Neural Computing and Applications (2020) 32:16723–16736
https://doi.org/10.1007/s00521-020-05349-w

ORIGINAL ARTICLE

R-CNN and wavelet feature extraction for hand gesture recognition with EMG signals

Vimal Shanmuganathan[1] · Harold Robinson Yesudhas[2] · Mohammad S. Khan[3] · Manju Khari[4] · Amir H. Gandomi[5]

Received: 25 March 2020 / Accepted: 7 September 2020 / Published online: 22 September 2020

Abstract

This paper demonstrates the implementation of R-CNN in terms of electromyography-related signals to recognize hand gestures. The signal acquisition is implemented using electrodes situated on the forearm, and the biomedical signals are generated to perform the signals preprocessing using wavelet packet transform to perform the feature extraction. The R-CNN methodology is used to map the specific features that are acquired from the wavelet power spectrum to validate and train how the architecture is framed. Additionally, the real-time test is completed to reach the accuracy of 96.48% compared to the related methods. This kind of result proves that the proposed work has the highest amount of accuracy in recognizing the gestures.

Keywords R-CNN · EMG signal · Wavelet power spectrum · Discrete wavelet transform · Gesture recognition · Validation

1 Introduction

Human–machine interfaces (HMI) [1] might potentially expand the quality of life of disabled persons who suffer from neuromuscular diseases. The interactions between a machine and a human will need a spontaneous, healthy and high information/transmission ratio for these kinds of defects [2]. An HMI framework normally performs the feedback- and control-based information exchange to implement the 2-directional communication between a machine and a human. Many bio-based signals are utilized, like MMG [3], ultrasound [4] and EEG [5], to implement this kind of control.

Within these kinds of bio-related signals, the electromyography (EMG) signal is mostly used for HMI. EMG-related systems [6] are commonly utilized in different applications. Moreover, the control mechanisms are constructed using simplified methods. The EMG signals [7] are stored in the antagonistic muscles for every channel to construct a degree of freedom. The signal is picked up as the amplified at the electrode field that the signal could be utilized to diminish the frequency noise; it is normally rectified into a standard format to identify the EMG amplitude. It contains a tool as it implements as a basic

✉ Mohammad S. Khan
adhoc.khan@gmail.com

Vimal Shanmuganathan
svimalphd@gmail.com

Harold Robinson Yesudhas
yhrobinphd@gmail.com

Manju Khari
Manjukhari@yahoo.co.in

Amir H. Gandomi
gandomi@uts.edu.au

1 Department of Information Technology, National Engineering College, Kovilpatti, India

2 School of Information Technology and Engineering, Vellore Institute of Technology, Vellore, India

3 Department of Computing, East Tennessee State University, Johnson City, USA

4 Department of CSE, Ambedkar Institute of Advanced Communication Technologies and Research, New Delhi, India

5 Faculty of Engineering and Information Technology, University of Technology Sydney, Sydney, NSW 2007, Australia

way of classifying and sensing the dissimilar movements in the body. A degree of freedom robotic technique can efficiently activate the human limb motion. The microcontroller-based technology has been used to increase the control actions for robotic technique. These methodologies are effective and utilized for long-time evaluation and training.

EMG signals have been the subject of study since the discovery of diseases caused by the neuromuscular system [8]. Therefore, an EMG signal provides the information necessary to diagnose a neuromuscular disease or some damage caused by an injury to any of the muscles of the body [9]. However, researchers and scientists have taken advantage of these signals by implementing them for the control of prostheses and robotic manipulators [10]. In other words, they are electrical signals [11] that are generated by the various ions that are present in the muscle during its flexion and contraction movements, which can be monitored by 2 techniques. The first technique is invasive through electrodes inserted inside the muscle, mainly used to study deep muscles; this consists of stimulating the muscle with small electrical impulses and observing the muscle's behavior before the discharges [12]. The second is through noninvasive electrodes using surface electrodes that allow the study of muscle activity in dynamic actions; in other words, it allows simultaneous analysis of different muscle movements [13].

Monitoring EMG signals then allows not only discrimination of myocardial and muscle failures but also characterization of a person's intention of movement [14]. It is necessary to use pattern recognition techniques for this [15], framed within artificial intelligence (AI) algorithms [16]. Currently, these have been focused on deep neural networks that correspond to a model inspired by the brain's functioning, which analyzes information as a deep framework [17], for example, 5–10 layers of processing for visual information. This technique, by its layered learning framework, allows learning framework to reach several levels of abstraction and to obtain remarkable results in the recognition of patterns. Convolutional neural networks (CNN) [18] are a specific algorithm of deep learning; these are defined as multi-layered, trainable structures that, with their convolution filters, guarantee the learning of specific characteristics throughout each level of the network.

The input features contain a classification-related accuracy to maintain the feature extractions, such as frequency for efficient gesture recognition [19]. The features are extracted using wavelet transformation [20] for effective pattern recognition. The features that are extracted from the EMG signals indicate that the neural derive is delivered to the muscles within the spinal cord [21]. The neural features are proposed to muscle-recognized EMG signals that are identified to perform EMG decomposition [22].

The myoelectric signal is used to control the powered upper limbs, the dimension reduction is analyzed using the principle component analysis that has the effective control. EMG records the muscle movement and it is according to the channel element that consists of the channel labels equivalent to the data elements [23]. The movement component contains the sampling frequency data that the qualitative signal evaluation holds the EMG signal parameters. The force channel identifies the finger movement with generation of quality signals. It indicates that the fine wires are migrated to the adjacent muscles using the channels, the signals were individually analyzed, and the EMG signals are classified as the good signals and bad signals. The results show that the myoelectric control has greater precision when four channels are implemented, but the increased channels could cause response delays [24].

The wavelet transformation is used for classification and utilized to decompose the EMG signals while measuring the similarity of the spectrum matching into the wavelet domain. The interference is developed to reduce the frequency noise. The muscle contraction for the spectral elements of the surface myoelectric signal is compressed through the low-level frequency. The muscle contraction is used to identify the surface EMG while obtaining the more precise data regarding the user's forearm movement using the accelerometer. The muscle contraction can be identified to any level of noise as the interface component on muscle tone is utilized to identify the moving objects.

However, the investigations focused on the processing of EMG signals using CNN for the recognition of patterns are just beginning to be explored. A decoding method adaptable to each user to obtain the intention of involvement is proposed. It would use EMG signals for the portable rehabilitation of upper limbs in which they use a CNN network to predict 6 different hand movements. By adopting a simple CNN framework consisting of 5 convolution layers, they conclude that the framework is successful, although with an accuracy of $66.59 \pm 6.4\%$, due to the large number of patterns they seek to recognize.

The important contribution of the proposed work is

- Using the R-CNN methodology for the recognition of gestures.
- Implementing a technique specialized in feature extraction with the purpose of improving the input of the network
- Improving the precision in the recognition of the network.

The remaining contents of the paper are constructed as follows: Sect. 2 provides the detailed description and methodology of the proposed work; Sect. 3 presents the results and discussion; and Sect. 4presents finally the conclusion.

2 Proposed work

The surface electromyography signals (EMGs) are one-dimensional patterns, so any technique of extraction features is applicable to this type of signal. The signal acquisition is performed using the My Signals HW V2 card, which obtains a continuous signal that expresses the behavior of the muscle by voltage variation before an involvement as a function of time. My Signals HW V2 has the myoelectric signals with hand gesture that generates the probability for controlling the hands using the extraction motion through EMG signals. It permits to measure the muscle electromyography signals, it has the sensor data acquisition through several devices, and it is able to interact with the devices [25].

The wavelet discrete transformation is utilized to identify the feature extraction behavior, and the types of gestures are classified as the close fingers, wave-in, fist and gun gesture. Thus, to develop an algorithm capable of recognizing hand gestures based on EMG signals, the electrodes should be placed in the muscles where hand involvements predominate using the methodology presented in Fig. 1, which shows the steps of capture of the signal and its digitalization to a computer. The signal acquisition process generates the EMG signals using muscular contradiction that the classification of the hand gestures has been completed. The signal preprocessing procedure eliminates the noise according to the real characteristics. The wavelet packet transform decomposes the low frequencies to provide the quality signal, the time frequency plane through the rectangles of dissimilar sizes. The pattern recognition has been successfully completed using the R-CNN and creates an effective framework. The training accuracy and confusion matrix are used to analyze the efficiency of the proposed technique.

Processing by MATLAB for feature extraction and the training of the R-CNN is evaluated by means of confusion matrices. These steps are described as follows.

2.1 Signal acquisition

The EMG signals are generated by muscular contraction, which leads to an analysis of the muscular region involved in the movements caused by the gestures to be classified. Therefore, it is necessary to make a correct placement of the electrodes, shown in Fig. 2, and it must have a reference electrode or ground that should be located in a place where it is unaffected by the gestures. It is important to recognize that the EMGs signals are gathered by non-invasive surface electrodes, made up of electrodes made up in electrolytic gel layer that is responsible for increasing the conductivity and achieving a better adhesion on the skin. The layer is called as the region of the electrode that generations are dissimilar from their initial value. It is important to recognize that the EMGs signals are gathered by noninvasive surface electrodes, made up of electrodes made up in electrolytic gel layer that is responsible for increasing the conductivity and achieving a better adhesion on the skin.

Fig. 2 The location of the 2 electrodes that acquire the signal

2.2 Hand gesture recognition

According to the muscular group chosen, a group of gestures are set that will be recognized and classified. A previous analysis of these is important to establish which movements are associated with the electrode locations that will allow a better extraction of characteristics. The 3-dimensional hand gestures are used to dynamically analyze the time-based hand skeleton generation. It demonstrates the hand gesture with the shape. For every frame in the sequence, the location in the camera space of coordinates is generated using the frames. The frame for 3-dimensional row vectors is computed in Eq. (1)

$$Fr_t = \left[\alpha_1(t)\beta_1(t)\gamma_1(t)\ldots\alpha_j(t)\beta_1(t)\gamma_j(t)\right] \tag{1}$$

where $\alpha_1(t), \beta_1(t), \gamma_1(t)$ are the vectors for the particular frame time t. The total amount of frames T_{fr} is represented within the matrix, and it is computed using Eq. (2)

$$T_{fr} = \begin{bmatrix} \alpha_1(t) \\ \vdots \\ \gamma_j(t) \end{bmatrix} \tag{2}$$

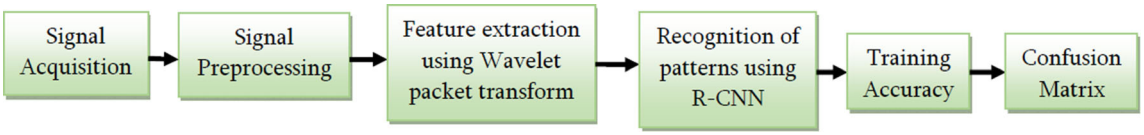


Fig. 1 Block diagram for the proposed methodology

The data are computed for the motion and the hand shape with the sequence of joints to rotate the hand with dissimilar features.

The gestures are demonstrated the hand movement in the space so that every frame with the set of location of the palm joint is denoted as $joint_{palm}$. The direction $\overrightarrow{dir_x}(t)$ is computed using Eq. (3)

$$\overrightarrow{dir_x}(t) = \frac{joint_{palm}(t) - joint_{palm}(t-\varphi)}{\left|\left|joint_{palm}(t) - joint_{palm}(t-\varphi)\right|\right|} \tag{3}$$

where φ is constant for the computed value. The vector direction is computed using the separation from the normalized form.

For generating total amount of frames T_{fr}, we need to find the generated direction function δ_{Dir} using Eq. (4)

$$\delta_{Dir} = \left\{\overrightarrow{dir_x}(t)\right\}_{[1<t<T_{fr}]} \tag{4}$$

The wrist rotation within the gesture is demonstrated by the way the hand is moving. For every frame, we calculate the write-based vector to the palm to generate the rotation data in 3 dimensions using Eq. (5)

$$\overrightarrow{dir_{rot}}(t) = \frac{joint_{palm}(t) - joint_{wrist}(t)}{\left|\left|joint_{palm}(t) - joint_{wrist}(t)\right|\right|} \tag{5}$$

To generate the total amount of frames T_{fr}, we need to find the generated rotation function δ_{Rot} using Eq. (6)

$$\delta_{Rot} = \left\{\overrightarrow{dir_{rot}}(t)\right\}_{[1<t<T_{fr}]} \tag{6}$$

2.3 Signal preprocessing

When analyzing EMG signals ESTR, it is necessary to perform preprocessing to eliminate noise in order to extract the real characteristics of the same for its classification. A case of success in the preprocessing stage is presented, applying the wavelet denoising algorithm technique that allows it to achieve an increase in the precision of recognizing EMG signals between 50 and 60. Next, the basic concepts of this technique that are basis for noise suppression in this research are presented.

2.4 Feature extraction using discrete wavelet transform

2.4.1 Wavelet packet transform (WPT)

The WPT, developed by a modification of the wavelet transform (WT), is responsible only for decomposing the signal approximations (low frequencies). In other words, if the most important features of the signal are in the details (high frequencies), this method could cause loss of information. The WPT consists of passing the signal through several levels, which is recommended to determine the signal quality.

It is important to emphasize that the first level of the approach obtains new details and approximations, as Fig. 2 presents, thus achieving a better resolution in frequency which, in turn, allows an increase in the signal features' quality.

The highly rated EMG signals were crumbled using the kernel-based reimbursement methodology. The creation of EMG signals is done with a multi-channel environment, a convolutional mixture of a group of signals. The mixture process in the form of a matrix is demonstrated in Eq. (7)

$$\omega(n) = \aleph t + \rho(n) \tag{7}$$

where $\omega(n) = [\omega_1(n), \ldots \omega_M(n)]^T$ is the EMG signals for M multi-channel, n is the value for generating discrete time in a particular point, $\rho(n)$ is additive noise $\aleph t$ is the vector value. The separation vector $Se_j(n)$ is computed using Eq. (8)

$$Se_j(n) = wt_j^T \omega(n) \tag{8}$$

The weighted value of the gradient descent vector wt_j^T is used to compute the separation vector. The decomposition process is implemented to identify every element. The elements are combined, and the pulse noise ratio is utilized to measure the performance with the correlated decomposition, based on which its accuracy is evaluated.

The Daubechies function [26] with every symmetric wavelets shares the common properties with the conventional base. The scaling function needs to construct the coif lets that the parameter is computed from the feature coefficients. The numerical solutions of the equations within the specified interval have been produced that the wavelet system contains the linear spine with similar parameters. The quadrature points are evaluated with all kinds of kernels, the convergence rates through the linear spine. The mother wavelet is a necessary function to implement the WPT. Its choice is made by identifying the best correlation that it has with the signal under study. It is associated with the mother wavelet function Daubechies (dB), which is defined from dB 1 to 45. This order is determined by the behavior of the function; for example, the behavior from dB 2 to 10. This function is mainly applied in low-order configurations (dB 1–20). According to the above, the Daubechies function is chosen with an order 7 (dB 7), according to the correlation that it has with the EMG signal.

The sequence of generated data is demonstrated using connected joints. Hand gestures are received from sensors containing 3-dimensional coordinates for the joint hand to

be represented in a coordinate methodology. Hence, the rotation of the hand is related to the camera. A description is used for the hand geometric transfer to demonstrate the hand shape, and we can analyze dissimilar hand sizes within the performers. The average size of every hand bone is performed in all hands from the dataset. In spite of maintaining the consistency for the rotation-based transformations, we generate a hand-based reference and skeleton-based rotation transformation with corresponding vector is used to compute the root joint. We demonstrate an origin of the root joint that embraces what has the wrist node vector $\overrightarrow{wn}$, and the cross-product is generated using Eq. (9)

$$\overrightarrow{node_B} = \overrightarrow{wn}\,\overrightarrow{xt} \tag{9}$$

The computed base is rotated to design a new reference base $Base_0$ from the hand gesture. The computation of the optimal rotation within the 2 bases $Base_0$ and $Base_1$ of the present skeleton is performed by the decomposition methodology. This entire procedure in a new hand has been measured while maintaining its shape facing the camera. For every gesture sequence, we calculate the rotation of the initial hand skeleton according to the hand formation and again apply the same kind of rotations to other hand sequences. This assures the invariance of the representation to the location of the hand.

This package is responsible for processing the coefficients provided by the WPT where it performs the following processes: first, it extracts the coefficients from the WPT terminals and converts them to absolute values. Second, it determines the length of time corresponding to each wavelet packet coefficient, and it fills the time slots between neighbors by creating a vector length equal to the first level of the wavelet packet tree object.

The WPT provides a time–frequency decomposition of a signal in the manner Fig. 3 illustrates. The time–frequency plane with rectangles of different sizes means that energy components of the signal within different rectangles of specific time and frequency coordinates can be discerned. Accordingly, the rectangles are an indication of the optimal resolution achieved by the time–frequency. The WPT stipulates an octave-based decomposition of the frequency domain and gives a good frequency resolution in the lower frequencies. The WPT also provides a useful time resolution in the higher frequencies that worsens as it passes through to lower frequencies.

2.4.2 Framework of an R-CNN

An R-CNN framework is mainly designed for the recognition of patterns in images. However, gestures from EMG signals present a greater difficulty in not having defined

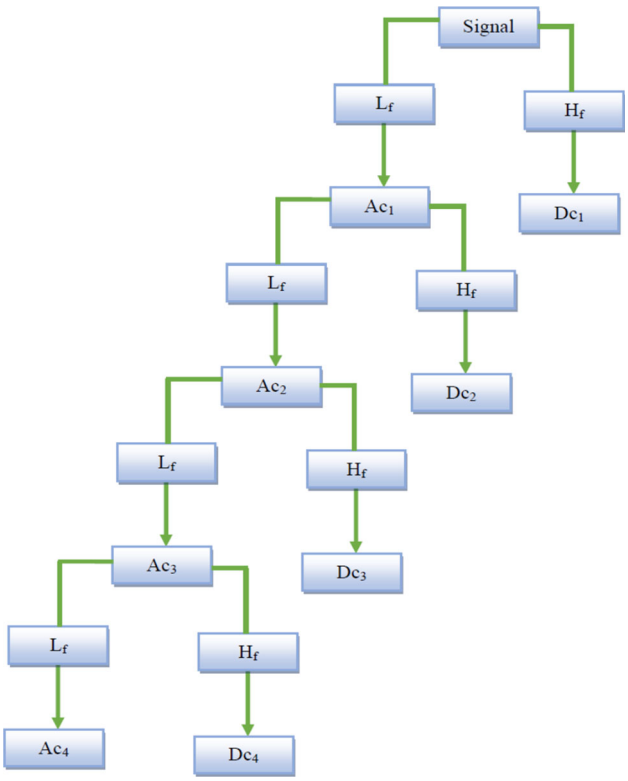


Fig. 3 Wavelet decomposition tree

parameters, such as dimension and object, to be recognized; instead, they require working with much more complex patterns that make it difficult to create an efficient framework in comparison with the high efficiency of the R-CNN oriented images.

2.4.3 Dataset

The network to be implemented must be able to recognize the 4 gestures, i.e., each one is a classification category in the network output for this. The EMG signal is acquired for a time interval of 1 s, which is sufficient time to contain the information necessary to test the performance of the CNN. The optical flow frames are computed from the adjacent frames to produce the RGB frames through the additional channels, and every frame is fed into RCNN to identify the feature vector generating the temporal segments. The extracted features are appended into the fusion layer and transformed into the fully connected layers to acquire the desired accuracy. Jester dataset [27] is the hand gesture recognition-based dataset; it is a huge collection of clips that demonstrate the identifying predefined hand gestures through the laptop camera or webcam through 30 fps frame rate, total amount of 138,082 gesture videos with 27 classes.

The dataset is composed of 10 dissimilar subjects with 5 women and 5 men, and the dataset elements are resized to

50 × 50 × 1 binary format for preprocessing data that the reduced amount of training data is utilized for data augmentation process. The hand gestures are performed as a sequence of specified period of time in a dynamic manner as the gestures are categorized as the shape, orientations, the flex angles of the finger, position and the motion speed with direction.

The different kinds of gesture recognition through a Jester dataset is that the neural networks are trained on that data samples. The end-to-end solution is constructed that performs on several camera-based platforms. This process is completed to generate the gesture recognition using RGB camera, and the dataset is also joining with several cases like different lighting backgrounds with user camera distance change, the taking process speed of 30 fps that the length of the frames up to 45.

2.4.4 Construction of R-CNN

R-CNN is the combined structure of a convolutional neural network with scalable detection methodology. The feature extraction for every process of the proposed methodology and the computation of R-CNN is very high in real time. VGG16 network is used to classify the objects and execute the feature extraction process quickly. The feature extraction is needed as the methodology of allocating convolution to minimize the required time period for detecting the objects. The region proposed network (RPN) provides the objects' location. The convolutional features are used to correctly detect and identify the objects in the proposed methodology. Figure 4 demonstrates the R-CNN [28] architecture model. VGG16 is the CNN model which achieves more than 93% of test accuracy from the dataset; it is the enhanced framework than AlexNet by using the kernel-sized filters. The Maxpooling is implemented through 2 × 2 pixels with stride 2. The fully connected layers are initially having 4096 channels, and final layer is called the Softmax layer having 1000 channels. The same level of configuration is performed in every network.

The RPN is the primary process for constructing an R-CNN combined with a fully convolutional network; the sliding window-related target identification is used to divide the RPN into smaller networks to develop the convolution feature map with the convolutional layer. Figure 5 demonstrates the output. The input for the network is specified using a spatial window for every position of the convolutional feature map. The spatial window size is described as 3 × 3, such that every sliding window has the feature with 256-dimensional ZFNet and 512-dimensional VGGNet-16 combined with linear rectification function (ReLU). The feature of the input is divided into layers of regression and classification. Hence, the minimum amount of network generates the output through the sliding window method; the fully connected layers could be distributed to every spatial position and produced as $n \times n$ convolutional layers. The multiple multi-scale parameters of every sliding window are positioned in the region.

The frame coordinates for producing the output for the classification layer to demonstrate every region of the proposed system in the object classification. The parameter-based reference boxes are used to simulate the scales in relevant ratios. A convolutional feature map is constructed with $n \times n$ dimensions to produce the variance of the object classification. Whenever the object is converted into an image, the function is used to identify the classification. The image conversion is used to predict the invariance of the convolutional features that are calculated for a given image using the R-CNN functionality. The multi-scale parameters are used to share the features within a specified time period. Figure 6 demonstrates the process of an R-CNN pertained network model.

According to the common concerns, the R-CNN is constructed with 3 kinds of layers: input, middle and final. While constructing the input layer, the lowest detection is measured for an RGB image using 32 × 32 pixels. The middle layers play a vital role in producing the convolutional layers with ReLU layers. The kernel size and filters are identified for pooling to discover the highest value of other layers using the fully connected layer and

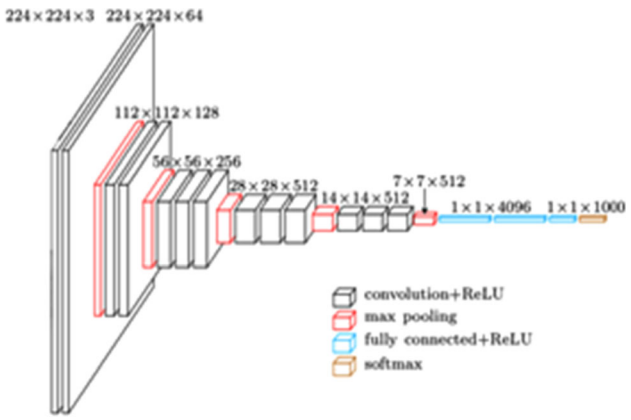


Fig. 4 R-CNN

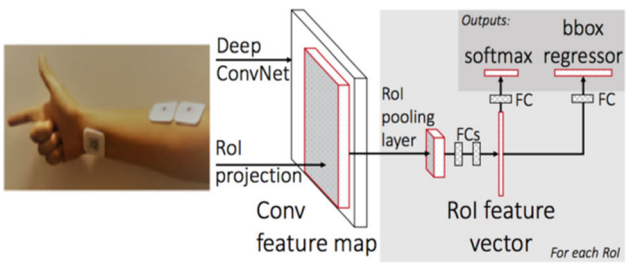


Fig. 5 R-CNN pertained network model

classification layer. The output layers have the nonlinear ReLU layers identified to confirm that the output image has the objects obtained from a Softmax layer with classification. The feature extraction for training the optimization methodology with an optimizer is constructed using the convolutional neural networks.

The proposed R-CNN is developed to utilize the large amount of regions using convolutional neural networks. The CNN has produced 4096-dimensional highly densed layer with the features extracted from the input image, and the output is classified using the regions. The proposed methodology is constructed to implement the process within the minimized amount of time. The ROI pooling layer is used to reconstruct the regions into a common fixed size with fully connected layers. The working process is demonstrated in Fig. 7.

Table 1 demonstrates the symbol description with meaning for this proposed methodology.

3 Results and discussion

The feature extraction for implementing the R-CNN is used to analyze the optimization methodology. The training-based method is used to find the gradient elements using convolutional neural networks. GPU cache is used to identify the coefficient for maximizing the optimization. The epochs for every learning method are used for fast convergence and stability. The proposed methodology is used to reduce the loss value and obtain the maximum accuracy in a limited period. The pretraining structure is implemented to initialize the RPN for every convolutional

Pseudocode for R-CNN

Input: Training data $\delta = \{\delta_1, \delta_2, \ldots\ldots, \delta_k\}, k \in class$

Feature vector $FV = \{FV_1, FV_2, \ldots\ldots, FV_n\}$, the extracted feature vector is collected from the high-dimensional space

For i: 1→ k do

$Dis_i \leftarrow FV_i$, distance within the 2 points in class

$\{Pt_{i1}, Pt_{i2}\} \leftarrow \underset{i}{\operatorname{argmin}}\, Dis_i$, the closest point from Dis_i

$FV_i \leftarrow FV_i - \{Pt_{i1}, Pt_{i2}\}$, remove the closest points

$Loc_i \leftarrow out_i + bias_i$, location is obtained from the bias value

$Tar_{obj} \leftarrow Softmax\ (dis_i)$, target object is computed using Softmax function

$X_1 \leftarrow \{\alpha \,|dX_1 < Th_i\}$, threshold value

$FV_i \leftarrow Find_{max}(FV_i, X_{i+1})$

End for

while $FV_i \neq \emptyset$ do

$$Pt_i \leftarrow \bigcup_{j=1}^{m} X_j$$

$Obj_e \leftarrow \sum_{i=1}^{N} \left(R_k^i - \rho_k^T \emptyset(IF_i)\right)^2$, objective function

$pv_j^m \leftarrow fn_2(\beta_j^m \sum_j^{m-1} \alpha_j + bias_j^m$, pooling vector

$of_j^l \leftarrow fn_i \alpha_i^{l-1}\, ck_{ij}^l + bias_j^l$

end while

Output: $Pt \leftarrow \{Pt_1, Pt_2, \ldots., Pt_k\}$, set of all classes

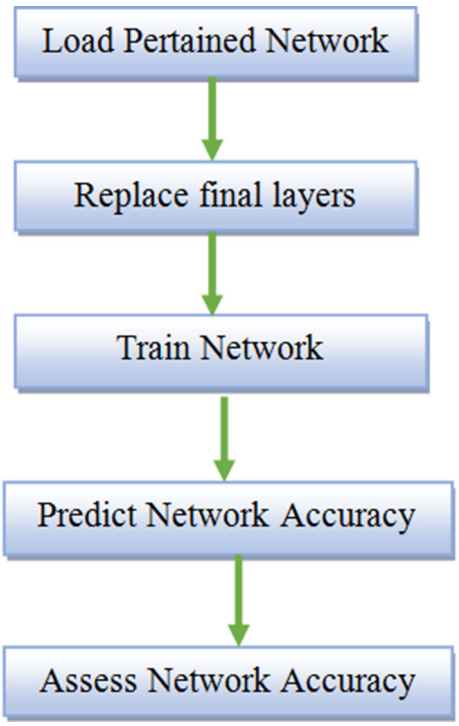


Fig. 6 Process of R-CNN pertained network model

layer to train the weight-based RPN. The proposed work is compared with the related methods of KNN [29], CNN [30] and SVM [31] with the parameter metrics of precision, recall, $F1_{score}$, accuracy [32].

The precision value is evaluated using the recall ratio that the proportion of objects within the image is detected using Eq. (10)

$$Pre_{val} = \frac{True_{Po}}{True_{Po} + False_{Po}} \tag{10}$$

The recall ratio demonstrates the proportion of images that consists of the objects which have been computed using Eq. (11)

$$Recall_{ratio} = \frac{True_{Po}}{True_{Po} + False_{Ne}} \tag{11}$$

The metric function is defined using Eq. (12)

$$F_1 = \frac{2Pre_{val} * Recall_{ratio}}{Pre_{val} + Recall_{ratio}} \tag{12}$$

The accuracy (Ay) is computed using the precision value and recall ratio in Eq. (13)

$$Ay = \frac{True_{Po} + True_{Ne}}{True_{Po} + True_{Ne} + False_{Po} + False_{Ne}} \tag{13}$$

The false discovery ratio demonstrates the incorrectly detected images within the total amount of images. It is computed using Eq. (14)

$$FD_{ratio} = 1 - Pre_{val} = \frac{False_{Po}}{False_{Po} + True_{Po}} \tag{14}$$

False omission ratio is computed using Eq. (15)

$$FO_{ratio} = \frac{False_{Ne}}{False_{Ne} + True_{Ne}} \tag{15}$$

First, wavelet discrete transform (WDT) is used to evaluate the behavior of the feature extraction. Applying the wavelet packet transform tree for each gesture produces graphs as illustrated in Fig. 8a: wavelet packet tree description for close fingers, Fig. 8b wavelet packet tree description for wave-in,

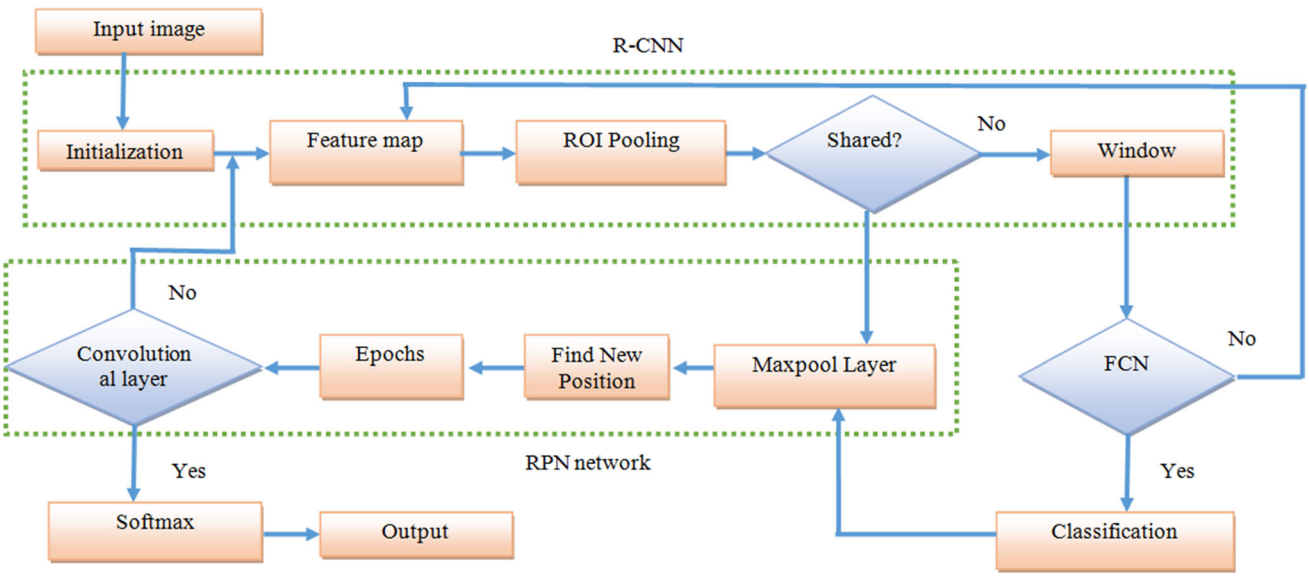


Fig. 7 R-CNN working process

Fig. 8c wavelet packet tree description for fist and Fig. 8d wavelet packet tree description for gun gesture.

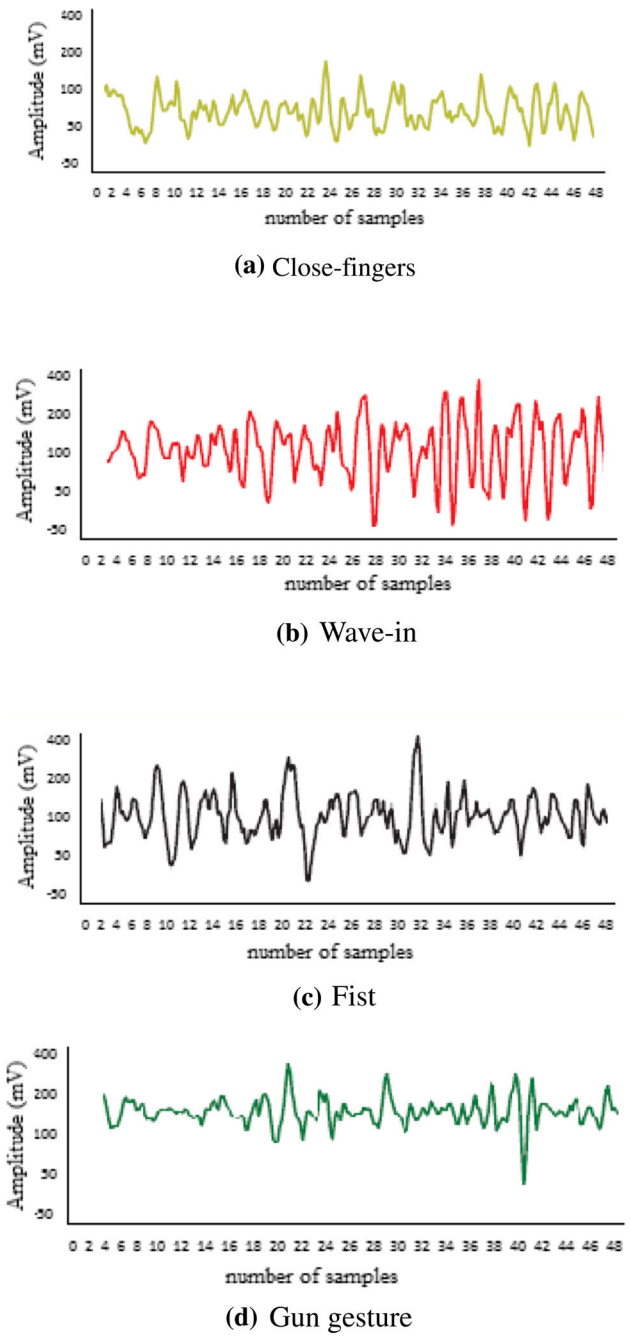
Second, after decomposing the signal and obtaining the approximation coefficients and the detailed coefficients, the packet spectrum is applied, which allows reconstruction of the signal from the coefficients and, in this way, determines the frequency bands in which the greatest number of features is accumulated. These are characterized in the range of purple tones, presenting a greater tone in the frequencies where there is greater activity. It also fades its color in the frequencies with lower activities until presenting a blue tonality that presents inactivity.

Third, the feature map is entered into the proposed network, and the network validation is performed with the "test" database, obtaining a 96.5% of accuracy, from where it is possible to observe the percentage of accuracy in the recognition of each one of the categories. According to the above, the lowest accuracy with a 91.8% for the "gun" gesture may be caused by 2 important factors: (1) with the "fist" gesture, by the similarity of the gesture in which simply a movement variation is done in 2 fingers of the hand, a similarity evidenced in the features map; (2) the gesture "Close-fingers" does not present similarity in the gesture, but there is activity of the same muscles with small variations of force when making each gesture. This similarity is also evidenced in the feature map.

Fourth, the performance of the framework is determined by evaluating the degree of membership with which the

Table 1 Symbol description

Symbol	Description
Fr_t	Frame for 3-dimensional row vectors
$\alpha_1(t), \beta_1(t), \gamma_1(t)$	Vectors for the particular frame time t
T_{fr}	Total amount of frames
$joint_{palm}$	Location of the palm joint
$\overrightarrow{dir_x}(t)$	Direction
φ	Constant for the computed value
δ_{Dir}	Direction function
$joint_{wrist}$	Location of the palm wrist
δ_{Rot}	Rotation function
$Signal$	Signal input
L_f	Low-pass filter
H_f	High-pass filter
Ac_1	Approximation coefficient
Dc_1	Detail coefficient
$\omega(n)$	EMG signals for multi-channel
n	Value for generating discrete time in a particular point
$\rho(n)$	Additive noise
$\aleph t$	Vector value
$Se_j(n)$	Separation vector
wt_j^T	Weighted value of the gradient descent vector
$\overrightarrow{wn}$	Wrist node vector
$Base_0$	New reference base
Pre_{val}	Precision value
$Recall_{ratio}$	Recall ratio
$True_{Po}$	True positive
$False_{Ne}$	False negative
F_1	Metric function
Ay	Accuracy
$True_{Ne}$	True negative
$False_{Po}$	False positive
FD_{ratio}	False discovery ratio
FO_{ratio}	False omission ratio



(a) Close-fingers

(b) Wave-in

(c) Fist

(d) Gun gesture

Fig. 8 Wavelet packet tree description

gesture was recognized, considering that the validation of the network is performed in real time by means of 5 repetitions per gesture, thus obtaining the percentage of accuracy and the recognition label of the gesture performed (Table 2).

Figure 9 demonstrates the confusion matrix for the four output classes—wave-in, gun, fist and close fingers—for the proposed methodology. The proposed framework is trained with a previously processed dataset composed of 180 samples per category for a total of 720 training samples. According to the above, it is important to note that the proposed frameworks reached 100% training accuracy. The main difference between the frameworks is that the configuration of the proposed framework does not use the linear rectification function (ReLU) among its convolution capable methods and, in turn, consists of three layers of fully connected networks at the end [33].

Considering the above performance analysis will use the successful proposed framework that shows the behavior of the training accuracy graph that begins to learn from the earliest epochs and reaches its first 100% accuracy around the 200th epoch, where it reached its stabilization in the 450 epoch having a suitable behavior to discard over fitting in training. However, training loss shows that effectively the network is learning with the passing of the epochs, and the behavior of the graph shows that the chosen learning rate is adequate. Figure 10 demonstrates the frequency for the wavelet packet indices.

Taking into account that it was obtained with an accuracy of $> 95\%$, the validation of the network is performed, i.e., a database is made under the same conditions with which the dataset training base was acquired. In this case it was with a smaller amount, assigned the name "test," and composed of 85 samples per category for a total of 340 samples.

The implemented validation technique corresponds to the confusion matrix, which is designed to evaluate the behavior of the network with a database different from that of training. This consists of rows corresponding to the predicted category (output class) and columns that show the corresponding real category (target class). The diagonal cells show the true positives of the category and finally, i.e., the correctly classified members of the category in the lower right comer, the overall accuracy of the network in the recognition of the database is shown.

The wavelet packet transformation utilizes the Daubechies function which is used to identify the exact correlation with the signal, and it is mainly applied with the behavior of the function. Figure 11 demonstrates the behavior from dB 2 to dB 7 according to the correlation with the EMG signal.

First, in order to obtain the best training behavior of each of the implemented networks, it is necessary to understand the graph training accuracy that allows us to determine in the first instance if the training can present over fitting to predict the essential elements. However, it is important to monitor the networks' behavior during training to discriminate training if the network is learning over the epochs. Figure 12a demonstrates this.

Table 2 Gesture recognition results

Tests	Close fingers (%)	Fist (%)	Wave-in (%)	Gun (%)
1	99.988	100	100	99.844
2	100	99.969	100	99.875
3	99.998	99.994	100	82.831
4	100	99.676	100	82.831
5	100	99.799	100	54.190

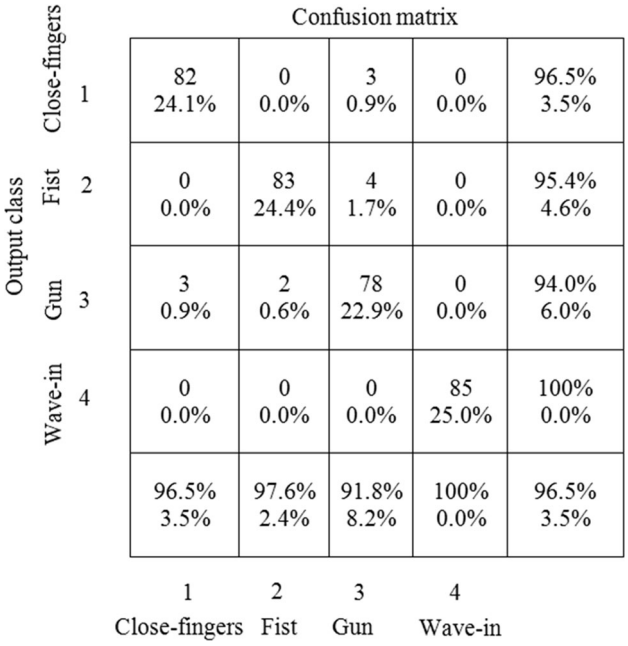


Fig. 9 Validation of the proposed methodology

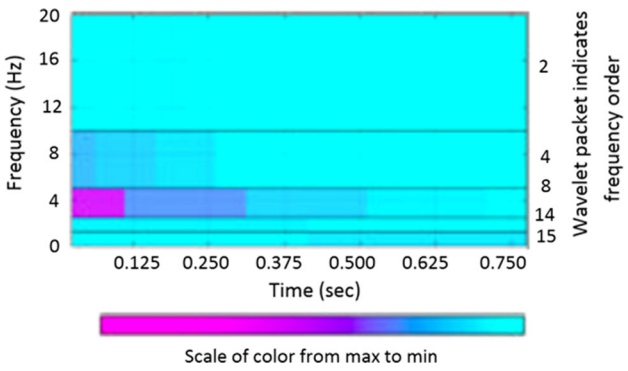


Fig. 10 Frequency for wavelet packet indices

Second, the behavior associated with the loss training should be analyzed. This allows us to determine in its first data if the chosen learning rate is adequate for the training.

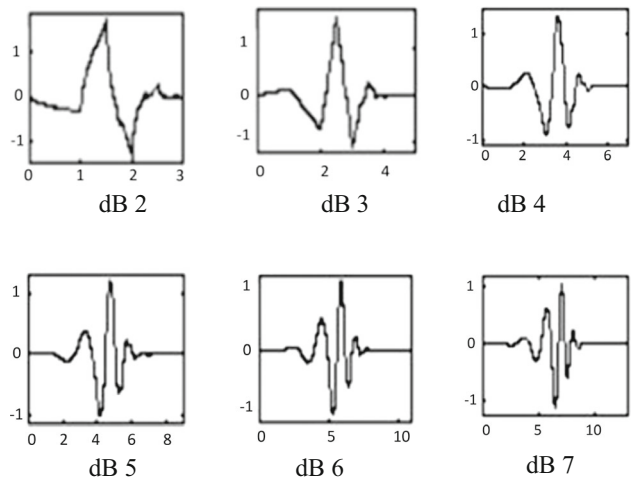


Fig. 11 Wavelet "Daubechies" family behavior

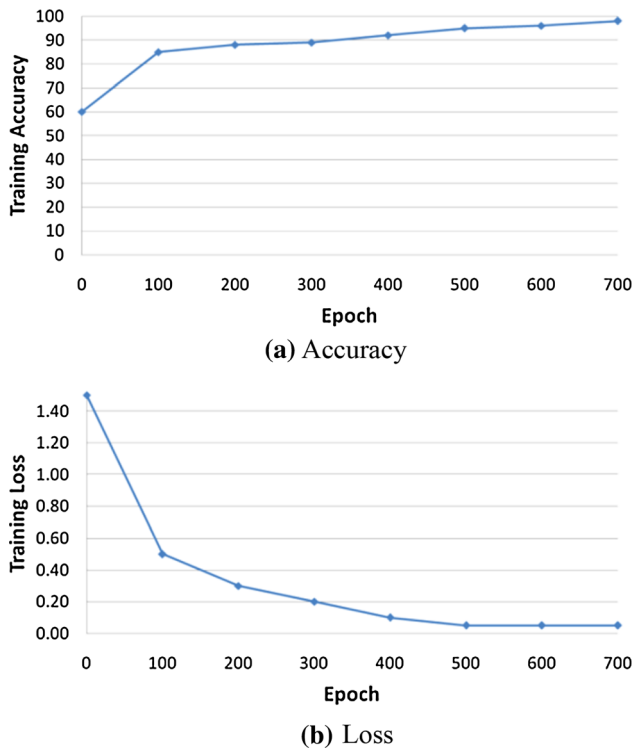
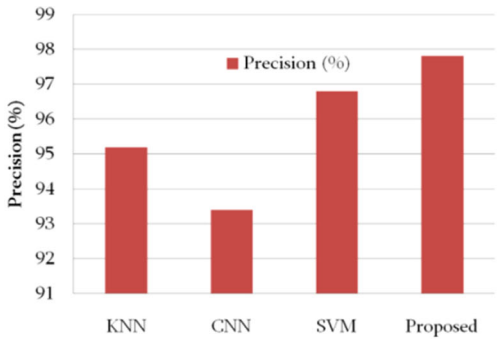


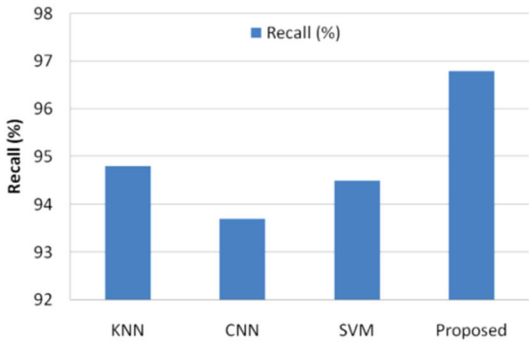
Fig. 12 Training accuracy and loss

This also allows us to validate if the network is actually learning or is experiencing a decrease in learning (loss). Figure 12b illustrates this.

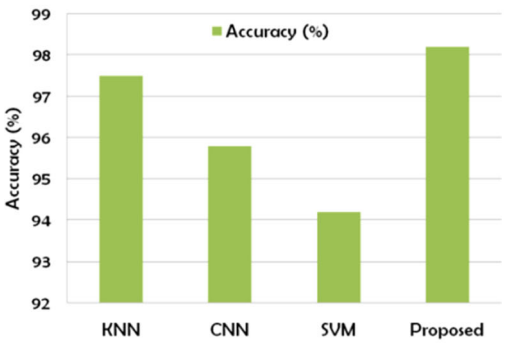
Figure 13a demonstrates precision %; Fig. 13b illustrates recall %; Fig. 13c denotes accuracy %; and Fig. 13d demonstrates $F1_{score}$ %. The performance results proved that the proposed methodology has improved in all the performance metrics. Precision is the rate of accurately predicted positive values to the total amount of positive



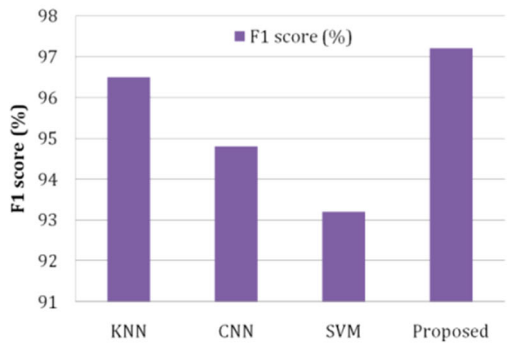
(a) Precision %



(b) Recall %



(c) Accuracy %



(d) $F1_{score}$ %

◀**Fig. 13** Performance evaluation

values. Recall is the rate of accurately predicted positive values to the entire values in the absolute class. Accuracy is the most important parameter measurement that the rate of accurately predicted values is from the total amount of values. The highest amount of accuracy will prove that the proposed work has the enhanced performance. F1 score is the weighted mean value of the precision and the recall value. Hence, there are values of the false-positive and false-negative values while performing the class distribution.

Considering the results obtained, convolutional neural network frameworks for the recognition of gestures can be implemented in applications such as limb rehabilitation, neuromuscular disease recognition, protein device construction or man–machine interaction for the control of a tele-operated robot.

4 Conclusion

The results obtained through techniques based on deep learning for the recognition of patterns have revolutionized several fields of engineering. For this reason, this technique is applied to the processing of signals, more specifically to EMG signals, allowing the development of much more efficient and functional devices. For example, it contributes to the development of protein devices or robots focused on tele-operation.

The proposed framework is novel, since it was not implemented in previous works, and satisfactory results were obtained for the proposed framework by reaching a $> 90\%$ percent accuracy. However, it is important to emphasize the technique of feature extraction that was implemented, since this technique allows evaluating in a more detailed way the behavior of the signal in different frequency bands, managing to capture the minimum details that allow differentiating one signal from the others.

However, the implementation of the proposed methodology has allowed the evaluation of 2 important points: First, the function ReLU theoretically converts the negative values to 0. In this case, for signals where its feature map does not contain negative values, but is very close to 0, it would be recognizing them as 0, generating a significant loss of the information evidenced in the network's efficiency. Second, there is importance of being fully connected in this type of framework, since the similarity of signal behavior is high and success is in the greater recognition of characteristics that differentiate one signal from the other. It was necessary to implement 2 fully

connected networks specifically to learn a greater number of characteristics and a third one for connection with the Softmax and final classification.

Authors contribution VS contributed to conceptualization, data curation, validation, formal analysis, writing—original draft, HRY helped in writing—original draft, writing—review and edition, supervision, conceptualization, MSK contributed to conceptualization, data curation, validation, supervision, MK involved in conceptualization, data curation, validation, supervision and AHG helped in conceptualization, data curation, validation, supervision.

Compliance with ethical standards

Conflict of interest The authors declare that they do not have any conflict of interests. All authors have checked and agreed the submission.

Human and animal rights statement This research does not involve any human or animal participation.

References

1. Shi Q, Zhang Z, Chen T, Lee C (2019) Minimalist and multifunctional human machine interface (HMI) using a flexible wearable triboelectric patch. Nano Energy 62:355–366
2. Sahin Y, Erdogan E, Gucluoglu T (2019) Performance analysis of two way decode and forward relay network with maximum ratio transmission. Phys Commun 32:75–80
3. Plewa K, Samadani A, Orlandi S, Chau T (2018) A novel approach to automatically quantify the level of coincident activity between EMG and MMG signals. J Electromyogr Kinesiol 41:34–40
4. Rubio-Acosta E, Solano-González J, FabiánGarcía-Nocetti D, Fuentes-Cruz M (2019) On the truncation of time frequency distributions to improve the computational performance in the estimation of fundamental parameters of a Doppler ultrasound blood flow signal. Biomed Signal Process Control 54:101608
5. Sotoodeh A, Weber JE (2019) Tele-EEG: Technik und Anwendung. Das Neurophysiologie-Labor 41(3):121–141
6. Jiang S, Gao Q, Liu H, Shull PB (2020) A novel co-located EMG-FMG-sensing wearable armband for hand gesture recognition. Sens Actuators A 301:111738
7. Zhang J, Ling C, Li S (2019) EMG signals based human action recognition via deep belief networks. IFAC PapersOnLine 52(19):271–276
8. Márquez-Figueroa S, Shmaliy YS, Ibarra-Manzano O (2020) Optimal extraction of EMG signal envelope and artifacts removal assuming colored measurement noise. Biomed Signal Process Control 57:101679
9. Pasiakos SM, Berryman CE, Philip Karl J, Lieberman HR, Rood JC (2019) Effects of testosterone supplementation on body composition and lower-body muscle function during severe exercise- and diet-induced energy deficit: a proof-of-concept, single centre, randomised, double-blind, controlled trial. EBioMedicine 46:411–422
10. Zhang L, Wang J, Chen J, Chen K, Fang X (2019) Dynamic modeling for a 6-DOF robot manipulator based on a centrosymmetric static friction model and whale genetic optimization algorithm. Adv Eng Softw 135:102684
11. Chen D, Li S, Qing W, Luo X (2020) Super-twisting ZNN for coordinated motion control of multiple robot manipulators with external disturbances suppression. Neurocomputing 371:78–90
12. Cipriani C, Zaccone F, Micera S, Carrozza MC (2008) On the shared control of an EMG-controlled prosthetic hand: analysis of user–prosthesis interaction. IEEE Trans Rob 24:170–184
13. Schinkel-Ivy A, Drake JD (2019) Interaction between thoracic movement and lumbar spine muscle activation patterns in young adults asymptomatic for low back pain: a cross-sectional study. J Manip Physiol Ther 42(6):461–469
14. Taya M, Amiya E, Hatano M, Maki H (2019) Inspiratory muscle training for advanced heart failure with lamin-related muscular dystrophy. J Cardiol Cases 20(6):232–234
15. De la Torre-Gutiérrez H, Pham D (2019) A control chart pattern recognition system for feedback-control processes. Expert Syst Appl 138:112826
16. Obuchowski NA, Bullen JA (2019) Statistical considerations for testing an AI algorithm used for prescreening lung CT images. Contemp Clin Trials Commun 16:100434
17. Atzori M, Cognolato M, Muller H (2016) Deep learning with convolutional neural networks applied to electromyography data: a resource for the classification of movements for prosthetic hands. Front Neurorob 10:1–10
18. Tian C, Yong X, Zuo W (2020) Image denoising using deep CNN with batch renormalization. Neural Netw 121:461–473
19. Ai H, Tang K, Han L, Wang Y, Zhang S (2019) DuG: dual speaker-based acoustic gesture recognition for humanoid robot control. Inf Sci 504:84–94
20. Pliinyomark A, Pliukpattaranont P, Limsakul C (2011) Wavelet-based denoising algorithm for robust EMG pattern recognition. Fluct Noise Lett 10:157–167
21. Li X (2020) Human–robot interaction based on gesture and movement recognition. Signal Process Image Commun 81:115686
22. Enoka RM (2019) Physiological validation of the decomposition of surface EMG signals. J Electromyogr Kinesiol 46:70–83
23. Englehart K, Hudgin B, Parker PA (2001) A wavelet-based continuous classification scheme for multifunction myoelectric control. IEEE Trans Biomed Eng 48(3):302–311. https://doi.org/10.1109/10.914793
24. Anonymous (2017) MySignals HWv2-eHealth and medical IoT development platform for arduino. Libeliwn Communications Distribuidas S.L., Zaragoza, Spain. https://www.cooking-hacks.com/mysignals-hw-ehealth-medical-biometric-iot-platform-arduino-tutorial/
25. Franţi E, Milea L, Buţu V, Cismas S, Lungu M, Şchiopu P, Barbilian A, Plâviţu A (2012) Methods of acquisition and signal processing for myoelectric control of artificial arms. Rom J Inf Sci Technol 15:91–105
26. Sarode TK, Agrawal P, Deshpande G, Jogeshwar A (2016) Hand gesture recognition by Daubechies wavelet transformation. In: International conference & workshop on electronics & telecommunication engineering (ICWET 2016), Mumbai, pp 1–6. https://doi.org/10.1049/cp.2016.1133
27. https://www.twentybn.com/datasets/jester/v1
28. Quan L, Feng H, Lv Y, Wang Q, Zhang C, Liu J, Yuan Z (2019) Maize seedling detection under different growth stages and complex field environments based on an improved faster R-CNN. Biosyst Eng 184:1–23
29. GuimarãesPedronette DC, Weng Y, Baldassin A, Hou C (2019) Semi-supervised and active learning through manifold reciprocal kNN graph for image retrieval. Neurocomputing 340:19–31
30. Lim KM, Tan AWC, Lee CP, Tan SC (2019) Isolated sign language recognition using convolutional neural network hand modelling and hand energy image. Multimed Tools Appl 78(14):19917–19944

31. Cao J, Wang S, Wang R, Zhang X, Kwong S (2019) Content-oriented image quality assessment with multi-label SVM classifier. Signal Process Image Commun 78:388–397
32. Velandia NS, Moreno RJ, Hernández RD (2017) CNN architecture for robotic arm control in a 3D virtual environment by means of by means of EMG signals. Contemp Eng Sci 10(28):1377–1390
33. Sampath P, Packiriswamy G, Pradeep Kumar N, Shanmuganathan V, Song O-Y, Tariq U, Nawaz R (2020) IoT Based health—related topic recognition from emerging online health community (med help) using machine learning technique. Electronics 9:1469