Computers in Biology and Medicine 112 (2019) 103376

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

A neural network-based software to recognise blepharospasm symptoms and to measure eye closure time

Gianpaolo F. Trotta[a], Roberta Pellicciari[b], Antonio Boccaccio[a,*], Antonio Brunetti[c], Giacomo D. Cascarano[c], Vito M. Manghisi[a], Michele Fiorentino[a], Antonio E. Uva[a], Giovanni Defazio[d], Vitoantonio Bevilacqua[c]

[a] Dipartimento di Meccanica, Matematica e Management, Politecnico di Bari, Bari, Italy
[b] Dipartimento di Scienze Mediche di Base, Neuroscienze ed Organi di Senso, Università degli Studi di Bari, Bari, Italy
[c] Dipartimento di Ingegneria Elettrica e dell'Informazione, Politecnico di Bari, Bari, Italy
[d] Dipartimento di Scienze Mediche e Sanità Pubblica, Università degli Studi di Cagliari, Cagliari, Italy

ARTICLE INFO

Keywords:
Blepharospasm
Neural network topology optimisation
Blepharospasm rating scale
Eye closure time
Focal dystonia

ABSTRACT

Blepharospasm (BSP) is an adult-onset focal dystonia with phenomenologically heterogeneous effects, including, but not limited to, blinks, brief or prolonged spasms, and a narrowing or closure of the eyelids. In spite of the clear and well-known symptomatology, objectively rating the severity of this dystonia is a rather complex task since BSP symptoms are so subtle and hardly perceptible that even expert neurologists can rate the gravity of the pathology differently in the same patients. Software tools have been developed to help clinicians in the rating procedure. Currently, a computerised video-based system is available that is capable of objectively determining the eye closure time, however, it cannot distinguish the typical symptoms of the pathology. In this study, we attempt to take a step forward by proposing a neural network-based software able not only to measure the eye closure, time but also to recognise and count the typical blepharospasm symptoms. The software, after detecting the state of the eyes (open or closed), the movement of specific facial landmarks, and properly implementing artificial neural networks with an optimised topology, can recognise blinking, and brief and prolonged spasms. Comparing the software predictions with the observations of an expert neurologist allowed assessment of the sensitivity and specificity of the proposed software. The levels of sensitivity were high for recognising brief and prolonged spasms but were lower in the case of blinks. The proposed software is an automatic tool capable of making objective 'measurements' of blepharospasm symptoms.

1. Introduction

Idiopathic blepharospasm (BSP) is an adult-onset focal dystonia that is commonly characterised by bilateral, synchronous, and symmetric dystonic spasms in the Orbicularis Oculi (OO) muscle [1–4]. Dystonic spasms can be phenomenologically heterogeneous, with either brief or prolonged spasms and a narrowing or closure of the eyelids [5]. In addition to spasms, patients affected by BSP might present a spectrum of additional signs/symptoms, including sensory symptoms in the eyes that indicate ocular diseases (e.g. dry eye syndrome) [6], an increased spontaneous blink rate [7], the presence of sensory tricks (stretching, massaging, or touching the eyebrow, the eyelid, or the forehead) to transiently improve eyelid spasms, apraxia of eyelid opening [8], and dystonia in other body parts. Clinical evaluation of BSP severity poses a number of challenges and several drawbacks might limit the widespread use of most of the existing severity scales. Since involuntary eye closure is the most disabling BSP feature, it would be reasonable to assume that objective measures of eye closure might be a good measure of BSP severity. Recently, algorithms have been developed that, based on the analysis of standard video recordings, measure the percentage of time the patients' eyes were closed while they were instructed to keep them open [9]. However, such a methodology allows only the total time of eye closure to be measured, without identifying the events contributing to the episodes of eye closure, that is, blinking, brief spasm with complete eye closure, and prolonged spasms with complete eye closure. The relevance of these phenomenological aspects in the evaluation of BSP severity is supported by recent evidence indicating that the type of spasm might identify BSP subtypes that are characterised by varying severity and the tendency of dystonia to spread to adjacent anatomic regions [1].

* Corresponding author. Dipartimento di Meccanica, Matematica e Management, Politecnico di Bari, Viale Japigia, 182 70126, Bari, Italy.
E-mail addresses: antonio.boccaccio@poliba.it, a.boccaccio@poliba.it (A. Boccaccio).

https://doi.org/10.1016/j.compbiomed.2019.103376
Received 11 April 2019; Received in revised form 30 July 2019; Accepted 30 July 2019

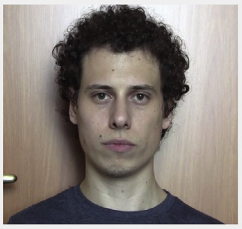
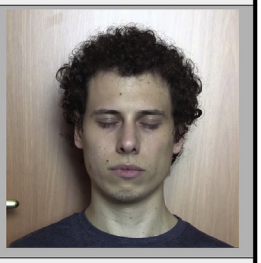
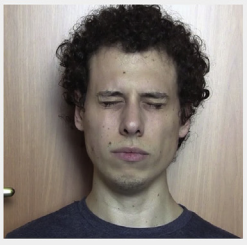
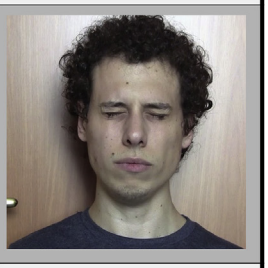
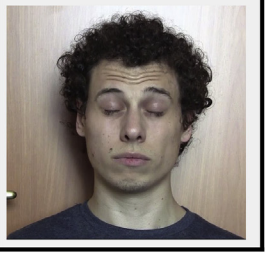
Symptom	Duration	Features	
Open eyes	-	-	
Blink	< 1s	transient eyelid drop without any lowering of the eyebrow	
Brief spasm	0.3 to 3 s	lowering of the eyebrow and narrowing/closure of the eyelid rim	
Prolonged spasm	> 3 s	lowering of the eyebrow and narrowing/closure of the eyelid rim	
apraxia of eyelid opening		raising of the eyebrow above the superior orbital margin	

Fig. 1. Typical symptoms observed in patients with blepharospasm.

Peterson et al. [9] utilised the Computer Expression Recognition Toolbox (CERT) to compare clinical rating scales of blepharospasm severity with involuntary eye closures. The software performs a frame-by-frame video analysis to evaluate the eye closure time and compares the results obtained implementing three commonly used clinical rating scales: the Burke-Fahn-Marsden Dystonia Rating Scale [10], the Global Dystonia Rating Scale [11] and the Jankovic Rating Scale (JRS) [12]. The results demonstrated that CERT has convergent validity with conventional clinical rating scales and, hence, can be used with video recordings to measure blepharospasm symptom severity automatically and objectively. However, the previous severity scales are limited by a number of potential drawbacks. In particular, they comprehensively measure dystonia severity in all body parts, and severity grading is based on the intensity of dystonic contractions merged with [10] or weighted [11] by duration and daily frequency of the spasms. In other words, the main problem with these scales is that they adopt the same grading modality for dystonia at different body sites, each of which ideally warrants a specific severity assessment system. To date, the only severity scale specifically developed to assess BSP severity is the Jankovic Rating Scale [12]. In fact, this scale includes two subscales that measure the intensity and frequency of eyelid spasms, both based on a 5-point grading system. However, this severity scale has not yet been validated in terms of reproducibility. In a study by some authors of the present paper [5], a novel scale for rating BSP severity—the Blepharospasm Severity Rating Scale (BSRS)—was developed and validated, which allows all the above-mentioned limitations to be overcome.

Since facial movements/expressions in patients with blepharospasm are subtle and hardly perceptible, the objectivation of blepharospasm

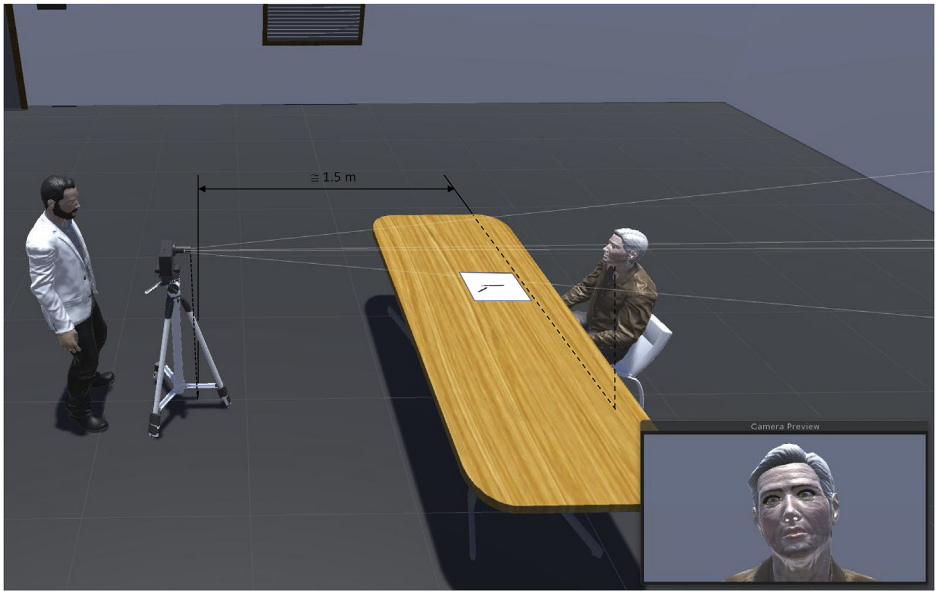


Fig. 2. Set-up utilised to acquire the facial expressions of the patient during the clinical test.

severity is such a difficult task that even expert neurologists can rate the severity of the pathology in the same patients differently. To address these issues, software tools have been recently developed [9]. However, in order to properly characterise BSP severity, facial recognition systems must be utilised that not only measure the eye closure time, but are also able to recognise the facial expressions/movements typically related to the events contributing to the episodes of eye closure.

In a previous study [13], a frame-by-frame video analysis was performed to measure the geometry of specific facial features [14,15]. The distances between reference points were measured, and investigations of how these distances changed in BSP-related facial expressions were carried out. Threshold values were then fixed for these distances to distinguish the pathological from the non-pathological facial movements [13]. However, the approach adopted showed important limitations. (i) The strategy for fixing the threshold values for the distance of facial landmarks cannot be generalised to all the patients; in other words, each patient requires the determination of specific threshold values for the distinction of pathological from non-pathologic conditions. (ii) In the clinical context, where the environmental conditions might change over time and from place to place, it is very difficult to standardise the settings of the acquisition system (i.e. the CCD camera with which patients are video-recorded), which is a requirement of crucial importance for adopting the strategy based on threshold values.

In this study, we propose a software tool capable of overcoming the main limitations of the previous approach, implementing Artificial Neural Networks (ANN) [16,17] with a topology optimised via Genetic Algorithms. The software, based on standard video-recordings from commonly available video cameras, is capable not only of measuring the percent time of eye closure, but also of recognising blinking, and brief and prolonged spasms, which are the typical facial movements that take place in patients with blepharospasm. The proposed software is a practical system very suited to the clinical context where the environmental conditions cannot be easily standardised; it is a promising tool for supporting/assisting physicians to rate blepharospasm severity according to the BSRS scale [5].

2. Material and methods

2.1. The protocol adopted to evaluate blepharospasm severity

Nine patients with BSP were recruited (3 women and 6 men, average age 69.55 ± 8.94 years) to be recorded with a digital video camera (Canon, Tokyo (Japan), Legria HFM306, 3.3 MP Full HD CMOS, HD Video Lens (up to 18× zoom), DIGIC DV III) at 29.97 frames per second. An experienced neurologist, reviewing the video-recordings, identified dystonic spasms and blinks and evaluated the overall Severity Index (*SIn*) of the recruited patients by applying the BSRS scale. The neurologist classified the BSP symptoms observed in the recruited patients as follows (Fig. 1).

1) A sudden Orbicularis Oculi (OO) muscle contraction causing lowering of the eyebrow and narrowing/closure of the eyelid rim was classified by the neurologist as a spasm. In turn, the spasm was classified as:
   1a) *brief spasm*, i.e. a spasm inducing a brief eyelid closure lasting 0.3–3 s;
   1b) *prolonged spasm*, i.e. a spasm inducing a prolonged eyelid closure with a duration of more than 3 s.
2) A bilateral, synchronous, short duration (< 1 s) OO muscle contraction causing a transient eyelid drop without any lowering of the eyebrow was classified by the clinician as a *blink*.
3) A delay in reopening the eyelids after involuntary closure associated with no overt OO contraction or raising of the eyebrow above the superior orbital margin was classified as *apraxia of eyelid opening* (Fig. 1).

The following clinical test was performed to determine the severity index *SIn*, according to the BSRS scale, of the nine recruited patients. The participants were seated on a chair placed in front of the video camera with their feet resting on the floor and their hands on their knees (Fig. 2). The camera objective was zoomed in so that the resulting field of view was entirely occupied by the patient's head and her/his shoulders. The video-recordings had a total duration of approximately 5 min and were performed according to the protocol described in a previous study [5]. In the first three minutes, a training phase for the clinician took place; the neurologist asked the patient to perform some tasks and made careful observations, thus acquiring a deep knowledge of the modalities with which the BSP symptoms took place in the patient. During the last two minutes of the test, the patient was asked to remain at rest with their eyes open and fixed on a specific point located in front of her/him. During this time interval, the clinician recognised and counted blinks, brief and prolonged spasms, and apraxia of the eyelid opening.

In detail, the adopted clinical test is articulated in the following

steps.

(i) In the first 10 s, the patient is at rest with their eyes open;
(ii) The patient is asked to voluntarily and forcefully close and open their eyes five times, approximately one cycle per second (in reality, the time necessary to perform the requested task very often depends on the patient and the severity of the dystonia);
(iii) Again, for 10 s, the patient is at rest with their eyes open;
(iv) The patient is asked to voluntarily and gently close and open their eyes five times, one cycle per second;
(v) Again, for another 10 s, the patient is at rest with their eyes open;
(vi) The neurologist poses the following questions: Are you able to avoid closing your eyes? How? With sheer will? Or, do you need to touch your eyes, face, or neck?
(vii) The patient has 50 s to reply to the questions posed;
(viii) The patient is asked to write on a sheet of paper a stereotyped sentence (e.g., "Today is a nice sunny day") three times;
(ix) The patient remains at rest for at least 150 s, with their eyes open and fixed on a specific point located in front of her/him. In the last 120 s, the neurologist "manually" counts the number of blinks, brief and prolonged spasms, and apraxia of eyelid opening. Patients were instructed to avoid antagonistic gestures (i.e. manuevers voluntarily adopted by the patient to minimise the spasm effects, such as touching the eyelid, the temple, etc.) so as not to falsify the evaluation of the proposed software.

The (approximately) first three minutes of the test included steps (i) to (viii), whereas during the last (approximately) two minutes, only the step (ix) was performed. It is worth noting that the tasks performed in steps (ii) and (iv) are of crucial importance for the neurologist. In fact, a voluntary and forceful closure of the eyes can provide useful information about how a spasm occurs in the patient. Similarly, a voluntary and gentle closure of the eyes is a type of simulation of a blink and, therefore, can instruct the clinician on how this event can take place in the patient.

2.2. The software tool: the face detector and face pose estimator

A software tool was developed, based on the *dlib* library [18] and implementing the face detector (A) and the face pose estimator (B) algorithms, to measure the duration of the time interval during which the patient's eyes were closed and to automatically recognise three of the four BSP symptoms described above, namely: 1a) blinks, 1b) brief, and 2) prolonged spasms; the last symptom, the apraxia of eyelid opening, was neglected in this study phase. The schematic in Fig. 3 briefly summarises the principal steps followed to develop and validate the software.

(A) Face detector algorithm. For each of the acquired frames (during the clinical test, Fig. 3, Blocks [1,2], and [3]), the face detector algorithm, which is based on the traditional Histogram of Oriented Gradients (HOG) feature combined with a linear classifier, and a sliding window detection scheme, identifies the bounding box in which the patient's head can be inscribed (Fig. 3, Block [4]). Preliminary analyses revealed that the face detector was significantly robust with respect to variable head movements and variable lighting conditions during the video recording. For each of the nine recruited patients, the percentage of frames ϵ_{FD} in which the face detector algorithm was capable of reliably identifying the face of the patient was computed with respect to the total number of frames acquired for the same patient. The average value of ϵ_{FD} computed over all the patients was 99.960%, whereas the lowest result was 99.948%. Table 1 lists, for each of the nine patients, the number of frames acquired during the last (approximately) two minutes of the clinical test, (i.e. during the time interval in which the patient remained at rest with their eyes open and fixed on a point) and the corresponding value of ϵ_{FD}.

(B) Face pose estimator algorithm. The face pose estimator detects the position of 68 facial landmarks distributed at different points on the patient's face, such as the edges of the mouth, the eyes, the eyebrows, the nose, etc. (Fig. 3, Block [5]; Fig. 4 (a)). In detail, the pose estimator algorithm first identifies the position of specific face points that allow the definition of the principal facial features, then predicts the location of the 68 facial landmarks in real-time [19]. The pose estimator was created by using dlib's implementation of the study by Kazemi and Sullivan [19] and trained on the iBUG 300-W face landmark dataset [20].

The two above described algorithms were implemented to process the video frames acquired from all the patients. For each of the analysed frames, first the face of the patient was detected via the face detector algorithm, then the location of the 68 facial landmarks using the face pose estimator was detected. Preliminary analyses revealed that, after a training phase, the face pose estimator algorithm was always capable of identifying the location of all the 68 landmarks in all the videos recorded. However, we observed that due to head movements and variable lighting conditions, the face pose estimator sometimes could not correctly predict the location of some facial landmarks (Fig. 4(b)). A correction tool, T_{corr}, was hence developed that allows the re-training of the pose estimator utilising the facial landmarks correctly re-located (Fig. 3, Block [6]). The core of T_{corr} is *imglab*: a dlib simple graphical tool for annotating images with object bounding boxes and optionally their part locations. T_{corr} requires the clinician to choose a random number of frames—at least thirty—recorded during the first three minutes of the clinical test, showing the patient with their eyes in several states (eyes closed during a spasm, eyes closed during a blink, and eyes open). For each frame, the clinician can drag and drop, in a more correct position, all the facial landmarks not correctly positioned. Implementing T_{corr} led to a better face pose estimator algorithm re-training, thus making it capable of predicting the location of the 68 facial landmarks with higher accuracy (Fig. 3, Block [7]; Fig. 4(b), (c), and (d)). To prove the higher accuracy obtained due to the implementation of the correction tool T_{corr}, further analyses were carried out that are described in Appendix A.

2.3. The software tool: extraction of features

To detect the BSP symptoms, an algorithm was developed to crop the Region Of Interest (ROI) around the eyes of each patient. To do this, specific facial landmarks predicted by the facial pose estimator algorithm were utilised to define the rectangular bounding box delimiting each eye (Fig. 3, Block [8]). Concerning the right eye, points 36 and 39 were utilised to determine the horizontal dimension of the rectangle (Fig. 5). In detail, the horizontal dimension of the ROI goes from point 36′, placed 30 pixels to the left of point 36, to point 39′, placed 30 pixels at the right of point 39 (Fig. 5(a)). The vertical dimension of the rectangle was defined by the points 38 and 41. Again, 30 pixels were added above and below points 38 and 41, respectively, thus identifying 38′ and 41'. The same procedure as that adopted for the right eye was also utilised for the left eye (Fig. 5(b)). The points utilised to define the horizontal and the vertical dimensions of the ROI are, in this case: 42 and 45, and 44 and 46, respectively, to which a 30-pixel padding margin was added, as for the case of the right eye. For each video frame, two ROIs were extracted, one for each eye, resized to 64 × 32 pixels and saved in gray scale. In general, the dimensions of each extracted ROI depends on the eye size and, therefore, is patient-dependent. The resizing procedure enables all the ROIs to have the same dimensions, which is an essential requirement for the successive computation procedures described below.

The algorithm to crop the ROI frame by frame was implemented on the video registered from all the patients during the first three minutes of the clinical test, i.e., during the time interval in which the clinician asks the patient to perform some tasks and deeply observes her/his "response". Two folders were then created: one to gather the ROIs with open eyes, the other, the ROIs with closed eyes (Fig. 3, Block [9]; Fig. 5

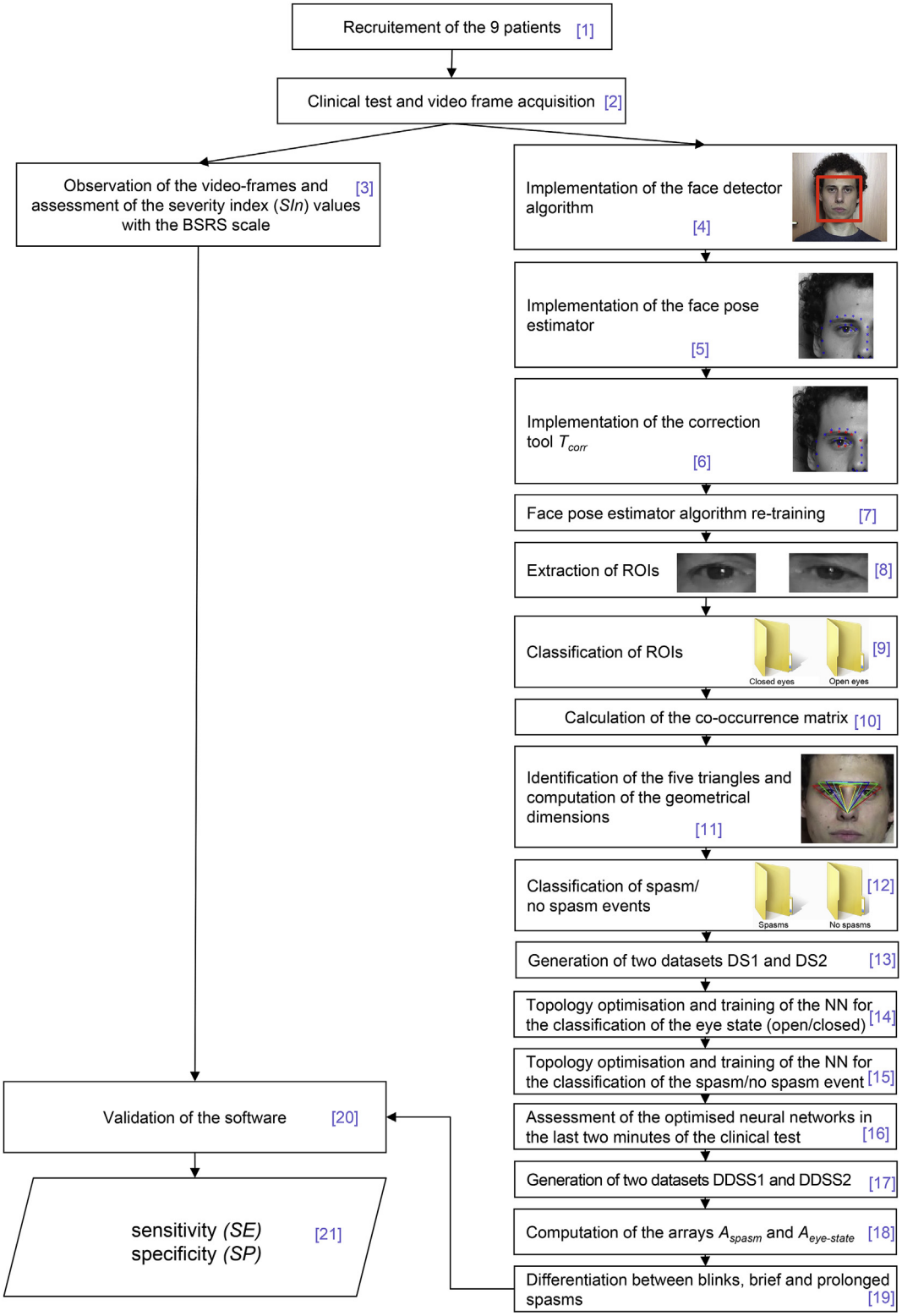


Fig. 3. Schematic of the steps followed to develop and validate the proposed software.

(c) and (d)). Then, the same neurologist that assessed the severity index *SIn* (according to the BSRS scale) of the nine recruited patients manually classified/labelled the obtained ROIs and moved each one into the correct folder. It is worth noting that, although this classification must be performed for all the ROIs extracted, the described procedure is rather easy to accomplish and requires relatively little time. In the first three minutes of the clinical test, the patient gently or forcefully closes their eyes; most of the frames (and hence ROIs) with open (or closed) eyes are close in time and, therefore, it is rather easy for the neurologist to gather frames with the same eye state. For all of the ROIs extracted and labelled, the co-occurrence matrices of oriented gradients were computed for the classification of the eye state (Fig. 3, Block [10]). Indeed, the typical descriptive feature implemented in the computer vision for the eye state classification is represented by the histogram of

Table 1
Values obtained for ϵ_{FD}, the percentage of video frames acquired during the (approximately) last two minutes of the clinical test when the face of the patient was detected by the face detector algorithm.

Patient	Total number of analysed video-frames	Number of video-frames in which the patient's head was detected	ϵ_{FD} [%]
P1	3897	3897	100.000
P2	3927	3925	99.949
P3	3927	3925	99.949
P4	3837	3835	99.948
P5	3867	3865	99.948
P6	3897	3895	99.949
P7	3927	3926	99.974
P8	3867	3865	99.948
P9	3897	3896	99.974
Average			99.960

oriented gradients (HOG), which is a useful and commonly utilised tool suffering from the limitation of local gradient information. The co-occurrence matrix of oriented gradients has been proven to enhance the capability to describe the global gradient information of eye images, thus allowing classification of the eye state with a higher accuracy than the classical HOG [21]. This matrix is a 4D array with dimensions *levels-of-gray* × *levels-of-gray* × *number-of-distances* × *number-of-angles*. The value that the co-occurrence matrix assumes, for example at the co-ordinates *xx*, *yy*, *zz,* and *ww,* is the number of times the grey level *yy* is present at the distance *zz* and at the angle *ww* starting from the grey level *xx*. Following Zhang et al. [21], we fixed the number of grey levels equal to 8, the distance equal to 1 pixel (hence, the number of distances is 1), and the angle equal to 0 radians (therefore, the number of angles is 1, further details on the computation of the co-occurrence matrix can be found in Zhang et al. [21]). Therefore, the dimensions of the computed co-occurrence matrices, which represent the number of features that will be given in input to the artificial neural network described below, are: 8 × 8 × 1 × 1 for a total number of features $n_{Feye\text{-}state}$ equal to $n_{Feye\text{-}state} = 64$.

In order to detect the eyebrow movements related to the spasm events, an algorithm was developed that measures the height of the five triangles with one of the vertices on the tip of the nose (i.e. at point 30) and the others defined by the pairs of facial landmarks symmetric with respect to the sagittal plane and located on the eyebrows (Fig. 3, Block [11]; Fig. 5 (e) and (f)). In detail, to avoid sudden changes in the height of the triangle due to possible rotations of the patient's head, the height of the triangles was normalised with respect to that of the nose, which is the distance between points 27 and 30 (Fig. 5(f)). For each acquired frame *k*, the normalised height *Y*(*k*,*j*) of the *j*th triangle (*j* = 1, 2, ..., 5), the average normalised height value $\overline{Y}(k)$, and the standard deviation $\sigma(Y(k))$, which is a measure used to quantify the amount of variation or dispersion of a set of data values, were computed as:

$$\begin{cases} Y(k,j) = \frac{T(k,j)}{L(k)} \qquad j = 1,\ 2,\ ...,\ 5 \\ \bar{Y}(k) = \frac{1}{5}\sum_{j=1}^{5} Y(k,j) \\ \sigma(Y(k)) = \sqrt{\frac{\sum_{j=1}^{5}[Y(k,j) - \bar{Y}(k)]^2}{5}} \end{cases} \tag{1}$$

where, *T*(*k*,*j*) is the height of the *j*th triangle and *L*(*k*) is the distance between points 27 and 30. Figs. 6 and 7 show the average normalised height of triangles $\bar{Y}(k)$ typically registered during a blink and a spasm, respectively. It is interesting to note how in the case of blinking (Fig. 6), the average height of the triangles remained practically constant, whereas a large decrease in $\bar{Y}(k)$ occurred during a spasm (Fig. 7).

Therefore, the number of features, n_{Fspasm}, given in input to the artificial network described below is $n_{Fspasm} = 7$, i.e. the five normalised heights *Y*(*k*,*j*), the average normalised height $\bar{Y}(k)$, and the standard deviation $\sigma(Y(k))$ (Eq. (1)). In addition, in this case, two folders were created, 'Spasms' and 'No spasms' (Fig. 3, Block [12]). All the frames acquired in step (ii) (in which the patient was asked to voluntarily and forcefully close and open their eyes five times) in which the patient's eyes are closed were labelled as 'spasm'; on the contrary, the frames acquired before or after step (ii) with the patient's eyes open were labelled as 'no spasm'. It is worth noting that, according to its definition, the spasm event requires the eyes to be closed, which is why we classified as 'spasm' the frames (acquired during the step (ii)) when the eyes were closed. The strategy of including in the folder 'No spasms' the frames with open eyes is justified by the fact that when the patient has their eyes open, the height of the five triangles does not change over time (which is the basic requirement for having the 'no spasm' event). For instance, preliminary analyses revealed that considering step (iv), when patients are asked to voluntarily and gently close and open their eyes five times, this is like a simulation of a blink, and the height of the triangles can change significantly. Some patients, in fact, due to the pathology, were not able to "gently" close and open the eyes and, having difficulties in re-opening the eyes, often moved their eyebrows. To avoid these issues, we decided to include in the folder 'No spasms' the only frames recorded in the first three minutes of the clinical test with the eyes open.

Two data sets (DS) were finally generated (Fig. 3, Block [13]): the first dataset, DS1, included 30160 entries and regarded the classification of the eye state (16576 entries were labelled as 'closed eyes' and 13584 as 'open eyes'); the second dataset, DS2, included 11266 entries and regarded the classification of spasm/no spasm events (4474 entries were labelled as 'spasm', the remaining 6792 as 'no spasm'). It is worth noting that, for each video frame, two entries can be obtained regarding the classification of the eye state, one entry for each extracted ROI and one entry regarding the classification of a spasm/no spasm event. Each entry of DS1, in turn, includes $n_{Feye\text{-}state} = 64$ features, whereas each entry of DS2 includes $n_{Fspasm} = 7$ features. Table 2 lists, for each patient, the number of entries obtained for each of the two datasets.

2.4. Topology optimisation and training of artificial neural networks

The two datasets DS1 and DS2 were given as input to Artificial Neural Networks (ANNs), which are models including a large number of processing units (neurons), used to solve specific classification or patterns recognition problems. Due to their capability "to learn" the hidden relationships between the input pattern and the output target, ANNs have been widely used to solve problems in different fields, from medical [22–28] to control [29] and manufacturing [30].

In this work, ANNs were utilised to automatically classify the blinks, and the brief and prolonged spasms observed in the nine recruited patients with BSP (Fig. 3, Blocks [14,15]). In general, the performance exhibited by such a classifier is strictly dependent on its topology, expressed in terms of the number of hidden layers, number of neurons per layer, and activation function per layer. In fact, the overall performance of the ANN depends on how the neurons in the hidden layers (the layers between the input layer and the output layer) process the weighted input information coming from the previous layers. Identifying the optimal ANN topology is a task of crucial importance, indeed, incoherent choices in the design phase can lead to unstable classification models with limited performance [16].

In this study, the optimal neural network topology was designed implementing evolutionary algorithms. According to Bevilacqua et al. [17], a mono-objective genetic algorithm (MOGA) can be used as an optimisation strategy to design ANNs with optimal topologies. For this purpose, a binary chromosome of 30 bits was assembled to describe the following features characterising an artificial neural network topology:

- first hidden layer; the number of neurons ranging in the interval [1–256], coded with 8 bit;
- second and third hidden layers; the number of neurons ranging in the interval [0–255], coded with 8 bit for each layer;
- first, second, and third hidden layers; activation function [0–3],

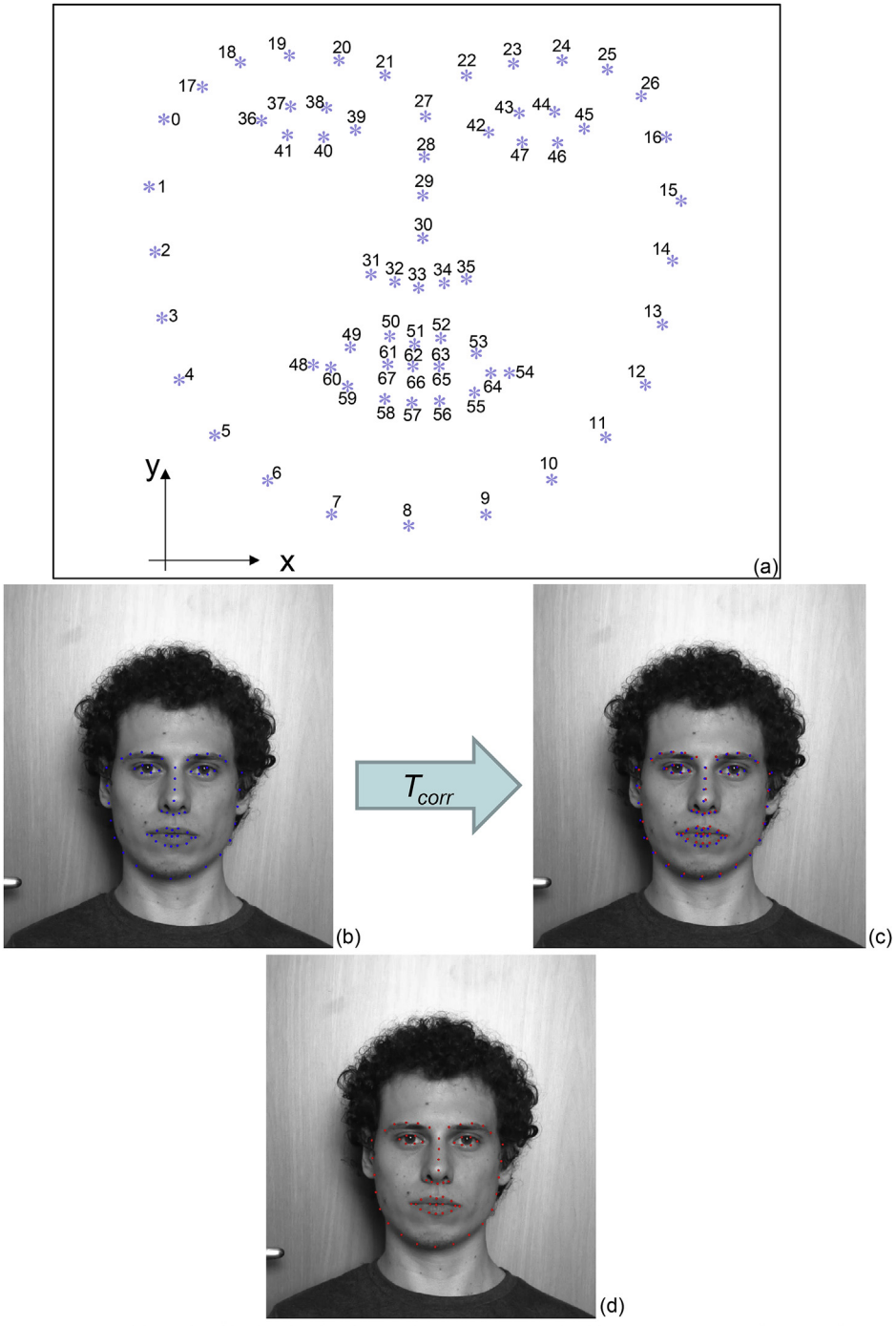


Fig. 4. (a) Schematic of the 68 facial landmarks placed by the face pose estimator algorithm. Implementation of the algorithm on one of the acquired frames before (b) and after (c) applying the correction tool T_{corr}. The final result of the implementation of T_{corr} tool (d) was represented by more stable and more correctly positioned facial landmarks. The red and blue points represent the facial landmarks correctly and not correctly positioned, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

coded with 2 bit for each activation function of each layer.

The four activation functions coded in the chromosome were: the log-sigmoid (*logsig*), the hyperbolic tangent sigmoid (*tansig*), the pure linear (*purelin*), and the symmetric saturating linear (*satlins*), whereas the activation function utilised in the output layer was the softmax function (*softmax*). The ANN features not encoded in the chromosome were fixed; the training algorithm for weights (W, Fig. 8) and bias (b, Fig. 8) update was the resilient backpropagation algorithm [31]. The parameters utilised in the genetic algorithm were: an initial population with 100 randomly generated individuals, where each individual corresponds, practically, to a candidate ANN topology; crossover with two points, with a probability of 0.8; mutation with a probability of 0.2; and the selection system elitism. The solution computed with the genetic algorithm was the optimal ANN topology, which, after training, validation, and testing for a given number of iterations n_{IT} of different permutations of the input dataset, showed the highest mean accuracy *av_ACC*. According to the confusion matrix shown in Table 3, if *TP(i)*, *TN(i)*, *FP(i),* and *FN(i)* are the number of true positives, true negatives, false positives, and false negatives, respectively, predicted with the

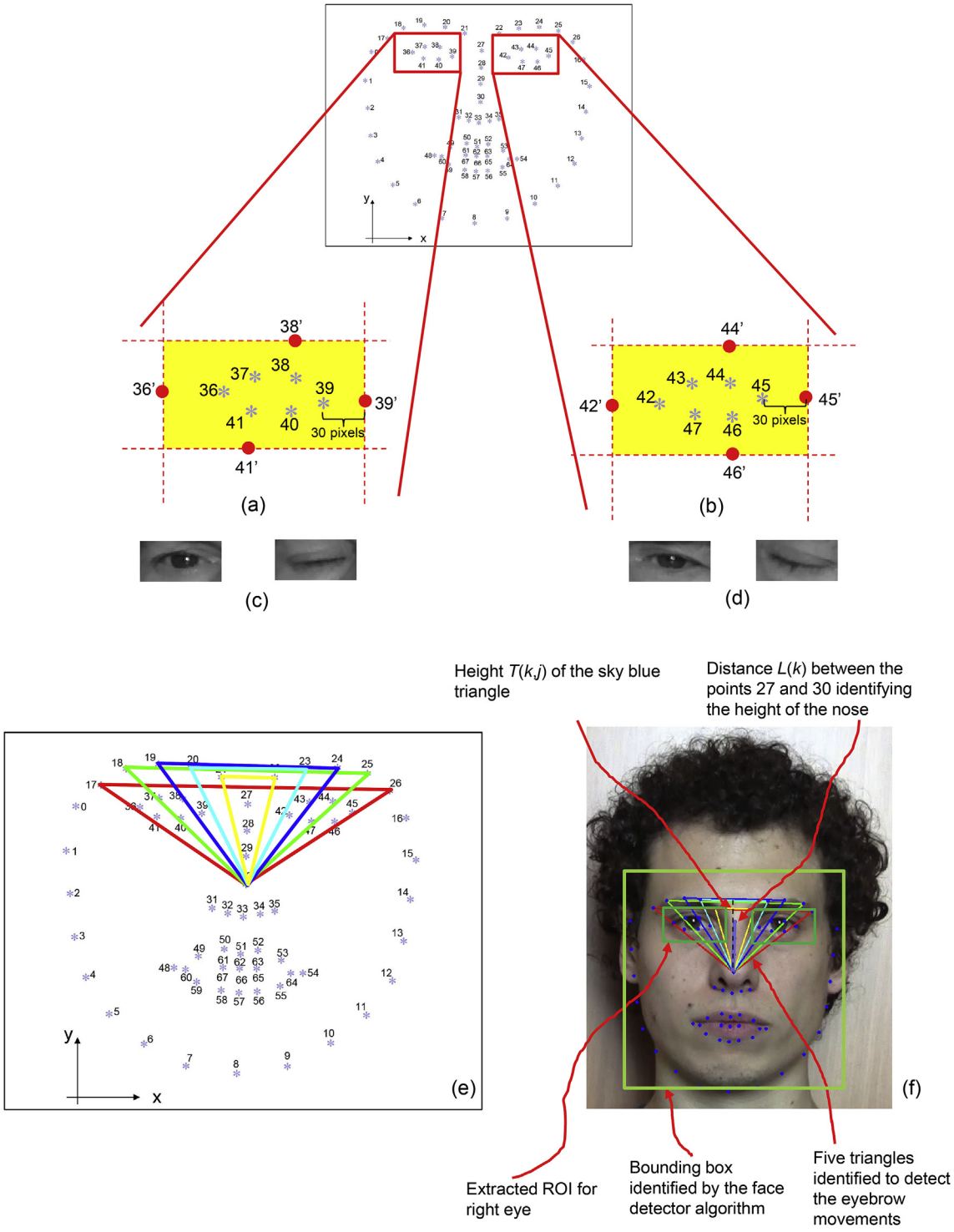


Fig. 5. (a) Regions of Interest (ROIs) extracted around the right (a) and the left (b) eye. Examples of ROIs extracted for the right (open and closed) (c) and the left open (e) and closed (d) eyes. The five triangles with one of the vertices on the tip of the nose and the others defined by the facial landmarks located on the eyebrows identified (on the schematic (e) and on the face of the patient (f)) to detect the eyebrow movements.

specific ANN topology for the *i*th permutation of the input dataset, the accuracy *ACC*(*i*) is defined as the ratio between the sum of the true predictions (i.e. *TP*(*i*) and *TN*(*i*)) and the sum of all the predictions (the true and the false ones, i.e. *TP*(*i*), *TN*(*i*), *FP*(*i*), *FN*(*i*)).

Therefore, the mean accuracy denoted as *av_ACC* can be computed as:

$$av\_ACC = \frac{\sum_{i=1}^{n_{IT}} ACC(i)}{n_{IT}} \quad (2)$$

In addition to *ACC*(*i*) and *av_ACC*, the specificity *SSPP*(*i*), and the sensitivity *SSEE*(*i*), for the *i*th permutation of the input dataset, and the average values of these quantities were also computed:

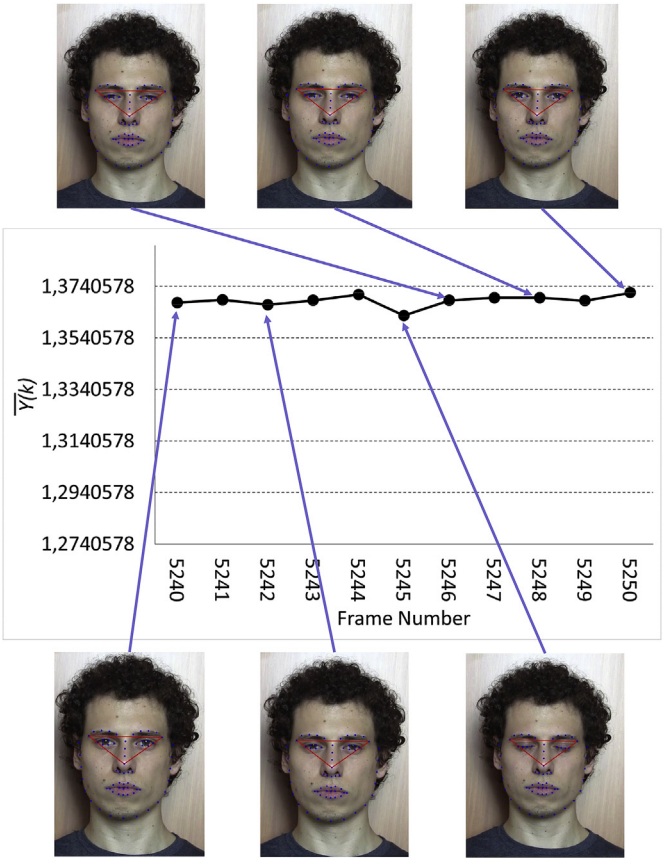


Fig. 6. Typical values of the average normalised height $\overline{Y}(k)$ of triangles registered during a blink.

$$\begin{cases} SSPP(i) = \frac{TN(i)}{FP(i)+TN(i)} \\ av\_SSPP = \frac{\sum_{i=1}^{n_{IT}} SSPP(i)}{n_{IT}} \\ SSEE(i) = \frac{TP(i)}{TP(i)+FN(i)} \\ av\_SSEE = \frac{\sum_{i=1}^{n_{IT}} SSEE(i)}{n_{IT}} \end{cases} \quad (3)$$

In this study, we set $n_{IT} = 200$. Training, validation, and test datasets were obtained from the two input datasets (i.e. DS1 and DS2); 60% of the samples were utilised for training, 20% for validation, and 20% for the test.

The genetic algorithm described above and successfully utilised in previous studies [27,28], was implemented to determine the optimal topology of two artificial neural networks: the ANN for classification of the eye state and the ANN for classification of the spasm/no spasm event. For both of them, the genetic optimisation algorithm predicted an optimal ANN topology including three hidden layers (Fig. 8).

In detail, the optimal ANN topology computed by the genetic algorithm for the eye state classifier included (Fig. 8(a)) 235, 108, 34, and 2 neurons for the first hidden, the second hidden, the third hidden, and the output layer, respectively. Furthermore, the genetic algorithm found the following activation functions: *tansig* for the first and the second hidden layer, and *logsig* for the third hidden layer. For the output layer, as stated above, the *softmax* function was utilised.

The optimal topology of the neural network for the classification of the spasm/no spasm event included (Fig. 8(b)) 71, 175, 24, and 2 neurons for the first hidden, second hidden, third hidden, and the output layer, respectively. The activation functions predicted by the genetic algorithm were *purelin*, *tansig*, and *logsig*, for the first, the second, and the third hidden layer, respectively. Again, *softmax* was utilised for the output layer.

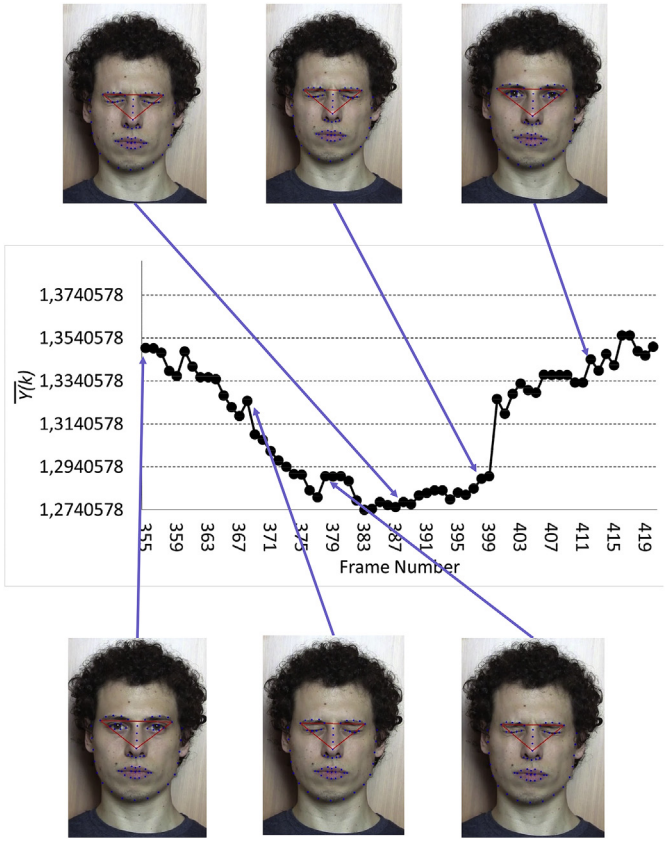


Fig. 7. Typical values of the average normalised height $\overline{Y}(k)$ of triangles registered during a spasm.

Table 2
Entries obtained from each patient and given in input to the neural network.

Patient	Entries 'closed eyes'	Entries 'open eyes'	Entries 'spasm'	Dataset eye state: total entries 'closed eyes' + 'open eyes'	Dataset spasm/no spasm: total entries 'spasm' + 'open eyes'
P1	2666	1950	978	4616	1953
P2	1544	1634	421	3178	1238
P3	2930	848	788	3778	1212
P4	1666	2112	785	3778	1841
P5	2634	784	341	3418	733
P6	1162	1596	301	2758	1099
P7	1068	1750	279	2818	1154
P8	1736	1022	325	2758	836
P9	1170	1888	256	3058	1200
				30160	**11266**

The average values of accuracy (*ACC*), specificity (*SSPP*), and sensitivity (*SSEE*) with the standard deviations computed over the $n_{IT} = 200$ iterations for the optimal ANN topologies are listed in Table 4.

2.5. Assessment of the optimised neural networks

The assessment of the two optimised neural networks was performed with the frames acquired in the last two minutes of the clinical test when patients were asked to remain at rest with their eyes open and staring at a specific point located in front of them (Fig. 3, Block [16]). For each acquired frame, the ROIs were cropped, the co-occurrence matrix was computed, and the heights of the five triangles were identified.

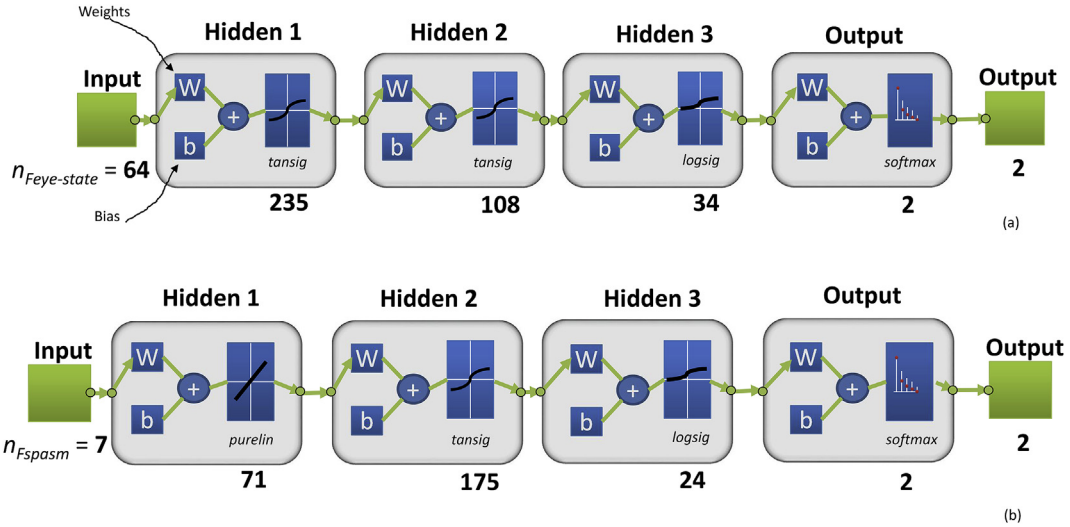


Fig. 8. Neural networks with optimised topology utilised to classify the eye state (a) (open or closed) and the spasm/no spasm event (b).

Table 3
Confusion matrix utilised to classify the predictions of the specific ANN topology for the *i*th permutation of the input dataset.

		True condition	
		Positive	Negative
Predicted condition	Positive	*TP(i)*	*FP(i)*
	Negative	*FN(i)*	*TN(i)*

Table 4
Values of *av_ACC*, *av_SSPP*, and *av_SSEE* computed over the $n_{IT} = 200$ iterations for the optimal ANN topologies.

	Eye state	Spasm/no spasm event
av_ACC ± *std*	0.9641 ± 0.0015	0.9290 ± 0.0051
av_SSPP ± *std*	0.9643 ± 0.0002	0.9507 ± 0.0007
av_SSEE ± *std*	0.9637 ± 0.0026	0.8743 ± 0.0136

In detail, two different data sets (DDSS) were created (Fig. 3, Block [17]): the first dataset, DDSS1, was given in input to the neural network for the classification of the eye state, the second, DDSS2, to the neural network for the classification of the spasm/no spasm event. DDSS1 included a number of entries equal to twice the number of frames acquired in the two minutes (for each frame, two ROIs can be extracted), and DDSS2 the number of frames acquired in the two minutes. Again, each entry of DDSS1 included $n_{Feye\text{-}state} = 64$ features (i.e. the dimensions of the co-occurrence matrix), whereas each entry of DDSS2 included $n_{Fspasm} = 7$ features (i.e. the number of triangle heights considered, Eq. (1)). Then, the following workflow was implemented:

- Giving DDSS1 as input to the first optimised neural network for the classification of the eye state, the frames containing closed eyes were first identified. Then, for each frame, the co-occurrence matrix, for both the right and the left eye, was computed. However, if one of the two co-occurrence matrices was predicted to be 'closed eye', the other one was automatically hypothesised to be 'closed eye'. This is since BSP is a focal dystonia with bilateral and synchronous symptoms that simultaneously affect the right and the left eyes [2–4]. Therefore, the output of this first classification was an array $A_{eye\text{-}state}$, with the length equal to the number of frames acquired in the two minutes of the clinical test, assuming, for each frame number, one of the following possible values: 0 in the case of open eyes and 1 in the case of closed eyes (Fig. 3, Block [18]; Fig. 9).

- Giving DDSS2 as input to the second optimised neural network, the frames where a lowering of the eyebrows took place were detected. The output of this second classification was an array A_{spasm}, with the same length s the previous array, which assumes, for each frame, one of the two possible values: 1 if an eyebrow narrowing occurs, and 0 otherwise (Fig. 3, Block [18]; Fig. 9).

- Finally, we distinguished the spasms from the blinks. It is worth noting that the requirements for a symptom to be classified as spasm are eye closure, the lowering of the eyebrows, and a duration of at least 300 ms, which corresponds to the time necessary to acquire approximately 10 frames with the camera utilised to video record the patients. A symptom with a duration shorter than this cannot be classified as a spasm but instead as a blink [32]. Furthermore, as stated above, a spasm lasting less than 3 s (i.e. the time to acquire approximately 100 frames) must be classified as a brief spasm, and a spasm lasting longer than this as a prolonged spasm. Therefore, the frames where $A_{eye\text{-}state}$ assumes the value 1 (i.e. 'eyes closed') were considered and, in correspondence with these frames, the values of A_{spasm} were also observed. If a set of less than 10 consecutive frames (i.e. a set with a number of consecutive frames $n_{c\text{-}frames} < 10$) is characterised by $A_{eye\text{-}state} = A_{spasm} = 1$, then the entire set is classified as a blink. Instead, if a set includes more than 10 and less than 100 consecutive frames (i.e. $10 \leq n_{c\text{-}frames} < 100$) with $A_{eye\text{-}state} = A_{spasm} = 1$, then the set is

Table 5
Spearman correlation coefficients between the scores (assigned to the different items) computed by the software and those determined by the expert neurologist.

Item	Features (further details are reported in Appendix B)	Spearman rho	*p-value*
A1	Type of eyelid spasm	0.793	0.019
A2	Apraxia of eyelid opening, not measurable by the proposed software	–	–
A3	Spasms occurring during the writing of the stereotyped sentence, not measurable by the proposed software	–	–
A4	Average duration of the prolonged spasms	0.806	0.009
B1	Frequency of blinks + brief spasms	0.676	0.046
B2	Frequency of prolonged spasms	0.756	0.030
Total 'measurable' severity index *SIn_m*	*SIn_m = S(A1) + S(A4) + S(B1) + S(B2)*	0.863	0.003

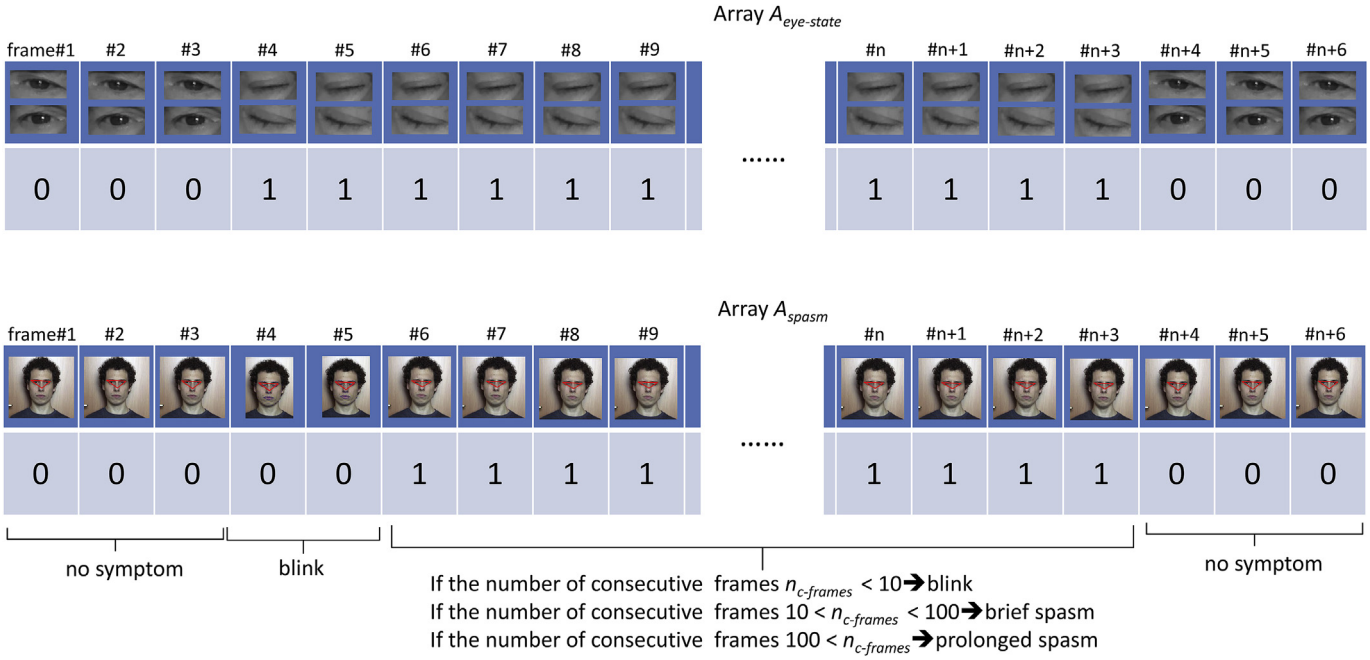


Fig. 9. Computation of the arrays $A_{eye\text{-}state}$ and A_{spasm} and classification of the blepharospasm symptoms.

classified as a brief spasm. If the condition $A_{eye\text{-}state} = A_{spasm} = 1$ is satisfied for a number of consecutive frames higher than 100 (i.e. $100 \leq n_{c\text{-}frames}$), then the set is classified as a prolonged spasm. Finally, all the sets of consecutive frames with $A_{eye\text{-}state} = 1$ and $A_{spasm} = 0$ were classified as a blink (Fig. 3, Block [19]; Fig. 9).

A short video-report shows how the proposed software works. A counter of symptoms was included in the video as well as a diagram showing the average height of the five triangles $\overline{Y}(k)$ as a function of the frame number.

Supplementary video related to this article can be found at https://doi.org/10.1016/j.compbiomed.2019.103376

2.6. Validation procedure

The last two minutes of the videos registered for the nine patients during the clinical test were manually segmented into a number of clips lasting 10–20 s. Each clip was trimmed to include just one of the following symptoms/events: brief spasm, prolonged spasm, blink, or no involuntary eye closure (Fig. 3, Block [20]). The segmentation was carried out by a (different) expert neurologist who did not participate to the other steps of the study. Four folders were then created and named: *Blinks*, *Brief_spasms*, *Blinks + brief_spasms*, *Prolonged_spasms,* which included the clips reproducing the symptom corresponding to the name of the folder. Furthermore, in the four folders, clips detected by the expert neurologist (who did not participate to the other steps of the study) showing no involuntary eye closure were also included. Therefore, for instance, the folder *Blinks* included all the clips of blinks as well as some of those showing no involuntary eye closures. Similarly, the folder *Blinks + brief_spasm* included the clips of brief spasms, those of blinks and, finally, some of those with no involuntary eye closures. These folders were then evaluated by both the proposed software and the expert neurologist that participated to the other steps of the study (i.e. the neurologist that determined the severity index *SIn* of the nine recruited patients). The severity index evaluated by the proposed software was correlated with the eye closure time and with the severity index measured by an expert neurologist. In particular, the Spearman rank correlation coefficient was computed, which is a nonparametric measure of rank correlation. It allows an evaluation of how well the relationship between two variables can be described using a monotonic function. Whereas Pearson's correlation assesses linear relationships, Spearman's correlation assesses, in general, monotonic relationships (that can be either linear or nonlinear). Furthermore, the Spearman correlation is particularly suited to evaluating relationships involving ordinal variables (such as, for instance, the variable severity index *SIn*). The sensitivity *SE* and the specificity *SP* (Fig. 3, Block [21]) of the software were then assessed. In detail, regarding the folder *Blinks*, if n_s is the number of clips correctly identified by the software as showing blinks (true positives) and n_n is the number of clips identified by the neurologist and showing the same symptom (positives), then the sensitivity is given by the ratio $SE = n_s/n_n \times 100$. The same procedure was adopted to compute the values of *SE* related to the other symptoms. Similarly, with reference to the folder *Blinks*, if n_{as} is the number of clips correctly detected by the software as showing no blink symptoms (true negatives) and n_{ans} is the correspondent number determined by the neurologist (negatives), then the specificity *SP* is given by: $SP = n_{as}/n_{ans} \times 100$. The same procedure was followed when assessing the specificity *SP* related to the other symptoms.

3. Results

The values of *SE* and *SP* computed for each detected symptoms are diagrammed in Fig. 10.

The clinimetric properties of the proposed software were assessed. For each patient, all the frames recorded in the last two minutes of the clinical test when the eyes were closed, were considered. Therefore, to determine the percentage of closure time for the investigated symptoms, the frames (with closed eyes) were distinguished depending on the symptom and counted. If f_{blink}, f_{bsp}, and f_{psp} is the number of frames showing blinks, brief spasms, and prolonged spasms, respectively, and f_{tot} the total number of frames registered in the last two minutes, the percentage of closure time for blinks t_{blink}, brief spasms $t_{bsp,}$ and prolonged spasms t_{psp} can be computed as follows:

$$\begin{cases} t_{blink} = \frac{f_{blink}}{f_{tot}} \times 100 \\ t_{bsp} = \frac{f_{bsp}}{f_{tot}} \times 100 \\ t_{psp} = \frac{f_{psp}}{f_{tot}} \times 100 \end{cases} \quad (4)$$

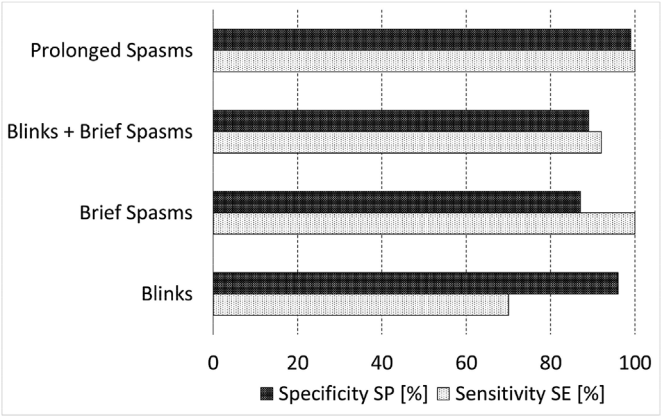


Fig. 10. Values of sensitivity *SE* and specificity *SP* obtained with the proposed software for the different investigated symptoms.

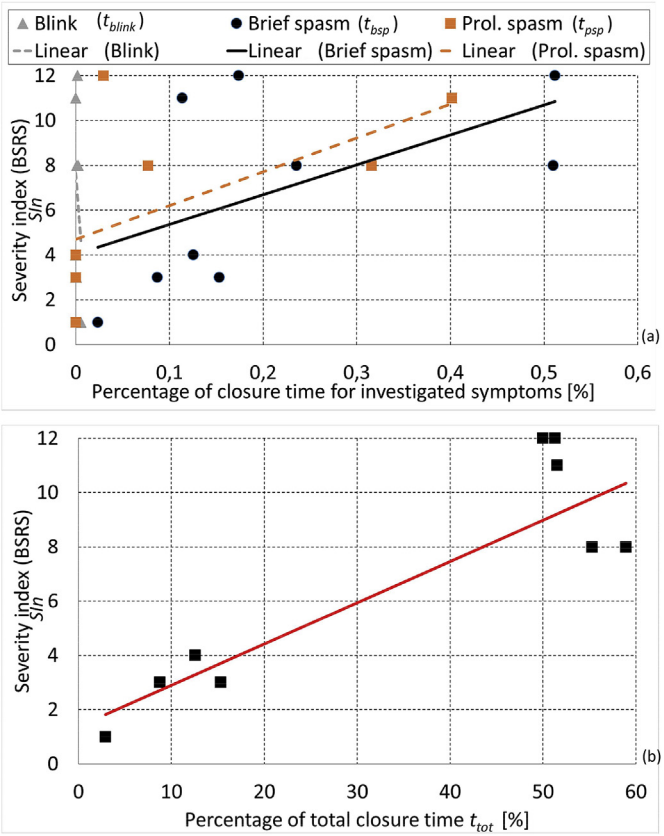


Fig. 11. Correlation between: (a) the severity index *SIn* determined by the expert neurologist and the percentages of closure time for the investigated symptoms and; (b) the severity index *SIn* determined by the expert neurologist and the total closure time.

The values of percentage of closure time were computed for all the patients and correlated with the severity index values *SIn* evaluated by the expert neurologist according to the Blepharospasm Severity Rating Scale (BSRS) scale (Fig. 11 (a)). Details of the BSRS scale and the modalities to compute *SIn* are given in Appendix B.

The values of the severity index (BSRS) were also correlated with the percentage of total closure time. If f_{totce} is the total number of frames registered in the last two minutes of the clinical tests and characterised from having closed eyes, the percentage of total closure time t_{tot} can be computed as:

$$t_{tot} = \frac{f_{totce}}{f_{tot}} \times 100 \quad (5)$$

The values of the severity index (BSRS) were diagrammed as a function of t_{tot}, and a linear regression line was also included in the diagram (Fig. 11 (b)).

It is worth noting that the proposed software is not capable of evaluating the index severity values *SIn* according to the Blepharospasm Severity Rating Scale (BSRS). The BSRS scale, in fact, includes six items, for each of which a score *S* must be assigned according to specific criteria (further details on the BSRS scale are reported in Appendix B). Among others, the BSRS includes item A2 which regards the apraxia of eyelid opening and item A3 which concerns the spasms occurring during the writing of the stereotyped sentence (i.e. the step (viii)). Due to how the system is designed, it is not capable of assigning a score for the items A2 and A3. However, we found that by only considering the 'measurable' items and summing up the scores given to each of the measurable items, the software gives output values of (measurable) severity index *SIn_m* (see the definition of the measurable severity index in Appendix B) that are consistent with the corresponding values determined by the expert neurologist (Spearman rho 0.863, *p-value* 0.003) (Fig. 12). Significant values of the Spearman correlation coefficients can also be found considering the score given to individual 'measurable' items computed by the software and the score determined by the expert neurologist (Table 5).

4. Discussion

The sensitivity *SE* of the software was excellent for prolonged spasms and satisfactory for brief spasms (Fig. 10). Lower values of sensitivity were, however, found in the case of blinks; a confusion related to the imperceptible difference between blinks and brief spasms was probably responsible for this result. Proof of this is given by the satisfactory level of sensitivity computed in the case when blinks and brief spasms are combined in the same folder. The high levels of specificity *SP* demonstrate the capability of the proposed software to distinguish the non-pathological conditions.

High Spearman correlation coefficients were computed for brief (Spearman rho 0.684, *p-value* 0.042) and prolonged (Spearman rho 0.783, *p-value* 0.022) spasms (Fig. 11(a)). A very low correlation coefficient was, by contrast, found in the case of the blinks (*p-value* n.s.), which indicates that no clear correlation exists between t_{blink} and the severity index *SIn* values. In reality, it is worth noting that the severity index values *SIn* reported in the diagram take into account, in addition to blinks, also other symptoms. Considering that the weight blinks have on this severity scale is very small compared to the weight of the other symptoms (practically, only item B1 partially depends on the number of

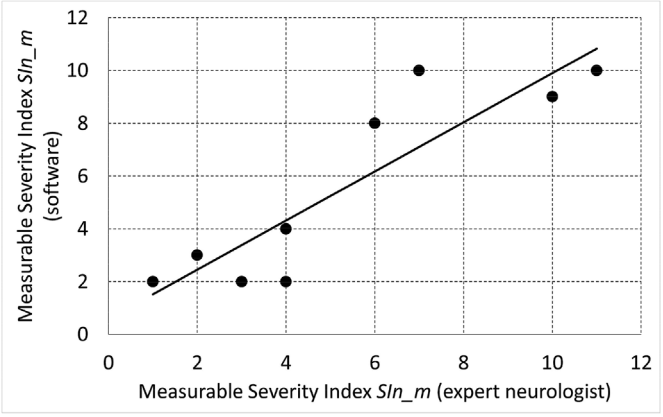


Fig. 12. Correlation between the measurable severity index *SIn_m* computed by the software and that determined by the expert neurologist.

blinks, please see item B1 in Appendix B), one can understand that blinks only in a very marginal way affect the severity index values, which justifies the absence of a correlation. A significant correlation coefficient was found between the severity index *SIn* values and t_{tot} (Spearman rho 0.735, *p-value* 0.038) (Fig. 11(b)).

The proposed software tool presents some limitations. First, although the software exhibits high values of sensitivity *SE* in distinguishing brief spasms, prolonged spasms, and blinks + brief spasms, the value of *SE* computed for blinks was small. This can be justified by the argument that the software sometimes confuses blinks with brief spasms. However, it is worth noting that the difference between the symptoms of the blink and the symptoms of the brief spasm is subtle and, as often occurs in clinical practice, the task of distinguishing the two symptoms is complex even for the neurologist. This is true especially in the case the brief spasms that have a duration close to the threshold value of 300 ms [32] which represents, de facto, the time duration that distinguishes brief spasms from blinks. During the clinical evaluation, the neurologist does not physically measure the time, thus, increasing the probability of confusing the two symptoms. The proposed software instead takes into account the exact number of frames included in the set showing the symptom under investigation and can measure the time with an accuracy of approximately 0.03 s. Furthermore, the distinction of blepharospasm symptoms is often subtle and imperceptible. It is commonly known, in fact, that different expert neurologists can assign different severity index *SIn* values to the same patients. Further investigations should be carried out on this topic.

The second limitation of the study is represented by the time necessary to implement the software on the specific patient, which includes: 1) the time necessary to utilise the correction tool T_{corr} (when the neurologist is required to drag and drop the incorrect facial landmarks on approximately 30 acquired frames); 2) the time necessary to re-train the face pose estimator after implementing T_{corr}; 3) the time necessary to extract, from the acquired frames, all the entries to give in input to the neural networks. Preliminary investigations revealed that for an experienced neurologist, all these tasks require approximately and on average 30 min. However, considering that an experienced neurologist spends approximately 60–90 min per patient observing all the acquired video clips, one can conclude that implementing the software allows the saving of more than 0.5 h of time per patient. Furthermore, it is worth noting that the proposed software was not conceived to 'replace' the neurologist but to 'assist/support' her/him in defining the severity of BSP. From this point of view, the limitation of the time necessary to implement the software is relevant and is abundantly counterbalanced by the important advantage of making the process of evaluating the BSP symptoms objective.

The third limitation of the study is represented by the fact that the datasets given in input to the ANNs for the optimisation of their topology are unbalanced (Table 2). A measure of the quality of binary classifications, even if the classes are of very different sizes, is represented by the Matthews Correlation Coefficient (*MCC*). *MCC* is practically a correlation coefficient, the value of which ranges in the interval [-1; 1]. An *MCC* coefficient of +1 indicates a perfect prediction, and −1 a total disagreement between the prediction and observation. The Matthews Correlation Coefficient *MCC(i)* for the *i*th permutation of the input dataset can be computed directly from the confusion matrix using the following relationship:

$$MCC(i) = \frac{TP(i) \times TN(i) - FP(i) \times FN(i)}{\sqrt{(TP(i) + FP(i)) \times (TP(i) + FN(i)) \times (TN(i) + FP(i)) \times (TN(i) + FN(i))}} \quad (6)$$

The value of *MCC(i)* averaged over n_{IT} iterations *av_MCC* obtained for the two optimised ANNs are: ANN for the classification of the eye state *av_MCC* = 0.9612 ± 0.0016; and ANN for the classification of the spasm/no spasm event *av_MCC* = 0.8481 ± 0.0055, which are rather high values of the correlation coefficient.

Despite these limitations, we found that a correlation exists between the total closure time t_{tot} and the severity index values *SIn* (Fig. 8), which is consistent with the results of Peterson et al. [9]. Furthermore, a very high level of ϵ_{FD}, the percentage of video frames when the face of the patient was detected with the face detector algorithm, was found. The lowest percentage of face-found frames was ϵ_{FD} = 99.948%, which is higher than that (93%) found by Peterson et al. [9]. Furthermore, it is worth noting that currently, the only computerised and automatic system capable of rating the blepharospasm severity is represented by the toolbox CERT [9], which is capable of measuring the eyes closure time but cannot recognise, and hence count, the specific BSP symptoms. Finally, the implementation of T_{corr} led to correctly determining the position of the facial landmarks, which allowed the definition of 'stable' triangles for the identification of the spasms.

5. Conclusions

A new software tool was developed which, starting from video frames registered on patients with blepharospasm, was capable of recognising the symptoms of this focal dystonia. In detail, the software, after detecting the state of the eyes (open or closed) and the movement of specific facial landmarks and properly implementing artificial neural networks with an optimised topology, can recognise blinking, and brief and prolonged spasms. Comparing the software predictions with the observations of an expert neurologist allowed assessment of the sensitivity *SE* and the specificity *SP* of the proposed software. Significant levels of *SE* were obtained for brief and prolonged spasms, whereas the values of sensitivity were lower in the case of blinks. Possible explanations were given for this result. Significant values of specificity were also found, which demonstrates the capability of the software to recognise the non-pathological facial movements. Significant Spearman correlation coefficients were found between the percentages of closure time related to the specific symptoms determined by the software and the severity index *SIn* values (BSRS scale) determined by an expert neurologist. Significant correlation coefficients were also found for other parameters (quantified in the BSRS scale) measured by both the software and an expert neurologist. The spread of classification systems based on deep learning strategies, working directly on the video frames, might help improve the classification performance in discriminating blinks from spasms which, as already mentioned, reports low levels of sensitivity.

The development of the software allowed a separate evaluation of the contribution of the individual symptoms to the global severity index, thus opening up new perspectives in the problem of evaluating/measuring BSP symptoms. The proposed software is an automatic tool capable of making the 'measurement' of BSP symptoms objective and, hence, assisting/supporting the neurologist in rating the severity of the dystonia.

Conflicts of interest

The authors declare that there is no conflict of interests.

Acknowledgments

This research was partially supported with the financial sources of Regione Puglia project INNOLABS – RECALL (project n. QIRYKE8).

Appendix A. Implementation of the correction tool T_{corr}

In order to test the reliability of the proposed correction tool T_{corr}, we developed an algorithm that extracts the coordinates x and y of the five landmarks located on the two eyebrows, i.e. points 17–21 for the right eyebrow and 22–26 for the left one (Fig. A1 (a)). For each patient, 10 portions of video with a duration (each) of 0.3 s were considered. Therefore, the total number of frames considered for each patient and included in each portion of video was 29.97 (acquisition frequency) × 0.3 (duration in seconds of the video portion) = 9. For all the nine patients, the standard deviation of the x and y coordinates of points 17–21 and 22–26 detected in the 9 frames included in each portion of video was computed (Fig. A1 (a)). If $\sigma_{x,i,j}$ (or, $\sigma_{y,i,j}$) is the value of the standard deviation obtained for the coordinate x (or, the coordinate y), in the ith video portion (i = 1, 2,...,10), and for the jth point (j = 17, 18, …26), the value of the standard deviation $\sigma_{x,\,j}$ (or, $\sigma_{y,\,j}$) obtained for the jth point (Fig. A1) was computed as:

$$\begin{cases} \sigma_{x,j} = \frac{1}{10}\sum_{i=1}^{10}\sigma_{x,i,j} \\ \sigma_{y,j} = \frac{1}{10}\sum_{i=1}^{10}\sigma_{y,i,j} \end{cases} \tag{A1}$$

Video portions with a duration of only 0.3 s were considered since this is the shortest time interval in which blinks of patients with blepharospasm and oromandibular dystonia can occur (brief and prolonged spasms, in fact, have a longer duration). Electromyographic analyses revealed, in fact, that bursts of cocontracting activity in the facial muscles accompanying the involuntary movements take place in patients with BSP and last from 200 to 300 ms to many seconds [32]. Therefore, one can conclude that in this short time, very small facial movements occur and hence, any changes experienced by the facial landmarks in this same time interval can be justified with only the stability/accuracy of the face pose estimator. In other words, if the coordinates of a given landmark change in this short time interval, this will principally depend on the accuracy with which the face pose estimator algorithm predicts the landmark location and only marginally on the possible BSP symptoms occurring in this same time interval. In general, the larger the standard deviation, the larger the changes in the landmark coordinates and hence the lower the stability/accuracy of the face pose estimator algorithm. Conversely, the smaller the standard deviation, the higher the accuracy with which the algorithm predicts the location of the landmarks. Interestingly, the values of the standard deviation computed after implementing the correction tool T_{corr} are significantly lower than those obtained without implementing T_{corr} (Figs. A1 (b)), which demonstrates that T_{corr} actually increases the predictive power of the face pose estimator algorithm. The values of standard deviation shown in the diagram of Fig. A1 (b) refer to video frames acquired in the first three minutes of the clinical test, and include the video frames “corrected” by the clinician by dragging and dropping points into correct locations. Interestingly, we found that by computing the standard deviation on portions of video frames registered in the last two minutes of the clinical test, the same conclusions can be drawn (Fig. A1 (c)). Again, the effect of T_{corr} consists of increasing the accuracy and predictive power of the face pose estimator.

Appendix B. Blepharospasm Severity Rating Scale (BSRS)

The blepharospasm severity rating scale includes six items [1], for each of which a score S must be assigned.
In detail:
Item A1 concerns the type of eyelid spasm occurring in the patient.

- If a brief spasm (duration $<$ 3 s) with complete rim closure take place → score $S(A1)$ = 1;
- If a prolonged spasm (duration ≥3 s) with partial rim closure take place → score $S(A1)$ = 2;
- If a prolonged spasm (duration ≥3 s) with complete rim closure take place → score $S(A1)$ = 3.

Item A2 concerns the apraxia of eyelid opening.

- If apraxia is present → score $S(A2)$ = 2;
- If apraxia is absent → score $S(A2)$ = 0;

Item A3 concerns the spasms occurring during the writing of the stereotyped sentence.

- If spasms of the orbicularis oculi occur → score $S(A3)$ = 1
- If spasms of the orbicularis oculi do not occur → $S(A3)$ = 0;

Item A4 concerns the average duration of the prolonged spasms.

- If the average duration is 3–4 s → score $S(A4)$ = 1;
- If the average duration is 4.1–5 s → score $S(A4)$ = 2;
- If the average duration is more than 5 s → score $S(A4)$ = 3;

Item B1 regards the frequency of blinks + brief spasms.

- If 1 to 18 blinks + brief spasms take place per minute → score $S(B1)$ = 1;
- If 19 to 32 blinks + brief spasms take place per minute → score $S(B1)$ = 2;
- If more than 32 blinks + brief spasms take place per minute → score $S(B1)$ = 3;

Item B2 regards the frequency of prolonged spasms.

- If 1 to 3 prolonged spasms take place per minute → score $S(B2)$ = 1;

- If 3.1 to 7 prolonged spasms take place per minute → score $S(B2) = 2$;
- If more than 7 prolonged spasms take place per minute → score $S(B2) = 3$.

The total score is given by the sum:

$$SIn = S(A1) + S(A2) + S(A3) + S(A4) + S(B1) + S(B2) \quad \text{(B1)}$$

Due to how the software was designed, it is not capable of evaluating Items A2 and A3. Therefore, the measurable severity index SIn_m that can be automatically determined by the software is given by:

$$SIn\_m = S(A1) + S(A4) + S(B1) + S(B2) \quad \text{(B2)}$$

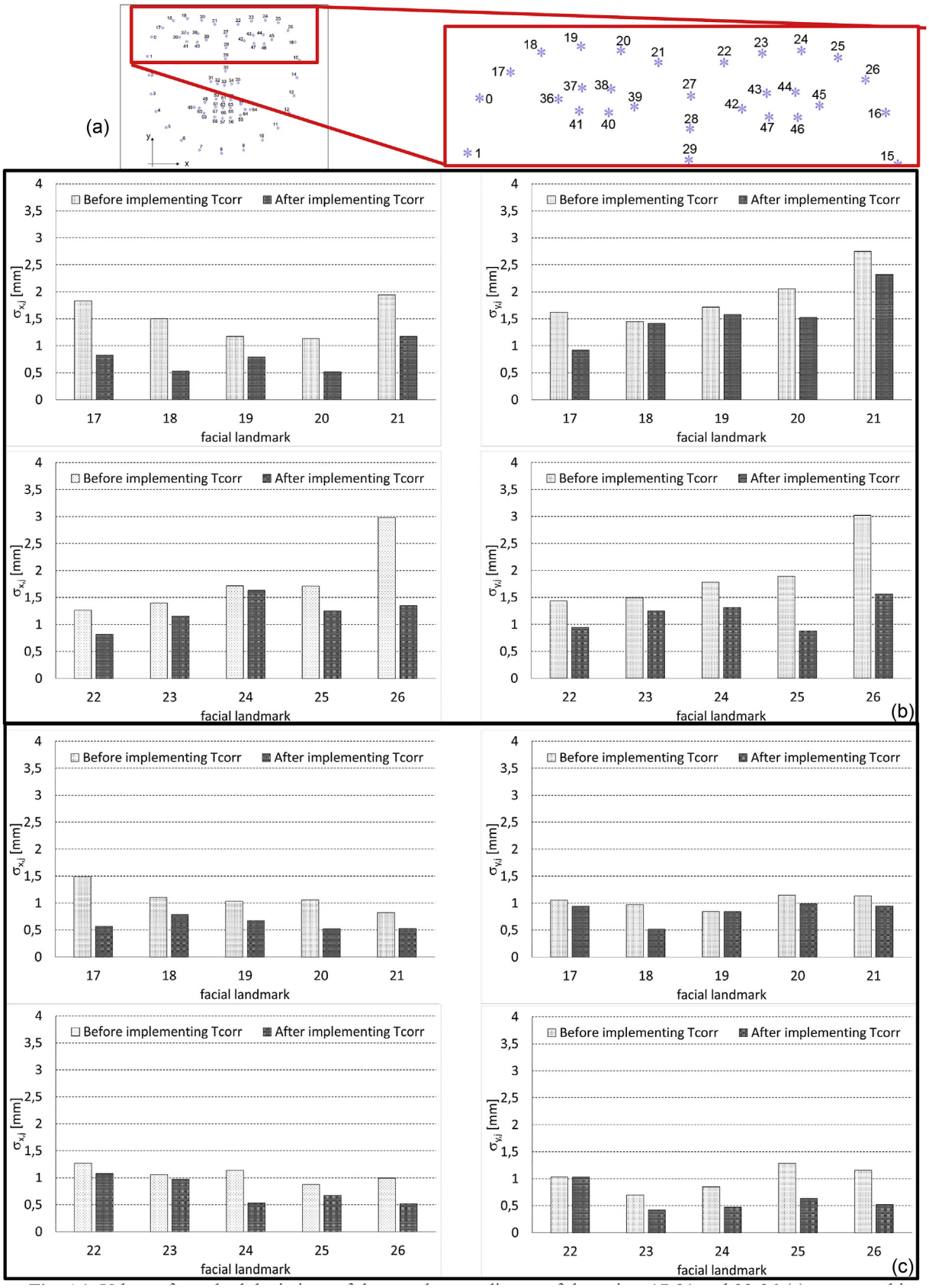


Fig. A1. Values of standard deviations of the x and y coordinates of the points 17–21 and 22–26 (a), computed in the first three minutes (b) and in the last two minutes (c) of the clinical test.

References

[1] G. Defazio, M. Hallett, H.A. Jinnah, A. Berardelli, Development and validation of a clinical guideline for diagnosing blepharospasm, Neurology 81 (2013) 236–240.

[2] F. Grandas, J. Elston, N. Quinn, C. Marsden, Blepharospasm: a review of 264 patients, J. Neurol. Neurosurg. Psychiatry 51 (1988) 767–772.

[3] M. Hallett, C. Evinger, J. Jankovic, M. Stacy, Update on blepharospasm report from the BEBRF international workshop, Neurology 71 (2008) 1275–1282.

[4] H. Jinnah, A. Berardelli, C. Comella, G. Defazio, M.R. DeLong, S. Factor, W.R. Galpern, M. Hallett, C.L. Ludlow, J.S. Perlmutter, others, The focal dystonias: current views and challenges for future research, Mov. Disord. 28 (2013) 926–943.

[5] G. Defazio, M. Hallett, H.A. Jinnah, G.T. Stebbins, A.F. Gigante, G. Ferrazzano, A. Conte, G. Fabbrini, A. Berardelli, Development and validation of a clinical scale for rating the severity of blepharospasm, Mov. Disord. 30 (2015) 525–530.

[6] D. Martino, G. Defazio, G. Alessio, G. Abbruzzese, P. Girlanda, M. Tinazzi, G. Fabbrini, L. Marinelli, G. Majorana, M. Buccafusca, others, Relationship between eye symptoms and blepharospasm: a multicenter case–control study, Mov. Disord. 20 (2005) 1564–1570.

[7] A. Conte, G. Defazio, G. Ferrazzano, M. Hallett, A. Macerollo, G. Fabbrini, A. Berardelli, Is increased blinking a form of blepharospasm? Neurology 80 (2013) 2236–2241.

[8] M. Ugarte, M. Teimory, Apraxia of lid opening, Br. J. Ophthalmol. 91 (2007) 854–854.

[9] D.A. Peterson, G.C. Littlewort, M.S. Bartlett, A. Macerollo, J.S. Perlmutter, H. Jinnah, M. Hallett, T.J. Sejnowski, Objective, computerized video-based rating of blepharospasm severity, Neurology 87 (2016) 2146–2153.

[10] R.E. Burke, S. Fahn, C.D. Marsden, S.B. Bressman, C. Moskowitz, J. Friedman, Validity and reliability of a rating scale for the primary torsion dystonias, Neurology 35 (1985) 73–73.

[11] C.L. Comella, S. Leurgans, J. Wuu, G.T. Stebbins, T. Chmura, Rating scales for dystonia: a multicenter assessment, Mov. Disord. 18 (2003) 303–312.

[12] J. Jankovic, J. Orman, Botulinum A toxin for cranial-cervical dystonia A double-blind, placebo-controlled study, Neurology 37 (1987) 616–616.

[13] V. Bevilacqua, A.E. Uva, M. Fiorentino, G.F. Trotta, M. Dimatteo, E. Nasca, A.N. Nocera, G.D. Cascarano, A. Brunetti, N. Caporusso, others, A comprehensive method for assessing the blepharospasm cases severity, International Conference on Recent Trends in Image Processing and Pattern Recognition, Springer, 2016, pp. 369–381.

[14] C. Loconsole, D. Chiaradia, V. Bevilacqua, A. Frisoli, Real-time emotion recognition: an improved hybrid approach for classification performance, International Conference on Intelligent Computing, Springer, 2014, pp. 320–331.

[15] V. Bevilacqua, D. D'Ambruoso, G. Mandolino, M. Suma, A new tool to support diagnosis of neurological disorders by means of facial expressions, Medical Measurements and Applications Proceedings (MeMeA), 2011 IEEE International Workshop on, IEEE, 2011, pp. 544–549.

[16] V. Bevilacqua, G. Mastronardi, F. Menolascina, P. Pannarale, A. Pedone, A novel multi-objective genetic algorithm approach to artificial neural network topology optimisation: the breast cancer classification problem, Neural Networks, 2006. IJCNN'06. International Joint Conference on, IEEE, 2006, pp. 1958–1965.

[17] V. Bevilacqua, A. Brunetti, M. Triggiani, D. Magaletti, M. Telegrafo, M. Moschetta, An optimized feed-forward artificial neural network topology to support radiologists in breast lesions classification, Proceedings of the 2016 on Genetic and Evolutionary Computation Conference Companion, ACM, 2016, pp. 1385–1392.

[18] D.E. King, Dlib-ml: a machine learning toolkit, J. Mach. Learn. Res. 10 (2009) 1755–1758.

[19] V. Kazemi, J. Sullivan, One millisecond face alignment with an ensemble of regression trees, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 1867–1874.

[20] C. Sagonas, E. Antonakos, G. Tzimiropoulos, S. Zafeiriou, M. Pantic, 300 faces in-the-wild challenge: database and results, Image Vis Comput. 47 (2016) 3–18.

[21] B. Zhang, W. Wang, B. Cheng, Driver eye state classification based on cooccurrence matrix of oriented gradients, Adv. Mech. Eng. 7 (2015) 707106.

[22] A. Saghafi, C.P. Tsokos, M. Goudarzi, H. Farhidzadeh, Random eye state change detection in real-time using EEG signals, Expert Syst. Appl. 72 (2017) 42–48.

[23] K.S. Rao, V. Saroj, S. Maity, S.G. Koolagudi, Recognition of emotions from video using neural network models, Expert Syst. Appl. 38 (2011) 13181–13185.

[24] T. D'Orazio, M. Leo, C. Guaragnella, A. Distante, A visual approach for driver inattention detection, Pattern Recognit. 40 (2007) 2341–2355.

[25] S. Raith, E.P. Vogel, N. Anees, C. Keul, J.-F. Güth, D. Edelhoff, H. Fischer, Artificial Neural Networks as a powerful numerical tool to classify specific features of a tooth based on 3D scan data, Comput. Biol. Med. 80 (2017) 65–76.

[26] B. Pandey, R. Mishra, Knowledge and intelligent computing system in medicine, Comput. Biol. Med. 39 (2009) 215–230.

[27] V. Bevilacqua, N. Pietroleonardo, V. Triggiani, A. Brunetti, A.M. Di Palma, M. Rossini, L. Gesualdo, An innovative neural network framework to classify blood vessels and tubules based on Haralick features evaluated in histological images of kidney biopsy, Neurocomputing 228 (2017) 143–153.

[28] A.I. Triggiani, V. Bevilacqua, A. Brunetti, R. Lizio, G. Tattoli, F. Cassano, A. Soricelli, R. Ferri, F. Nobili, L. Gesualdo, others, Classification of healthy subjects and Alzheimer's disease patients with Dementia from cortical sources of resting state EEG rhythms: a study using artificial neural networks, Front. Neurosci. 10 (2017) 604.

[29] S. Soyguder, Intelligent control based on wavelet decomposition and neural network for predicting of human trajectories with a novel vision-based robotic, Expert Syst. Appl. 38 (2011) 13994–14000.

[30] O. Duran, J. Maciel, N. Rodriguez, Comparisons between two types of neural networks for manufacturing cost estimation of piping elements, Expert Syst. Appl. 39 (2012) 7788–7795.

[31] M. Riedmiller, H. Braun, A direct adaptive method for faster backpropagation learning: the RPROP algorithm, Neural Networks, 1993., IEEE International Conference on, IEEE, 1993, pp. 586–591.

[32] A. Berardelli, J. Rothwell, B. Day, C. Marsden, Pathophysiology of blepharospasm and oromandibular dystonia, Brain 108 (1985) 593–608.