al.*, "From local explanations to global understanding with explainable AI for trees," *Nat. Mach. Intell.*, vol. 2, no. 1, pp. 56–67, 2020.

[35] D. Zhang *et al.*, "Research on drivers' hazard perception in plateau environment based on visual characteristics," *Accid. Anal. Prev.*, vol. 166, p. 106540, 2022.

[36] L. Ābele, S. Haustein, L. M. Martinussen, and M. Møller, "Improving drivers' hazard perception in pedestrian-related situations based on a short simulator-based intervention," *Transp. Res. Part F Traffic Psychol. Behav.*, vol. 62, pp. 1–10, Apr. 2019.