Eye Movement Classification using Feature Engineering and Ensemble Machine Learning

Hassanein Riyadh Mahmood
Department of Renewable Energy Techniques, Technical Institute-Kut, Middle Technical University, Baghdad, Iraq
hassanein-riyadh@mtu.edu.iq

Dhurgham Kareem Gharkan
Department of Cyber Security Techniques, Technical Institute-Kut, Middle Technical University, Baghdad, Iraq
dhurgham-kareem@mtu.edu.iq (corresponding author)

Ghusoon Ismail Jamil
Department of Renewable Energy Techniques, Technical Institute-Kut, Middle Technical University, Baghdad, Iraq
ghusson.ismaeel@mtu.edu.iq

Asmaa Ali Jaish
Electric Engineering Technology Department, Kut University College, Al Kut, Wasit, Iraq
asmaa.jaish@alkutcollege.edu.iq

Sarah Taher Yahya
Electric Engineering Technology Department, Kut University College, Al Kut, Wasit, Iraq
sarah.yahya@alkutcollege.edu.iq

Received: 26 September 2024 | Revised: 15 October 2024 | Accepted: 19 October 2024

 DOI: https://doi.org/10.48084/etasr.9115

ABSTRACT

This paper explores the classification of gaze direction using electrooculography (EOG) signals, integrating signal processing, deep learning, and ensemble learning techniques to enhance accuracy and reliability. A complex technique is proposed in which several feature types are derived from EOG data. Spectral properties generated from power spectral density analysis augment basic statistical characteristics such as mean and standard deviation, revealing the frequency content of the signal. Skewness, kurtosis, and cross-channel correlations are also used to represent intricate nonlinear dynamics and inter-channel interactions. These characteristics are then reformatted into a two-dimensional array imitating picture data, enabling the use of the pre-trained ResNet50 model to extract deep and high-level characteristics. Using these deep features, an ensemble of bagging-trained decision trees classifies gaze directions, lowering model variance and increasing prediction accuracy. The results show that the ensemble deep learning model obtained outstanding performance metrics, with accuracy and sensitivity ratings exceeding 97% and F1-score of 98%. These results not only confirm the effectiveness of the proposed approach in managing challenging EOG signal classification tasks but also imply important consequences for the improvement of Human-Computer Interaction (HCI) systems, especially in assistive technologies where accurate gaze tracking is fundamental.

Keywords-electrooculography (EOG); gaze direction classification; deep learning; ensemble learning; ResNet50; feature extraction

I. INTRODUCTION

Electrooculography (EOG) is a biometric method that quantifies the corneo-retinal standing potential that exists between the front and the back of the human eye [1, 2]. Captured with superficially placed silver electrodes on the face, the EOG signal follows eye movements along both horizontal (left-right) and vertical (up-down) channels [3]. EOG is useful,

especially for people with motion problems who use eye movement to communicate. Many everyday systems use this technology [4]. Users may, for example, write by choosing words from a restricted list to create sentences, operate electric wheelchairs, and move cursors or even guide robots solely with their eye movements. Focusing on average and maximum speeds and voltage ranges, these applications use EOG data, including saccadic movements and ocular reflexes [5, 6]. EOG signals are usually observed in a low-frequency bandwidth of 0.5 to 50 Hz. Thus, the classification of eye movements depends on algorithms that compute signal derivatives. To precisely characterize motions, these algorithms focus on several criteria, including velocity, acceleration, amplitude, and signal fit [7, 8].

Eye motions within the EOG data have been effectively detected using classification techniques such as K-Nearest Neighbor (KNN), Support Vector Machines (SVM), and Decision Trees (DT) [9, 10]. Combining these techniques, especially KNN and SVM, improves classification efficiency and achieves notable accuracy in identifying unique characteristics, including states conveyed by eye movements. Deep learning's ability to extract significant, deep features from challenging datasets has transformed the method academics take to classification challenges in many different fields. When paired with sophisticated classifiers that can use these characteristics for higher accuracy, this capacity is very strong. In this context, the conversion of manufactured characteristics into picture-shaped forms becomes a crucial step that allows the use of Convolutional Neural Networks (CNNs), including ResNet50, which are intrinsically designed to analyze image input [11-13]. CNN's spatial hierarchical feature extraction capacity allows the use of either signal-based or numerical data in 2D picture arrays. This method improves the feature representation and fits the architectural strengths in capturing important patterns that are less obvious in conventional, non-image data formats.

This study proposes a hybrid method to improve the classification of gaze directions in EOG data using an integrative approach that includes signal processing, deep learning, and ensemble learning techniques. To ensure data quality, EOG signals are preprocessed. This preprocessing consists of filtering and normalizing the signals to eliminate noise and normalize the dataset for subsequent investigations. An advanced feature extraction phase follows, which extracts features from the EOG data. Although spectral data produced by power spectral density analysis disclose the frequency content, statistical features such as mean and standard deviation offer fundamental signal insights. Skewness, kurtosis, and cross-channel correlations, among other aspects, help to represent the nonlinear dynamics and interactions among signal channels. The features obtained are converted into a 2D array form, suitable for image processing using deep learning methods. These images are then sent to a pre-trained ResNet50 model to produce high-level deep features. Using the bagging technique, an ensemble of decision trees is trained to categorize gaze directions after feature extraction with ResNet50. Using this combined technique, model variance decreases and prediction accuracy increases. Model performance is evaluated using accuracy, sensitivity, specificity, F1-score, and AUC. The robustness and dependability of the model are ensured using validation methods such as k-fold cross-validation and holdout methods.

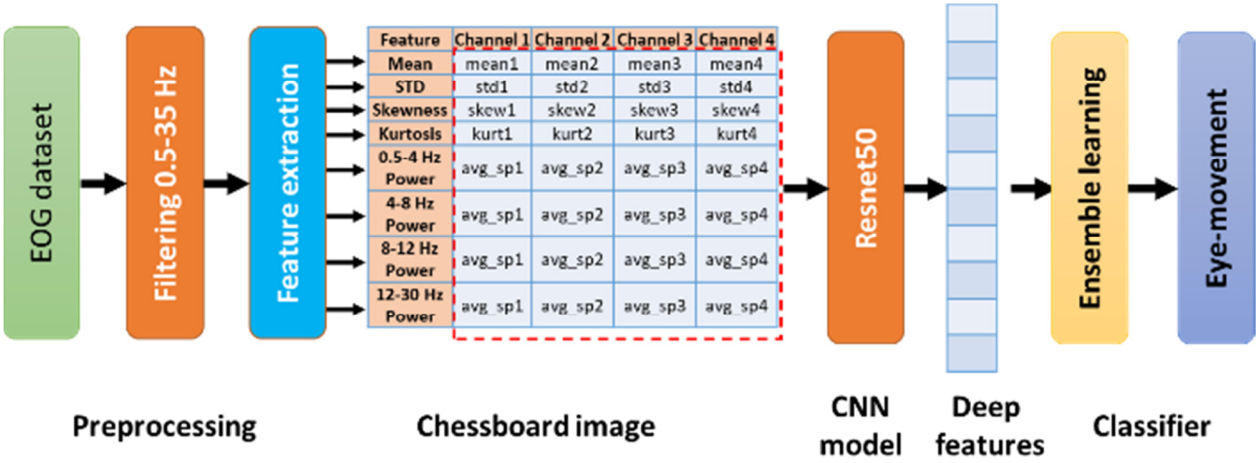

Fig. 1. The proposed approach to classify eye movements based on EOG.

II. RELATED WORKS

Various classification methods have been applied in Human-Computer Interaction (HCI) using EOG signals. In [14], 16 parameters were extracted from EOG movement data in the time domain from a sample of 12 test subjects. Using DT, KNN, and SVM techniques, 95.4%, 99.6%, and 99.1% classification accuracy was obtained. The ROC curve showed that the KNN method was especially successful. The study in [15] focused on the time domain implementation. It was observed that with six parameters, the SVM algorithm was able to obtain a classification accuracy of 60%. In the larger framework of EOG signal classification, several studies have underlined the simplicity and good performance of KNN, SVM, and DT. In [16], Power Spectral Density (PSD) was used for EOG signal classification, in which SVM paired with neural networks achieved a classification accuracy of 69.75%.

In [9], the cooperative use of KNN and SVM in image classification was investigated, using spectral feature parameters to improve accuracy and classification speed. In [10], emotional classification was performed using 23 distinct emotional states, including wrath, fear, happiness, and sadness. This study showed an accuracy of 75.15% for KNN and 80% for SVM, highlighting the need to evaluate several methods to choose the best classifier.

In [17], applications were customized using digital EOG signals as control inputs for HCI. In [18], several classifiers, including DT, KNN, and SVM, were used for effective signal classification, obtaining excellent accuracies. EOG signal analysis has progressed significantly due to feature extraction methods. The study in [19] focused on class separability and feature extraction robustness. In [20], frequency-based AR models were coupled with template matching and wavelet decomposition to improve the analytical process. In [21], ensemble methods were used, utilizing bagging and adaptive boosting to classify eye movements and achieving an accuracy of 92.27%. Meanwhile, in [22], a pulse detection method was integrated into a Human-to-Television Interface (HTI), demonstrating an average performance of 93.41%. EOG signals have been used to help people with severe disabilities. The studies in [23] and [24] focused on aiding patients with ALS and severe mobility impairments, respectively, through innovative EOG signal processing for eye-writing systems and wheelchair control. Real-time applications, such as the asynchronous EOG-based virtual keyboard in [25] and a wearable forehead EOG measurement system in [26], showcased practical implementations of EOG technology with high classification accuracy. Furthermore, the advanced processing algorithms in [27] and [28] have refined the detection of eye movements. The study in [27] combined derivative and amplitude level inputs to effectively filter noise and detect signal edges, while in [28], Independent Component Analysis (ICA) was used with SVM for saccadic signal recognition, achieving a classification accuracy of 99.57%.

In [29], a cutting-edge Recurrent Neural Network (RNN) was introduced, utilizing Gated Recurrent Unit (GRU) layers in a bidirectional configuration followed by dense layers. This model was developed to classify four directional eye movements, up, down, left, and right, across both vertical and horizontal channels of EOG signals. This classifier achieved accuracies of 99.77% and 99.74% for vertical and horizontal channels, respectively, setting a new benchmark in classification performance for HCI applications in rehabilitation and beyond. In [30], EOG signals generated during various eye movements, including blinking, were classified. This approach modeled the eye as a dipole and used the wavelet transform to extract features in the frequency domain, within a bandwidth of 0.5 to 50 Hz. KNN, SVM, and DT were implemented, achieving classification accuracies of 69.4%, 76.9%, and 60.5%, respectively. The effectiveness of these classifiers was evaluated using the Jaccard index, confusion matrix, and ROC curve, with SVM emerging as the superior classifier for this set of tasks.

III. METHOD

A. Data Collection

The dataset [31] used in this study includes EOG data along with head posture and position information recorded from eight healthy participants (six males and two females, mean age of 25.8 ± 5.8 years), all of whom had normal or corrected-to-normal vision. Data collection was carried out according to an approved protocol by the University Research Ethics Committee (UREC) of the University of Malta, and all subjects provided their informed consent before participating in recording sessions. The participants were seated at a distance of approximately 60 cm from a 24-inch LCD screen. During the sessions, they were instructed to focus on visual cues presented at various screen positions while allowing natural head movements. Each session consisted of 200 trials, each lasting four seconds, with specific eye movements being recorded. At the start of each trial, a visual cue was displayed in a position (P1j), which then shifted to a different location (P2j) after one second. During the next two seconds, participants were instructed to blink as the cue color changed, marking the end of the trial. The initial position (P20) was centered on the screen, while the subsequent positions (P1j and P2j) were randomized. The EOG signals were recorded using a standard setup with four electrodes, including reference and ground electrodes, and amplified using a biosignal amplifier with a sampling frequency of 256 Hz. Signal processing involved applying a 50 Hz notch filter to remove electrical noise and a bandpass filter with a range of 0-30 Hz. Additionally, the head posture and position data were captured using the 3D Guidance TrakSTAR system, which utilizes an electromagnetic transmitter and a sensor mounted on the participant's head via a headband. This system allows six degrees of freedom tracking, capturing both the position (3D coordinates) and orientation (yaw, pitch, and roll angles) of the head throughout the trials. The data were provided in MATLAB format, with separate files for each subject containing EOG signals, control signals, gaze angles, inter-pupillary distance, distance from the screen, head pose, and head position.

B. Preprocessing and Features

Target gaze angles aid in segment contextualization, while control signals mark each sample of the EOG data for precise segmentation. The segmentation approach ensures no information loss at segment borders by including a pre-segmentation phase having 51 samples before each detected change in the control signal. After that, the signals are filtered using a bandpass FIR filter with a filter order of 20 and cutoff frequencies set between 0.5 and 35 Hz, matching the frequency characteristics relevant to EOG signals.

After filtering, the EOG segments are normalized using mean subtraction and standard deviation division. This stage helps improve attention to waveform patterns and lower model sensitivity to amplitude fluctuations. These processed sections are converted to a format suitable for deep learning analysis with the ResNet architecture. This transformation, along with spectral elements produced from the normalized power spectral density over specified frequency bands, entails the computation of temporal characteristics such as mean, standard deviation, skewness, and kurtosis. Spectral characteristics are

fundamental as the Welch technique computes the power spectrum density of every segment. The frequency spectrum is segmented using predetermined sub-bands ranging from 0.5-4 Hz, 4-8 Hz, 8-12 Hz, to 12-30 Hz, enabling a focused study on particular frequency ranges. The overall power throughout the wider spectrum 0.5–35 Hz is calculated for every channel within a segment, and then the power within each sub-band is found. The proportion of the total power occupied by each sub-band is computed from these data, normalizing the spectral data and emphasizing the dominating frequencies for every segment. These elements are combined into one complete feature vector per segment.

After normalizing and reshaping this feature vector into an image structure, shrinking it to meet the input dimensions needed by ResNet that are typically 224×224 pixels, the single-layer feature picture is repeated across three channels to meet the three-channel input need of the network, therefore producing a chessboard-like pattern that faithfully captures the features of the original EOG data.

C. Resnet 50

ResNet50 is a powerful deep learning model derived from the family of residual networks (ResNets), especially created to solve the challenge of training very deep neural networks [32]. Residual blocks consist of a few layers where the input to the block is added to the output, enabling gradients to flow straight through these connections throughout the training phase. ResNet50 consists of 50 layers, including fully connected, convolutional, normalizing, and ReLU activation layers. Starting with a convolutional layer, then a batch normalizing layer, and finally a max-pooling layer, the network reduces spatial dimensions and enhances feature extraction. A series of residual blocks follows, with several convolutional, batch normalized, and ReLU activation layers. These blocks are designed so that, by using a shortcut connection to bypass one or more levels, the input to each block is connected to its output, providing a direct route for the data flow across the layers. This design can mitigate the vanishing gradient problem, which is a common issue in conventional deep networks where gradients could become too small to have any appreciable impact during backpropagation as the depth increases. ResNet50 accelerates training convergence by allowing gradients to escape numerous layers at once via shortcut connections, conserving the gradient's strength. Typically, at the end of the network, ResNet50 pulls features from the avg_pool layer. This layer performs global average pooling on the feature maps derived from the last convolutional block, transforming each map into a single scalar value. By using pooling, the spatial information of the feature maps is efficiently condensed into a small feature vector, preserving the most important information required for classification tasks. The avg_pool layer produces a highly distilled representation of the input data that captures the most essential patterns and features to separate between various classes in a classification task.

D. Ensemble Learning Classifier

Ensemble learning combines many models or the same model with varying initializations to detect eye movement directions based on deep data. In ensemble learning, bagging or bootstrap aggregating is a typical method where many models are trained on separate subsets of the training data and subsequently aggregated by voting or averaging. This method is advantageous in complicated feature spaces, such as deep learning outputs, because it lowers the variance of the predictions, reducing the ensemble's overfitting susceptibility.

To implement ensemble learning in classifying deep features, a specific type of ensemble model known as Random Forest (RF) or an ensemble of decision trees can be typically used [33]. These are trained on random subsets of the training data using a technique called bootstrap aggregating. Each tree in the ensemble votes on the class, and the class receiving the majority of votes becomes the model's prediction. When applying ensemble methods to classify eye movement directions, the model processes labeled data that correspond to specific directions, quantified from EOG signal characteristics such as horizontal and vertical components of gaze angles.

E. Performance Metrics

The key performance metrics used in this study are:

- Accuracy: Measures the overall correctness of the model and is calculated as the ratio of correct predictions to the total number of cases examined. It provides a quick indication of how well the model performs across all classes.

$$\text{Accuracy} = \frac{\text{Number of correct predictions}}{\text{Total number of predictions}} \quad (1)$$

- Confusion matrix: A table used to describe the performance of a classification model on a set of test data with known true values. It allows the visualization of the model's predictions, including the number of True Positives (TP), False Positives (FP), True Negatives (TN), and False Negatives (FN) for each class.

- Sensitivity (true positive rate or recall): Measures the proportion of actual positives that are correctly identified as such (e.g., the percentage of 'up' directions that are correctly predicted as 'up'). It is crucial for applications where missing a positive prediction is costly.

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP+FN}} \quad (2)$$

- Specificity (true negative rate): Measures the proportion of actual negatives that are correctly identified (e.g., the percentage of non-up directions that are correctly identified as not up). High specificity reduces false alarms.

$$\text{Specificity} = \frac{\text{TN}}{\text{TN+FP}} \quad (3)$$

- Precision: Indicates the accuracy of positive predictions. Formulated as the ratio of TP to the sum of TP and FP, shows how many of the positively labeled predictions were actually correct.

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \quad (4)$$

- F1-score: It is a weighted average of precision and recall, taking both FP and FN into account. It is especially useful

when the class distribution is uneven. The F1-score is the harmonic mean of precision and recall.

$$\text{F1} - \text{score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (5)$$

- Area Under the ROC curve (AUC-ROC): A performance measurement for classification problems at various threshold settings. ROC is a probability curve that plots the true positive rate (sensitivity) against the false positive rate (1-specificity) at various threshold settings. AUC represents the degree or measure of separability, indicating how much the model is capable of distinguishing between classes.

IV. RESULTS

The dataset involves eight subjects with a total of 1600 trials that were meticulously analyzed to understand the distribution of eye movement types. Each subject underwent 200 trials, focusing on their gaze behavior across various directional stimuli. This research categorizes eye movements into nine distinct cases, eight representing specific directional movements and one indicating no movement.

The distribution of these eye movements is quantified and presented systematically. As outlined in Table I, the samples used for classification reveal a relatively balanced distribution among the nine classes, albeit with slight variations that may reflect natural differences in individual gaze patterns or the experimental setup. For instance, eye movements towards the lower-left direction (↙) accounted for the highest percentage of the total samples (12.22%), represented by 220 instances. This was closely followed by down (↓) and down-right (↘) movements, with 206 (11.44%) and 209 (11.61%) instances, respectively. In contrast, the least represented were right (→) and left (←) movements, each constituting slightly more than 10% of the dataset, indicating a possibly less frequent occurrence of these movements under experimental conditions.

TABLE I. DISTRIBUTION OF DIRECTIONS AND SAMPLES USED FOR CLASSIFICATION

Eye movement type	Percentage of total	Count
1: top left (↖)	10.61%	191
2: top (↑)	10.72%	193
3: top right (↗)	11.33%	204
4: left (←)	10.17%	183
5: no movement	11.78%	212
6: right (→)	10.11%	182
7: bottom left (↙)	12.22%	220
8: bottom (↓)	11.44%	206
9: bottom right (↘)	11.61%	209

A structured training and testing approach was used, assigning 70% of the data to training and the remaining 30% to testing. This division is standard practice in machine learning to ensure both robust training and an unbiased evaluation of the model's generalization capabilities on unseen data. The performance metrics, as presented in Table II, demonstrate ResNet50's effectiveness in learning and predicting the classification of eye movements based on the generated features from EOG data. The training and testing accuracy reflect the model's consistency across various eye movements, capturing the inherent variability and complexity of eye movement patterns. For instance, the model achieved a training accuracy of 86.52% and a testing accuracy of 88.93% for the 1 - top left (↖) eye movement, suggesting a commendable generalization beyond the training set. Similar trends are observed in other types, such as 9 - bottom right (↘), where the training accuracy was 87.28%, with a slightly higher testing accuracy of 89.74%, indicating effective learning and an impressive capture of underlying patterns within the test data. However, some types, such as 5 - no movement and 7 - bottom left (↙), showed a decrease from training to testing accuracy, which may indicate potential overfitting or inadequacy in the model's ability to generalize from training data to real-world variability in these specific types of eye movements. Specifically, 5 - no movement showed a substantial drop from 88.06% in training to 85.23% in testing, prompting a need for further analysis and possible adjustments in the model's training phase or parameter tuning to enhance its performance.

TABLE II. TRAINING AND TESTING PERFORMANCE OF RESTNET50 USING CHESSBOARD IMAGES

Eye movement type	Training accuracy (%)	Testing accuracy (%)
1: top left (↖)	86.52%	88.93%
2: top (↑)	87.62%	86.00%
3: top right (↗)	87.16%	87.57%
4: left (←)	86.46%	87.96%
5: no movement	88.06%	85.23%
6: right (→)	85.70%	88.04%
7: bottom left (↙)	86.46%	85.85%
8: bottom (↓)	86.83%	85.33%
9: bottom right (↘)	87.28%	89.74%

As shown in Table III, the deployment of ensemble learning methods using deep features extracted by ResNet50 has demonstrated significant efficacy in classifying eye movements. Including accuracy, sensitivity, and F1-score, the performance evaluations provide a full view of training and testing phases across several eye motion patterns. With accuracy rates typically near or over the 95% threshold, the models demonstrate great performance in the training phase, therefore proving their potential to effectively learn and adapt to the complexity of the input data. For 1 - top left (↖), for example, the accuracy reached almost 100%, stressing the precision with which the ensemble model could identify this sort of movement dependent on the acquired traits. Similarly, high sensitivity values, such as 97.73% for 3 - top right (↗), showcase the model's responsiveness to the existence of certain classes, therefore ensuring that the movements are precisely detected with low false negatives. In the testing phase, the models maintain great accuracy with few deviations, such as in motions like 8 - bottom (↓), where accuracy falls significantly. Still, the test sensitivity and F1-scores, such as a 99.55% sensitivity for 2 - top (↑) and a 99.93% F1-score for 6 - right (→), showcase the models' consistent capacity to extend the learned patterns to fresh data. These findings indicate the model's balanced capacity to effectively regulate both recall and accuracy, hence lowering FP and FN. Furthermore, differences in training and testing performance, such as the lower testing accuracy for 9 - bottom right (↘) relative to its training equivalent, require improvements. Improving model

resilience and ensuring constant performance across several situations and datasets need further investigation.

TABLE III. TRAINING AND TESTING PERFORMANCE OF ENSEMBLE LEARNING USING DEEP FEATURES.

Eye movement type	Training			Testing		
	Accuracy (%)	Sensitivity (%)	F1-score (%)	Accuracy (%)	Sensitivity (%)	F1-score (%)
1 (↖)	99.83	96.52	98.42	99.04	95.49	97.20
2 (↑)	95.61	95.17	96.29	97.48	99.55	98.31
3 (↗)	96.56	97.73	99.85	97.60	95.92	98.88
4 (←)	99.70	97.99	95.44	99.47	99.61	95.98
5 -	95.23	96.94	99.14	96.63	96.36	96.78
6 (→)	96.40	95.70	95.37	97.71	99.01	99.93
7 (↙)	98.86	95.03	98.53	95.99	99.08	98.65
8 (↓)	98.86	96.79	99.32	95.37	95.58	98.12
9 (↘)	96.65	96.55	98.65	95.32	96.63	98.19
Average	**97.52**	**96.49**	**97.89**	**97.18**	**97.47**	**98**

Figure 2 shows the confusion matrix, providing a comprehensive overview of the classifier's performance across nine distinct eye movement classes. Each green-highlighted cell on the diagonal quantifies the correct predictions for each class, showcasing accuracy rates predominantly above 96%, which underscores the classifier's effectiveness. Off-diagonal pink cells, indicating misclassifications, are minimal, with most error rates below 0.3%, suggesting a high degree of precision in class differentiation.

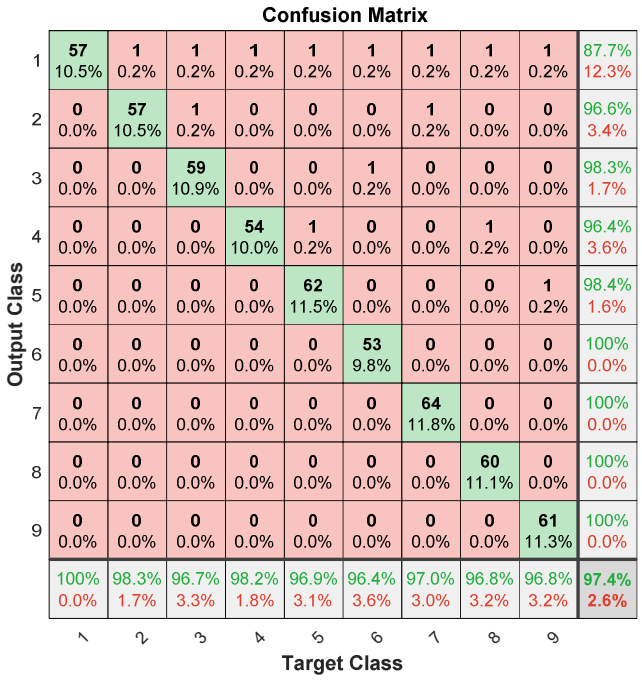

Fig. 2. Confusion matrix of classifying testing data.

The ROC graph in Figure 3 presents a clear and effective illustration of the classifier performance in comparison to random chance. The ROC curve, represented by the solid blue line, hugs the upper left corner of the plot, which indicates an exceptionally high true positive rate while maintaining a low false positive rate across all thresholds. This characteristic is indicative of excellent classifier performance, as the ROC curve nearly reaches the perfect classification point at (0,1), where the FP rate is zero and the TP rate is one.

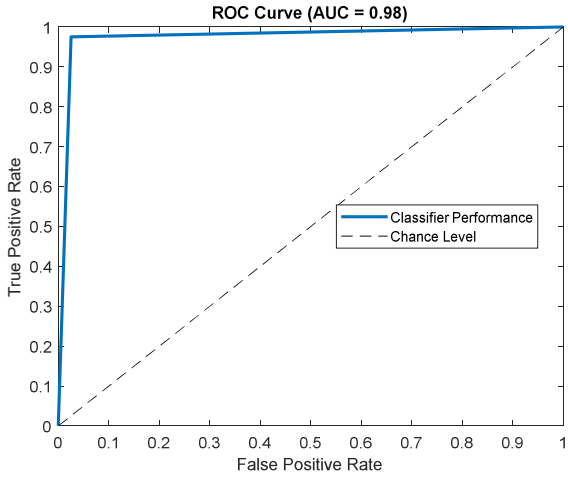

Fig. 3. AUC curve of training data.

Table IV shows a detailed comparison of test performance across several classification models based on accuracy, sensitivity, and F1-score. These measures are essential for evaluating a model's general performance in not only accurate data classification but also in ensuring that classifications are balanced among several categories. SVM achieved 79.44% accuracy, 72.39% sensitivity, and 84.26% F1-score, reflecting quite modest performances. The lower sensitivity suggests a propensity to overlook more true positive classifications than in other models. With an accuracy of 85.22%, a sensitivity of 81.23%, and an F1-score of 85.42%, the DT classifier demonstrated superior results, implying greater overall performance in collecting pertinent data features. RF showed a possible compromise between precision and recall, with a lower F1-score of 78.55 %. However, its accuracy and sensitivity were comparable to those of DT. The simplest of the models, KNN, had the lowest performance on all measures. Although with somewhat less sensitivity, Logistic Regression (LR) offered a decent mix between accuracy (82.73%) and F1-score (80.17%). With an accuracy of 87.54%, a sensitivity of 82.64%, and an F1-score of 86.01%, a second DT model showed remarkably improved performance, perhaps indicating adjustments in tree parameters or training subsets that might favorably affect performance measures. With excellent metrics, such as 97.18% accuracy, 97.47% sensitivity, and 98% F1-score, the ensemble deep learning model outperformed all other models. This better performance emphasizes the benefits of applying sophisticated ensemble methods to increase classification accuracy and robustness. Combining several learning algorithms helps the ensemble model to efficiently use deep characteristics, allowing it to shine in performance criteria and proving its ability to offer thorough and consistent classifications.

TABLE IV. COMPARISON OF TESTING PERFORMANCE OF DIFFERENT MODELS

Classifier	Accuracy (%)	Sensitivity (%)	F1-score (%)
SVM	79.44	72.39	84.26
DT	85.22	81.23	85.42
RF	79.88	80.45	78.55
KNN	70.51	72.16	70.63
LR	82.73	76.29	80.17
DT	87.54	82.64	86.01
Ensemble deep learning	97.18	97.47	98

V. DISCUSSION

Conventional models such as SVM, DT, RF, KNN, and LR displayed a spectrum of results, thereby stressing a fundamental fact: no model consistently performed better than others on all measures. DT and RF, for example, may not always ensure increased sensitivity, even if they are often excellent at processing nonlinear data. They may not be suitable for applications where failure to recognize true positives may have substantial effects. Conversely, models such as KNN lagged significantly on all metrics, even if they are straightforward and user-friendly, suggesting their inadequacy in handling complex or large data. The remarkable performance of the ensemble deep learning model highlighted the need to apply many techniques to increase prediction accuracy and robustness. From medical diagnoses to financial projections, this model's potential to achieve high scores in accuracy, sensitivity, and F1-score illustrates its usefulness in balancing precision and recall, which is a fundamental aspect in many pragmatic purposes. Combining numerous models or iterations helps the ensemble approach to clearly use the strengths of every component model and minimize their particular weaknesses. This performance disparity fuels a broader discussion of application-specific variables that influence the choice of the model. High accuracy might be given top attention in some circumstances, depending on the cost of false negatives or the necessity of balanced precision-recall trade-offs. In others, sensitivity or F1-score may be more crucial. Thus, the model choice should be driven not only by performance criteria but also by a whole understanding of the behavior of the model under several conditions. Furthermore, the results of this study reinforce the ongoing discussion of the evolution of algorithmic sophistication and its implications. As models become more complex, transparency and interpretability become more critical, highlighting a sector where simpler models may still be useful. The balance between simplicity for interpretability and complexity for speed will always determine how machine learning systems evolve and find applications.

In this framework, this study makes a significant contribution by adopting an ensemble learning strategy combined with deep features retrieved with ResNet50. This approach not only accepts the complexity inherent in EOG data but also utilizes the resilience provided by ensemble techniques, establishing a new standard in the classification accuracy and dependability of eye movement predictions. With performance measures well above those attained by traditional classifiers such as SVM, DT, and KNN, the proposed ensemble deep learning model exhibits extraordinary capability. For example, while works such as [30] show SVM with an accuracy of up to 76.9%, the ensemble model in this study mostly achieves higher accuracy, sensitivity, and F1 scores, all in the upper 90^{th} percentile. This performance jump emphasizes the success of combining deep learning with ensemble techniques by using the capabilities of several sophisticated algorithms to minimize the shortcomings of single models. Furthermore, using ResNet50 for deep feature extraction leverages the model's capacity to detect complex patterns in EOG data that less advanced methods could ignore. Especially in applications that require great accuracy, including assistive technologies where the precise interpretation of eye movements may greatly improve user engagement, this capacity is very vital. Furthermore, the application of ensemble techniques, such as bagging and boosting, improves the model stability and reduces the risk of overfitting. This feature is especially crucial in medical and HCI applications, where mistakes can be somewhat costly. Apart from offering protection against such hazards, the ensemble model ensures consistency in performance over several datasets and in different operational environments. Unlike the literature cited, which mostly addresses single parts of EOG signal processing or the effectiveness of particular classifiers, this study combines various elements into a coherent framework. In terms of classification accuracy and model dependability, this all-encompassing strategy not only pushes the edge but also provides a basis for further investigations into increasingly difficult practical uses of EOG technology.

Although the results of this study show great promise, certain constraints require attention and open the way for further investigation. The intrinsic complexity of ensemble learning and deep feature extraction methods is one major constraint that can make model interpretability difficult. Understanding the particular contributions of various elements or decisions inside a model becomes more challenging as the models become more complicated, which might provide difficulties in therapeutic or sensitive settings when explainability is absolutely important. Future research should focus on improving the scalability and efficiency of these models. This can entail the creation of more simplified deep learning architectures, lowering processing requirements while preserving high accuracy. Moreover, studies might look at ways to make deep learning models more interpretable, perhaps by combining ideas from the discipline of explainable artificial intelligence (XAI). This would not only help to make the models more transparent but also more trustworthy, particularly in applications requiring significant decision-making.

VI. CONCLUSION

This study introduced a novel ensemble deep learning model for classifying EOG signals, demonstrating significant advances in biometric signal processing. The proposed ensemble deep learning model outperformed traditional machine learning classifiers, achieving an accuracy of 97.18%, a sensitivity of 97.47%, and an F1-score of 98%. In comparison, other models such as DT (87.54% accuracy), SVM (79.44% accuracy), and RF (79.88% accuracy) showed notably lower performance. This better performance

underscores the advantages of applying sophisticated ensemble methods that combine multiple learning algorithms to enhance predictive accuracy and robustness. The superior performance of the ensemble deep learning model highlights its ability to efficiently leverage deep features, providing thorough and consistent classifications across various testing scenarios. Unlike traditional classifiers, which often struggle with the complexity and variability of EOG signals, the ensemble approach capitalizes on the strengths of multiple models, reducing variance and improving the overall stability of the predictions. Despite challenges with model complexity and computational requirements, the benefits of this approach emphasize its potential to expand HCI technologies, particularly for people with mobility limitations. Future work should aim to streamline the model to reduce computational load and improve interpretability, making it more accessible for real-time applications. Additionally, efforts to integrate other biometric signals could further amplify the model's utility in clinical and consumer technology domains. In general, this study sets a new standard in EOG signal classification, offering both technical advances and a practical framework for future research in this field.

REFERENCES

[1] P. A. Constable, "The Clinical Electro-Oculogram," in *Ophthalmic Diagnostics: Technology, Techniques, and Clinical Applications*, T. Das and P. Satgunam, Eds. Singapore: Springer Nature, 2024, pp. 449–461.

[2] K. S. Park, "Electrical Signals from the Muscles and Nerves," in *Humans and Electricity: Understanding Body Electricity and Applications*, K. S. Park, Ed. Cham, Switzerland: Springer International Publishing, 2023, pp. 199–222.

[3] F. D. Pérez-Reynoso, L. Rodríguez-Guerrero, J. C. Salgado-Ramírez, and R. Ortega-Palacios, "Human–Machine Interface: Multiclass Classification by Machine Learning on 1D EOG Signals for the Control of an Omnidirectional Robot," *Sensors*, vol. 21, no. 17, Aug. 2021, Art. no. 5882, https://doi.org/10.3390/s21175882.

[4] C. Belkhiria, A. Boudir, C. Hurter, and V. Peysakhovich, "EOG-Based Human–Computer Interface: 2000–2020 Review," *Sensors*, vol. 22, no. 13, Jun. 2022, Art. no. 4914, https://doi.org/10.3390/s22134914.

[5] A. Fischer-Janzen, T. M. Wendt, and K. Van Laerhoven, "A scoping review of gaze and eye tracking-based control methods for assistive robotic arms," *Frontiers in Robotics and AI*, vol. 11, Feb. 2024, https://doi.org/10.3389/frobt.2024.1326670.

[6] H. Mulam, M. Mudigonda, B. P. S. Kumar, and H. Kuchulakanti, "Electrooculogram Based Wheelchair Control in Real-Time," presented at the Second International Conference on Emerging Trends in Engineering (ICETE 2023), Nov. 2023, pp. 55–67, https://doi.org/10.2991/978-94-6463-252-1_8.

[7] Ľ. Hládek, B. Porr, and W. O. Brimijoin, "Real-time estimation of horizontal gaze angle by saccade integration using in-ear electrooculography," *PLOS ONE*, vol. 13, no. 1, 2018, Art. no. e0190420, https://doi.org/10.1371/journal.pone.0190420.

[8] M. Toivanen, K. Pettersson, and K. Lukander, "A probabilistic real-time algorithm for detecting blinks, saccades, and fixations from EOG data," *Journal of Eye Movement Research*, vol. 8, no. 2, Jun. 2015, https://doi.org/10.16910/jemr.8.2.1.

[9] G. Alimjan, T. Sun, Y. Liang, H. Jumahun, and Y. Guan, "A New Technique for Remote Sensing Image Classification Based on Combinatorial Algorithm of SVM and KNN," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 32, no. 07, Jul. 2018, Art. no. 1859012, https://doi.org/10.1142/S0218001418590127.

[10] R. A. Nugrahaeni and K. Mutijarsa, "Comparative analysis of machine learning KNN, SVM, and random forests algorithm for facial expression classification," in *2016 International Seminar on Application for Technology of Information and Communication (ISemantic)*, Semarang, Indonesia, Aug. 2016, pp. 163–168, https://doi.org/10.1109/ISEMANTIC.2016.7873831.

[11] P. K. Mall *et al.*, "A comprehensive review of deep neural networks for medical image processing: Recent developments and future opportunities," *Healthcare Analytics*, vol. 4, Dec. 2023, Art. no. 100216, https://doi.org/10.1016/j.health.2023.100216.

[12] R. Archana and P. S. E. Jeevaraj, "Deep learning models for digital image processing: a review," *Artificial Intelligence Review*, vol. 57, no. 1, Jan. 2024, Art. no. 11, https://doi.org/10.1007/s10462-023-10631-z.

[13] R. Sharma and H. K. Meena, "Emerging Trends in EEG Signal Processing: A Systematic Review," *SN Computer Science*, vol. 5, no. 4, Apr. 2024, Art. no. 415, https://doi.org/10.1007/s42979-024-02773-w.

[14] Babita, P. Syal, and P. Kumari, "Comparative Analysis of KNN, SVM, DT for EOG based Human Computer Interface," in *2017 International Conference on Current Trends in Computer, Electrical, Electronics and Communication (CTCEEC)*, Mysore, India, Sep. 2017, pp. 1023–1028, https://doi.org/10.1109/CTCEEC.2017.8455122.

[15] Y. Paul, V. Goyal, and R. A. Jaswal, "Comparative analysis between SVM & KNN classifier for EMG signal classification on elementary time domain features," in *2017 4th International Conference on Signal Processing, Computing and Control (ISPCC)*, Solan, Sep. 2017, pp. 169–175, https://doi.org/10.1109/ISPCC.2017.8269670.

[16] L. J. Qi and N. Alias, "Comparison of ANN and SVM for classification of eye movements in EOG signals," *Journal of Physics: Conference Series*, vol. 971, no. 1, Nov. 2018, Art. no. 012012, https://doi.org/10.1088/1742-6596/971/1/012012.

[17] L. Y. Deng, C.-L. Hsu, T.-C. Lin, J.-S. Tuan, and S.-M. Chang, "EOG-based Human–Computer Interface system development," *Expert Systems with Applications*, vol. 37, no. 4, pp. 3337–3343, Apr. 2010, https://doi.org/10.1016/j.eswa.2009.10.017.

[18] A. G. A. Abdel-Samei, A. S. Ali, F. E. A. El-Samie, and A. M. Brisha, "Efficient Classification of Horizontal And Vertical EOG Signals For Human Computer Interaction." Research Square, Jun. 02, 2021, https://doi.org/10.21203/rs.3.rs-471385/v1.

[19] S. Aungsakul, A. Phinyomark, P. Phukpattaranont, and C. Limsakul, "Evaluating Feature Extraction Methods of Electrooculography (EOG) Signal for Human-Computer Interface," *Procedia Engineering*, vol. 32, pp. 246–252, Jan. 2012, https://doi.org/10.1016/j.proeng.2012.01.1264.

[20] T. Wissel and R. Palaniappan, "Considerations on Strategies to Improve EOG Signal Analysis," *International Journal of Artificial Life Research (IJALR)*, vol. 2, no. 3, pp. 6–21, Jul. 2011, https://doi.org/10.4018/jalr.2011070102.

[21] A. Banerjee, M. Pal, S. Datta, D. N. Tibarewala, and A. Konar, "Eye movement sequence analysis using electrooculogram to assist autistic children," *Biomedical Signal Processing and Control*, vol. 14, pp. 134–140, Nov. 2014, https://doi.org/10.1016/j.bspc.2014.07.010.

[22] F. E. Samann and M. S. Hadi, "Human to television interface for disabled people based on EOG," *Journal of Duhok University*, vol. 21, no. 1, pp. 54–64, 2018.

[23] J. Tsai, C. Lee, C. Wu, J. Wu, and K. Kao, "A feasibility study of an eye-writing system based on electro-oculography," *Journal of Medical and Biological Engineering*, vol. 28, no. 1, pp. 39–46, 2008.

[24] R. Barea, L. Boquete, L. M. Bergasa, E. López, and M. Mazo, "Electro-Oculographic Guidance of a Wheelchair Using Eye Movements Codification," *The International Journal of Robotics Research*, vol. 22, no. 7–8, pp. 641–652, Jul. 2003, https://doi.org/10.1177/02783649030227012.

[25] N. Barbara, T. A. Camilleri, and K. P. Camilleri, "EOG-based eye movement detection and gaze estimation for an asynchronous virtual keyboard," *Biomedical Signal Processing and Control*, vol. 47, pp. 159–167, Jan. 2019, https://doi.org/10.1016/j.bspc.2018.07.005.

[26] J. Heo, H. Yoon, and K. S. Park, "A Novel Wearable Forehead EOG Measurement System for Human Computer Interfaces," *Sensors*, vol. 17, no. 7, Jul. 2017, Art. no. 1485, https://doi.org/10.3390/s17071485.

[27] M. Merino, O. Rivera, I. Gomez, A. Molina, and E. Dorronzoro, "A Method of EOG Signal Processing to Detect the Direction of Eye Movements," in *2010 First International Conference on Sensor Device*

Technologies and Applications, Venice, Italy, Jul. 2010, pp. 100–105, https://doi.org/10.1109/SENSORDEVICES.2010.25.

[28] Z. Lv, Y. Wang, C. Zhang, X. Gao, and X. Wu, "An ICA-based spatial filtering approach to saccadic EOG signal recognition," *Biomedical Signal Processing and Control*, vol. 43, pp. 9–17, May 2018, https://doi.org/10.1016/j.bspc.2018.01.003.

[29] K. S. Roy and S. Md. R. Islam, "An RNN-based Hybrid Model for Classification of Electrooculogram Signal for HCI," *International Journal of Computing*, pp. 335–344, Oct. 2023, https://doi.org/10.47839/ijc.22.3.3228.

[30] S. N. Hernández Pérez, F. D. Pérez Reynoso, C. A. G. Gutiérrez, M. D. los Á. Cosío León, and R. Ortega Palacios, "EOG Signal Classification with Wavelet and Supervised Learning Algorithms KNN, SVM and DT," *Sensors*, vol. 23, no. 9, Jan. 2023, Art. no. 4553, https://doi.org/10.3390/s23094553.

[31] N. Barbara, T. A. Camilleri, and K. P. Camilleri, "Real-time continuous EOG-based gaze angle estimation with baseline drift compensation under non-stationary head conditions," *Biomedical Signal Processing and Control*, vol. 90, Apr. 2024, Art. no. 105868, https://doi.org/10.1016/j.bspc.2023.105868.

[32] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 770–778, https://doi.org/10.1109/CVPR.2016.90.

[33] A. Miltiadous *et al.*, "An Ensemble Method for EEG-based Texture Discrimination during Open Eyes Active Touch," *Engineering, Technology & Applied Science Research*, vol. 14, no. 1, pp. 12676–12687, Feb. 2024, https://doi.org/10.48084/etasr.6455.