Eyes-based Detection of Neurological Disorders in Videos

Journal:	*IEEE Transactions on Circuits and Systems for Video Technology*
Manuscript ID	TCSVT-03491-2019
Manuscript Type:	Transactions Papers - Regular Issue
Date Submitted by the Author:	23-Jul-2019
Complete List of Authors:	Kunwar, Sharik; Graphic Era University JERRIPOTHULA, Koteswar; Graphic Era University, Department of Computer Science and Engineering; Nanyang Technological University, Interdisciplinary Graduate School Nagpal, Pragya; Graphic Era Hill University Mittal, Ankush; Graphic Era University
EDICS:	2 ☐ IMAGE/VIDEO ANALYSIS AND COMPUTER VISION, Medicine < 2.11.1 ☐ Vision-for-X < 2.11 ☐ Vision Applications < 2 ☐ IMAGE/VIDEO ANALYSIS AND COMPUTER VISION, Eyes Tracking < 2.5.4 ☐ Human Tracking < 2.5 ☐ Tracking < 2 ☐ IMAGE/VIDEO ANALYSIS AND COMPUTER VISION, 2.6.1 ☐ Feature Extraction and Description < 2.6 ☐ Detection, Recognition and Classification < 2 ☐ IMAGE/VIDEO ANALYSIS AND COMPUTER VISION, 7.10.8 ☐ Computer-Assisted Screening and Diagnosis < 7.10 ☐ Biomedical and Biological Imaging < 7 ☐ IMAGE/VIDEO APPLICATIONS

Eyes-based Detection of Neurological Disorders in Videos

Sharik Kunwar, Koteswar Rao Jerripothula, Pragya Nagpal, Ankush Mittal

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 just from eyes of the subjects in the videos. Although some previous works have utilized visual information to detect such disorders, none of them have yet especially explored eyes to do the same, and for a multiple of them. To the best of our knowledge, this work is first in establishing a relationship between eyes-based features and these neurological disorders as far as video analytics research is concerned. Specifically, we develop novel features, namely blink similarity, CD-like iris appearance, and blink normalness, with the help of CNNs and build machine learning models that can predict these neurological disorders. Since such models are data-driven, we needed to collect numerous images/videos and developed novel benchmark datasets for these disorders. Our exhaustive experiments demonstrate that the proposed method obtains accuracies of above 90% while detecting these disorders individually and an accuracy of about 83% while detecting these disorders collectively. Our proposed eyes-based detection approach is much cheaper and more convenient compared to the existing detection approaches having complicated hardware requirements, as the proposed approach requires only a video of the subject.**

Index Terms—**eye, neurological, video, disorders, features, blink, Bell's Palsy, Cervical Dystonia, Blepharospasm.**

I. Introduction

The neurological disorders are diseases of the nervous system. A nervous system comprises of brain, spinal cord, and nerves. The nerves not only connect the brain and the spine mutually but also connect them to different parts of the body. The nervous system is highly complex: it coordinates our actions and sensory information by transmitting signals to and from different parts of our body via its nerves. As a result, when there are some disorders in this system, we can have trouble moving, speaking, swallowing, breathing, blinking, learning, etc. These troubles usually manifest in the form of visible symptoms that can be recorded in videos and analyzed for diagnosis. Considering it as a video analytics [1]–[5] problem, we employ computer vision [6]–[12] and machine learning [13]–[16] algorithms to build a computer-assisted diagnosis system. Given a video of subject, as shown in Fig. 1, our main goal is to automatically detect if the subject in the video has any of the neurological disorders. Particularly, we exploit symptoms related to eyes and detect the following three neurological disorders.

Sharik Kunwar, Koteswar Rao Jerripothula, and Ankush Mittal are with Graphic Era University, India. Pragya Nagpal is with Graphic Era Hill University, India. Email: kunwar.sharik@gmail.com, koteswar001@e.ntu.edu.sg, nagpalpragya85@gmail.com, and dr.ankush.mittal@gmail.com. Contact author: Koteswar Rao Jerripothula

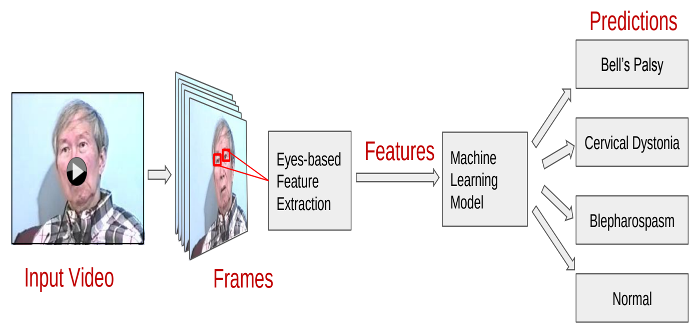

Fig. 1. Given a video of subject, the objective is to automatically detect, using the eye-based features, if the subject has any of the three neurological disorders: Bell's Palsy, Cervical Dystonia, and Blepharospasm.

Bells Palsy: It's a condition where muscles on one side of the face become weak or paralyzed. It's caused by some kind of trauma to the seventh cranial nerve. According to [17], it affects one in every 60 persons once in a lifetime at least. And [18] reports that 7% of those patients are likely to have it again. This makes half of one's face appear to droop. The smile becomes one-sided, and eye muscles on the paralyzed side become so weak that eye resists closing, causing its blink rate to go down.

Cervical dystonia: It's a condition where neck muscles contract involuntarily causing the head to turn sideways. It's believed that it occurs due to an abnormality in the basal ganglia or other brain regions that control movement. However, in most cases, the exact cause is difficult to be known. [19] reports the presence of depression and anxiety in large percentages of such patients. It has four forms: the most common is Torticollis, and the other three are Laterocollis, Anterocollis, and Retrocollis, as mentioned in [20]. It's Torticollis where head gets turned to sideways. With the turned head, patients strain their eyes as well, trying to adjust the eye-gaze for conversing, because they now have to push their iris to an extreme quite often.

Blepharospasm: It's a rare condition where one experiences an involuntary repeated forcible contraction of eyelid and forehead muscles, as mentioned in [21]. Although the exact cause is difficult to point out, just like Cervical Dystonia, it is believed that it's also due to the malfunctioning in Basal Ganglia, a region of the brain. It affects only 20 to 133 people per million according to [22]; however, it's a disorder for an individual and can also lead to accidents for the kind of condition it is.

Each of the above disorders has some eye-related visible symptom or other. Visible symptoms, in general, are very important signals which the body gives to alarm us, and most importantly they can be easily captured in the videos for computer-assisted diagnosis. Even doctors give so much importance to such symptoms at the time of diagnosis, although the real disease may be hidden deep within. We attempt to automate the process of detecting these three neurological disorders, trying to assist in the diagnosis. This makes diagnosis cheap and convenient: it just requires capturing a video. Nevertheless, videos captured may not always be focused on the area of concern where the symptom appears. However, eyes are easy to locate using existing computer vision algorithms, for a particular pattern they possess. Therefore, we study eyes to detect the three neurological disorders discussed above. In Bell's Palsy, we observe that the blinking rate of two eyes will differ because it slows down for the paralyzed side. In Cervical Dystonia, we observe that iris normally moves to an extreme to adjust the eye-gaze to counter the abnormal facial gaze while conversing. In Bleprosphasm, we observe that due to the tendency of blinking tightly distance between eye center and eyebrows varies a lot while blinking. These localized observations can be modeled into predictors (features) of videos of the subjects for building an eye-based detection system of these neurological disorders with the help of existing machine learning algorithms.

However, there are some challenges while building such neurological disorders detection system: (i) There are no publicly available video datasets for these disorders. (ii) There are no ready-made video features to represent these eye-based observations. (iii) To the best of our knowledge, there are fewer works done before on the videos to detect neurological disorders compared to images. It's done only for Bell's Palsy. There are works done on images, but they can be misleading at times because same visual symptoms can be manifested in a normal person temporarily; however, in a video, such manifestation cannot occur all the time for normal persons, like a patient. For example, a wink of a normal person can be regarded as a tight blink of Blepharospasm if it's an image, but he cannot keep winking throughout the video. (iv) The features designed for videos should be such that they take into account varying video-durations.

We address the above challenges in the following manner: (i) We build separate video datasets for each of the disorder comprising of both patients and normal persons. (ii) We develop three eye-based features for a video, namely blink similarity (for Bell's Palsy), CD-like-iris appearance (for Cervical Dystonia) and blink normalness (for Blepharospasm). To develop them, we had to build three other image datasets to perform intermediate detections for monitoring eye-related phenomenons like blink, CD-like iris (iris in the extreme to adjust the eye-gaze) and blink including the tight ones. (iii) We base our system on entirely on videos instead of possibly misleading images. (iv) We device several counters and take their ratios for feature generation to avoid effects of video-durations. The features generated in this way are used to build models (using machine learning algorithms) that can predict the possibility of these neurological disorders; and thus, we create an eyes-based detection system of neurological disorders in videos. Our exhaustive experiments on the datasets developed demonstrate that (i) our features are superior and more meaningful compared to the ready-made features already available, (ii) proposed system achieves accuracies of about 90% while detecting these disorders individually, and (iii) proposed system achieves accuracy of about 83% while detecting these disorders collectively.

Our contributions through this work are as follows: (1) We develop benchmark datasets for neurological disorders and eye-monitoring (blink, iris position, etc.). (2) We develop eye-based features of videos that independent of video-durations. (3) We develop intermediate detectors and final detectors to monitor eyes and to detect neurological disorders, respectively. (4) We are first to establish a relationship between eye-related features and a few neurological disorders from videos. The remaining paper is structured in the following order: there is a discussion on related works; then, there is a discussion on the proposed method; and lastly, there is a discussion on the experiments conducted.

II. Related Work

In this section, we discuss different related works of the three neurological disorders.

A. *Bells Palsy:*

In [23], the video clips captured with webcam of PC is analyzed for diagnosis by measuring the asymmetry index around the mouth region. However, there are people with one-sided smiles; therefore, relying on the mouth only could be misleading. So, our work focused on eyes can act as another detection tool for a conclusive diagnosis. In [24], 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, patterns in the facial movement are observed, and videos are classified. However, this method requires videos of the subjects to be taken in a controlled environment (see [24] for more details) and need specific movements to be carried out for accurate detection. Similarly, [25], [26] also have constraints with video recording environment, video lengths, etc. In contrast, we attempt to detect the disorder in the wild, even the videos we collect are from the web. Interestingly, in [27], 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 [28] also that proposes incremental face alignment method to handle videos in the wild for this problem. However, the problem with such facial points-based analysis is that it's quite possible that facial points may not exactly match because of the challenges involved. Therefore, as far as Bell's Palsy is concerned, we use facial points only for detection (localization) of eye region; once detected, we just rely entirely on the dissimilarity in blinks of the two sides, which is a novel way of looking at detecting Bell's Palsy. Recently, a deep learning work [29] was proposed; but again, it depends upon the accuracy of facial landmarks entirely whereas we can afford to handle its inaccuracy by cropping more region surrounded by the eye detected from these landmarks.

B. Cervical Dystonia and Blepharospasm

Unlike Bell's Palsy where there are works done using visual information, there is hardly any such work for Cervical Dystonia and Blepharospasm. The focus mainly has been to provide prospective studies on the use of different drugs for treatment purpose and how these disorders affect daily living such as low work productivity [30] and increased level of depression [31]. For example in Cervical Dystonia, in [19], authors provided a survey on using multiple injections of botulinum toxin A while authors in [20] provided a survey of using abobotulinum toxin A (Abo-BTX A) and neubotulinum toxin A (Neu-BTX A). In a study provided in [26], authors conclude that Cervical Dystonia is reported to be 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.

III. Proposed Method

We propose methodologies to extract three eye-related features, namely blink similarity, CD-like iris appearance and blink normalness for detecting Bell's palsy, Cervical Dystonia and Blepharospasm, respectively, from the videos. While developing these features, we require intermediate detection models to monitor eye-blinking and iris-position. After extracting the features, we build final detection models to classify the videos into normal or a particular disorder.

A. Blink similarity (especially for Bell's Palsy)

Bells Palsy is a unilateral condition in which muscles in 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 regular closure of eyelids for a brief period. This eye-blinking phenomenon can get severely affected in the paralyzed side of a patient suffering from Bells Palsy. It leads to an abnormal difference in the blinking of the two eyes of a person. As shown in the Fig.2, there is a significant difference between 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 eye of normal side blinks once for a patient suffering from Bell's Palsy. So, the blinking of a normal person will usually be simultaneous, whereas there will be 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 due to some instantaneous agitation, etc. Therefore, these blinks need to be monitored for 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 will be much lower than that of the normal side because of difficulty in closing it. And for a normal patient ECT of the two sides should be almost equal.

Assuming eye-blinking is a periodic affair, there will be two phases: eye-closure time (EC) and eye-open time (EO). Considering the assumption of periodicity, these phases will have a constant value for a given eye of a given person. So, they still vary across the eyes and vary across the persons. The blink time-period can be computed as $EC+EO$. If L denotes length (total time duration) of a video, the total eye-closure time (ECT) now can be computed 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) because an eye will be either closed or opened in any frame. So, multiplying the length of video with such a fraction will yield the total time duration during which the eye was closed, i.e. ECT. Although L can be extracted 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 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 left eye is closed (ECF_L) and the number of frames in which 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. And, 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's a closed eye or an open eye. From any frame, we extract two eyes of the person and flip left one sideways because the CNN is trained on all right-types (even left-types are flipped). Although for a normal person the ratio value would be close to 1, it can deviate from 1

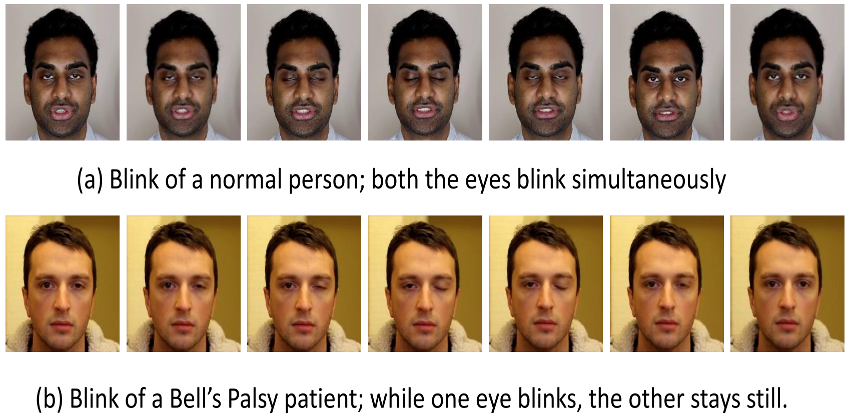

Fig. 2. An illustration demonstrating the difference between the blink of a normal person and that of a Bell's Palsy patient.

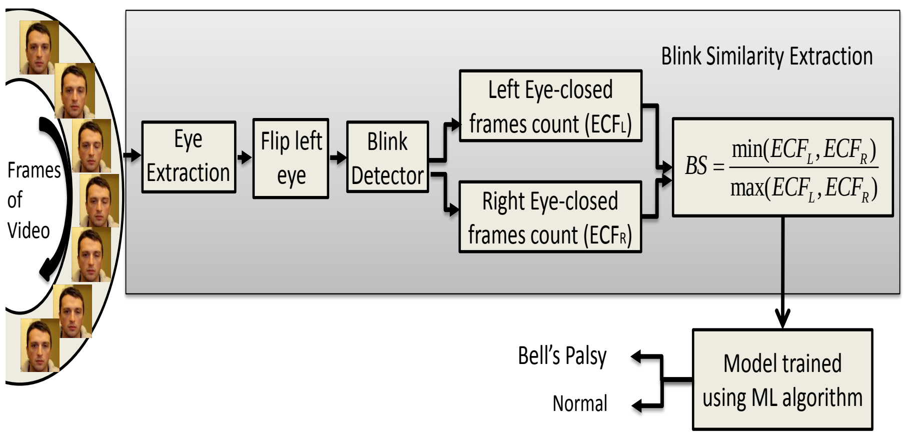

Fig. 3. Proposed approach for the extraction of blink-rate similarity feature and the end-to-end workflow for the detection of Bell's Palsy.

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 will be 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:

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

which will automatically ensure the deviation of patients' feature scores from 1 towards one side only; that is, towards 0. Thus, we can have a single decision boundary. And, 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's difficult to find the optimal decision boundary between feature values of normal persons and the 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. The entire end-to-end work-flow process is given in Fig. 3.

B. CD-like iris appearance (especially for Cervical Dystonia)

A person suffering from Cervical Dystonia has an abnormal face gaze as their head gets turned; as a result, one eye comes to the front and another goes back. Due to this abnormal face gaze, the patient is forced to adjust his eye gaze to converse. If the head is turned towards left, the eye-gaze is adjusted to move towards right, and if the head is turned to the right, the eye-gaze is adjusted to move towards left. More one does like that more is the chance of his having 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 get captured in the video at all. Also, since gazes of both the eyes can be assumed to parallel, the gaze of the other eye 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 right and extreme right if the head is turned to left. We call this CD-like iris appearance phenomenon, and the idea is to monitor how frequently this

Fig. 4. An illustration differentiating what is CD-like iris position and what's not.

happens. This phenomenon is well illustrated in Fig. 4. The head of the person is turned to his right, and when the iris position in his EOI is his extreme left, we call it CD-like-iris position. A normal human also may have such an appearance occasionally, but a patient will have it quite frequently to look at the person or into the camera while conversing.

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 to neglect tolerable face gaze changes that occur while conversing. We use the classic solution of PnP (Perspective-n-Point) problem, i.e., an iterative method based on Levenberg-Marquardt algorithm [32] (implemented in OpenCV library) to compute FGA. After selecting such frames, we now extract our EOI based on FGA, i.e. for the right rotated face (positive FGA) the eye of interest (EOI) will be the left eye, and for the left rotated face (negative FGA) the eye of interest (EOI) will be the right one. We again extract eye-region of EOI as discussed in the previous section pass it through the blink detector (discussed in the previous section) to filter out frames of 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 this detector is also trained on the right eye. So, if the EOI is left one, it's flipped as done earlier. It's clear that as the feature value increases, there is more chance that the subject is a patient. However, it's difficult to find the optimal decision boundary between feature values for the normal case and the abnormal one. Therefore, we need to develop a model trained using machine learning algorithms to classify well any video into these two categories based on our feature. The entire end-to-end work-flow process is given in Fig.5.

C. Blink Normalness (especially for Blepharospasm)

In Blepharospasm, due to forceful eye contractions, when a patient blinks, it appears that he is squeezing his eyes as shown in Fig.6. Note the decrease in the eye-eyebrow distance (between landmarks 2 and 6) in fourth and fifth frames compared to others. For a normal person, such squeezing of eyes doesn't occur; consequently, there will be less variation in the distances. On the contrary, there will be 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, first, we detect longer blinks and compute the min-max ratio of this distance for such long blinks. Since a video may comprise of many blinks, we compute the average of these ratios as the blink-normalness feature. If the person is normal, it will be high; if the person is suffering from this disorder, it will be on the lower side. Therefore, if $B = \{B_1, B_2, ...\}$ is set of all the blinks in a video, $\min \delta(\cdot)$ and $\max \delta(\cdot)$denote minimum and maximum eye-eyebrow distances for a blink B_i and $\lambda(\cdot)$ denotes duration of a blink, the blink normalness feature can be computed as

$$BN = Avg\Big(\{\frac{\min \delta(B_i)}{\max \delta(B_i)} | \lambda(B_i) > 200ms\}\Big). \tag{7}$$

where the condition on λ is to check if the blink is long enough. Note that we do this process is carried out only on the right eye because if it happens it happens to both the eyes.

D. Intermediate Detection Models

We develop three intermediate detection models [Blink Detector, CD-like Iris Detector and Blink Detector (incl. tight)] in the process of designing the three features discussed above. For creating the blink detector, we build blink detection dataset (details in next section) comprising of open and closed eye images to train a convolutional neural network (CNN). For creating the CD-like iris detector, we build CD-like-iris detection dataset (details in next section) comprising of eye-images

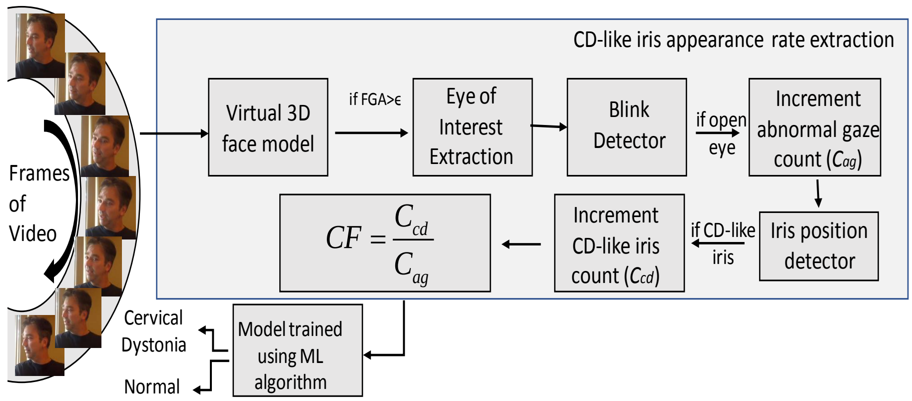

Fig. 5. Proposed approach for the extraction of CD-like iris appearance rate feature and the end-to-end workflow for detection of Cervical Distonia.

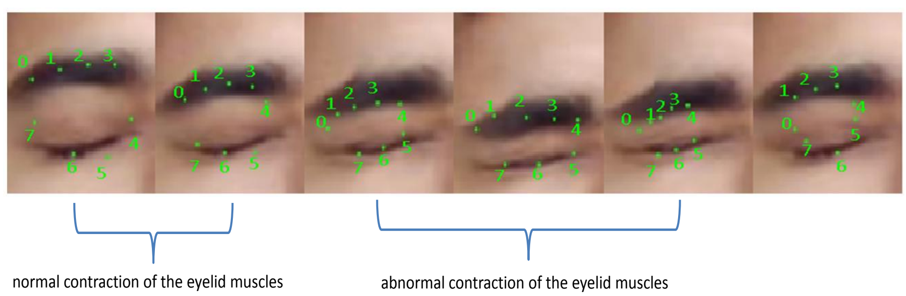

Fig. 6. An illustration differentiating normal and abnormal contraction of eyelids during a blink. Due to such abnormality, the distance between eyebrows (0-4) and eyes (5-7) considerably decreases compared to the normal condition.

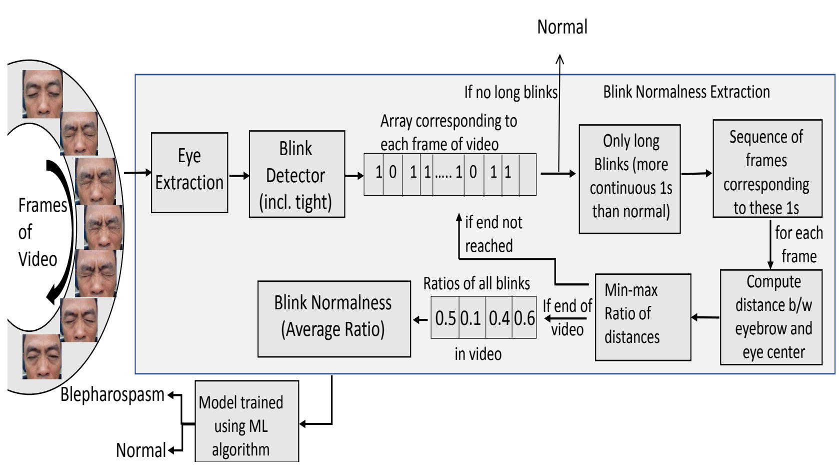

Fig. 7. Proposed approach for the extraction of blink normalness feature and the end-to-end workflow for the detection of Blepharospasm.

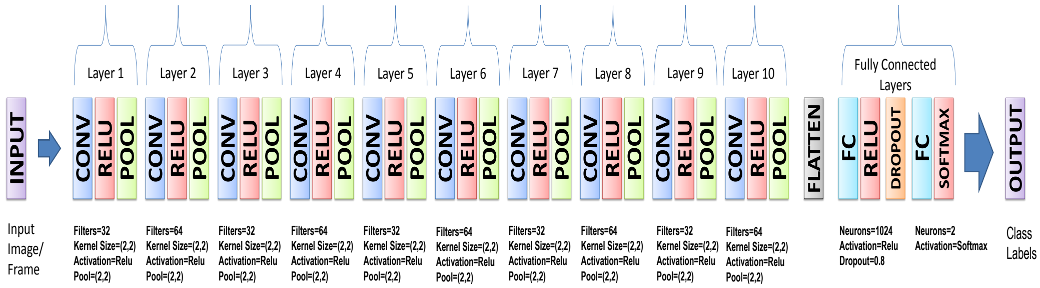

Fig. 8. CNN architecture used for developing our intermediate detectors.

TABLE I
COMPARISON OF OUR INTERMEDIATE DETECTORS WITH THOSE BUILT USING VGG AND INCEPTION-RESNET FEATURES

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%**

TABLE II
CONFUSION MATRICES IN THE TESTING PHASE WHILE BUILDING OUR INTERMEDIATE DETECTORS

Blink Detector		Predicted Open	Predicted Close
	Actual Open	3538	15
	Actual 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

having CD-like-iris and Non-CD-like-iris to train another convolutional neural network (CNN). And for creating Blink Detector (incl. tight), we build blink (incl. tight) detection dataset (details in next section) comprising of open and closed (incl. tightly closed) eye-images to train another convolutional neural network (CNN). We use the same architecture given in Fig. 8 for building all these three CNN-based detectors. The convolutional base of the architecture consists of 10 convolutional layers followed by ReLU (Rectified Linear Unit) and Max-pooling layers. The output of the convolutional base is flattened and fed to the final fully-connected layers to generate the final output.

E. Final Detection Models

Since features for videos have been generated, we now tune over several supervised learning algorithms to build the best final detection models in terms of 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 the models of 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 collective-final detector. We found that best such multi-class classifier was built using Support Vector Machine (SVM).

IV. EXPERIMENTAL RESULTS

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

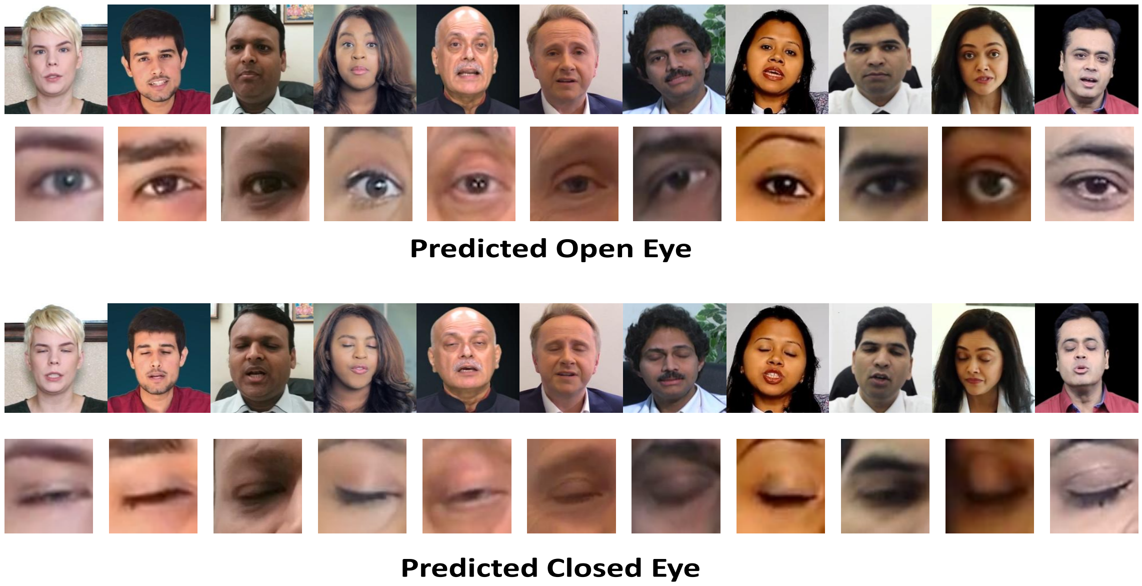

Fig. 9. Sample results of our blink detector

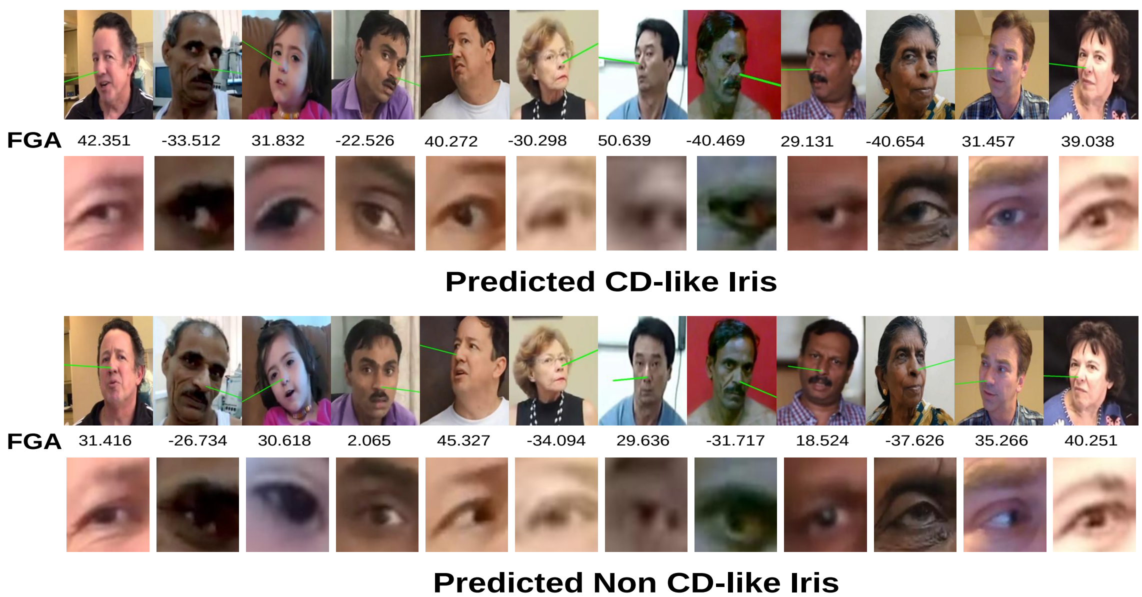

Fig. 10. Sample results of our CD-like-iris detector. Number below images indicate FGA (face gaze angle)

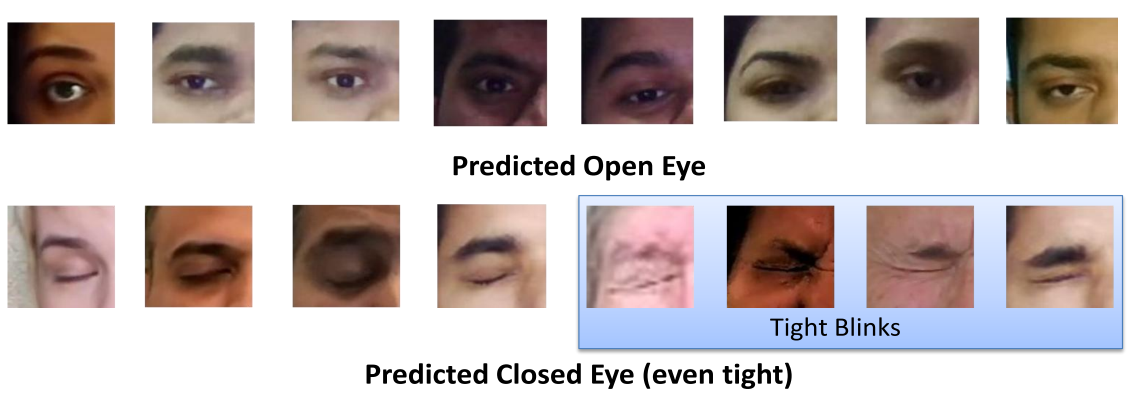

Fig. 11. Sample results of our blink detector (incl. tight)

TABLE III
CLASSIFICATION ACCURACIES OBTAINED DURING CROSS VALIDATION USING DIFFERENT MACHINE LEARNING ALGORITHMS WHILE BUILDING THE INDIVIDUAL FINAL DETECTORS.

	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	

TABLE IV
CLASSIFICATION ACCURACIES OBTAINED DURING CROSS-VALIDATION AND TRAINING OF OUR INDIVIDUAL FINAL DETECTORS.

	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

TABLE V
CONFUSION MATRICES OBTAINED DURING CROSS-VALIDATION OF OUR INDIVIDUAL FINAL DETECTORS.

		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

A. Datasets

We develop two types datasets: one type for developing intermediate detection models [i.e. blink detector, CD-like-iris detector and blink detector (incl. tight) detector] and 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 video as input. While the output of the intermediate detectors is if events like blink and CD-like-iris have occurred in a particular frame, the output of the final detectors is if the patient has a neurological disorder.

The details of the datasets we created for developing intermediate detectors are as follows. Note that these consist of all right-type eye images; that is, left ones are flipped. **Blink Detection Dataset**: It consists of 21738 open eye images and 17704 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 where iris is either near the center or left side of the eye and 1261 eye images in which iris was in the extreme right side of the eye. 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 3019 open eye images and 5220 closed eye images, some of which are tightly closed. The purpose again is to detect if a given eye is closed, including tightly closed.

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 normal persons. **Cervical Dystonia Video Dataset**: It consists of 14 videos of patients suffering from Cervical Dystonia, and 36 videos are of normal persons. **Blepharospasm Video Dataset**: It consists of 21 videos of patients suffering from Blepharospasm, and 30 videos are of normal persons. **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 Blephrospasm), and 30 videos are of normal persons. The purpose of this dataset is to predict the disorders collectively.

B. 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 approaches for evaluation has been taken.

For intermediate detectors, since we have labeled data in abundance, we can afford to divide the labeled data into three parts: for training, for 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. With the help of validation scores, we manage to set best hyperparameters while training the CNNs for building these intermediate detectors. Once they are set, the trained CNN is tested on the training dataset.

In contrast, for the final detector, since we have very limited labeled data, we avoid the testing phase altogether and just 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 of each of the parts and take 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 9 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 the best learning algorithm is chosen, we train on the entire labeled data to create our final detectors. Note that we use the default setting of hyper-parameters for the algorithms available in the Orange package of Python.

In addition to the classification accuracy, we also report appropriate confusion matrices while reporting the performances of our detectors.

TABLE VI
COMPARISON OF PROPOSED FEATURES WITH THAT OF 3D-CNN [33] 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.

	AB	LR	NB	NN	RF	SGD	SVM	DT	kNN	Average
3DCNN(avg)	**77.4**	**83.0**	64.2	74.5	80.2	76.4	67.0	**77.4**	72.6	74.7
3DCNN(var)	70.8	70.8	70.8	65.1	71.7	65.1	67.0	68.9	66.0	68.5
3DCNN(avg+var)	**77.4**	**83.0**	63.2	75.5	73.6	74.5	70.8	72.6	75.5	74.0
3DCNN(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 VII
CONFUSION MATRIX GENERATED WHILE PERFORMING CROSS-VALIDATION TO BUILD OUR COLLECTIVE FINAL DETECTOR.

	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

C. Intermediate Detectors Results

The quantitative results of our intermediate detectors in terms of classification accuracy are given is Table I 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 results compared to those detectors in the challenging validation and testing phases. The datasets used here are split into three parts: 70% for the training, 15% for validation, and 15% for testing. The confusion matrices obtained finally during testing phase using our intermediate detectors are given in Table II. The sample detection results of our intermediate detectors can be seen in Figures 9-11: the blink detector is able to detect if an eye is open or closed; the CD-like-iris detector is able to detect if the eye of interest has cd-like-iris or not; and the blink detector (incl. tight) is able to detect if an eye is open or closed, including the ones closed tightly.

D. Individual Final Detectors Results

The individual-final detectors are built using only the dedicated features for different disorders to predict the presence of that particular disorder. The quantitative results of these individual-final detectors in terms of classification accuracy during cross-validation are given in Table III. It's clear from the table that SGD, SVM, and LR are best algorithms for the Bell's Palsy, Cervical Dystonia, and Blephrospasm detectors, respectively, using our features. Overall SVM appears to be the best. Moreover, the high average accuracies across the algorithms show that our features are robust. We summarize our individual-final detectors in terms of the algorithm used and different accuracies obtained in Table IV. 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 confusion matrices obtained while performing the cross-validation in Table V.

E. Collective Final Detector Results

The collective-final detectors are built using all the three features to predict the presence of any of the three disorders. The quantitative results such a detector in terms of classification accuracy during cross-validation are given in Table VI. We also compare our features with that of 3D-CNN. First, the 3D-CNN features [33] are extracted 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. Also, they are added (+) and concatenated ($\oplus$) to generate more features per video for comparison. It can be seen that our features obtain best results in 6 out of 9 algorithms. It's clear from the table that best performance of 83.0% is given by SVM; therefore, we use it to build our collective final detector on the entire labeled to produce 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 VII.

V. CONCLUSION

We attempted eye-based detection of the three neurological disorders: Bell's Palsy, Cervical Dystonia, and Blepharospasm. From the subjects' videos, we extract novel eye-based features and develop the final models to detect these disorders both individually and collectively. Specifically, we design three features: 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 7 datasets (3 for intermediate detection + 4 for final detection) developed by us demonstrate competitive performance. While the proposed approach can detect the disorders with 90% accuracy individually, it can detect them collectively with 83% accuracy.

References

[1] J. Shi, I. Alikhani, X. Li, Z. Yu, T. Seppänen, and G. Zhao, "Atrial fibrillation detection from face videos by fusing subtle variations," *IEEE Transactions on Circuits and Systems for Video Technology*, 2019.

[2] X. Zhang, Z. Zhu, Y. Zhao, and D. Chang, "Learning a general assignment model for video analytics," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 28, no. 10, pp. 3066–3076, 2017.

[3] H.-C. Shih, "A survey of content-aware video analysis for sports," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 28, no. 5, pp. 1212–1231, 2017.

[4] Z. Luo, P.-M. Jodoin, S.-Z. Su, S.-Z. Li, and H. Larochelle, "Traffic analytics with low-frame-rate videos," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 28, no. 4, pp. 878–891, 2016.

[5] Z. Shao, L. Wang, Z. Wang, W. Du, and W. Wu, "Saliency-aware convolution neural network for ship detection in surveillance video," *IEEE Transactions on Circuits and Systems for Video Technology*, 2019.

[6] A. Vedaldi and B. Fulkerson, "Vlfeat: An open and portable library of computer vision algorithms," in *Proceedings of the 18th ACM international conference on Multimedia*. ACM, 2010, pp. 1469–1472.

[7] T. B. Moeslund and E. Granum, "A survey of computer vision-based human motion capture," *Computer vision and image understanding*, vol. 81, no. 3, pp. 231–268, 2001.

[8] E. Murphy-Chutorian and M. M. Trivedi, "Head pose estimation in computer vision: A survey," *IEEE transactions on pattern analysis and machine intelligence*, vol. 31, no. 4, pp. 607–626, 2008.

[9] K. R. Jerripothula, J. Cai, and J. Yuan, "Efficient video object co-localization with co-saliency activated tracklets," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 29, no. 3, pp. 744–755, March 2019.

[10] A. Pereira, P. Carvalho, G. Coelho, and L. Côrte-Real, "Efficient ciede2000-based color similarity decision for computer vision," *IEEE Transactions on Circuits and Systems for Video Technology*, 2019.

[11] K. R. Jerripothula, J. Cai, J. Lu, and J. Yuan, "Object co-skeletonization with co-segmentation," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2017, pp. 3881–3889.

[12] D. Beymer, P. McLauchlan, B. Coifman, and J. Malik, "A real-time computer vision system for measuring traffic parameters," in *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. IEEE, 1997, pp. 495–501.

[13] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg *et al.*, "Scikit-learn: Machine learning in python," *Journal of machine learning research*, vol. 12, no. Oct, pp. 2825–2830, 2011.

[14] T. G. Dietterich, "Ensemble methods in machine learning," in *International workshop on multiple classifier systems*. Springer, 2000, pp. 1–15.

[15] T. Mitchell, B. Buchanan, G. DeJong, T. Dietterich, P. Rosenbloom, and A. Waibel, "Machine learning," *Annual review of computer science*, vol. 4, no. 1, pp. 417–433, 1990.

[16] W. Kuang, Y.-L. Chan, S.-H. Tsang, and W.-C. Siu, "Machine learning based fast intra mode decision for hevc screen content coding via decision trees," *IEEE Transactions on Circuits and Systems for Video Technology*, 2019.

[17] F. Sullivan, F. Daly, and I. Gagyor, "Antiviral agents added to corticosteroids for early treatment of adults with acute idiopathic facial nerve paralysis (bell palsy)," *Jama*, vol. 316, no. 8, pp. 874–875, 2016.

[18] D. B. Pitts, K. K. Adour, and R. L. Hilsinger Jr, "Recurrent bell's palsy: analysis of 140 patients," *The Laryngoscope*, vol. 98, no. 5, pp. 535–540, 1988.

[19] B. Leplow, A. Eggebrecht, and J. Pohl, "Treatment satisfaction with botulinum toxin: a comparison between blepharospasm and cervical dystonia," *Patient preference and adherence*, vol. 11, p. 1555, 2017.

[20] S. Kongsaengdao, N. Maneeton, and B. Maneeton, "Long-term quality of life in cervical dystonia after treatment with abobotulinum toxin a: a 2-year prospective study," *Neuropsychiatric disease and treatment*, vol. 14, p. 1119, 2018.

[21] T. J. Broadbent, R. E. Wesley, and L. A. Mawn, "A survey of current blepharospasm treatment patterns among oculoplastic surgeons," *Ophthalmic Plastic & Reconstructive Surgery*, vol. 32, no. 1, pp. 24–27, 2016.

[22] G. Defazio, M. Hallett, H. A. Jinnah, A. Conte, and A. Berardelli, "Blepharospasm 40 years later," *Movement Disorders*, vol. 32, no. 4, pp. 498–509, 2017.

[23] M. Park, J. Seo, and K. Park, "Pc-based asymmetry analyzer for facial palsy study in uncontrolled environment: A preliminary study," *Computer methods and programs in biomedicine*, vol. 99, no. 1, pp. 57–65, 2010.

[24] T. Wang, J. Dong, X. Sun, S. Zhang, and S. Wang, "Automatic recognition of facial movement for paralyzed face," *Bio-medical materials and engineering*, vol. 24, no. 6, pp. 2751–2760, 2014.

[25] S. He, J. J. Soraghan, B. F. O'Reilly, and D. Xing, "Quantitative analysis of facial paralysis using local binary patterns in biomedical videos," *IEEE Transactions on Biomedical Engineering*, vol. 56, no. 7, pp. 1864–1870, 2009.

[26] T. H. Ngo, M. Seo, N. Matsushiro, W. Xiong, and Y.-W. Chen, "Quantitative analysis of facial paralysis based on limited-orientation modified circular gabor filters," in *2016 23rd International Conference on Pattern Recognition (ICPR)*. IEEE, 2016, pp. 349–354.

[27] H. Kim, S. Kim, Y. Kim, and K. Park, "A smartphone-based automatic diagnosis system for facial nerve palsy," *Sensors*, vol. 15, no. 10, pp. 26 756–26 768, 2015.

[28] A. Asthana, S. Zafeiriou, S. Cheng, and M. Pantic, "Incremental face alignment in the wild," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 1859–1866.

[29] G.-S. J. Hsu, J.-H. Kang, and W.-F. Huang, "Deep hierarchical network with line segment learning for quantitative analysis of facial palsy," *IEEE Access*, vol. 7, pp. 4833–4842, 2018.

[30] E. S. Molho, M. Stacy, P. Gillard, D. Charles, C. H. Adler, J. Jankovic, M. Schwartz, and M. F. Brin, "Impact of cervical dystonia on work productivity: an analysis from a patient registry," *Movement disorders clinical practice*, vol. 3, no. 2, pp. 130–138, 2016.

[31] J. R. Bedarf, S. Kebir, J. P. Michelis, B. Wabbels, and S. Paus, "Depression in blepharospasm: a question of facial feedback?" *Neuropsychiatric disease and treatment*, vol. 13, p. 1861, 2017.

[32] K. Levenberg, "A method for the solution of certain non-linear problems in least squares," *Quarterly of applied mathematics*, vol. 2, no. 2, pp. 164–168, 1944.

[33] D. Tran, L. Bourdev, R. Fergus, L. Torresani, and M. Paluri, "Learning spatiotemporal features with 3d convolutional networks," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 4489–4497.