Eye Blink Rate Based Detection of Cognitive Impairment Using In-the-wild Data

Fatimah Alzahrani[1,4], Bahman Mirheidari[1], Daniel Blackburn[2], Steve Maddock[1], and Heidi Christensen[1,3]

[1]*Department of Computer Science,* [2]*Sheffield Institute for Translational Neuroscience (SITraN),* [3]*Centre for Assistive Technology and Connected Healthcare (CATCH), University of Sheffield, Sheffield, UK*

[4]*Department of Computer Science, Umm Al-Qura University, Mecca, SA*

{faalzahrani1, heidi.christensen, s.maddock}@sheffield.ac.uk

***Abstract*—Investigating automatic methods for the early detection of dementia and related conditions that cause cognitive impairment is an area of growing interest. Video processing could play a role by providing a non-invasive and low-cost alternative to current expensive assessments. For this to be successful it is crucial that approaches are robust to in-the-wild challenges. In this paper, visual cues, related to the eye blink rate (EBR), are investigated to quantify the early phase of neurodegenerative disorder (ND) and mild cognitive impairment (MCI) as well as functional memory disorder (FMD; problems with memory not related to neurodegenerative disorder). This paper aims to improve the detection of ND and MCI by investigating a novel approach to calculating the EBR that is more robust to in-the-wild challenges. An in-house dataset with 18 participants is used. The EBR is calculated from eye landmarks extracted using two libraries (Dlib and Openface). To mitigate issues observed in the noisy, in-the-wild recordings, a multiple threshold approach for EBR detection is proposed. It involves generating multiple thresholds for identifying a blink, where a threshold is used to determine whether an eye is open or closed. Several supervised machine learning approaches are used for automatic classification. The results show that accuracy measures of 89% and 78% are achieved using Dlib and OpenFace data, respectively, when distinguishing between three conditions with ND, MCI and FMD.**

***Index Terms*—cognitive impairment, dementia, eye blink rate**

I. Introduction

Dementia is a growing socio-economic challenge. Currently 50 million people have dementia, and this is predicted to rise to 152 million by 2050 [1]. Diagnosing dementia is a complex and costly process requiring neurological expertise. There is therefore growing interest in investigating automatic methods for detecting early signs of dementia and other memory problems. With the increasing prevalence of devices with cameras, exploring the use of video-based cues such as eye blink rate (EBR) could help pave the way for easy-to-use, low-cost, home-based assessments.

Neurodegenerative disorder (ND), including dementia, is a deterioration in cognitive function caused by cell death in a particular area of the brain. It significantly influences people's ability to communicate and express their feelings [2]. Moreover, it affects their personality, and people may lose independence in daily life. Mild cognitive impairment (MCI) is a deterioration in cognitive function but without the major functional problems associated with dementia. It does not affect everyday activities but exhibits a high probability of progression to dementia [2]. This paper investigates the use of EBR to help distinguish people with ND and MCI from those with FMD (people with memory problems not related to dementia) because each group has different way of treatments. For instance, ND, which is often Alzheimer's disease (AD), needs treatment and care while MCI could develop into AD and thus needs regular monitoring appointments. However, FMD could be dealt with as anxiety or depression because it is not related to cognitive impairment problem.

Diagnosing early signs of ND and MCI is complex and typically involves a range of methods, such as medical and collateral history, physical examination, cognitive assessment tests (e.g. MoCA and MMSE), MRI, lab tests, and neurological examinations [3], [4]. Traditional testing is subjective and dependent on specialists who already have significant training and practice. Therefore, previous studies have investigated ND and MCI characteristics that might lead to an efficient objective detection system that could hopefully aid diagnosis. This study's approach uses data recorded by common consumer devices, such as webcams. These kinds of devices are available in people's homes and may be advantageous to use as people would need fewer visits to clinics or hospitals and a reduced number of invasive diagnostic procedures. Establishing an early diagnosis could allow for treatment before a condition reaches an irreversible phase. However, for approaches to be successful they need to be able to handle the less homogeneous nature of the collected data, typically not found in datasets recorded in controlled lab settings.

Facial behaviour is intrinsically linked to cognitive processes. Some studies have examined social communication skills to evaluate mental state [5] using facial behaviour signs such as facial expressions, gaze, and eye blinks [6]. Research has shown that EBR can be a sign of MCI and ND [7], motivating work on developing automatic diagnostic aids based on visual features (e.g., [8], [9]; more details in Section II). An advantage of using visual symptoms is that it could be language independent. People who have migrated to another country often forget a new language when suffering from cognitive impairment, and they generally revert to their mother tongue [10]. In these cases, using visual cues instead of language-based approaches could help with diagnosing certain

978-1-6654-0019-0/21/$31.00

health conditions.

This paper investigates the use of EBR for distinguishing between different types of cognitive impairment: ND, MCI and FMD. We use a dataset that has been recorded in a real-world scenario (*in-the-wild*). As a result, participants are seen to change their distance to the camera and sit in non-optimal positions, the illumination changes, and there is background noise as other people appear in the view. This makes the detection of EBR more challenging as it relies on measuring the ratio between the eye's width and the openness of the eye (Eye Aspect Ratio, EAR) and detecting, as a blink, regions where the EAR is below a certain threshold. For controlled datasets, where the participant is relatively still in the frame, identifying what the threshold level should be is more straightforward, however, this will not be sufficient for in-the-wild recordings like those in the dataset used, and obtaining an accurate EBR is challenging. We present a solution that uses a novel approach based on applying multiple thresholds for EBR calculation. This results in every participant having a vector of blink rates calculated corresponding to a given range of thresholds. This is subsequently used as a feature vector to train a classifier for cognitive impairment detection. We investigate the use of two eye tracking libraries and demonstrate classification accuracies comparable to those achieved using one-modal (audio/video) or multi-modal approaches [8], [9], [11]–[13].

The remaining sections of this paper are structured as follows. Section II discusses related work. Section III describes the dataset and the proposed method and how this method addresses realistic conditions such as those in our dataset. Section IV presents the experiments and the analysis of the results. Finally, Section V presents the conclusions.

II. Related Work

Several studies have investigated automatic methods for detecting health conditions, such as early-stage of dementia or MCI, using visual signs [7], [8], [11], [14].

A. Eye Blink and Cognition

In general, non-verbal behaviour is considered a type of human communication and a continuous signal that provides important information about people's feelings, personalities, and mental state [5]. The eyes contribute significantly to nonverbal behaviour in social communication, for instance, eye blink has received significant attention. A relation has been found between eye blink and cognitive state in terms of brain activity [15], [16]. Spontaneous eye blinks reflects cognitive states. For instance, a person's blink rate increases during certain activities, such as speaking (for adults) [17], conversation [18], memorising [19], stress, positive mood and emotions, fatigue, pain, physical activity, disease, and when expressing anger or excitement [20]–[23]. In contrast, decreases in a person's blink rate can be seen during visual tracking and reading activities [19], [24]. Spontaneous eye blink is an unconscious expression that is related to attention and transmitted by frontal, parietal, and temporal cortical structures [25]. Moreover, a person's EBR plays a significant role in eye movements, fixations, emotional expressions, and visual cognition [26], [27]. Eyelid movements, which can be either voluntary or involuntary, occur as a reflex to sensory stimulation. Aging affects the average blink rate by increasing from about 24/minute at age 40 to 49 years to 32/minute at age 80 to 89 years [22]. In addition, environment-related factors may affect the blink rate and duration, such as temperature, brightness, air conditions, and relative humidity [22]. Several studies have focused on studying eye blink in the context of conditions affecting cognition [23], [28].

Investigations on EBR show that it is correlated with a person's cognitive state, and it can change based on various cognitive conditions. Therefore, a person's EBR could be an early indicator of certain cognitive conditions, including dementia and MCI [7], [29]. A reduced rate has also been observed in patients with Parkinson's disease, which also affects cognition and may lead to a type of dementia [30]. Reference [7] showed that people with MCI have a higher rate of eye blink than healthy people do, and an increase in their EBR may indicate a transition from MCI to an early stage of dementia.

B. Visual Dementia Data

Previous studies have attempted to investigate dementia detection by extracting audiovisual features [8], [11], [31], [32]. In [11], they recorded 18 participants and found some significant acoustic features, but the smile ratio feature was not a significant feature. However, in another study, [8] recorded their data from 29 participants finding significant features, such as smile ratio and some acoustic features, for the detection system. A study by [9] investigated using facial features from OpenFace to detect dementia with 24 participants and found important features, such as lip activity, facial action units, and eye gaze. Notably, they used advanced technology to record the audio and video in a controlled environment. Such recording processes increase the cost, effort, and recording time needed and do not represent real-world scenarios.

C. Eye Blink Detection Techniques

A range of approaches to detecting eye status (open or closed) have been proposed. For example, [33] used features that were extracted by a scale-invariant feature transform and the histogram of oriented gradients with a classification stage. Their research showed a lower accuracy score in some conditions, such as skin colour variations, illumination and head movements. Convolutional neural networks have been investigated to overcome previous limitations by extracting effective features [34], [35] and estimating the EBR, even when the person is not facing forward [36]. Several methods have been investigated the detection and tracking through videos rather than processing a single image to overcome previous methods' limitation [33], [37], [38]. An eye blink detection approach using a state machine was proposed with a scheme for merging the left and right eye blink frames [39]. They further improved the performance by extracting features using dense optical flow and sending them to a recurrent neural network [40]. In [38], an algorithm was developed for finding the eye region landmarks and computing the EAR, which was then used with a support vector machine to determine whether the eye was open or closed. In addition, a study by [41] was carried out to detect whether the eye was open or closed in each frame by using a pre-trained convolutional neural network to label the eye region on a few hundred annotated images.

To validate if the eye closure is a blink or not based on the length of the closure, previous studies have used a state machine (SM) [39], [40], [42], [43]. A number of studies have reported that eye blink usually lasted from 100ms to 400ms [44] and 50ms to 400ms [21]. When the eye blink is partially closed it is considered an incomplete blink [45]. Fully closed eyes that last from 70ms to 1s are called extended blinks [46]. People may have multiple blinks in the same sequence. In this study, the SM parameters used will be described in Section III-D.

D. Eye Blink Datasets

To evaluate the performance of eye blink detection approaches, datasets recorded in-the-wild should include many samples of different eye status cases with a high variation in head orientation, a changeable distance from the camera over time, low illumination and very low resolution images or videos. Several common datasets, such as ZJU [47], TalkFace[1] and EyeBlink8 [42], were not recorded in-the-wild. Most of the available datasets were recorded with the participant facing the camera. The Researcher's Night dataset [40] is considered challenging because it includes some head movements. Those head movements do not show high head-angle variations, whereas this study's data does.

In summary, research has shown that EBR is a robust and relevant cue for ND/dementia and MCI detection. This paper proposes an approach to distinguish between ND, MCI, and FMD automatically by using EBR extracted from data that were recorded in a real-world scenario (i.e., in-the-wild). Such data involve a number of challenges which will be described in more detail in Section III-A2.

III. Methods

This section presents the full experimental setup, including the data and its inherent challenges, as well as describing the classification pipeline and the proposed new method for extracting a more robust measure for EBR.

A. Data

1) Task and Participants: This study uses data that were recorded at Hallamshire Hospital Memory Clinic in Sheffield, UK. This data include videos, audio recordings, and diagnostic details of people with different types of cognitive impairment (FMD, MCI and ND) as they answer memory-probing questions by an intelligent virtual agent (IVA). The questions were different types: open questions, closed questions, and combined questions to assess participant's long and short term memory. Ethical approval for collecting and using this data was given by the National Research Ethics Service (NRES) Committee South West-Central Bristol (Rec number 16/LO/0737) in May 2016. This data cannot be shared due to the ethical guidelines. For full details of the data, please see [13], [48], [49].

A total of 24 participants took part, and 18 participants out of 24 were used, split equally into 6 with ND, 6 with MCI, and 6 with FMD. We excluded 4 with depressive pseudodementia and 2 for whom the diagnosis was not clear. The total duration of all the videos is 208 minutes, and the average is 11 minutes and 56 seconds. The dataset is small, but this is a common issue in studies that involve human participants in clinical settings, as described in Section II. The participants were told that they could bring someone with them and, as a result, 6 of the 18 participants brought a caregiver/partner with them (4 ND, 1 FMD, and 1 MCI). Therefore, some videos contain four people: the participant, the accompanying person, the neurologist, and the person who operates the laptop. Although participants were not given any specific instructions as to where to look, the talking head on the screen will have been the most salient point on which to look. This obviously poses a challenge for video-based processing. The study under which the recordings were done was mostly focusing on speech processing, and this paper presents the first research done using the videos.

2) Data Challenges: Data recorded in-the-wild can contain a high level of noise due to the lack of restrictions on the participants and the environment with respect to the webcam position. In-the-wild conditions include a semi-dark or dark, noisy room. In addition, spontaneous behaviour means that participants may act as they would in their natural environment, such as moving about freely. A participant may continually change the orientation of their face, rotate their body, and move closer to and further away from the camera. Other people may also appear with the participant and move around too. In addition, participants who wear glasses sometimes have their eyes obscured by the frames or a reflection from the laptop on the glasses. The majority of the IVA data recordings were recorded at 30fps. However, five of the recordings were done at 24fps, which also produced a different resolution recording. These issues cause complications for automatic methods to extract visual information from the data.

Some of the standard datasets recorded in lab-controlled environments try to address certain aspects of real-world scenarios. For instance, the ZJU [47] corpus was developed to address certain situations, such as sitting in front of the camera, both with and without glasses, and with an upward view. The recordings had a resolution of 320x240, and the videos were mostly captured at 30fps. Fig. 1 shows the plot of the EAR for two participants: one in the ZJU dataset (left) and one in the IVA dataset (right). The details of how the relevant feature is extracted to produce the figure are given in Section III-C. From Fig. 1 (left), it is clear that this participant had three blinks where the EAR drops down from the steady level around 0.3.

[1]http://www-prima.inrialpes.fr/FGnet/data/01-TalkingFace/talking_face.html

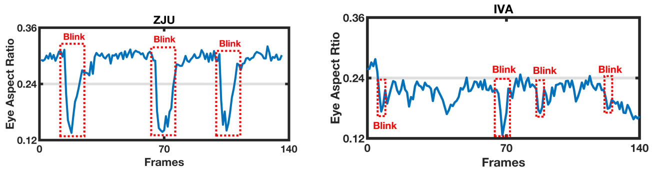

Fig. 1: EAR values plotted for two recordings of participants in ZJU and the IVA data, respectively.

As such, using one threshold to detect an eye blink should work reliably. For example, assuming a threshold of 0.2, three blinks would be correctly identified. In comparison, Fig. 1 (right) shows the calculated EAR for a random participant from the IVA data. Red boxes are used to indicate the location of real eye blinks. However, several additional dips in the EAR values can also be seen, representing false blinks. These look like a blink due to the video's low resolution, particularly in the eye region, low illumination, and because an animation was played on a monitor positioned behind the participant. Another difference to the more homogenous ZJU data is the fact that the base mean of the EAR value does not appear to have a fixed value, but is instead seen to fluctuate. The challenges of the in-the-wild dataset can affect visual feature detection, such as the EBR. In addition to the aforementioned challenges, the fact that these are recordings of people with health conditions also means that factors such as the blink speed, length and frequency may vary dependent on the participant. In people with health conditions, such as ND and MCI, false blinks may be more prevalent due to the increased head and body movements and the challenges mentioned earlier [39], [40]. This may lead to an increase in the variability of the EAR values through the video.

B. Data Pre-processing

Before extracting the EAR features, the videos need pre-processing due to some people's appearance with the participant in front of the camera. The height and width of the video frames are cropped. To this end, two different computer vision techniques implemented in the Dlib[2] and the OpenFace[3] libraries are investigated to evaluate the performance of the proposed method on the IVA data. OpenFace is known to be very efficient in predicting facial landmarks for data recorded in-the-wild [50], whereas Dlib shows high accuracy for recordings with faces that are frontal, slightly non-frontal, and have small occlusions [51].

C. Feature Extraction

This section describes the EAR and the EBR extraction in detail, including determining whether an eye closure is a blink or not, based on a state machine.

$$EAR = \frac{\|p2 - p6\| + \|p3 - p5\|}{2\|p1 - p4\|} \tag{1}$$

[2]http://dlib.net

[3]https://github.com/TadasBaltrusaitis/OpenFace/

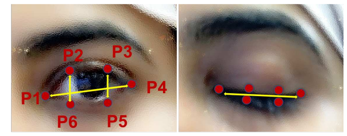

Fig. 2: Eye landmarks detected by Dlib.

Fig. 2 shows the six eye (x,y) coordinates. These are used to calculate the EAR using (1) and the approach in [38]. Equation (1) is used to calculate the EAR for both eyes for each frame. The average of both eyes' EAR is used. The conventional approach is to compare this averaged EAR with a particular threshold to determine whether it is a frame where the eye is closed (EAR is lower than the threshold) or not. As is common in similar approaches, a state machine is then used to determine if this signals a true blink (defined as the eye closure being longer than a certain number of consecutive frames) or a false one.

D. State Machine

In the context of investigating the use of EBR as a feature for classifying different types of cognitive impairment, it is not clear what would make for a good choice of eye closure length in the state machine. In this study, different values for the blink duration were therefore explored:

- Type 1 - one frame or any number of consecutive frames having EAR values below the threshold will be considered a blink.
- Type 2 - a sequence of two frames or more being below the threshold.
- Type 3 - a sequence of between two and 30 frames, inclusive, being below the threshold. The range corresponds to approximately 60ms to 1s due to patients who may have a long eye blink. When frames per second are lower than 30, they may not show a complete blink.

Two different approaches to calculating the EBR are investigated: i) calculating the EBR by an automatic setting of a single threshold (*baseline*) and ii) a novel approach that uses multiple thresholds. Each approach is tested using the two different facial landmark libraries (Dlib and OpenFace) and the three state machines (Type 1, 2 and 3).

E. Automatic Calculation of a Threshold (Baseline System)

To construct a baseline system, a simple approach that determines an appropriate threshold and is participant-dependent is used. The approach depends on the mean (μ) and the standard deviation (σ). A blink is detected when the averaged EAR is below the μ of the EAR minus half the σ. This method leads to one threshold or one EBR feature for the whole video for each participant.

This approach has drawbacks in the case of data recorded in-the-wild due to several challenges, as described in Section III-A2. These issues also exist in the used IVA data as illustrated in Fig. 3. They show the EAR values through the entire video of two random participants. For example, for participant P13, the EAR values ranged between 0.05–0.5 on the

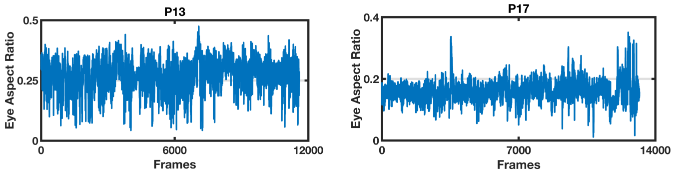

Fig. 3: EAR calculated for ND participant (P13) and FMD participant (P17).

y-axis, whereas the values for participant P17 ranged between 0–0.3. Both of the participants' EAR values show a large degree of noise in the signal as the up and down fluctuations that corrupt the quality of the signal make detecting the blinks very difficult, and the mean changes for each participant over time, especially for the ND participant P13.

F. Multiple Thresholds Approach

We propose a novel approach that can detect the EBR using multiple thresholds regardless of the noise level. This approach helps to address the impact of tiredness or dryness changes on the calculations of the EAR and the EBR values. The investigation of this approach involves generating many thresholds of the whole video for each participant. The range of the thresholds is determined by finding the maximum and minimum EAR for each of the participants. Two different facial landmark detection libraries – Dlib and OpenFace – are used in evaluating the method. Fig. 4 shows the processing pipeline.

Dlib Landmarks: Multiple thresholds are generated within a particular range. With a minimum of 0.0 and a maximum of 0.7, a step size of 0.1 gives 0.0, 0.1, 0.2, ..., 0.6 which gives 7 thresholds. A step size of 0.01 between 0.0 and 0.7 gives 70 thresholds. A step size of 0.001 between 0.0 and 0.7 gives 700 thresholds. These different number of thresholds are resulted in 7, 70, 700 blink rates features.

OpenFace Landmarks: The multiple threshold range starts from 0.0 to 33.85, which is rounded to 34. As for Dlib, the range 0.0–34 is divided into 7, 70, and 700 thresholds and results in 7, 70, and 700 features for each participant. When OpenFace outputs high values, these are assumed to indicate head turns or movements, occluded faces, or any issue from the above-mentioned challenges, with Dlib subsequently losing facial landmarks tracking. Fig. 5 confirms this assumption from the calculated EAR using Dlib (orange) and OpenFace (blue). Two high values are indicated with a red rectangle, and it can be seen that Dlib does not detect landmarks during a number these frames in the middle of the rectangle.

These extreme high values are considered outliers. We apply the standard deviation (σ) method to identify these. In many data mining applications, outlier detection is commonly used to detect and remove or ignore the anomalous data points from the data [52]. This method considers any point greater than $(\mu + (3 \times \sigma)$ as an outlier. After calculating the outlier for each participant, the minimum outlier is 0.65, which is rounded to 0.7, making the new range for all participants to generate multiple thresholds from 0.0–0.7 (previously 0.0–34). Any number above 0.7 is considered an outlier.

G. Classification

Different supervised machine learning classifiers are investigated: a support vector machine with linear (L-SVM) and rbf (rbf-SVM) kernels, logistic regression (LR), k-nearest neighbour (kNN), and decision tree (DT). A grid-search approach is used to improve the classification accuracy by tuning the parameters C of each used classifier with Scikit-learn by trying a combination of C and other parameters of each classifier using subject-independent-Stratified cross-validation (CV) with 3-fold. The parameters with highest CV accuracy are estimated in each fold, then the average is taken across all folds for SVM with linear and rbf kernels, C=2000, and C=10, respectively. For LR parameters, multi-class is auto and C=2000. For kNN, the number of neighbours is 8 with weight = uniform. For DT, the min-sample-split = 3 and 11 for 3-class and 2-class classification, respectively. The classification involves a binary classification (ND/MCI, ND/FMD, and MCI/FMD) and a multi-class classification (ND/MCI/FMD). For system performance evaluation, we use the accuracy metric and confusion matrix to analyse the predictions.

IV. Experimental Results

A. Baseline System

The single automatic threshold approach was first evaluated on ZJU, a standard dataset with ground-truth annotations of blinks. Due to the known issues around achieving a common ground-truth set of annotations for this dataset, we used the annotations provided by [39] and visually inspected the files in addition, resulting in a further 4 blinks (at the beginning and ends of files) being added to this annotation, resulting in a total of 265 blinks. The eye blink is calculated by taking the average of both eyes. The blinks at the beginning and the end of the videos are counted as two different blinks. In addition, two double blinks are considered as two different blinks. This follows the approaches in [39], [42]. The detected-blinks =271, true-positive = 265 blinks, false-positive = 6 blinks, and false-negative = 0, which give 0.925 as f-measure.

Table I shows a comparison between our baseline method and related work using F-measure as a metric to measure the approach's performance on ZJU. However, any comparison is complicated by the fact that the evaluation procedure differs in each study. Our baseline score is based on 79 out of the 80 videos because one of the participant's face angle is up in the video, which makes it difficult for Dlib to detect facial landmarks from the frames. The motivation behind the one-threshold approach is to find a simple approach with good performance to construct a baseline and compare it with the multiple threshold approach. The performance of the method used is 0.925, which is considered acceptable to be used on the IVA data as a baseline system.

Three types of SM parameters are investigated to distinguish between ND, MCI, and FMD classes. The results of testing the baseline and multiple thresholds approaches on the IVA dataset are shown in Tables II and III, respectively. From

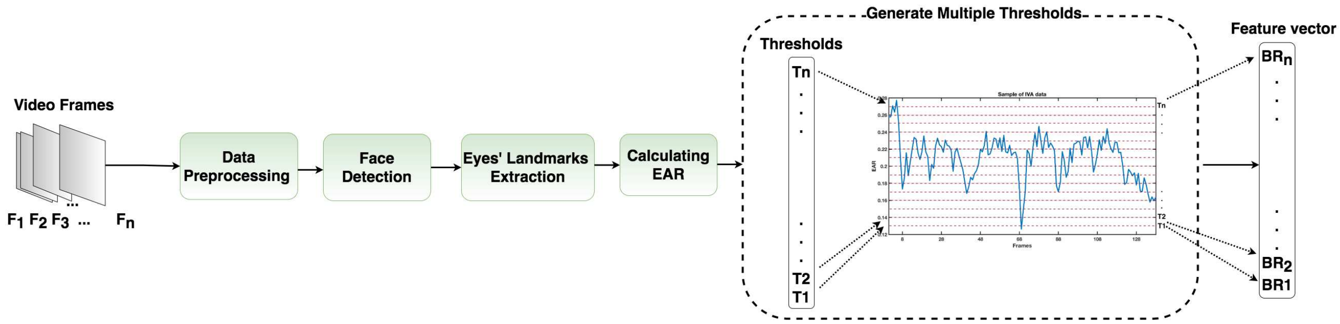

Fig. 4: Pipeline of the multi-threshold EBR extraction approach for each participant. Multiple thresholds (T_n) of the whole video for a participant are calculated, together with a blink rate for each threshold (BR_n).

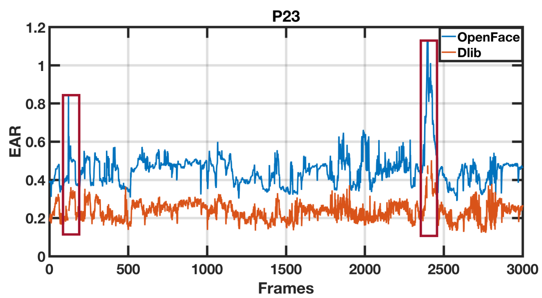

Fig. 5: EAR calculated for FMD participant (P23) using Dlib (orange) and OpenFace (blue).

TABLE I: EBR detection performance on ZJU (F-measure) for baseline method and related work.

Study	F-measure
[53]	0.992
[38]	0.952
[35]	0.937
[40]	0.976
Baseline	0.925

Table II, when only one threshold is used to calculate the EBR (baseline), the 2-class classification gives better results than the 3-class classification in Dlib and OpenFace. Distinguishing the ND/MCI and MCI/FMD groups using one threshold gives only the chance-level, 58% and 50%, respectively. These two diagnostic classes are challenging, and it is not easy to distinguish between them even in the clinic [54]. For the 3-class classification, the achieved accuracy is 60% using Dlib landmarks compared to 50% using OpenFace landmarks. The type of state machine does not show a big difference in the results.

B. Classification Results

Then, the classification of the proposed multiple thresholds approach is carried out on multi-class and binary classifications: ND/MCI/FMD, ND/MCI, ND/FMD, and MCI/FMD as shown in Table III. It can be seen from the table that when the multiple thresholds approach is applied, the highest obtained accuracy of Dlib and OpenFace landmarks are 88% and 78%, respectively. In Dlib, types 2 and 3 of the state machine and 70 features achieve the highest classification accuracy with

TABLE II: Classification accuracy in percentage (%) when using the baseline threshold approach using different classifiers: ($Linear{-}SVM^1, rbf{-}SVM^2, kNN^3, LR^4$, and DT^5) for IVA data.

SM	Technique	ND/MCI/FMD	ND/MCI	ND/FMD	MCI/FMD
Type 1	Dlib	56^{4}	$58^{1,4}$	$75^{1,3,4}$	$83^{1,2,4}$
	OpenFace	50^{4}	75^{4}	$75^{1,2,4}$	$50^{1,4,5}$
Type 2	Dlib	61^{5}	58^{4}	83^{3}	$83^{1,2,3,4}$
	OpenFace	$50^{3,4}$	67^{4}	$75^{2,4}$	$50^{1,3,4,5}$
Type 3	Dlib	$56^{3,5}$	$50^{1,2,3,4,5}$	83^{3}	$83^{1,3,4}$
	OpenFace	50^{4}	75^{4}	$75^{1,2,3,4}$	$50^{1,3,4,5}$

89%, 83%, 100%, and 92% for ND/MCI/FMD, ND/MCI, ND/FMD, and MCI/FMD classes using $L{-}SVM^1$. These results show a significant improvement from the baseline results. The DT^5 performed better using one EBR, whereas $L{-}SVM^1$ performed better when using multiple EBRs for each participant. There is a significant difference ($p<0.05$) between the baseline and the proposed method results. In OpenFace, type 3 of the state machine and 700 thresholds give the highest classification accuracy with 72%, 100%, 92%, and 92% for ND/MCI/FMD, ND/MCI, ND/FMD, and MCI/FMD classes, respectively, using DT^5. Interestingly, removing outliers based on STD improves only the 3-class classification by 78% with 7 thresholds for each participant. Comparing the proposed method's results with the baseline results in Table II, they show a significant enhancement of the accuracy scores. In the baseline results, the LR^4 achieves the highest accuracy. However, the DT^5 achieves better performance on multiple thresholds. There is a significant difference ($p<0.05$) between the baseline and the proposed method results of OpenFace. The multiple thresholds methodology is intuitively simple, but it provides efficient results.

In this study, the EBR is considered a very efficient feature on its own compared to the results of prior studies that have used visual and multi-modal features on limited data size, which were recorded in a controlled environment. They achieved 94% using the area under curve, 93% using SVM, and 82% using LR in dementia diagnosis after combining language, speech, and visual features [8], [9], [11]. We focus on distinguishing between people with ND, MCI, and FMD, unlike previous work that considered ND and MCI as one class. In comparison, [13] used the audio parts of this dataset

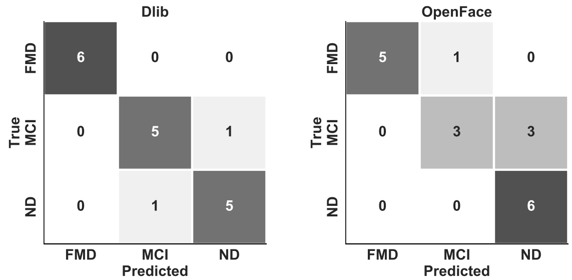

Fig. 6: Confusion matrix using DLib and OpenFace.

on ND/MCI/FMD, ND/MCI, ND/FMD, MCI/FMD and obtained 70%, 75%, 100%, and 81.25% accuracy using LR, respectively.

C. Discussion

As previously stated, we are interested in comparing the performance of two facial landmarks tracking techniques to handle data recorded in-the-wild. When only frontal or semi-frontal frames are detected, and the noisy frames are removed in Dlib, the detection of the cognitive impairment using EBR has improved the system's performance more than using all the frames in OpenFace when the end-goal is classification and not in itself an accurate EBR estimation. The high values of the EAR are assumed to be an indication of head movements, turns, or any one of the challenges in Section III-A2. Our results indicate that ignoring those frames using the outlier detection method improves only the 3-class classification, which is less than the obtained accuracy using Dlib landmarks.

A confusion matrix is used to analyse Dlib and OpenFace prediction results, as shown in Fig. 6. It shows that for both approaches, ND and FMD are predicted correctly. However, half of the MCI participants are predicted as ND when the OpenFace library is used for facial landmark tracking. These results demonstrate that classifying MCI from ND is the most challenging task in 2-class and 3-class classification. This may be due to the fact that people with MCI have an increased blink rate related to memory issues, which is even higher when the person is in the early phase of dementia [7].

V. Conclusions

This paper has presented a novel multiple threshold approach for EBR data that could be used for the early detection of dementia and related conditions that cause cognitive impairment. This has been applied to a dataset containing ND, MCI, and FMD participants. The results show that EBR is a significant cue to differentiate ND, MCI, and FMD from each other. As part of the approach, two different facial landmark tracking libraries – Dlib and OpenFace – were used. Dlib gave better results in a 3-class classification: 89% using Dlib data and 78% using OpenFace data. One limitation of this study is the limited data size. The aim is to address this in a future study. Further work will also be done on feature fusion, for example incorporating other visual signals such as head movements and facial expressions, and possibly speech, to create a multi-modal system.

References

[1] W. H. Organization *et al.*, "Risk reduction of cognitive decline and dementia: Who guidelines," 2019.

[2] A. Mavrodaris, J. Powell, and M. Thorogood, "Prevalences of dementia and cognitive impairment among older people in sub-saharan africa: a systematic review," *Bull. World Health Organ.*, vol. 91, pp. 773–783, 2013.

[3] S. T. Farias, D. Mungas, B. R. Reed, D. Harvey, and C. DeCarli, "Progression of mild cognitive impairment to dementia in clinic-vs community-based cohorts," *Arch. Neurol.*, vol. 66, no. 9, pp. 1151–1157, 2009.

[4] Alz.org, "Types of Dementia," 2018. [Online]. Available: https://www.alz.org/dementia/types-of-dementia.asp

[5] V. P. Richmond, J. C. McCroskey, and M. Hickson, *Nonverbal behavior in interpersonal relations*. Allyn & Bacon, 2008.

[6] A. Vinciarelli, M. Pantic, and H. Bourlard, "Social signal processing: Survey of an emerging domain," *Image Vis. Comput.*, vol. 27, no. 12, pp. 1743–1759, 2009.

[7] A. Ladas, C. Frantzidis, P. Bamidis, and A. B. Vivas, "Eye blink rate as a biological marker of mild cognitive impairment," *IEEE Int. J. Psychophysiol.*, vol. 93, no. 1, pp. 12–16, 2014.

[8] H. Tanaka, H. Adachi, N. Ukita, M. Ikeda, H. Kazui, T. Kudo, and S. Nakamura, "Detecting dementia through interactive computer avatars," *IEEE J. Transl. Eng. Health Med.*, vol. 5, pp. 1–11, 2017.

[9] H. Tanaka, H. Adachi, H. Kazui, M. Ikeda, T. Kudo, and S. Nakamura, "Detecting dementia from face in human-agent interaction," in *ICMI'19 Adjunct*, 2019, pp. 1–6.

[10] D. Murphy, A. Ni Loingsigh, I. Bernie, T. H. Bak *et al.*, "Bilingualism and dementia: How some patients lose their second language and rediscover their first," *Languages Victoria*, vol. 23, no. 2, p. 81, 2019.

[11] H. Tanaka, H. Adachi, N. Ukita, T. Kudo, and S. Nakamura, "Automatic detection of very early stage of dementia through multimodal interaction with computer avatars," in *Proc. ICMI'18*, 2016, pp. 261–265.

[12] B. Mirheidari, D. Blackburn, T. Walker, A. Venneri, M. Reuber, and H. Christensen, "Detecting signs of dementia using word vector representations." in *INTERSPEECH*, 2018, pp. 1893–1897.

[13] B. Mirheidari, D. Blackburn, T. Walker, M. Reuber, and H. Christensen, "Dementia detection using automatic analysis of conversations," *Comput. Speech Lang.*, vol. 53, pp. 65–79, 2019.

[14] J. Beltrán, M. S. García-Vázquez, J. Benois-Pineau, L. M. Gutierrez-Robledo, and J.-F. Dartigues, "Computational techniques for eye movements analysis towards supporting early diagnosis of alzheimer's disease: a review," *Comput. Math. Methods. Med.*, vol. 2018, 2018.

[15] T. Nakano, "Blink-related dynamic switching between internal and external orienting networks while viewing videos," *Neurosci. Res.*, vol. 96, pp. 54–58, 2015.

[16] B. J. Jongkees and L. S. Colzato, "Spontaneous eye blink rate as predictor of dopamine-related cognitive function—a review," *Neurosci. Biobehav. Rev.*, vol. 71, pp. 58–82, 2016.

[17] D. von Cramon and U. Schuri, "Blink frequency and speech motor activity," *Neuropsychologia*, vol. 18, no. 4-5, pp. 603–606, 1980.

[18] A. R. Bentivoglio, S. B. Bressman, E. Cassetta, D. Carretta, P. Tonali, and A. Albanese, "Analysis of blink rate patterns in normal subjects," *Mov. Disord.*, vol. 12, no. 6, pp. 1028–1034, 1997.

[19] P. J. De Jong and H. Merckelbach, "Eyeblink frequency, rehearsal activity, and sympathetic arousal," *Int. J. Neurosci.*, vol. 51, no. 1-2, pp. 89–94, 1990.

[20] S. Chermahini and B. Hommel, "Kreativniji kroz pozitivno raspoloženje. nitko," *Ispred. Pjevušiti. Neurosci*, vol. 6, no. 319.10, p. 3389, 2012.

[21] V. De Padova, G. Barbato, F. Conte, and G. Ficca, "Diurnal variation of spontaneous eye blink rate in the elderly and its relationships with sleepiness and arousal," *Neurosci. Lett.*, vol. 463, no. 1, pp. 40–43, 2009.

[22] W. S. Sun, R. S. Baker, J. C. Chuke, B. R. Rouholiman, S. A. Hasan, W. Gaza, M. W. Stava, and J. D. Porter, "Age-related changes in human blinks. passive and active changes in eyelid kinematics." *Investig. Ophthalmol. Vis. Sci.*, vol. 38, no. 1, pp. 92–99, 1997.

[23] E. Ponder and W. Kennedy, "On the act of blinking," *Q. J. Exp. Physiol.*, vol. 18, no. 2, pp. 89–110, 1927.

[24] M. Argilés, G. Cardona, E. Pérez-Cabré, and M. Rodríguez, "Blink rate and incomplete blinks in six different controlled hard-copy and electronic reading conditions," *Investig. Ophthalmol. Vis. Sci.*, vol. 56, no. 11, pp. 6679–6685, 2015.

TABLE III: Classification accuracy in percentage (%) when using the novel approach (multiple threshold approach) using different classifiers: ($Linear-SVM^1, rbf-SVM^2, kNN^3, LR^4$, and DT^5) for IVA data.

SM	Technique	Threshold No.	ND/MCI/FMD	ND/MCI	ND/FMD	MCI/FMD
Type 1	Dlib	7	$58^{1,3,4}$	58^{4}	$83^{1,4}$	83^{1}
		70	72^{1}	67^{1}	$92^{4,5}$	$\mathbf{92}^{4}$
		700	72^{1}	67^{1}	92^{4}	$83^{1,4}$
	OpenFace	7	33^{3}	50^{3}	$50^{2,3}$	$42^{1,2,3,4}$
		70	$50^{1,3,4}$	$75^{1,4}$	75^{4}	$50^{1,2,3,4}$
		700	56^{1}	$58^{2,3,5}$	83^{3}	$\mathbf{92}^{1}$
	OpenFaceSTD	7	$44^{1,4}$	$58^{1,2,3}$	$83^{1,3}$	75^{5}
		70	56^{1}	58^{3}	83^{3}	$\mathbf{92}^{1}$
		700	$44^{1,2,3}$	58^{3}	83^{3}	$\mathbf{92}^{5}$
Type 2	Dlib	7	67^{4}	67^{2}	92^{1}	83^{1}
		70	$\mathbf{89}^{1}$	83^{1}	$\mathbf{100}^{1,5}$	$\mathbf{92}^{1,4}$
		700	78^{4}	$75^{4,5}$	$\mathbf{100}^{5}$	$\mathbf{92}^{4}$
	OpenFace	7	33^{3}	$50^{3,4}$	$50^{2,3,4}$	$42^{1,3,4}$
		70	$44^{1,2,4}$	$67^{1,3,4}$	$75^{1,3,4}$	$50^{1,2,3,4}$
		700	50^{3}	$50^{3,4,5}$	$83^{1,4}$	$67^{1,3,5}$
	OpenFaceSTD	7	$50^{3,5}$	$50^{3,4}$	$83^{2,4}$	67^{5}
		70	$50^{3,4}$	58^{5}	83^{24}	$\mathbf{92}^{5}$
		700	61^{4}	50^{3}	83^{2}	$\mathbf{92}^{5}$
Type 3	Dlib	70	$\mathbf{89}^{1}$	83^{1}	$\mathbf{100}^{1,5}$	$\mathbf{92}^{1,4}$
	OpenFace	700	72^{5}	$\mathbf{100}^{5}$	92^{1}	$\mathbf{92}^{5}$
	OpenFaceSTD	7	78^{5}	67^{4}	$83^{1,2}$	$\mathbf{92}^{1}$

[25] I. A. Mota and O. G. Lins, "Bereitschaftspotential preceding spontaneous and voluntary eyelid blinks in normal individuals," *Clin. Neurophysiol.*, vol. 128, no. 1, pp. 100–105, 2017.

[26] J. M. Delgado-García, A. Gruart, and A. Múnera, "Neural organization of eyelid responses," *Mov. Disord.*, vol. 17, no. S2, pp. S33–S36, 2002.

[27] J. M. Delgado-Garcia, A. Gruart, and J. A. Trigo, "Physiology of the eyelid motor system," *Ann. N. Y. Acad. Sci.*, vol. 1004, pp. 1–9, 2003.

[28] D. C. King and K. M. Michels, "Muscular tension and the human blink rate." *J. Exp. Psychol.*, vol. 53, no. 2, p. 113, 1957.

[29] D. S. Woodruff-Pak, "Eyeblink classical conditioning differentiates normal aging from alzheimer's disease," *IPBS*, vol. 36, no. 2, pp. 87–108, 2001.

[30] W. Chen, T. Chiang, M. Hsu, and J. Liu, "The validity of eye blink rate in chinese adults for the diagnosis of parkinson's disease," *Clin. Neurol. Neurosurg.*, vol. 105, no. 2, pp. 90–92, 2003.

[31] K. C. Fraser, K. Lundholm Fors, M. Eckerström, F. Öhman, and D. Kokkinakis, "Predicting mci status from multimodal language data using cascaded classifiers," *FAN*, vol. 11, p. 205, 2019.

[32] O. Barral, H. Jang, S. Newton-Mason, S. Shajan, T. Soroski, G. Carenini, C. Conati, and T. Field, "Non-invasive classification of alzheimer's disease using eye tracking and language," in *MLHC*. PMLR, 2020, pp. 813–841.

[33] M. Lalonde, D. Byrns, L. Gagnon, N. Teasdale, and D. Laurendeau, "Real-time eye blink detection with gpu-based sift tracking," in *IEEE 4th Canadian Conf. on CRV*. IEEE, 2007, pp. 481–487.

[34] L. Schillingmann and Y. Nagai, "Yet another gaze detector: An embodied calibration free system for the icub robot," in *IEEE-RAS 15th International Conf. on HR*. IEEE, 2015, pp. 8–13.

[35] E. R. Anas, P. Henriquez, and B. J. Matuszewski, "Online eye status detection in the wild with convolutional neural networks," in *VISAPP*, vol. 7. SciTePress, 2017, pp. 88–95.

[36] D. W. Hansen and Q. Ji, "In the eye of the beholder: A survey of models for eyes and gaze," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 3, pp. 478–500, 2009.

[37] T. Appel, T. Santini, and E. Kasneci, "Brightness-and motion-based blink detection for head-mounted eye trackers," in *Proc. of the UbiComp '16*, 2016, pp. 1726–1735.

[38] T. Soukupová and J. Cech, "Real-time eye blink detection using facial landmarks," in *CVWW'16*, 2016.

[39] A. Fogelton and W. Benesova, "Eye blink detection based on motion vectors analysis," *Comput. Vis. Image Underst.*, vol. 148, pp. 23–33, 2016.

[40] ——, "Eye blink completeness detection," *Comput. Vis. Image Underst.*, vol. 176, pp. 78–85, 2018.

[41] K. Cortacero, T. Fischer, and Y. Demiris, "Rt-bene: A dataset and baselines for real-time blink estimation in natural environments," in *Proc. of the IEEE ICCV*, 2019, pp. 0–0.

[42] T. Drutarovsky and A. Fogelton, "Eye blink detection using variance of motion vectors," in *ECCV*. Springer, 2014, pp. 436–448.

[43] N. Alsaeedi and D. Wloka, "Real-time eyeblink detector and eye state classifier for virtual reality (vr) headsets (head-mounted displays, hmds)," *Sensors*, vol. 19, no. 5, p. 1121, 2019.

[44] J. A. Stern, L. C. Walrath, and R. Goldstein, "The endogenous eyeblink," *IEEE Int. J. Psychophysiol.*, vol. 21, no. 1, pp. 22–33, 1984.

[45] J. K. Portello, M. Rosenfield, and C. A. Chu, "Blink rate, incomplete blinks and computer vision syndrome," *Optom. Vis. Sci.*, vol. 90, no. 5, pp. 482–487, 2013.

[46] J. D. Rodriguez, G. W. Ousler III, P. R. Johnston, K. Lane, and M. B. Abelson, "Investigation of extended blinks and interblink intervals in subjects with and without dry eye," *Clin. Ophthalmol. (Auckland, NZ)*, vol. 7, p. 337, 2013.

[47] G. Pan, L. Sun, Z. Wu, and S. Lao, "Eyeblink-based anti-spoofing in face recognition from a generic webcamera," in *2007 IEEE 11th ICCV*. IEEE, 2007, pp. 1–8.

[48] B. Mirheidari, D. Blackburn, K. Harkness, T. Walker, A. Venneri, M. Reuber, and H. Christensen, "Toward the automation of diagnostic conversation analysis in patients with memory complaints," *J. Alzheimer's Dis.*, vol. 58, no. 2, pp. 373–387, 2017.

[49] ——, "An avatar-based system for identifying individuals likely to develop dementia," in *Interspeech 2017*. ISCA, 2017, pp. 3147–3151.

[50] T. Baltrusaitis, A. Zadeh, Y. C. Lim, and L.-P. Morency, "Openface 2.0: Facial behavior analysis toolkit," in *13th IEEE ICFG 2018*. IEEE, 2018, pp. 59–66.

[51] G. B. Huang, M. Mattar, T. Berg, and E. Learned-Miller, "Labeled faces in the wild: A database for studying face recognition in unconstrained environments," in *Workshop on Faces in 'Real-Life' Images: Detection, Alignment, and Recognition*, 2008.

[52] T. Jw, "Exploratory data analysis," *Reading: Addison-Wesley*, 1977.

[53] K. Radlak and B. Smolka, "Blink detection based on the weighted gradient descriptor," in *Proc. of the 8th Inter. Conf. on CORES 2013*. Springer, 2013, pp. 691–700.

[54] S. J. Wakefield, D. J. Blackburn, K. Harkness, A. Khan, M. Reuber, and A. Venneri, "Distinctive neuropsychological profiles differentiate patients with functional memory disorder from patients with amnestic-mild cognitive impairment," *Acta Neuropsychiatr.*, vol. 30, no. 2, pp. 90–96, 2018.