Multimedia Tools and Applications

Eyes-based Features for Detection of Neurological Disorders in Videos

--Manuscript Draft--

Manuscript Number:	MTAP-D-20-00779R2
Full Title:	Eyes-based Features for Detection of Neurological Disorders in Videos
Article Type:	1155T: Advanced machine learning algorithms for biomedical data and imaging
Keywords:	eye; neurological; Video; disorders; features; blink; Bell's Palsy; Cervical Dystonia; Blepharospasm
Corresponding Author:	Koteswar Rao Jerripothula, Ph.D Indraprastha Institute of Information Technology Delhi Delhi, INDIA
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	Indraprastha Institute of Information Technology Delhi
Corresponding Author's Secondary Institution:	
First Author:	Sharik Ali Ansari
First Author Secondary Information:	
Order of Authors:	Sharik Ali Ansari
	Koteswar Rao Jerripothula, Ph.D
	Pragya Nagpal
	Ankush Mittal
Order of Authors Secondary Information:	
Funding Information:	

Eyes-based Features for Detection of Neurological Disorders in Videos

Managing Editor's comments:

Please note the statement above regarding inclusion of citations suggested by editors and reviewers. They should only be considered IF they are directly relevant to the specific topic of the paper.

Response: *There were no such suggestions.*

Editor's comments:

Reviewers are agreed with the modifications. However, the current version is not improved significantly and lack several issues. Authors are advised to re-look the Editor's concerns in the last revision (most of the comments unanswered). Please note this is the FINAL re-submission opportunity for the authors, the paper will be automatically rejected in the next round if it does not fully address the reviewers/editorial comments and/or meet the journal's expected, consistently high presentation and publication standards.

Response: *Thanks for pointing out that we missed responding to some of the concerns last time. We have now addressed each concern one-by-one, as shown below:*

Concern 1: Please stress on the novelty aspects and put your paper in proper context by contrasting against the latest ML approaches for Eye-based detection.
Response: *The novelty issue was already addressed in the last revision. We had stressed upon the following aspects of novelty:*

1) This is the first work on detecting Cervical dystonia and Blepharospasm from the videos.

2) We have developed three novel eyes-based features to detect Bell's Palsy, Cervical Dystonia, and Blepharospasm. For accomplishing this, we also had to develop novel video datasets for these disorders.

3) Since we use deep learning to detect certain eye-events like blink, CD-like-iris, and blink (incl. tight), we also had to develop eye-images datasets specifically to detect these events.

4) Our work can perform well in the low-data setting, thanks to the hand-crafted features we have developed. One of the reasons our features work so well is that they have been specifically designed to detect the disorders under consideration.

In this round, we add the following lines in our introduction to put our paper in a proper context of Eye-based detection, as suggested:

“Previously as well, some of the disorders have been detected based on the eyes, such as computer vision syndrome [12], neurodegenrative disorders [13], Autism Spectrum Disorders [14] and so on. However, none of the works specifically deal with neurological disorders using eyes-based detection. Therefore, in this paper, we study how we can detect some of the neurological disorders just from the eyes.”

Concern 2: Try to validate your method by using more challenging datasets.
Response : *We have already addressed this in our previous revision by testing the proposed method on Youtube Facial Palsy (YFP) dataset. We achieved 90.3% accuracy for Bell's Palsy detection. Note that there are no publicly available datasets for Cervical Dystonia and Blepharospasm. Ours is the first work on diagnosing these two disorders using visual data. See Sec 4.6 for more details.*

Concern 3: Kindly note that this SI focusses on Advanced ML algorithms. You need to justify how is your work closely related to SI?
Response : *Last time, we had justified only how it is contributing towards biomedical applications. This time, we also present how ours is an advanced machine learning algorithm by adding the following to our introduction:*

"In summary, we have developed three novel eyes-based features for biomedical applications. These features are specifically helpful in detecting Bell's Palsy, Cervical Dystonia, and Blepharospasm. Note that ours is the first work on detecting Cervical Dystonia and Blepharospasm from the videos, to the best of our knowledge. We explore nine machine learning algorithms to build the best models for detecting these disorders individually and collectively. We develop a total of seven datasets, which can be used for many other biomedical applications. For example, more eye-related features can be designed in the future with the help of such data. We propose a hybrid machine learning approach (as an advanced machine learning strategy) that efficiently blends deep learning and machine learning (ML). We attempt to detect neurological disorders (for which data is difficult to get) based on eye-events (for which lots of data can be obtained). It is well known that deep learning requires loads of data to solve a problem. Therefore, due to the data constraints, instead of building an end-to-end deep learning model for neurological disorders detection, we break the problem into two stages: one is eye-event detection, and another is subsequent neurological disorders detection. In this way, while we could solve the eye-event detection problem with loads of data using deep learning, we could also solve the subsequent neurological detection problem by designing disorder-specific features for machine learning. Our experiments demonstrate that proposed features perform exceptionally well, thanks to our efficient hybrid machine learning approach."

Concern 4: The English language of this paper too week, please send the paper to the native speaker, there are several sentences which too poor.
Response : *We addressed this thoroughly last time.*

Concern 5: Please remove all the lumped references. This is not the practice of MTAP. This is to prevent inappropriate citation and to provide more information to the authors. If necessary, please provide a short description for each of the reference use.
Response: *This was already addressed last time.*

Concern 6: The literature review of this paper must be expanded and rewritten considerably by including more recent advanced classification algorithms, see for example papers like: 10.1109/TPAMI.2009.30, https://doi.org/10.1016/j.eswa.
Response: *Thanks for the reference, we have now added a subsection titled "Eyes-events detection" to report the related research done that employ advanced classification algorithms.*

Concern 7: The literature review should evaluate the sources and advise the reader on the most pertinent or relevant literature in particular subject areas. The depth and breadth of the literature review must emphasize the credibility of the authors in their fields.
Response: *Since our work can be divided into two parts, namely interemediate detection (for eye-events) and final detection (for neurorogical disorders), We have now divided our literature review into two parts. (I) Eyes-events detection: we illustrate here how none of such existing works have*

focused explicitly on neurological disorders. (II) Neurological disorders detection: we illustrate here how none of the three neurological disorders have been explicitly detected from the eyes using videos.

Concern 8: The literature sources used must be as current as possible.
Response: *We have taken this into account now and would like to report that about 50% of our cited sources have been published in the 2017 or later.*

Concern 9: The authors should summarize or critique their sources by discussing a common theme or issue. The authors should not simply list their sources and some details of each one of them.
Response: *Thanks for the suggestion. As you can see in our updated related works section, we have discussed the existing literature and contrasted with them using the "eyes-events detection" and "neurological disorders detection" themes, which are basically the two major themes our proposed work addresses.*

Concern 10: Need to highlight the novelty of study in the introduction.
Response: *Please see our response to Comment 1. We have added the points mentioned there in our introduction.*

Concern 11: Introduction should be clearly stated research questions and targets first. Then answer several questions: Why is the topic important (or why do you study on it)? What are the research questions? What has been studied? What are your contributions? Why is to propose this particular method?
Response: *Thanks for the suggestions on improving our introduction. As suggested, we have restructured our introduction in the following manner.*
Our first paragraph now clearly defines the targets and gives a glimpse of the research questions (more on this in the third paragraph) that need to be answered to achieve them.
Our second paragraph presents the motivation of our work, answering why this study is important.
Our third paragraph discusses various challenges (research questions) in details.
Our fourth paragraph gives the details of our approach while also answering why we take a particular approach.
Our fifth paragraph establishes how our work is closely related to this SI
Our sixth paragraph is explicitly dedicated to listing all the contributions this work makes.

Concern 12: Results are good, but you need to add more discussion of the results.
Response: *This was already addressed last time. Note that Section 4.6 was specifically added for this purpose.*

Concern 13: More future directions should be presented.
Response: *The current work can be lay path to (1) discovery of further advanced features, (2) development of methods for neurological rehabilitation where we need to measure severity of the disorder, and (3) addition of security features in such studies because we demonstrated that all we needed was eyes to detect these disorders.*

Reviewer #2:

Comment: The manuscript has been revised based on the comments and suggestions. It is recommended to accept and publish through the journal.

Response: *Thanks for acknowledging our efforts and your recommendation.*

Reviewer #3:

Comment: I think the authors have done a good job in revisiting the manuscript, especially considering the other reviewers' comments.

Response: *Thanks for your encouragement.*

Eyes-based Features for Detection of Neurological Disorders in Videos

Sharik Ali Ansari[1,4], Koteswar Rao Jerripothula[2⋆], Pragya Nagpal[3], and Ankush Mittal[4]

[1] College of Engineering Roorkee, India
[2] Indraprastha Institute of Information Technology Delhi (IIIT-Delhi), India
[3] University of Texas at Dallas, USA
[4] Raman Classes, India

Abstract. In this paper, we study how some of the neurological disorders, namely Bell's Palsy (BP), Cervical Dystonia (CD), and Blepharospasm, can be detected from the eyes of the subjects in the videos. Although previous works also utilized visual information to detect such disorders, none of them have yet especially explored eyes to do so. Also, unlike them, we detect multiple disorders. To the best of our knowledge, this work is first in such solely eyes-based exploration of the neurological disorders, as far as video analytics research is concerned. Specifically, we develop novel hybrid (hand-crafted+deep-learning) features, namely blink similarity, CD-like iris appearance, and blink normalness. We use deep learning to detect crucial instances in a video frame-by-frame to develop hand-crafted features for the entire video later on. Using these video-level features, we build machine learning models that can predict these neurological disorders. Since such models are data-driven, we needed to collect numerous images/videos for developing novel benchmark datasets for these disorders. Our exhaustive experiments demonstrate that our proposed method detects Bell's Palsy, Cervical Dystonia, and Blepharospasm with 94.7%, 90.0%, and 94.1% accuracies while detecting these disorders individually. Also, we achieve 83.0% cross-validation accuracy while detecting these disorders collectively. Our proposed eyes-based detection approach is much cheaper and more convenient than the existing detection approaches, which have complicated hardware requirements, for the proposed approach requires only a video of the subject.

Keywords: eye · neurological · video · disorders · features · blink · Bell's Palsy · Cervical Dystonia · Blepharospasm · spasmodic torticollis.

1 Introduction

Neurological disorders[1,2] are diseases of the nervous system, comprising the brain, spinal cord, and nerves. The nerves connect the brain with the spine and

⋆ Corresponding author: K. R. Jerripothula (email: koteswar001@e.ntu.edu.sg)

both to different parts of the body. The nervous system is highly complex as it coordinates all our actions and sensory information by transmitting signals through nerves to different parts of the body. Thus, when there is a neurological disorder, a person may have difficulty in various bodily movements, such as walking, speaking, swallowing, breathing, blinking, learning, etc. These difficulties are often visible and can serve as symptoms that can be leveraged to diagnose neurological disorders [3]. Our target is to build a computer-assisted neurological disorder diagnosis system using computer vision and machine learning algorithms. Specifically, if the system is given a video of a subject, as shown in Fig. 1, the system should be able to automatically detect the neurological disorder with which the subject is suffering. While attempting to build such systems requiring a video dataset for learning, concerns like privacy and insufficient patient data are common. However, if we focus on the eyes for building such a system, these concerns can be addressed. It is because we do not need other body parts, which may reveal the identity, for analysis anymore, and also eyes have a much simpler structure and movements than the entire face or whole body. Therefore, to achieve our goal of building a computer-assisted neurological disorders diagnosis system, we have identified the following three neurological disorders, which could be detected just from the eyes, to work upon:

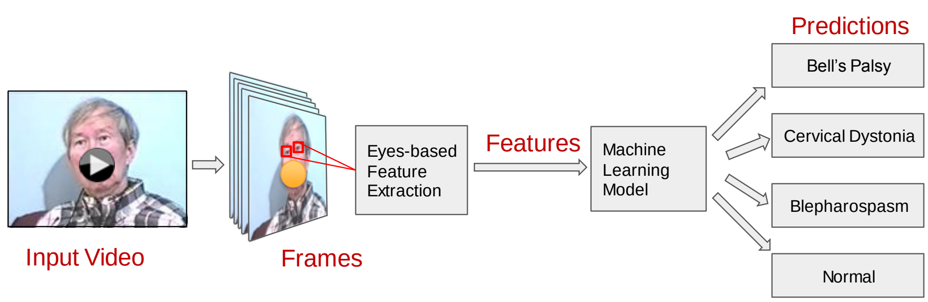

Fig. 1. Given a video of the subject, our objective is to automatically detect, using the eye-based features, if the subject has any of the three neurological disorders: Bell's Palsy, Cervical Dystonia, and Blepharospasm.

1. **Bell's Palsy:** It is a condition where muscles on one side of the face become weak or paralyzed. It could be caused due to trauma to the seventh cranial nerve. According to [4], it affects at least one in every 60 persons once in a lifetime. Moreover, [5] reports that it gets repeated in 7% of them. This disorder makes one half of the face appear to droop. The smile becomes one-sided, and the eye muscles on the paralyzed side become weak and resist closing, causing the blink rate to go down.

2. **Cervical dystonia:** It is a condition where neck muscles contract involuntarily, causing the head to turn to one of the sides. It is believed that it occurs due to an abnormality in the basal ganglia or other brain regions that control our movement. However, in most cases, the exact cause is difficult to be known. In [6], it has been reported that large percentages of such patients suffer from depression and anxiety. It has four forms: the most common is Torticollis, and the other three are Laterocollis, Anterocollis, and Retrocollis, as mentioned in [7]. It is Torticollis where the head gets turned to one of the sides. With the turned head, patients strain their eyes as they try to adjust their eye-gaze[8] while conversing because they now have to push their iris to an extreme quite often.
3. **Blepharospasm:** It is a rare condition where one experiences an involuntary repeated forcible contraction of eyelid and forehead muscles, as mentioned in [9]. Although the exact cause is difficult to point out, just like Cervical Dystonia, it can also happen due to the malfunctioning in Basal Ganglia, a brain region. It affects only 20 to 133 people per million, according to [10].

Each of the above disorders has one or more visible symptoms in the eye characteristics or behavior. In general, visible symptoms are vital signals which the body gives to alert us. Although the real disorder may be lying deep within the nervous system, these symptoms are indeed useful because even doctors sometimes rely on such symptoms while diagnosing. Such an approach makes diagnosis highly cost-effective and convenient as it only requires a video of the subject. However, videos[11] captured may not always focus well on the area of concern where the symptom appears. Interestingly, eyes are easy to locate using existing computer vision algorithms for a simple pattern they possess. Previously as well, some of the disorders have been detected based on eyes, such as computer vision syndrome [12], neurodegenerative disorders [13], Autism Spectrum Disorders [14], etc. However, none of the works specifically deal with neurological disorders through eyes-based detection. Therefore, in this paper, we study how we can detect the above three neurological disorders just from the eyes. In Bell's Palsy[15], we observe that blinking rates of two eyes differ because the blink in the paralyzed side slows down considerably. In Cervical Dystonia, we observe that the iris moves typically to an extreme to adjust the eye-gaze to counter the abnormal facial gaze while conversing. In Bleprosphasm, we observe that, during a blink, the distance between the eye center and eyebrows varies a lot while blinking due to the tendency of blinking tightly. These localized observations can be modeled into predictors (features) at the video-level for building an eye-based detection system of these neurological disorders.

However, there are some challenges while building such a neurological disorder detection system: (i) There are no publicly available video datasets for these disorders. (ii) There are no ready-made video-level features to represent the eye-based observations discussed in the above paragraph. Additionally, the features suitable for a low data setting (since neurological disorders are rare) are hard to find. (iii) Although several works are done on images, to the best of our knowledge, there are very few prior works on the videos to detect neurological disorders. Mostly, they have been done for detecting Bell's Palsy, not on Cervical Dystonia and Blepharospasm. Note that images[16,17] can be misleading at times because the same visual symptoms can be manifested in a healthy person as well, at least temporarily. However, in a video, such symptoms cannot manifest all the time for healthy persons like a patient. For instance, a healthy person's wink can be regarded as a blink of a patient from Bell's Palsy if it is just an image, but he cannot keep winking throughout the video. (iv) The system should be as natural and robust as possible. There should not be conditions attached, such as the video to be of a particular length, and the subject to perform particular expressions like smile, as demonstrated in some previous works (e.g., [18]).

In summary, we have developed three novel eyes-based features for biomedical applications. These features are specifically helpful in detecting Bell's Palsy, Cervical Dystonia, and Blepharospasm. Note that ours is the first work on detecting Cervical Dystonia and Blepharospasm from the videos, to the best of our knowledge. We explore nine machine learning algorithms to build the best models for detecting these disorders individually and collectively. We develop a total of seven datasets, which can be used for many other biomedical applications. For example, more eye-related features can be designed in the future with the help of such data. We propose a hybrid machine learning approach (as an advanced machine learning strategy) that efficiently blends deep learning and machine learning (ML). We attempt to detect neurological disorders (for which data is difficult to get) based on eye-events (for which lots of data can be obtained). It is well known that deep learning requires loads of data to solve a problem. Therefore, due to the data constraints, instead of building an end-to-end deep learning model for neurological disorders detection, we break the problem into two stages: one is eye-event detection, and another is subsequent neurological disorders detection. In this way, while we could solve the eye-event detection problem with loads of data using deep learning, we could also solve the subsequent neurological detection problem by designing disorder-specific features for machine learning. Our experiments demonstrate that proposed features perform exceptionally well, thanks to our efficient hybrid machine learning approach.

In summary, we have developed three novel eyes-based features for biomedical applications. These features are specifically helpful in detecting Bell's Palsy, Cervical Dystonia, and Blepharospasm. Note that ours is the first work on detecting Cervical Dystonia and Blepharospasm from the videos, to the best of our knowledge. We explore nine machine learning algorithms to build the best models for each disorder and also for combined detection. We develop total seven datasets, which can be used for many other biomedical applications. For example, more eye-related features can be designed in the future with the help of such data. We propose a hybrid machine learning approach (as an advanced machine learning strategy) that efficiently blends deep learning and machine learning (ML). We attempt to detect neurological disorders (for which data is difficult to get) based on eye-events (for which lots of data can be obtained). It is well known that deep learning requires loads of data to solve a problem.

Therefore, due to the data constraints, instead of building an end-to-end deep learning model for neurological disorders detection, we break the problem into two stages: one is eye-event detection, and another is subsequent neurological disorders detection. In this way, while we could solve the eye-event detection problem with loads of data using deep learning, we could also solve the subsequent neurological detection problem by designing disorder-specific features for machine learning. Our experiments demonstrate that proposed features perform exceptionally well, thanks to our efficient hybrid machine learning approach.

We make the following contributions in this paper: (1) We develop novel benchmark datasets for detecting neurological disorders (Bell's Palsy, Cervical Dystonia, and Blepharospasm) and eye-monitoring (detecting blinks, iris position, and blinks (incl. tight)). (2) We develop three novel eye-based features (blink similarity, CD-like iris appearance, and blink normalness) for videos, without any constraints on the video-lengths. Our disorder-focused feature designing allows us to work with even a low-data setting, which is quite typical for rare neurological disorders. Thanks to disorder-focused feature design [19,20], our features have the potential to indicate the severity degree of the disorder. (3) We build novel DL-based intermediate detectors to monitor eyes and traditional ML-based final detectors to detect neurological disorders. (4) We are first to propose an eyes-based diagnosis system for multiple neurological disorders from video footage.

2 Related Work

Since our work can be divided into two parts of eyes-events detection and neurological disorders detection, we discuss the prior works on these two problems in this section.

2.1 Eyes-events Detection

Thanks to the distinct texture, shape, and color, eyes are easily detectable using HOG [21,22,23] feature, Haar [21,22] feature, facial landmarks [24], GANs [25] and so on. Furthermore, [26] provides a detailed review of the existing video-based eye detection and tracking techniques and various issues involved. We specifically take the facial landmarks approach for localizing the eyes. Since we localization may not be accurate, we extract some extra region along the boundary of detection provided by such landmarks. Once eyes are detected, we detect three eye-events: normal blinks, CD-like-iris, and blinks (incl. tight).

Before our work, the blink detection problem has been explored quite extensively. There are heuristic methods such as [27], where the brightness feature has been exploited. Then, there are traditional ML-based methods using HOG [28] feature, motion [29] feature, etc. With the rise of deep learning, several CNN-based approaches have been explored recently, such as [29,30,31]. We also use a CNN-based architecture for this problem. However, our main contribution to this problem is the dataset we provide, which is at the magnitude of nearly

40k images, which is the largest dataset available so far for this particular problem, to the best of our knowledge. None of the current works deal with the tight blinks problem, which became necessary for us to solve the Blepharospasm problem later on. Also, none of the existing works specifically used blink detection to solve Bell's Palsy problem. So, although blink detection research has been there for quite a while, they have not been used explicitly to detect neurological disorders like us, to the best of our knowledge.

For CD-like-iris detection, we combine iris-position problem and gaze detection [32,33,34] problem to detect what we call as CD-like-iris, where a person's gaze is not straight, and the iris is at extreme, trying to meet the gaze of the other person. Although these two problems have been studied as two different problems earlier, they have not been combined in this way to detect Cervical Dystonia similar to us, to the best of our knowledge. [35,36] used CNNs similar to us for iris position detection. Recently, [37] used supervised descent and isophote curvature methods to detect the iris. We take the CNN-based approach to this problem. However, note that our main contribution here is the dataset that is specifically tailor-made for detecting CD-like-iris, where we are interested only in detecting the iris's outer extreme position. The CD-like iris is detected only when the face is not straight, which can be accomplished only using gaze estimation [38,39,40]. We use the publicly available generic 3d model[5] that is created using Dlib points for finding the face angle.

2.2 Neurological Disorders Detection

Bell's Palsy: In [41], authors analyze the video clips captured with webcam of PC for diagnosis by measuring the asymmetry index around the mouth region. In contrast, our work focuses on the eyes to detect Bell's Palsy. In [42], authors propose a method named ASMLBP (Active Shape Models plus Local Binary Patterns), where the face is divided into eight local regions to describe each region using the facial points extracted with the help of Active Shape Models (ASM) and region descriptors with the help of Local Binary Patterns(LBP). In this way, the authors observe the patterns in the facial movement and classify the videos. However, this method requires videos of the subjects to be taken in a controlled environment (see [42] for more details) and needs specific movements to be carried out for accurate detection. Similarly, [43,44] also have constraints with video recording environment, video lengths, etc. In contrast, we attempt to detect the disorder in the wild, without any constraints attached. In [45], the authors propose a smartphone-based facial palsy diagnostic system that can work in the wild. They localize and track facial landmarks with the help of a modified linear regression method and then compute the displacement ratio between the left and right side of facial landmarks located on the forehead and mouth region. There is another work [46] that proposes an incremental face alignment method to handle videos in the wild for this problem. However, there is always a possibility of facial points mismatching in such analysis, thanks to

[5] https://www.learnopencv.com/head-pose-estimation-using-opencv-and-dlib/

different challenges involved. Therefore, as far as Bell's Palsy is concerned, we use facial key-points only for detection [47] (localization [48]) of the eye region. Upon detecting such regions, we rely entirely on the dissimilarity in blinks of the two eyes, a novel approach to detecting the Bell's Palsy. Recently, [49] proposed a method to analyze facial Palsy using a deep hierarchical network by detecting asymmetric regions. It depends upon the accuracy of facial landmarks, whereas we can afford to handle its inaccuracy by extracting more regions around the eye region detected from these landmarks. Most close to our work, [50] also attempted facial palsy detection using videos by applying transfer learning on the 3D-CNN of [51]. Note that our blink similarity feature is more intuitive and interpretable than 3D-CNN features, for it has been specifically designed for Bell's Palsy.

Cervical Dystonia and Blepharospasm: Unlike Bell's Palsy, where there are works done using visual information, there is hardly any such work done for Cervical Dystonia or Blepharospasm. The focus mainly has been to provide prospective studies on the use of different drugs for treatment purposes and how these disorders affect daily living, such as low work productivity [52] and increased level of depression [53]. For example, in Cervical Dystonia, in [6], authors provided a survey on using multiple injections of botulinum toxin A while authors in [7] provided a survey of using abobotulinum toxin A (Abo-BTX A) and neubotulinum toxin A (Neu-BTX A). In a study provided in [44], authors conclude that Cervical Dystonia is more frequent than Blepharospasm. To the best of our knowledge, we are first to propose visual information based diagnosis system to these disorders. While we depend upon face gaze and iris position for detecting Cervical Dystonia, we rely on blink and eye landmarks to detect Blepharospasm.

3 Proposed Method

In Fig. 2, we give an overview of our proposed method. After frames extraction, we identify crucial instances using our deep learning networks (Blink Detector, CD-like-iris Detector, and Blink (incl. tight) Detector). Using these instances, we extract hand-crafted features through counters, and then these features are fed to machine learning models for the final detection. Notably, we develop three eye-related features, namely blink similarity, CD-like iris appearance, and blink normalness for detecting Bell's Palsy, Cervical Dystonia, and Blepharospasm, respectively. Deep learning usage for detecting crucial instances and then hand-crafting done on the counters of those instances makes these features hybrid (hand-crafted + deep-learning).

3.1 Blink Similarity (especially for Bell's Palsy)

Bell's Palsy is a unilateral condition in which muscles on one side of one's face become weak or paralyzed. In such a state, natural phenomenons like eye-blinking also get affected. Eye-blinking is the regular closure of eyelids for a brief period.

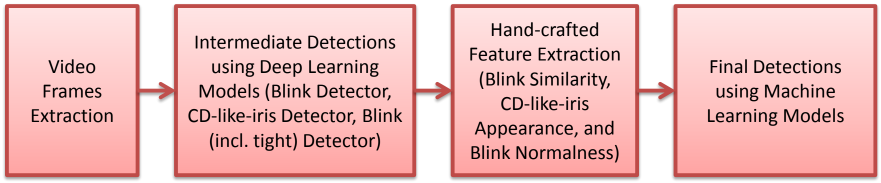

Fig. 2. Overview of the proposed method.

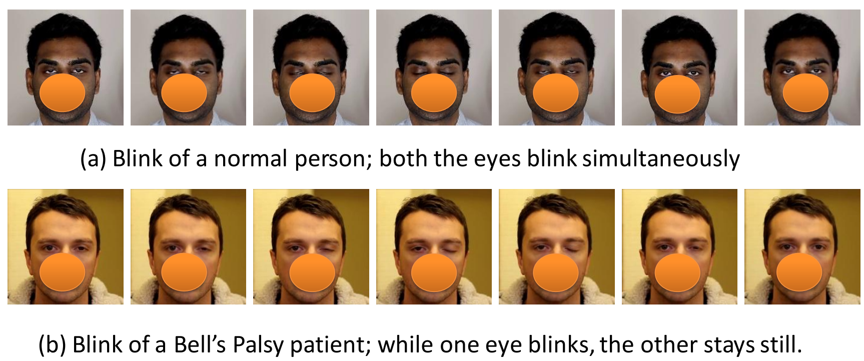

Fig. 3. An illustration demonstrating the difference between the blink of a normal person and that of a Bell's Palsy patient.

Table 1. The list of different abbreviations used and their full-forms

Abbreviation	Full Form	Abbreviation	Full Form
BP	Bell's Palsy	CF	CD like Iris Feature
CD	Cervical Dystonia	SGD	Stochastic Gradient Descent
DL	Deep Learning	LR	Logistic Regression
PC	Personal computer	AB	AdaBoost
ECT	Eye Close Time	NB	Naive Bayes
ECF	Eye Close Frame	NN	Neural Network
EC	Eye Close	RF	Random Forest
EO	Eye Open	DT	Decision Tree
L	Length of Video	3D	three-dimensional
F	Total number of Frames	avg	Average
BS	Blink Similarity Feature	var	Variance
EOI	Eye of Interest	CEL	Cross Entropy Loss
FGA	Face Gaze Angle	CNN	Convolutional Neural Network

This eye-blinking phenomenon gets severely affected in the paralyzed side of a patient suffering from Bell's Palsy. It leads to an abnormal difference in the blinking of the two eyes of a person. As shown in Fig.3, there is a significant difference between the blinking of (a) a normal person and (b) a Bell's Palsy patient. While both the eyes blink once for a normal person, only the normal side's eye blinks once for a patient suffering from Bell's Palsy. So, the blinking of a normal person is usually simultaneous, whereas there is a difference in the blinking for a Bell's Palsy patient. Hence, we can exploit this observation related to the eye-blinks for detecting Bell's Palsy neurological disorder in a person. However, even a normal person may blink just one eye during winking or instantaneous agitation. Therefore, we need to monitor these blinks need for a substantial time duration, and video-based diagnosis becomes necessary instead of just an image. We can consider the total time duration during which an eye remains closed, denoted as ECT, in a given period for computing our blink similarity feature. For a patient, the ECT of the paralyzed side is much lower than the normal side because of difficulty in closing it. Moreover, for a normal patient, ECT of the two sides should be almost equal. Note that we have given a list of abbreviations and full-forms to facilitate readability in Table 1.

Assuming eye-blinking is a periodic activity, there are two phases: eye-closure time (EC) and eye-open time (EO). If we consider the periodicity's assumption, these phases have a constant value for a given person's eye. So, they still vary across the eyes and vary across the persons. We can compute the blink time-period using $EC + EO$. If L denotes length (total time duration) of a video, we can now compute the total eye-closure time (ECT) as

$$ECT = \frac{EC}{EC + EO} \times L \tag{1}$$

where $(\frac{EC}{EC+EO})$ is the fraction coefficient of L to compute ECT (eye-closed time). So, multiplying the length of video with such a fraction yields the total time duration during which the eye was closed, i.e., ECT. Although we can extract L from the meta-data of a video, EC and EO are variables not only across different persons but also possibly between two eyes of the same person (particularly if patient). Therefore, they are difficult to compute. Nevertheless, we can also compute the number of eye-closed frames, denoted as ECF, in the same way, ie.

$$ECF = \frac{EC}{EC + EO} \times F \tag{2}$$

where F is the total number of frames. By dividing Eqn.(1) with Eqn.(2), we get the relationship between ECT and ECF as follows:

$$ECT = ECF \times \frac{L}{F}. \tag{3}$$

Now, considering both the eyes, with the subscription of L (for left) and R (for right) to ECT and ECF, and considering $\frac{L}{F}$ to be video constant, something independent of the eyes, we can conclude that

$$\frac{ECT_L}{ECT_R} = \frac{ECF_L}{ECF_R}. \tag{4}$$

Thus, the ratio of the total time duration for the two eyes is the same as the total number of frames with closed eyes. So, if we can compute the number of frames in which the left eye is closed (ECF_L) and the number of frames in which the right eye is closed (ECF_R), we can very determine the required ratio, which should be close to 1 for normal and significantly deviating from it otherwise.

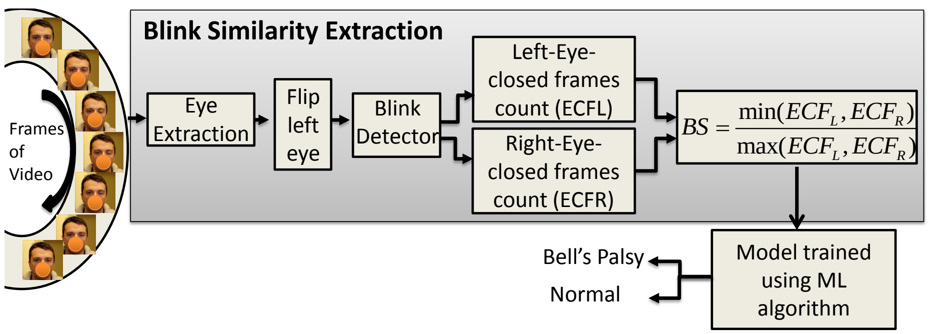

Fig. 4. Proposed approach for the extraction of blink-rate similarity feature and the end-to-end workflow for the detection of Bell's Palsy.

To detect the closed eye, we build an intermediate detector named blink detector using a Convolutional Neural Network (CNN). It takes eye images as input and outputs, whether it is a closed eye or an open eye. From any frame, we extract the two eyes of the person and flip left one sideways because we train the CNN on all right-types (by flipping the left-types). Although for a normal person, the ratio value would be close to 1, it can deviate from 1 either side (< 1 or > 1) for a patient depending upon which side is affected. Despite it being a good indicator in terms of detail (whether the right eye is paralyzed or left), note that since we are targeting only whether the person is affected or not, it needs some modification because we cannot find the single decision boundary (there are two). To have a single decision boundary, we compute our final blink similarity feature (BS) as the min-max ratio of the two counters, as shown below:

$$BS = \frac{min\big(ECF_L, ECF_R\big)}{max\big(ECF_L, ECF_R\big)}, \tag{5}$$

which automatically ensures the deviation of patients' feature scores from 1 towards one side only; that is, towards 0. Thus, we can have a single decision boundary. Furthermore, the ratio also automatically becomes $\frac{ECF_L}{ECF_R}$ or $\frac{ECF_R}{ECF_L}$ depending upon whether the paralyzed side is the left one or the right one, the ratios which we desire to use for demonstrating blinking similarity. However, it is difficult to find the optimal decision boundary between normal persons' feature values and abnormal ones. Therefore, we need to develop a final detection model using machine learning algorithms to classify videos into these two categories based on our BS feature. We give the entire end-to-end work-flow process in Fig. 4.

3.2 CD-like Iris Appearance (especially for Cervical Dystonia)

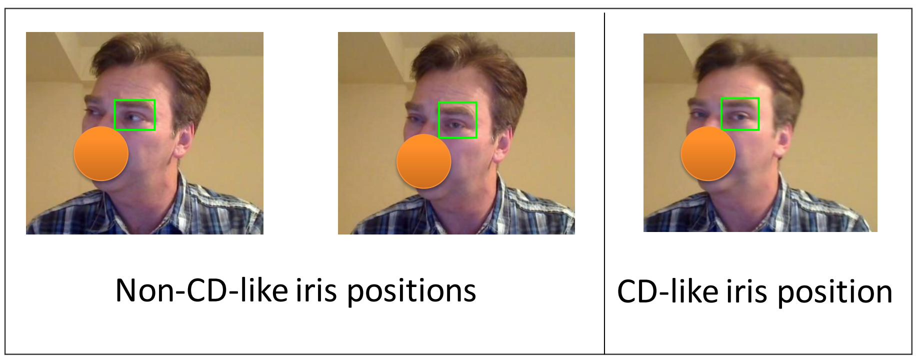

Fig. 5. An illustration differentiating what is CD-like iris position and what's not.

A person who has Cervical Dystonia has an abnormal face gaze as their head gets turned; thus, one eye comes to the front, and another goes back. Due to this abnormal face gaze, the patient has to adjust his eye gaze to converse. With head turned towards the left, the person adjusts his eye-gaze to move towards the right. Furthermore, with the head turned to the right, the person adjusts his eye-gaze to move towards the left. The more times a person does that, the more likely the person has Cervical Dystonia. In particular, the eye that comes to front due to the turned head is of particular interest because the other eye may not be captured in the video. Also, since both the eyes' gazes are usually parallel, the other eye's gaze need not be monitored. The most important clue for eye-gaze in visual analytics is iris-position, and in such circumstances, iris-position in the eye-of-interest (EOI) would be at extreme left if the head is turned to the right and extreme right if the head turns to the left. We call this CD-like iris appearance phenomenon, and the idea is to monitor how frequently this happens. We illustrate this phenomenon in Fig. 5. When the person has turned his right, and when the iris position in his EOI is his extreme left, we call it a CD-like-iris position. A healthy human may occasionally have such an appearance, but a patient has it quite frequently to look at the person or into the camera while conversing.

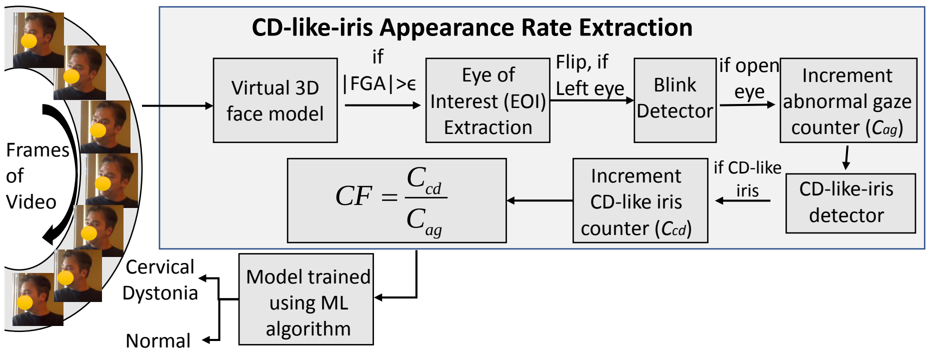

Fig. 6. Proposed approach for the extraction of CD-like iris appearance rate feature and the end-to-end workflow for detection of Cervical Distonia.

Specifically, we first detect if there is an abnormal gaze, i.e., $|FGA| > \epsilon$ where FGA is the face gaze angle and ϵ is tolerance level, set ϵ as $45°$, to neglect tolerable face gaze changes that occur while conversing. We use the classic solution of the PnP (Perspective-n-Point) problem, i.e., an iterative method based on the Levenberg-Marquardt algorithm [54] (implemented in OpenCV library) to compute FGA. After selecting such frames, we extract our EOI based on FGA, i.e., for the right rotated face (positive FGA), the eye of interest (EOI) is the left eye. Similarly, for the left rotated face (negative FGA), the eye of interest (EOI) is the right one. Again, we extract EOI's eye-region as discussed in the previous section, passing it through the blink detector (discussed in the previous section) to filter out the frames with closed EOI. We finally compute our feature named CD-like iris appearance as a fraction of the frames in which such CD-like iris occurs out of the total frames having abnormal FGA, as mentioned below.

$$CF = \frac{C_{cd}}{C_{ag}} \tag{6}$$

where CF is our computed feature, and C_{cd} and C_{ag} denote frame count for CD-like iris occurrence and frame count for the abnormal gaze of the face. To detect the CD-like iris, we build another Convolutional Neural Network (CNN) and call it a CD-like iris detector; it takes eye images as input and outputs whether it has CD-like-iris or not. Note that we train this detector on the right eye. So, if the EOI is left one, it is flipped, as done earlier. It is clear that as the feature value increases, the subject is more likely to be a patient. However, finding the optimal decision boundary between feature values for the normal and the abnormal cases is difficult. Therefore, we need to develop a model trained using machine learning algorithms to classify any video well into these two categories based on our feature. We give the entire end-to-end work-flow process in Fig.6.

3.3 Blink Normalness (especially for Blepharospasm)

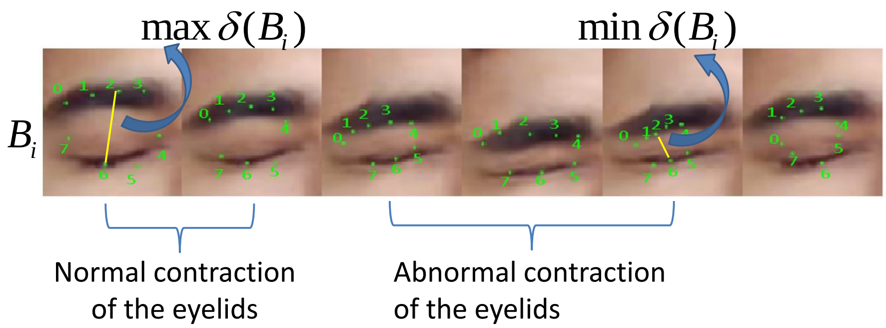

Fig. 7. This is an illustration to differentiate normal and abnormal contraction of eyelids during a blink. Due to the abnormal condition of Blepharospasm, the distance δB_i between the eyebrows (0-4) and the eyes (5-7) considerably decreases compared to the normal condition. The min-max ratio will during a blink will be significantly less compared to 1(for a perfectly normal condition) in such cases.

In Blepharospasm, when a patient blinks, it appears as if he is squeezing his eyes, as shown in Fig.7, which is due to forceful eye contractions. Note the decrease in the eye-eyebrow distance (between landmarks 2 and 6) in the fourth and fifth frames compared to others. For a normal person, such squeezing of eyes does not occur; consequently, there is less variation in those distances. On the contrary, there is a significant variation of this distance in a blink. Another noteworthy thing about Blepharospasm blink is that it lasts longer than normal. Considering these two factors, we first detect longer blinks and then compute the min-max ratio of these distances occurring in those long blinks. Since a video may comprise many blinks, we compute these ratios' average as the blink-normalness feature. If the person is normal, it is high; if the person is suffering from this disorder, it is lower. Let us assume $B = \{B_1, B_2, ...\}$ be the set of all the blinks in a video. Let $\delta(B_i)$ be the set of eye-eyebrow distances collected from the frames of the blink B_i, and let $\lambda(B_i)$ denotes duration of the blink B_i. Then, the blink normalness feature can be computed as

$$BN = avg\Big(\{\frac{\min \delta(B_i)}{\max \delta(B_i)} | \lambda(B_i) > \tau\}\Big), \tag{7}$$

where the condition on λ is to check if the blink is long enough. τ denotes the threshold time for a blink to be considered as a long blink. We set τ as 200ms, which is a bit more than the average blink duration of 150 ms, to cover some normal blinks that are little more than the average. We give the entire end-to-end work-flow process in Fig. 8. As the blink (incl. tight) detector detects blinks, a continuous stream of 1s will appear during the blink. The size of the continuous stream of 1s can be converted into the time domain using the following:

$$\lambda B_i = |B_i| \times \frac{L}{F}, \tag{8}$$

where L and F are the video's length and the total number of frames in the video. $|B_i|$ is the total number of frames across which the blink B_i occurred. After such conversion, we have the blink's duration in time-domain, and we can now easily compare it with τ.

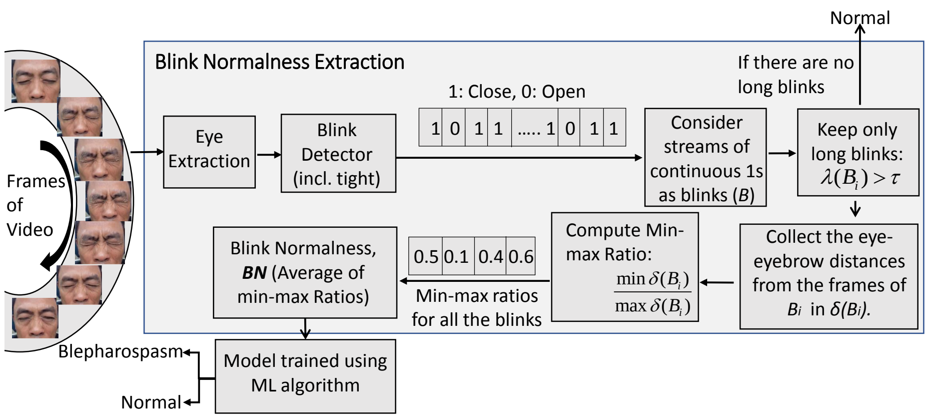

Fig. 8. Proposed approach for the extraction of blink normalness feature and the end-to-end workflow for the detection of Blepharospasm.

3.4 Intermediate Detection Models

We develop three intermediate detection models [Blink Detector, CD-like Iris Detector, and Blink Detector (incl. tight)] when designing the three features discussed above. For creating the blink detector, we build a blink detection dataset (details in the next section) comprising open and closed eye images to train a convolutional neural network (CNN). To create the CD-like iris detector, we build a CD-like-iris detection dataset (details in the next section) comprising eye-images with CD-like-iris and Non-CD-like-iris to train another convolutional neural network (CNN). Furthermore, for creating Blink Detector (incl. tight), we build a blink (incl. tight) detection dataset (details in the next section) comprising of open and closed (incl. tightly closed) eye-images to train another convolutional neural network (CNN). We use the same architecture shown in Fig. 9 for building all these three CNN-based detectors. The architecture's convolutional base consists of 10 convolutional layers, each followed by ReLU (Rectified Linear Unit)

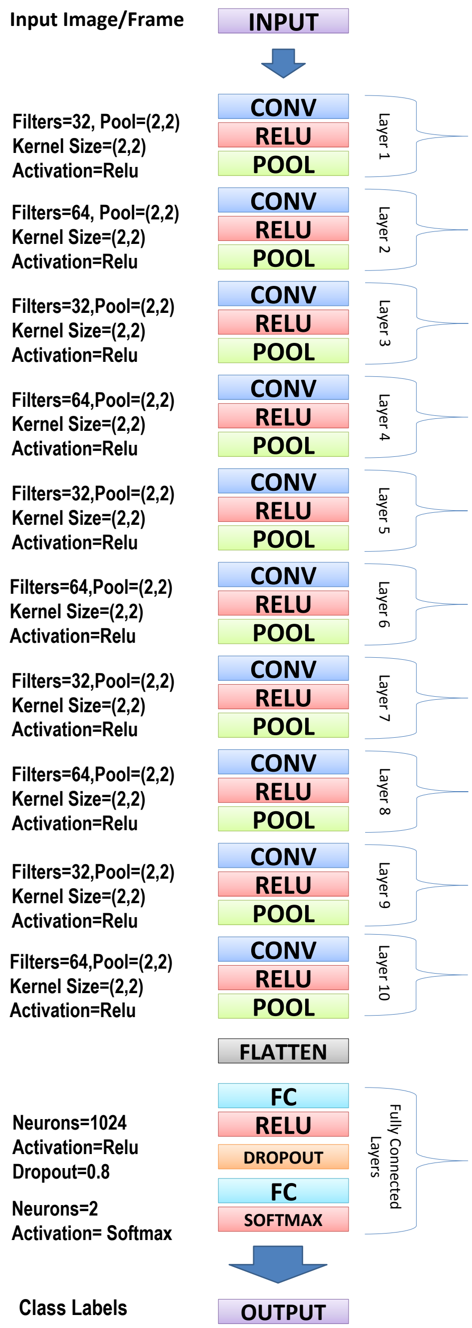

Fig. 9. CNN architecture used for developing our intermediate detectors.

and Max-pooling layers. The convolutional base's output is flattened and fed to the final fully-connected layers to generate the final output. The last layer for each uses softmax activation function, while the intermediate layer uses ReLU as the activation function. A dense layer of 1024 nodes was used at the end of each CNN followed with Dropout (0.8) to prevent any overfitting. While the input images for blink detectors are of resolutions 50x50, the CD-like iris detector has 100x100 resolution inputs because this requires more details to capture iris well. All our CNNs use the categorical cross-entropy loss (CEL) as loss function given by the equation below:

$$CEL(y, \hat{y}) = -\sum_{j=1}^{C}\sum_{i=1}^{|I|}\Big(y_{ij} * \log(\hat{y}_{ij})\Big) \tag{9}$$

where y_{ij} is the ground-truth label value of i^{th} image with respect to j^{th} class, and $\hat{y}_{ij}$ is the predicted label value of i^{th} image with respect to j^{th} class. C is the number of classes, which is two (2) in this case. $|I|$ is the total number of images in a mini-batch. We use the Adam optimization algorithm to train the CNN.

3.5 Final Detection Models

Since we have now generated the features for videos, we now tune over several supervised learning algorithms to build the best final detection models according to the cross-validation classification accuracy. We build three individual final detectors and one collective final detector. We found that Bell's Palsy disorder is best-detected individually using a model of Stochastic Gradient Descent (SGD). Similarly, the Cervical Dystonia disorder and Blepharospasm are best detected using Support Vector Machine (SVM) and Logistic Regression (LR), respectively. These models are all binary classifiers detecting if a subject has a particular disorder or not, and we call them individual-final detectors. We also build a multi-class classifier using all the three features to detects if a person has any of the neurological disorders considered in this work, and we call it a collective-final detector. We found that Support Vector Machine (SVM) helped us to build the best such multi-class classifier.

4 Experimental Results

In this section, we discuss different datasets created, the evaluation process employed, and our detectors' performance along with the comparisons.

4.1 Datasets

We develop two types of datasets: (i) One for developing intermediate detection models [i.e., blink detector, CD-like-iris detector, and blink detector (incl. tight) detector]. (ii) Another for final detection models (Bell's palsy detector, Cervical Dystonia detector, Blepharospasm detector, Neurological Disorders detector). While the intermediate ones take image/frame as input, the final ones take our video-level features as inputs. While the intermediate detectors' output is whether crucial events like blink and CD-like-iris have occurred in a particular frame, the final detectors' output is whether the person has a neurological disorder or not.

The details of the datasets we created for developing intermediate detectors are as follows: **Blink Detection Dataset**: It consists of 21896 open eye images and 19753 closed eyes images. The purpose is to detect blinks, i.e., to detect if a given eye image contains a closed eye. **CD-like-iris Detection Dataset**: It consists of 1481 eye images with iris at the center or on the left side, and 1261 eye images with iris on the right side. The purpose is to detect if the iris position is CD-like (extreme right side for the right eye). **Blink (incl. tight) Detection Dataset**: It consists of 17760 open-ey images and 30700 closed-eye images, some of which are tightly closed. The purpose again is to detect if a given eye is closed, including tightly closed. Note that these datasets consist of the only right eye images; we flip the left ones. Most of these images were collected from YouTube videos by manually cropping out the eyes frame-by-frame. However, we only clicked some images of persons performing tight blinks due to the lack of tight-blink images online.

The details of the datasets we created for developing the final detectors are as follows: **Bell's Palsy Video Dataset**: It consists of 41 videos of patients suffering from Bell's Palsy, and 34 videos are of healthy persons. **Cervical Dystonia Video Dataset**: It consists of 14 videos of patients who have Cervical Dystonia, and 36 videos are of healthy persons. **Blepharospasm Video Dataset**: It consists of 21 videos of patients suffering from Blepharospasm, and 30 videos are of healthy persons. The purpose of these three datasets so far is to detect neurological disorders individually. **Neurological Disorders Video Dataset**: It consists of the videos of patients from the above three datasets (41 Bell's Palsy, 14 Cervical Dystonia, and 21 Blepharospasm), and 30 videos are of healthy persons. The purpose of this dataset is to predict the disorders collectively. Most of these videos have been collected from YouTube. Only for the normal case, we recorded some of the videos ourselves.

All the above datasets have been made publicly available on the corresponding author's homepage: https://sites.google.com/site/koteswarraojerripothula/. Besides, we also run our experiments on publicly available YouTube Facial Palsy(YFP) [49] dataset. The YFP dataset is actually for detecting Bell's Palsy syndrome. So, this dataset contains videos of only patients.

4.2 Evaluation

We report classification accuracy throughout this work. Classification accuracy is defined as the percentage of correct predictions. Given the disparity in terms of availability of the labeled data for the intermediate and final detectors, different evaluation approaches have been taken.

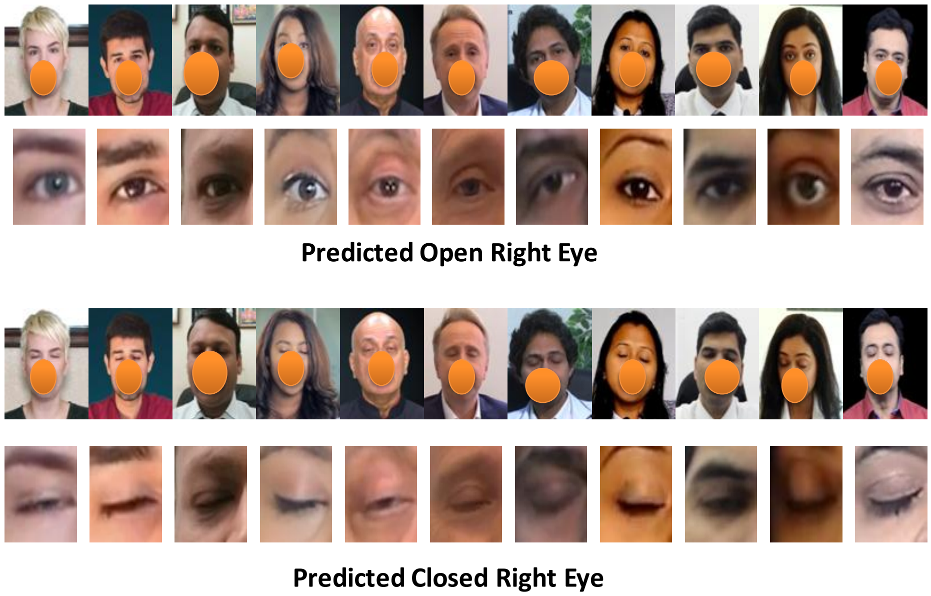

Fig. 10. Sample results of our blink detector.

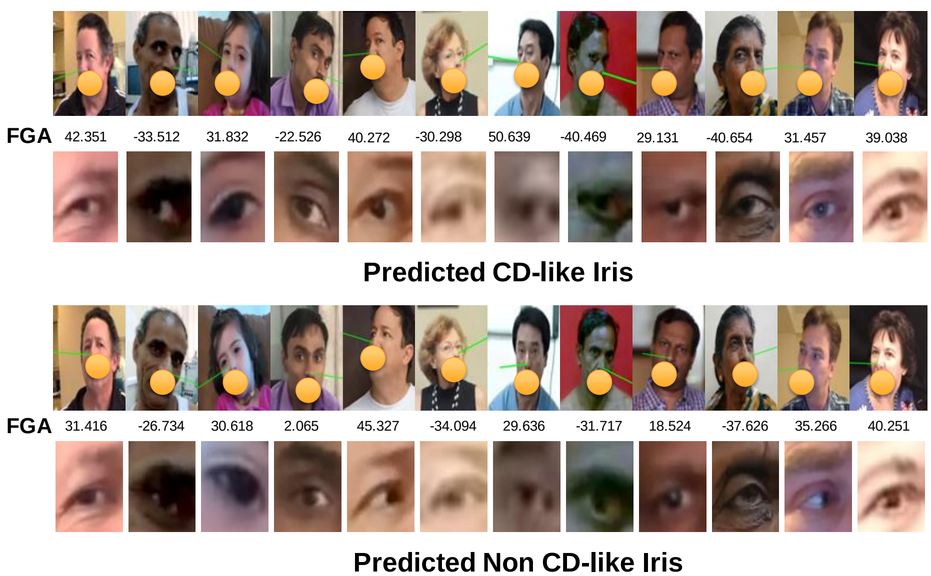

Fig. 11. Sample results of our CD-like-iris detector. The numbers below images indicate FGA (face gaze angle)

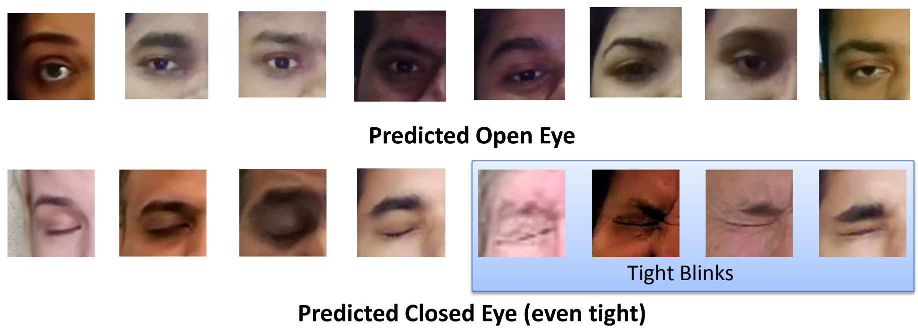

Fig. 12. Sample results of our blink detector (incl. tight).

For intermediate detectors, since we have labeled data in abundance, we can divide the labeled data into three parts: training, validation, and testing. Specifically, we use the hold-out approach to validation. In the hold-out approach, a portion of the dataset is kept aside for validation. We manage to set the best hyperparameters while training the CNNs for building these intermediate detectors with the help of validation scores. Once we set them, we test the trained CNN on the testing dataset.

In contrast, since we have minimal labeled data for the final detector, we avoid the testing phase altogether and employ the k-fold cross-validation approach on the entire labeled dataset. In k-fold cross-validation, the labeled data available for learning is divided into k parts to train on k-1 parts and validate on the remaining one part, and in this way, validated each of the parts and took the average of the accuracies obtained. Note that we set k to 3 for individual disorder datasets and to 10 for the collective disorders dataset. The cross-validation helps us in identifying the best algorithm to use among nine basic machine learning algorithms, namely AdaBoost (AB), Logistic Regression (LR), Naive Bayes (NB), Neural Networks (NN), Random Forests (RF), Stochastic Gradient Descent (SGD), Support Vectors Machine (SVM), Decision Tree (DT) and k-Nearest Neighbors (kNN). Once we choose the best learning algorithm, we train on the entire labeled dataset to create our final detectors. Note that we use the default setting of hyper-parameters for the algorithms available in Python's Orange package. In addition to the classification accuracy, we also report appropriate confusion matrices while reporting our detectors' performances.

4.3 Intermediate Detectors Results

We give the quantitative results of our intermediate detectors in terms of classification accuracy in Table 2 while comparing with other detectors built using Random Forest upon the features extracted from the pre-trained deep learning models like VGG19 and Inception-ResNet. While our detectors obtain classification accuracy of above 99% in all the three phases, they also obtain superior

Table 2. Comparison of our Intermediate detectors with those built using VGG and Inception-ResNet features on our three respective intermediate detection datasets.

Detectors	Training	Validation	Testing
VGG + Random Forest Blink Detector	**99.96**%	98.28%	98.09%
Inception-ResNet + Random Forest Blink Detector	99.90%	97.80%	97.56%
Proposed Blink Detector	99.80%	**99.72**%	**99.57**%
VGG + Random Forest CD-like-iris	99.89%	98.30%	97.86%
Inception-ResNet + Random Forest CD-like-iris	99.84%	97.38%	96.35%
Proposed CD-like-iris Detector	**99.95**%	**99.61**%	**99.55**%
VGG + Random Forest Blink incl. tight Detector	**99.96**%	98.47%	98.46%
Inception-ResNet + Random Forest Detector	99.88%	96.57%	96.34%
Proposed Blink (incl. tight) Detector	99.82%	**99.76**%	**99.06**%

results compared to those in the challenging validation and testing phases. We split the datasets into three parts: 70% for the training, 15% for validation, and 15% for testing. The confusion matrices obtained finally during the testing phase using our intermediate detectors are given in Table 3. We give the sample detection results of our intermediate detectors in Figures 10-12. Our blink detector detects well if an eye is open or closed; the CD-like-iris detector detects well if the eye of interest has cd-like-iris or not; and the blink detector (incl. tight) detects well if an eye is open or closed, including the ones closed tightly.

Table 3. Confusion matrices in the testing phase while building our intermediate detectors on our three respective intermediate detection datasets.

Blink Detector		Predicted Open	Predicted Close
	Actually Open	3538	15
	Actually Close	7	2849
CD-like Iris Detector		Predicted Non-CD-like Iris	Predicted CD-like Iris
	Actually Non-CD-like Iris	250	2
	Actually CD-like Iris	2	212
Blink Detector (incl. tight)		Predicted Open	Predicted Close
	Actually Open	2998	21
	Actually Close	16	5204

4.4 Individual Final Detectors Results

We build our individual-final detectors using only the dedicated features for different disorders to predict the presence of that particular disorder. The quantita-

Table 4. Classification Accuracies obtained during Cross Validation using different machine learning algorithms while building the individual final detectors on our three respective individual final detection datasets.

	AB	LR	NB	NN	RF	SGD	SVM	DT	kNN	Average
Bell's Palsy	90.7	92.0	90.7	93.3	93.3	**94.7**	93.3	90.7	93.3	92.4
Cervical Dystonia	90.0	80.0	86.0	90.0	88.0	90.0	**90.0**	88.0	88.0	87.8
Blephrospasm	92.2	**94.1**	90.2	92.2	92.2	86.3	92.2	90.2	92.2	91.3
Average	91.0	88.7	89.0	91.8	91.2	90.3	**91.8**	89.6	91.2	

tive results of these individual-final detectors in terms of classification accuracy during cross-validation are given in Table 4. It is clear from the table that SGD, SVM, and LR are the best algorithms for Bell's Palsy, Cervical Dystonia, and Blephrospasm detectors, respectively, using our features. Overall, SVM appears to work best on our features. Moreover, the high average accuracy across algorithms even in the low data setting shows that our features are robust, thanks to our disorder-specific design. We summarize our individual-final detectors in terms of the algorithm used and different accuracies obtained in Table 5. The proposed detectors obtain classification accuracy of above 90% for both cross-validation and training. Also, the closeness of validation accuracies with training accuracies show that models can generalize well. Like intermediate detectors, for these detectors also, we provide the cross-validation confusion matrices in Table 6.

Table 5. Classification Accuracies obtained during cross-validation and training of our individual final detectors on our three respective individual final detection datasets.

	Model	Cross-Validation	Training
Bell's Palsy	SGD	94.7	94.7
Cervical Dystonia	SVM	90.0	92.0
Blepharospasm	LR	94.1	92.2

4.5 Collective Final Detector Results

We build our collective-final detector using all three features to predict the presence of any of the three disorders. The dataset used here is our Neurological Disorders Video Dataset. We give the quantitative results of our collective final detection in terms of classification accuracy during cross-validation in Table 7. We also compare our features with that of 3D-CNN in the same table. First, we extract the 3D-CNN features [55] in the batches of 16 consecutive frames; their average (avg) and variance (var) are computed across the batches to generate features at the video level. They are added (+) and concatenated ($\oplus$) to generate

Table 6. Confusion Matrices obtained during cross-validation of our individual final detectors on our three respective individual final detection datasets.

		Predicted No	Predicted Yes
Bell's Palsy	Actual No	32	2
	Actual Yes	2	39
		Predicted No	Predicted Yes
Cervical Dystonia	Actual No	33	3
	Actual Yes	2	12
		Predicted No	Predicted Yes
Blephrospasm	Actual No	29	1
	Actual Yes	2	19

Table 7. Comparison of proposed features together with that of 3D-CNN [55] in terms of classification accuracy using different machine learning algorithms while building the collective final detector. Different features/video are built using 3D-CNN features extracted from the batches of 16 consecutive frames like average, variance, their addition and concatenation. The dataset used here is our Neurological Disorders Video Dataset.

	AB	LR	NB	NN	RF	SGD	SVM	DT	kNN	Average
3DCNN[55](avg)	**77.4**	**83.0**	64.2	74.5	80.2	76.4	67.0	**77.4**	72.6	74.7
3DCNN[55](var)	70.8	70.8	70.8	65.1	71.7	65.1	67.0	68.9	66.0	68.5
3DCNN[55](avg+var)	**77.4**	**83.0**	63.2	75.5	73.6	74.5	70.8	72.6	75.5	74.0
3DCNN[55](avg⊕var)	72.6	**83.0**	65.1	73.6	78.3	70.8	66.0	75.5	74.5	73.3
Ours	72.6	77.4	**76.4**	**81.1**	**81.1**	**80.2**	**83.0**	72.6	**76.4**	**77.9**

Table 8. Confusion Matrix generated while performing cross-validation to build our collective final detector (SVM). The dataset used here is our Neurological Disorders Video Dataset.

	Predicted Bell's Palsy	Predicted Cervical Dystonia	Predicted Blepharospasm	Predicted Normal
Actually Bell's Palsy	38	0	2	1
Actually Cervical Dystonia	2	6	5	1
Actually Blepharospasm	2	0	18	1
Actually Normal	1	1	2	26

more features per video for comparison. We can see that our features obtain the best results in 6 out of the 9 learning algorithms. It is clear from the table that SVM gives the best performance of 83.0%; therefore, we use it to build our collective final detector on the entire labeled dataset to obtain the training accuracy of 84.0%. The closeness of the validation accuracy with the training accuracy shows that the model can generalize well. Moreover, our features achieve 4.3% relative improvement over the best of 3D-CNN's features here (avg.) in terms of the average performance. For our collective final detector also, we provide the confusion matrix obtained while performing the cross-validation in Table 8. We also show ROC curves obtained while performing the cross-validation in Fig. 13. It can be seen that we could not get a very good ROC curve for the Cervical Dystonia class, as compared to other classes. It is because of having fewer samples than other classes.

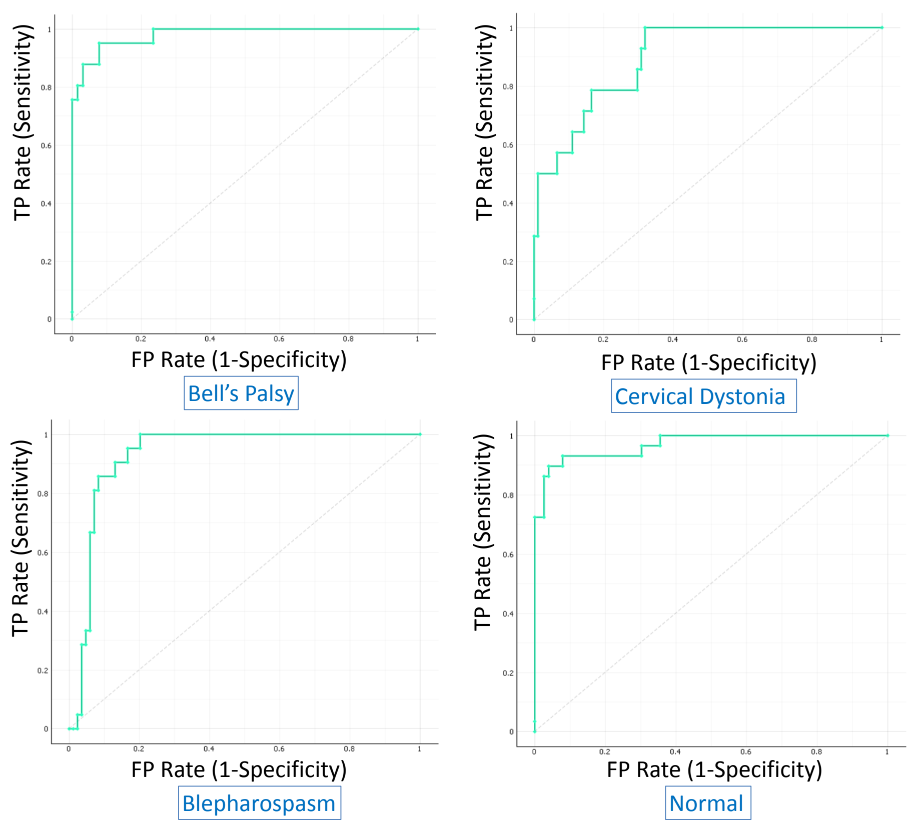

Fig. 13. ROC curves obtained during the cross-validation of our collective final detector.

4.6 Discussion and Limitations

Note that there are no publicly available datasets or prior works for detecting Cervical Dystonia and Blepharospasm. However, since there is a publicly available video dataset name YouTube Facial Palsy (YFP) dataset, we test our Bell's Palsy detector (trained on our Bell's Palsy Video dataset) on it. Out of the 31 videos of Bell's Palsy patients, our SVM-based Bell's Palsy detector could detect 28 videos as Bell's Palsy videos. It means the detector is 90.3% accurate. Since there is prior work[50] on detecting Bell's Palsy from videos, we extract the features from the network presented in [50] and perform the same experiments as we did with our Blink Similarity feature on our Bell's Palsy Video dataset. The results are compared with ours in Table 9. It can be seen that our blink similarity feature outperforms features of [50] no matter which machine learning algorithm is used. [50] achieves the best cross-validation accuracy of 81.6% using NN, whereas our feature achieves the best cross-validation accuracy of 94.7% using SGD. In terms of average accuracy across the machine learning algorithms, while [50] obtains 70.8%, our feature obtains 92.4%. This comparison shows that our blink normalness feature is superior to that of [50], thanks to our disorder-focused design.

Table 9. Comparison of 3DPalsyNet [50] features with our Blink Similarity feature in terms of classification accuracies obtained during cross-validation using different machine learning algorithms on our Bell's Palsy Video dataset.

	AB	LR	NB	NN	RF	SGD	SVM	DT	kNN	Average
3DPalsyNet [50]	62.2	73.8	66.8	81.6	68.2	71.0	70.8	63.0	79.7	70.8
Blink Similarity (Ours)	**90.7**	**92.0**	**90.7**	**93.3**	**93.3**	**94.7**	**93.3**	**90.7**	**93.3**	**92.4**

Our features can also provide the severity degree of a disorder. For example, the farther the blink similarity feature value from 1, the severer is the Bell's Palsy. Similarly, the farther the CD-like appearance feature value from 0, the severer is the Cervical Dystonia. Also, the farther the blink normalness feature value from 1, the severer is the Blepharospasm. Thus, the proposed features can assist in the rehabilitation [56] process, as well.

Note that our proposed method entirely relies on detecting certain eye events (blinks, iris position, and blinks(incl. Tight)). It might be possible that these events may not get adequately captured in some videos. For example, there could be a so short a video that not even one blink gets captured. We also require the subject to look at the camera at least once for accurately detecting Cervical Dystonia. These are some of the limitations of our proposed method.

Conclusion

We attempted the eyes-based detection of three neurological disorders: Bell's Palsy, Cervical Dystonia, and Blepharospasm. We design three novel eye-based features at the video-level and develop the machine learning models to detect these disorders individually and collectively. The three features are blink similarity, CD-like-iris appearance, blink normalness. While designing them, we also had to develop three intermediate detection models (CNNs): blink detector, CD-like-iris detector, blink detector (incl. tight). Our extensive experiments on the seven datasets (3 for intermediate detection + 4 for final detection) developed by us demonstrate competitive performance. While the proposed approach can detect the disorders with at least 90% accuracy individually, it can detect them collectively with 83% accuracy. Also, since all we need is eye regions, all other facial features can be masked. We will explore this security feature in greater detail in our future works. In the future, we plan to develop more advanced features for the detection of neurological disorders and work on rehabilitation processes by measuring the severity of these disorders.

References

1. Singh, D.N., Hashmi, M.F., Sharma, S.K.: Predictive analytics & modeling for modern health care system for cerebral palsy patients. Multimedia Tools and Applications (2019) 1–23
2. Yolcu, G., Oztel, I., Kazan, S., Oz, C., Palaniappan, K., Lever, T.E., Bunyak, F.: Facial expression recognition for monitoring neurological disorders based on convolutional neural network. Multimedia Tools and Applications **78**(22) (2019) 31581–31603
3. Goyal, D., Jerripothula, K.R., Mittal, A.: Detection of gait abnormalities caused by neurological disorders. In: 2020 IEEE 22nd International Workshop on Multimedia Signal Processing (MMSP). (2020) 1–6
4. Sullivan, F., Daly, F., Gagyor, I.: Antiviral agents added to corticosteroids for early treatment of adults with acute idiopathic facial nerve paralysis (bell palsy). Jama **316**(8) (2016) 874–875
5. Pitts, D.B., Adour, K.K., Hilsinger Jr, R.L.: Recurrent bell's palsy: analysis of 140 patients. The Laryngoscope **98**(5) (1988) 535–540
6. Leplow, B., Eggebrecht, A., Pohl, J.: Treatment satisfaction with botulinum toxin: a comparison between blepharospasm and cervical dystonia. Patient preference and adherence **11** (2017) 1555
7. Kongsaengdao, S., Maneeton, N., Maneeton, B.: Long-term quality of life in cervical dystonia after treatment with abobotulinum toxin a: a 2-year prospective study. Neuropsychiatric disease and treatment **14** (2018) 1119
8. Huang, Y., Cai, M., Li, Z., Sato, Y.: Predicting gaze in egocentric video by learning task-dependent attention transition. In: Proceedings of the European Conference on Computer Vision (ECCV). (2018) 754–769
9. Broadbent, T.J., Wesley, R.E., Mawn, L.A.: A survey of current blepharospasm treatment patterns among oculoplastic surgeons. Ophthalmic Plastic & Reconstructive Surgery **32**(1) (2016) 24–27

10. Defazio, G., Hallett, M., Jinnah, H.A., Conte, A., Berardelli, A.: Blepharospasm 40 years later. Movement Disorders **32**(4) (2017) 498–509
11. Jerripothula, K.R., Cai, J., Yuan, J.: Cats: Co-saliency activated tracklet selection for video co-localization. In Leibe, B., Matas, J., Sebe, N., Welling, M., eds.: Computer Vision – ECCV 2016, Cham, Springer International Publishing (2016) 187–202
12. Divjak, M., Bischof, H.: Eye blink based fatigue detection for prevention of computer vision syndrome. In: MVA. (2009) 350–353
13. Anderson, T.J., MacAskill, M.R.: Eye movements in patients with neurodegenerative disorders. Nature Reviews Neurology **9**(2) (2013) 74–85
14. Pusiol, G., Esteva, A., Hall, S.S., Frank, M., Milstein, A., Fei-Fei, L.: Vision-based classification of developmental disorders using eye-movements. In Ourselin, S., Joskowicz, L., Sabuncu, M.R., Unal, G., Wells, W., eds.: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016, Cham, Springer International Publishing (2016) 317–325
15. Wang, T., Zhang, S., Dong, J., Liu, L., Yu, H.: Automatic evaluation of the degree of facial nerve paralysis. Multimedia Tools and Applications **75**(19) (2016) 11893–11908
16. Jerripothula, K.R., Cai, J., Yuan, J.: Quality-guided fusion-based co-saliency estimation for image co-segmentation and colocalization. IEEE Transactions on Multimedia **20**(9) (2018) 2466–2477
17. Jerripothula, K.R., Cai, J., Lu, J., Yuan, J.: Object co-skeletonization with co-segmentation. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). (2017) 3881–3889
18. Song, A., Wu, Z., Ding, X., Hu, Q., Di, X.: Neurologist standard classification of facial nerve paralysis with deep neural networks. Future Internet **10**(11) (2018) 111
19. Jerripothula, K.R., Rai, A., Garg, K., Rautela, Y.S.: Feature-level rating system using customer reviews and review votes. IEEE Transactions on Computational Social Systems (2020) 1–10
20. Jerripothula, K.R., Cai, J., Yuan, J.: Image co-segmentation via saliency co-fusion. IEEE Transactions on Multimedia **18**(9) (2016) 1896–1909
21. Chen, S., Liu, C.: Precise eye detection using discriminating hog features. In: International Conference on Computer Analysis of Images and Patterns, Springer (2011) 443–450
22. Kim, H., Jo, J., Toh, K.A., Kim, J.: Eye detection in a facial image under pose variation based on multi-scale iris shape feature. Image and Vision Computing **57** (2017) 147–164
23. Sharma, R., Savakis, A.: Lean histogram of oriented gradients features for effective eye detection. Journal of Electronic Imaging **24**(6) (2015) 063007
24. Amos, B., Ludwiczuk, B., Satyanarayanan, M., et al.: Openface: A general-purpose face recognition library with mobile applications. CMU School of Computer Science **6**(2) (2016)
25. Luo, B., Shen, J., Cheng, S., Wang, Y., Pantic, M.: Shape constrained network for eye segmentation in the wild. In: The IEEE Winter Conference on Applications of Computer Vision. (2020) 1952–1960
26. Hansen, D.W., Ji, Q.: In the eye of the beholder: A survey of models for eyes and gaze. IEEE transactions on pattern analysis and machine intelligence **32**(3) (2009) 478–500

27. Appel, T., Santini, T., Kasneci, E.: Brightness-and motion-based blink detection for head-mounted eye trackers. In: Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct. (2016) 1726–1735
28. Schillingmann, L., Nagai, Y.: Yet another gaze detector: An embodied calibration free system for the icub robot. In: 2015 IEEE-RAS 15th International Conference on Humanoid Robots (Humanoids), IEEE (2015) 8–13
29. Anas, E.R., Henriquez, P., Matuszewski, B.J.: Online eye status detection in the wild with convolutional neural networks. In: International Conference on Computer Vision Theory and Applications. Volume 7., SciTePress (2017) 88–95
30. Kim, K.W., Hong, H.G., Nam, G.P., Park, K.R.: A study of deep cnn-based classification of open and closed eyes using a visible light camera sensor. Sensors **17**(7) (2017) 1534
31. Cortacero, K., Fischer, T., Demiris, Y.: Rt-bene: A dataset and baselines for real-time blink estimation in natural environments. In: Proceedings of the IEEE International Conference on Computer Vision Workshops. (2019) 0–0
32. Zhang, X., Sugano, Y., Fritz, M., Bulling, A.: Mpiigaze: Real-world dataset and deep appearance-based gaze estimation. IEEE transactions on pattern analysis and machine intelligence **41**(1) (2017) 162–175
33. Palmero, C., Selva, J., Bagheri, M.A., Escalera, S.: Recurrent cnn for 3d gaze estimation using appearance and shape cues. arXiv preprint arXiv:1805.03064 (2018)
34. Fischer, T., Jin Chang, H., Demiris, Y.: Rt-gene: Real-time eye gaze estimation in natural environments. In: Proceedings of the European Conference on Computer Vision (ECCV). (2018) 334–352
35. Chinsatit, W., Saitoh, T.: Cnn-based pupil center detection for wearable gaze estimation system. Applied Computational Intelligence and Soft Computing **2017** (2017)
36. Fuhl, W., Santini, T., Kasneci, G., Kasneci, E.: Pupilnet: Convolutional neural networks for robust pupil detection. arXiv preprint arXiv:1601.04902 (2016)
37. Xia, Y., Lou, J., Dong, J., Qi, L., Li, G., Yu, H.: Hybrid regression and isophote curvature for accurate eye center localization. Multimedia Tools and Applications **79**(1) (2020) 805–824
38. Park, S., Mello, S.D., Molchanov, P., Iqbal, U., Hilliges, O., Kautz, J.: Few-shot adaptive gaze estimation. In: Proceedings of the IEEE International Conference on Computer Vision. (2019) 9368–9377
39. Kothari, R., Yang, Z., Kanan, C., Bailey, R., Pelz, J.B., Diaz, G.J.: Gaze-in-wild: A dataset for studying eye and head coordination in everyday activities. Scientific reports **10**(1) (2020) 1–18
40. Lee, K.F., Chen, Y.L., Yu, C.W., Chin, K.Y., Wu, C.H.: Gaze tracking and point estimation using low-cost head-mounted devices. Sensors **20**(7) (2020) 1917
41. Park, M., Seo, J., Park, K.: Pc-based asymmetry analyzer for facial palsy study in uncontrolled environment: A preliminary study. Computer methods and programs in biomedicine **99**(1) (2010) 57–65
42. Wang, T., Dong, J., Sun, X., Zhang, S., Wang, S.: Automatic recognition of facial movement for paralyzed face. Bio-medical materials and engineering **24**(6) (2014) 2751–2760
43. He, S., Soraghan, J.J., O'Reilly, B.F., Xing, D.: Quantitative analysis of facial paralysis using local binary patterns in biomedical videos. IEEE Transactions on Biomedical Engineering **56**(7) (2009) 1864–1870

44. Ngo, T.H., Seo, M., Matsushiro, N., Xiong, W., Chen, Y.W.: Quantitative analysis of facial paralysis based on limited-orientation modified circular gabor filters. In: 2016 23rd International Conference on Pattern Recognition (ICPR), IEEE (2016) 349–354
45. Kim, H., Kim, S., Kim, Y., Park, K.: A smartphone-based automatic diagnosis system for facial nerve palsy. Sensors **15**(10) (2015) 26756–26768
46. Asthana, A., Zafeiriou, S., Cheng, S., Pantic, M.: Incremental face alignment in the wild. In: Proceedings of the IEEE conference on computer vision and pattern recognition. (2014) 1859–1866
47. Jerripothula, K.R., Cai, J., Yuan, J.: Qcce: Quality constrained co-saliency estimation for common object detection. In: 2015 Visual Communications and Image Processing (VCIP). (2015) 1–4
48. Jerripothula, K.R., Cai, J., Yuan, J.: Efficient video object co-localization with co-saliency activated tracklets. IEEE Transactions on Circuits and Systems for Video Technology **29**(3) (2019) 744–755
49. Hsu, G.S.J., Kang, J.H., Huang, W.F.: Deep hierarchical network with line segment learning for quantitative analysis of facial palsy. IEEE Access **7** (2018) 4833–4842
50. Storey, G., Jiang, R., Keogh, S., Bouridane, A., Li, C.T.: 3dpalsynet: A facial palsy grading and motion recognition framework using fully 3d convolutional neural networks. IEEE Access **7** (2019) 121655–121664
51. Hara, K., Kataoka, H., Satoh, Y.: Can spatiotemporal 3d cnns retrace the history of 2d cnns and imagenet? In: Proceedings of the IEEE conference on Computer Vision and Pattern Recognition. (2018) 6546–6555
52. Molho, E.S., Stacy, M., Gillard, P., Charles, D., Adler, C.H., Jankovic, J., Schwartz, M., Brin, M.F.: Impact of cervical dystonia on work productivity: an analysis from a patient registry. Movement disorders clinical practice **3**(2) (2016) 130–138
53. Bedarf, J.R., Kebir, S., Michelis, J.P., Wabbels, B., Paus, S.: Depression in blepharospasm: a question of facial feedback? Neuropsychiatric disease and treatment **13** (2017) 1861
54. Levenberg, K.: A method for the solution of certain non-linear problems in least squares. Quarterly of applied mathematics **2**(2) (1944) 164–168
55. Tran, D., Bourdev, L., Fergus, R., Torresani, L., Paluri, M.: Learning spatiotemporal features with 3d convolutional networks. In: Proceedings of the IEEE international conference on computer vision. (2015) 4489–4497
56. Avola, D., Cinque, L., Foresti, G.L., Marini, M.R., Pannone, D.: Vrheab: a fully immersive motor rehabilitation system based on recurrent neural network. Multimedia Tools and Applications **77**(19) (2018) 24955–24982

PDF file

Click here to access/download
Video Clip
neuro_minor_revision2.pdf