Pupil Diameter Classification using Machine Learning During Human-Computer Interaction

Parastoo Azizinezhad, Hamidreza Ghonchi and Anirban Chowdhury, *Member, IEEE*

***Abstract*— This study investigates the classification of pupil diameter data to differentiate between decision-making and focus time in a mobile robot navigation task. Data were collected from 19 healthy participants utilizing an eye-tracking-based user interface to control the robot's movements along pre-set paths. The significance of eye tracking and pupillary responses spans various disciplines, especially for people with severe disabilities. Effortful decision-making is marked by pupil dilation, reflecting increased cognitive load and can be used as a potential measure for system adaptation to users' mental states. This paper presents a deep learning and SVM-based classification approach to distinguish focus and decision-making from pupil diameter patterns, offering insights for future system improvements. On average, the Deep Learning model has an average accuracy of over 82% in classifying the data for participants using either the right, left, or average pupil diameter data.**

I. INTRODUCTION

This study explores the classification of pupil diameter data to discern decision-making and focus time, gathered from 19 healthy participants engaged in a mobile robot navigation task. The robot's movements are controlled by the participant's gaze using an eye-tracking-based user interface containing control buttons and a live video feed of the arena.

Gaze-controlled assistive devices have been used for people with severe disabilities [1], [2]. Eye tracking and pupillary responses have been the focus of researchers from various disciplines and for diverse applications, including education [3], psychology [4], and marketing [5]. Studies have found that higher cognitive load leads to an increase in pupil diameter [6]. Not only mental load but changes in pupil diameter can also indicate task difficulty [7]. Conversely, literature has discussed for some time the rapid responsiveness of pupil diameter to brain activities. Therefore, it may serve as a valuable measure for adjusting systems to meet users' needs and mental states.

Various sources report an increase in pupil dilation during effortful decision-making [8], [9] which is also indicated the increase of cognitive load during decision making. The extent of cognitive load can profoundly affect users' performance [10] therefore minimizing cognitive load is a key consideration in system design [11]. This is especially beneficial for assistive devices, as users often depend on these systems as their primary communication tools for prolonged periods, particularly while coping with underlying health issues. Despite considerable progress in human-computer interaction (HCI), a notable gap persists in adaptive settings and workload-aware interaction, particularly in the context of assistive technology users.

The possibility of distinguishing focus and decision-making from patterns in pupil diameter during a navigation task has been discussed in this paper. The data was collected from 19 healthy participants during a mobile robot navigation task through pre-set paths using an eye-tracking-based user interface. pupil data collected from the experiment has been labeled into two categories of focus and decision making. A deep learning and a SVM based classification models has been used. This detection capability holds promise for enhancing future iterations of the system, potentially enabling the replacement of the fixed dwell time with a more adaptive approach.

In this paper, we detail the experimental procedures, data collection methods, and pre-processing techniques. Additionally, we outline the parameters used in the deep learning-based classification approach and propose a comparative analysis of its performance with SVM classification. The models has been tested with various sets of inputs of data.

II. METHODOLOGY

A. Data Collection

The data were collected from 19 healthy participants as they completed a navigation task along two predetermined paths. Ethical approval for the experiment was granted by the University of Essex Ethics Subcommittee 3 (ERAMS Reference code: ETH2223-2300) and all subjects gave informed consent.

There has been two pre-set paths for this experiment. Participants completed the first, shorter path twice, followed by the longer route, resulting in a total of three experimental rounds. Prior to each round, the eye-tracking system was calibrated to mitigate posture-related issues. Throughout the experiment, the pupil diameter of both eyes and the point of gaze on the screen were recorded at a sampling rate of 60Hz.

The user interface comprised four control buttons for directing the movement of the mobile robot, along with a video feed displaying the path and the robot's movement which is shown in Figure.1. The mouse cursor was manipulated in accordance with the gaze point on the screen, with clicks initiated by maintaining the cursor on a button for a continuous period of 2 seconds.

P. Azizinezhad and A. Chowdhury are with the School of Computer Science and Electronic Engineering, University of Essex, Colchester, United Kingdom. H. Ghonchi is with School of Mathematics, Statistics and Actuarial Science, University of Essex, Colchester, United Kingdom. (Email: p.azizinezhad@essex.ac.uk; h.ghonchi@essex.ac.uk; a.chowdhury@essex.ac.uk)

Upon button press, the robot executed a pre-defined step and remained stationary until the next command was issued. The selection of a 2-second dwell time was deliberate, aimed at minimizing the likelihood of command conflicts and allowing the robot sufficient time to complete its preceding action.

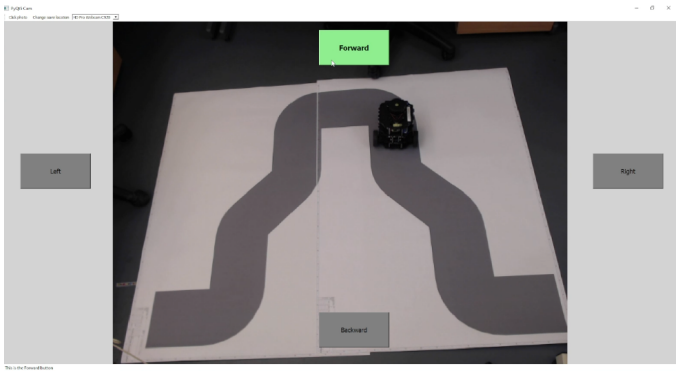

Fig. 1. A screenshot of the interface while navigating the robot through first path.

B. Pre-Processing

The pupil measurements have been divided into two main labels decision and focus. Focus duration is the pupil readings recorded during the 2 seconds spent on the button. As mentioned in II-A, participants have the option to select a button by maintaining the pointer over it for a duration of 2 seconds, known as the "Focus event." Within the context of this paper, we characterize the decision event as the phase of data wherein participants search for and comprehend the button they intend to focus on and activate. For this purpose, several steps have been developed.

- Find and Replace NaN values: Handling NaN values is essential in datasets, particularly in scenarios where tracking devices may fail to record or recognize participants' pupils due to head or body movements, or instances of blinking. To address this issue, we implement a method to replace NaN values with the average values from adjacent data points both before and after each occurrence. By employing this strategy, we ensure continuity in the dataset while mitigating the impact of missing data points caused by such factors.
- Find Gaze Position: The gaze position refers to the point where a participant directs their gaze using both eyes simultaneously. To determine this position, we examine the coordinates of both the left and right eyes at each time point. Subsequently, we calculate the average of these two positions, thereby establishing the gaze position.
- Find Exit Time: To identify the segment of data corresponding to the decision event, it is necessary to determine the exit time. The exit time signifies the moment when the participant's gaze shifts away from the previous button. This occurrence may arise under two circumstances: firstly, when the subsequent button remains the same, but the participant needs to glance at the robot's position before making a decision; or secondly, when the following button differs, prompting the participant to redirect their gaze towards a new location.
- Find Last Enter Time: The entry time marks the instant when participants make a decision and shift their eye positions towards the new button they intend to focus on. It's worth noting that participants' eyes may experience minor movements, potentially causing intermittent shifts away from the button. According to the explanation given in the section II-A, button activation occurs when the pointer remains steadily positioned over the button for a duration of two seconds; otherwise, it undergoes a reset. Consequently, the decision event encompasses the time interval between the initial instance of the participant's eyes leaving the previous button and the final instance of the participant's eyes fixating on the next button and maintaining focus on it.

C. Classification Methods

This paper employs two distinct classification methods to evaluate our study. The first method utilizes a deep learning model, while the second method employs a Support Vector Machine (SVM) [12] algorithm. These algorithms are applied to categorize two classes, namely decision and focus data, extracted from the pupil data collected from participants (as mentioned at section II). The utilization of these classification techniques facilitates a comprehensive analysis of the collected data and enables a deeper understanding of the underlying patterns and trends within the dataset.

1) Deep learning based classification: Recently, deep learning models have emerged as powerful tools for analyzing time series data [13], [14], a type of data which our study deals with, as outlined in Section II. To tackle the classification task inherent in our dataset, we leverage a convolutional neural network (CNN) architecture [15]. This model configuration consists of a series of layers tailored to extract pertinent patterns from the collected time series data related to participants' eye-pupil behaviour. Specifically, our architecture integrates three one-dimensional convolution layers, followed by corresponding max-pooling and dropout layers. Additionally, a fully connected layer and a classifier layer utilizing the softmax activation function are incorporated to facilitate robust classification.

Delving into the model's architecture, the one-dimensional convolution layers play a pivotal role in identifying localized patterns within the time series data. By convolving filters across the input sequences, these layers discern temporal dependencies and extract features at varying temporal scales. The initial two convolutional layers are outfitted with F_1 and F_2 filters, employing kernel sizes of K_1 and K_2 respectively. By harnessing the power of Rectified Linear Unit (ReLU) activation functions, these layers delve deep into the temporal intricacies of the data, allowing the network to discern a diverse array of local features present in the input data. Building upon the foundation laid by its predecessors, the

third one-dimensional convolutional layer further fine-tunes the model's comprehension of the features extracted in the preceding stages. This layer utilizes F_3 filters, doubling the quantity employed in the previous layer, and a kernel size of K_3 to capture even more nuanced temporal features.

Following each convolutional layer, max-pooling layers are employed to downsample the extracted features, preserving the most salient information while simultaneously reducing computational complexity. The kernel sizes for these max-pooling layers are denoted as M_1, M_1, and M_2 respectively.

2) Regularization: In order to mitigate overfitting, a comprehensive regularization strategy is implemented. Each convolutional layer undergoes kernel regularization using both $L1$ and $L2$ methods, along with bias regularization and activity regularization utilizing the $L2$ method. Additionally, Dropout layers are incorporated after each convolutional layer, employing dropout rates denoted as D_1, D_2, and D_3 respectively. This holistic approach to regularization stabilizes the training process and enhances the model's ability to generalize by discouraging reliance on specific features or patterns within the data.

Before classification layer, there is a fully connected layer with N_1 neurons which extract final temporal features. Finally, the classifier layer, employing the softmax function, assigns probabilities to the different classes, enabling the model to categorize the time series data into distinct focus and decision classes. Through this meticulously crafted architecture, our model exhibits the capability to discern intricate patterns within the time series data, ultimately enhancing the accuracy and reliability of the classification task at hand.

3) SVM based classification: In addition to deep learning models, this paper employs Support Vector Machines (SVM), a traditional yet powerful method for classification tasks, particularly in analyzing eye pupil data. SVMs leverage the concept of identifying the optimal hyperplane to effectively distinguish between various classes of data points, making them particularly adept at discerning patterns within datasets. This characteristic renders SVMs invaluable for tasks requiring a nuanced understanding of pupil behavior, shedding light on cognitive processes, attentional states, and decision-making dynamics.

In our study, the SVM kernel utilized is the Radial Basis Function (RBF), chosen for its flexibility in capturing complex relationships within the data. Additionally, the regularization parameter, denoted as R_1, is carefully selected to strike a balance between model complexity and generalization performance. Through the selection of appropriate parameters such as the kernel function and regularization parameter, the SVM framework is tailored to effectively capture and classify intricate patterns inherent in eye pupil data, further enhancing our understanding of cognitive processes and behavioral dynamics.

A comprehensive summary of all parameters and configurations employed in this study is provided in Table I.

TABLE I

PARAMETERS AND SETUPS USED FOR TRAINING MODEL.

	Parameter Name	Value
Data preparation	Normalization	Z-score algorithm
Deep Learning	F_1	16
	F_2	32
	F_3	64
	K_1	60
	K_1	30
	K_1	5
	Activations Functions	ReLU
	D_1	50%
	D_2	40%
	D_3	30%
	N_1	128
	Classifier	Softmax
	Loss	Binary Cross Entropy
	Optimizer	Adam
	Epochs	250
	Batch size	32
	K-fold	5
SVM	Kernel	Radial Basis Function
	R_1	1

III. RESULTS

The effectiveness of our proposed deep learning (DL) model and SVM is assessed through the obtained results. For the model configuration, we utilized binary cross-entropy as the loss function along with the Adam optimizer. The training was conducted over 250 epochs, employing k-fold cross-validation with $k = 5$. This approach ensures robust evaluation and enhances the reliability of our findings in classifying pupil diameter.

This paper utilized a deep learning and SVM-based classification model to differentiate between focusing and decision-making based on recorded pupil data. The accuracy of the model was assessed and compared using various sets of inputs, including the right, left, and average pupil diameters for each participant and each round of the experiment.

The initial comparison involved employing pupil data from all three rounds for each participant as separate inputs for both the SVM and deep learning models. Figure.2 presents the accuracies achieved by the SVM and deep learning models using the right eye pupil data for each participant, while Figure.3 displays the results for the left eye data. The average of both eye's pupil diameter is the input used for the results shown in Figure.4. The figures depict a substantial accuracy rate of approximately 90% for both the deep learning and SVM algorithms. These findings emphasize a notable distinction between decision and focus labels. The high accuracy achieved by both algorithms highlights their efficacy in discerning between these cognitive states during the task.

Table.II presents a comparison of the accuracies achieved by SVM and deep learning model. Overall, the deep learning approach demonstrates superior accuracy, reaching a maximum of 95.83% with right eye data. However, it is noteworthy that the classification accuracies of SVM and DL vary for each set of inputs and individuals. In some cases, SVM outperforms deep learning, while in others,

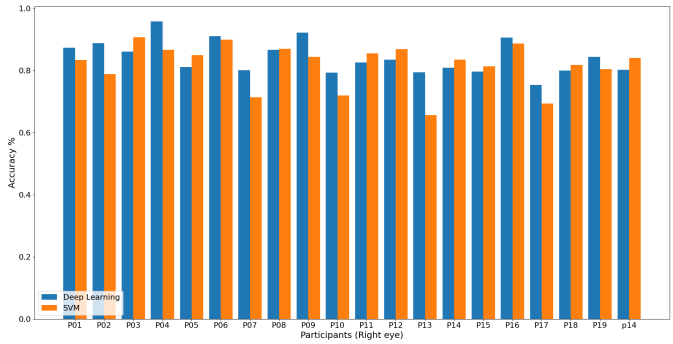

Fig. 2. Right eye accuracy results for SVM and DL for individual participants.

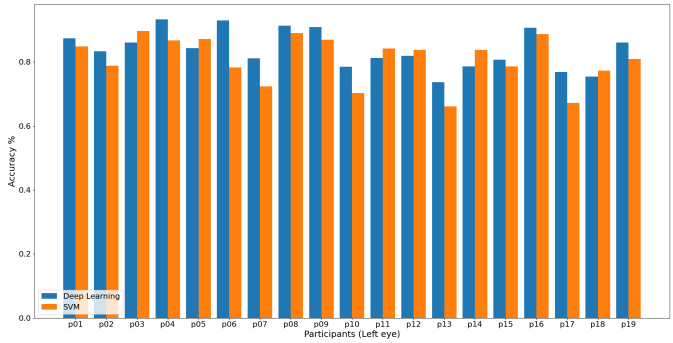

Fig. 3. left eye accuracy results for SVM and DL for individual participants.

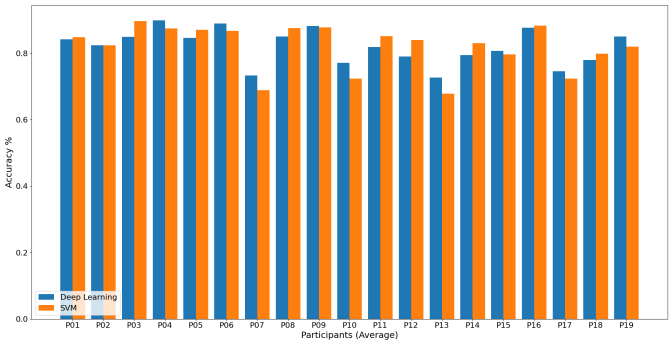

Fig. 4. Accuracy of average of right and left eye pupil diameters for SVM and DL for individual participants.

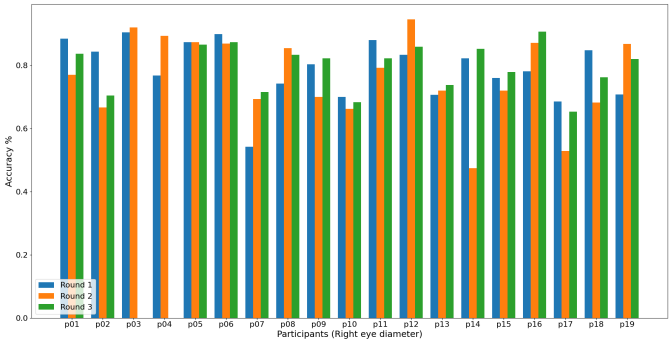

Fig. 5. Right eye accuracy results using SVM for three rounds.

the opposite is observed, and this discrepancy also vary depending on the input used. In general, the total accuracy for both algorithms are close to each other in all different inputs.

The subsequent comparisons were conducted using data collected from each round separately with the SVM model. Deep learning was not employed in this comparison due to an insufficient amount of data for the model in some cases. Figure.5 illustrates the comparison of accuracy across different rounds using right pupil data, while Figure.6 presents the results for the left eye. Additionally, Figure.7 depicts the accuracy comparison based on the average pupil data.

As mentioned before, Round 1 and 2 are done on the same path while round 3 is the new longer path. Some participants the accuracy is lower in round 2 in comparison with other rounds. This can be due to lower efforts in decision making due to familiarity with the path. Another contributing factor to this phenomenon may be because of the consecutive running of rounds 1 and 2, which likely induced tiredness among participants. Consequently, this tiredness could have adversely impacted their performance outcomes. The sustained engagement in multiple rounds of the task may have led to increased mental and physical exertion, thereby diminishing participants' cognitive capacities and overall effectiveness in executing the task. Participants number 3 and 4 haven't completed round three in this study therefore the data is missing from the figures.

Table.III has compared the accuracy achieved with separated rounds data. The highest accuracy of 96% is achieved in the first round by using the average pupil diameter data.

IV. CONCLUSION

The paper presents a comparative analysis of model accuracy in classifying focus and decision-making based on pupil diameter data collected from 19 healthy participants during a navigation task. Various sets of inputs were examined to assess model performance. The deep learning model achieved an accuracy exceeding 82%, with an average accuracy of 84.24% across all participants using right eye data, and a maximum accuracy of 95.83% for one participant. In contrast, SVM exhibited an average accuracy of 81%, indicating superior performance of the deep learning model overall. While there were variations in model performance among individual participants, with a minimum accuracy of 72.26% observed when utilizing average pupil diameter data for one participant, no significant differences were observed in the effectiveness of using right, left, or average pupil data.

The outcomes of this study hold implications for interface design enhancement and potential replacement of fixed dwell time with an adaptive setting. By leveraging the deep learning model's robust performance in pupil diameter classification, future iterations of the interface can be tailored to better accommodate users' cognitive states, thereby enhancing overall usability and effectiveness.

TABLE II

DEEP LEARNING AND SVM ACCURACY FOR EACH PARTICIPANTS.

	Left			Right			Average		
	Max	Min	Average	Max	Min	Average	Max	Min	Average
DL	93.33%	73.66%	83.89%	95.83%	75.32%	84.24%	90%	72.76%	82.03%
SVM	89.68%	66.09%	80.75%	90.68%	65.62%	81.81%	89.71%	67.84%	82%

TABLE III

SVM ACCURACY FOR EACH ROUND.

	Left			Right			Average		
	Max	Min	Average	Max	Min	Average	Max	Min	Average
Round 1	91.61%	65.55%	79.56%	90.51%	54.22%	78.90%	96%	60.04%	78.97%
Round 2	89.64%	46.66%	73.71%	94.66%	47.50%	76.37%	91.21%	40.77%	75.91%
Round 3	90.24%	61.16%	77.83%	90.73%	65.38%	79.60%	90.75%	63.16%	78.85%

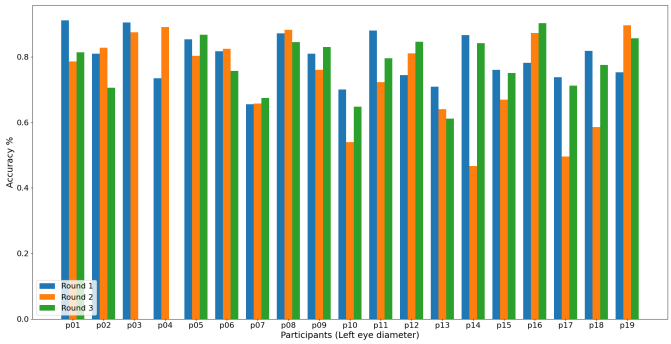

Fig. 6. left eye accuracy results using SVM for three rounds.

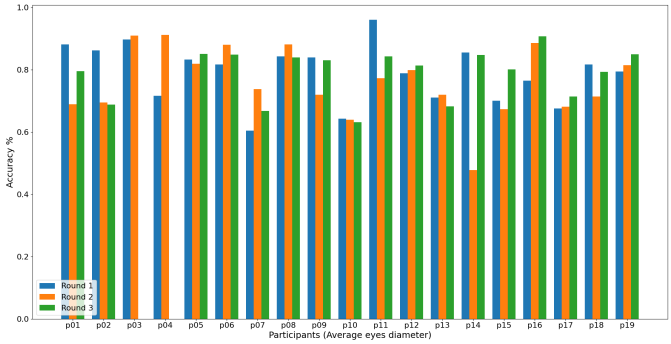

Fig. 7. average of right and left eye pupil diameters accuracy result using SVM for three rounds.

REFERENCES

[1] M. Borgestig, J. Sandqvist, R. Parsons, T. Falkmer, and H. Hemmingsson, "Eye gaze performance for children with severe physical impairments using gaze-based assistive technology—a longitudinal study," *Assistive technology*, vol. 28, no. 2, pp. 93–102, 2016.

[2] C.-S. Hwang, H.-H. Weng, L.-F. Wang, C.-H. Tsai, and H.-T. Chang, "An eye-tracking assistive device improves the quality of life for als patients and reduces the caregivers' burden," *Journal of motor behavior*, vol. 46, no. 4, pp. 233–238, 2014.

[3] J. L. Rosch and J. J. Vogel-Walcutt, "A review of eye-tracking applications as tools for training," *Cognition, technology & work*, vol. 15, pp. 313–327, 2013.

[4] J. F. Cavanagh, T. V. Wiecki, A. Kochar, and M. J. Frank, "Eye tracking and pupillometry are indicators of dissociable latent decision processes.," *Journal of Experimental Psychology: General*, vol. 143, no. 4, p. 1476, 2014.

[5] G. van Loon, F. Hermsen, and M. Naber, "Predicting product preferences on retailers' web shops through measurement of gaze and pupil size dynamics," *Journal of Cognition*, vol. 5, no. 1, 2022.

[6] R. Mitra, K. S. McNeal, and H. D. Bondell, "Pupillary response to complex interdependent tasks: A cognitive-load theory perspective," *Behavior Research Methods*, vol. 49, pp. 1905–1919, 2017.

[7] E. H. Hess and J. M. Polt, "Pupil size in relation to mental activity during simple problem-solving," *Science*, vol. 143, no. 3611, pp. 1190–1192, 1964.

[8] H. Simpson and S. M. Hale, "Pupillary changes during a decision-making task," *Perceptual and Motor Skills*, vol. 29, no. 2, pp. 495–498, 1969.

[9] J. W. De Gee, T. Knapen, and T. H. Donner, "Decision-related pupil dilation reflects upcoming choice and individual bias," *Proceedings of the National Academy of Sciences*, vol. 111, no. 5, E618–E625, 2014.

[10] J. Engström, G. Markkula, T. Victor, and N. Merat, "Effects of cognitive load on driving performance: The cognitive control hypothesis," *Human factors*, vol. 59, no. 5, pp. 734–764, 2017.

[11] S. Chen and J. Epps, "Using task-induced pupil diameter and blink rate to infer cognitive load," *Human–Computer Interaction*, vol. 29, no. 4, pp. 390–413, 2014.

[12] M. A. Hearst, S. T. Dumais, E. Osuna, J. Platt, and B. Scholkopf, "Support vector machines," *IEEE Intelligent Systems and their applications*, vol. 13, no. 4, pp. 18–28, 1998.

[13] N. Mohammadi Foumani, L. Miller, C. W. Tan, G. I. Webb, G. Forestier, and M. Salehi, "Deep learning

for time series classification and extrinsic regression: A current survey," *ACM Computing Surveys*, vol. 56, no. 9, pp. 1–45, 2024.

[14] J. Wang, W. Du, W. Cao, *et al.*, "Deep learning for multivariate time series imputation: A survey," *arXiv preprint arXiv:2402.04059*, 2024.

[15] K. O'shea and R. Nash, "An introduction to convolutional neural networks," *arXiv preprint arXiv:1511.08458*, 2015.