EMG- BASED HAND GESTURE RECOGNITION USING DEEP LEARNING AND SIGNAL-TO-IMAGE CONVERSION TOOLS

SABRINA AKTER, BIMAL KUMAR PRAMANIK, MD EKRAMUL HAMID*

Department of Computer Science and Engineering, University of Rajshahi, Rajshahi-6205, Bangladesh

Abstract: In this paper, deep learning-based hand gesture recognition using surface EMG signals is presented. We use Principal component analysis (PCA) to reduce the data set. Here a threshold-based approach is also proposed to select the principal components (PCs). Then the Continuous wavelet transform (CWT) is carried out to prepare the time-frequency representation of images which is used as the input of the classifier. A very deep convolutional neural network (CNN) is proposed as the gesture classifier. The classifier is trained on 10-fold cross-validation framework and we achieve average recognition accuracy of 99.44%, sensitivity of 97.78% and specificity of 99.68% respectively.

Keywords: EMG, deep learning, CWT, PCA, hand gesture recognition

1. INTRODUCTION

For human-to-human communication hand gestures are considered one of the most important modalities. Hand gesture recognition becomes a particular interest in human-computer interaction (HCI) research [1]. Recognition of hand gestures using surface electromyographic (sEMG) signals has also been employed in a wide range of other applications, such as powered electric, power wheelchairs [2], upper-limb prostheses [3], and diagnoses in clinical applications [4]. Electromyography (EMG) is considered the best modality of the various sensory modalities that have recently been used to record hand gesture information. This is because EMG captures a muscle's electrical activity; the physical phenomenon that produces hand movements. For hand gesture recognition many applications use noninvasive EMG signals. But the classification of EMG for gesture recognition is challenging because of their similar characteristics and intensities. Many studies use machine learning to classify EMG signals for gesture classification but their results are not very satisfactory. The reason for insignificant results in previous studies can be characterized as the lack of using appropriate and sophisticated techniques.

The analysis of surface EMG signals is more challenging compared to other well-studied biomedical signals [5]. For instance, for upper-limb myoelectric prosthesis control, many indicating issues have been presented that greatly influence the characteristics of the EMG signal and the performance of EMG-based pattern recognition systems. Very few of these challenges are the changes in the signal characteristics over time, fatigue muscle, electrode location shifts during data acquisition, and inter-subject variability. Muscle contraction intensity may vary over the same gesture which also makes EMG signal processing challenging [6]. Some studies show that a properly positioned smaller set of electrodes can improve or provide comparable recognition performance [7, 8]. Thus, there is little need to use all EMG channels to get perfect classification accuracy. However, proper positioning is difficult because the subject performs dynamic movements and the structure of the muscle patterns may vary with subjects. Therefore, there is, no consensus on the global optimum number of electrodes that can provide the maximum performance of gesture recognition. Even within the same experimental protocol, the

* Corresponding author, email: ekram_hamid@ru.ac.bd doi.org/10.29081/jesr.v29i2.001

optimal EMG electrode sub-set is highly subject-dependent [9] and this motivated us to think about how to reduce the number of electrodes and not worry about the electrode position. For that, we apply principal component analysis (PCA) to solve both the data reduction and electrode positioning problems.

Despite the fact that neural network-based machine learning algorithms have been implemented in EMG analysis for many decades, deep learning algorithms have also recently been applied to EMG-based gesture recognition. The reason for using traditional machine learning algorithms in the earlier years of the EMG-based gesture recognition field can be characterized as the lack of sufficient EMG data availability to train these deep neural networks. With the recent advancements in deep learning techniques and the availability of the shared bigger EMG data sets for addressing overfitting problems, the very recently developed deep learning algorithms and architectures have now been applied in EMG-based gesture recognition systems [10].

In machine learning and EMG-based gesture recognition technique, the raw EMG signal are segmented by the short window for feature extraction with the aim of improving the recognition rate. Several studies have reported that the quality and quantity of the manually extracted features have a great impact on the performance of EMG pattern recognition [6, 10-14]. The performance may be bad or good by chance. However, in deep learning algorithms, features are automatically created and unlike traditional machine learning algorithms explicit transformation of the raw EMG signals is not required. The deep learning algorithm, therefore, change the feature extraction approach from hand-craft feature engineering to automated feature engineering or learning. For this reason, in this study, we have employed a deep learning algorithm called “Inception v3”. This method has recently shown great performance on various image classification-based applications. Recently, in some studies, the manually extracted features are combined with the input data to feed into the input layer of a deep learning model and some benefits have been reported [15]. Here, we first convert the time domain sEMG signal into time-frequency represented images before putting them into the input layer of the convolutional neural network. For the time-frequency representation of EMG signal, we have used continuous wavelet transform (CWT).

In the hope of a better hand gesture recognition rate, in this paper, a very deep learning algorithm inception v3 was used which is mainly used in computer vision applications. Our contribution is to generate an idea for making input of the deep learning algorithm that can capture both temporal and spatial information with a reduced dataset. Principal component analysis (PCA) has been used for dimension reduction. Here we have introduced a technique on how to select principal components for the subsequent analysis. Later, to prepare the input image for the deep learning-based classifier we employed continuous wavelet analysis (CWT). Since, in practice, it is really difficult to obtain a high volume of EMG signal to properly train a deep learning-based model, in this work we have shown that using the transfer learning technique for the model’s parameter initialization and fine-tuning could be a good solution of model overfitting.

2. LITERATURE REVIEWS AND MOTIVATION

For prosthetic controlling and rehabilitation applications, the automatic recognition of hand gestures based on sEMG signals is considered a promising approach. Most of the work on sEMG-based gesture recognition is reported in the literature based on user-specific training of the classifiers. However, Inter-individual hand gesture recognition using sEMG signals is difficult due to differences in sEMG sensor placement, the noisy nature of sEMG signals, and inter-individual differences in gesture performance. To address the problem associated with gesture recognition using sEMG, so many researches are there in the literature using various approaches. In this section, some previous studies and works from the literature in the context of sEMG-based gesture recognition are addressed.

2.1. Deep learning in gesture recognition

A study shows that despite precisely controlling for electrode placement, sEMG signals can vary considerably between subjects [16]. It is found that regardless of classifiers that have been trained by a subject can be applied to new subjects achieving somewhat better than random performances with an accuracy of 85 % over 6 gestures. So deep learning can therefore be expected as a sophisticated technique for leveraging information between users for gesture recognition. A number of research has been done to find a projection of the feature space that bridges the hole between an original subject used in the training phase and a new subject [17]. Some studies have also proposed leveraging a pre-trained model prepared using data from some subjects that removes the need to simultaneously work with data from multiple subjects, especially for the model training purpose [18]. However, these techniques present important performance improvement with respect to the classification accuracies compared to their non-augmented versions.

Recently, convolutional neural network along with a single array and matrix of electrodes [19] has been started to be employed for hand gesture recognition. Other studies applied deep learning algorithms along with domain adaptation techniques [20] for inter-session classification but did not conduct experiments for inter-subject gesture recognition.

2.2. Fourier based features in gesture recognition

In the last decades, very few studies [21] used Short-Time Fourier Transform (STFT) and surface EMG data for gesture classification. A possible reason behind this little interest in Fourier-based features is that much of the research on sEMG-based gesture recognition focuses on model designing using feature ensembles. A large number of features are generated by STFT and adding this feature to other features makes the feature volume very high. Calculation of the Fourier-based features is also relatively computationally expensive. All these factors made it challenging to integrate STFT-based with other feature types. In addition, the Short-term Fourier Transform-based feature has also been proven to be less accurate than Wavelet Transforms-based feature [21] for the classification of biomedical (sEMG) data. However, the time-frequency representation of the STFT features, called spectrograms, have been applied as input data for the classification of surface EMG signals with the convolutional neural network [19]. Another study uses continuous wavelet transform features for EMG signal analysis [22], they only used it for EMG collected from the lower limbs [23]. Though, in the past, for sEMG-based hand gesture recognition wavelet-based features have been used, some studies tried to use the features based on the Discrete Wavelet Transform [24] and the Wavelet Packet Transform (WPT) [21] instead of the continuous wavelet transform. Discrete wavelet transform and wavelet packet transform are less computationally expensive than continuous wavelet transforms which makes DWT and WPT usable for fast classification. But none of the studies presented good classification accuracies. Besides, the computational power of the system has been growing exponentially which also makes high computation but more accurate methods popular.

However, similarly to STFT-based spectrograms, continuous wavelet transform presents an attractive time-frequency representation that makes convolutional neural network usable for sEMG-based gesture classification and can now be efficiently implemented on medical devices and prosthetic systems.

2.3. Motivation

The desired control of a prosthetic hand is complex and the effective control of multiple joints in a prosthetic limb and finger needs much attention and sophisticated usage of EMG. The ability to estimate dexterous motor intent is hindered by muscle co-activation, EMG cross-talk, and limited sampling depth. The complexities of muscle co-activation, EMG cross talk, and the contribution of deep muscle make the simple one-muscle one-function approach inappropriate to implement. This has motivated the use of a deep learning approach to gesture recognition which might be used for myoelectric control.

One can achieve control of many more classes of gestures by multidimensional classifiers and using multiple EMG sites and effective feature extraction techniques. The spatial arrangement in the multielectrode pattern-recognition approach should be taken special care of while cross-talk severely compromises conventional control. If the interfering EMG is not measured by another electrode it may even produce additional irrelevant information. The existence of muscle co-activation characterizes the natural synergistic behavior of muscles during a given contraction by providing spatial information to a pattern classifier. Thus, we are motivated to use PCA to calculate the principal components of the multielectrode EMG signal that can capture the co-activation characteristics. The overall flowchart of the analysis is provided in Figure 1.

3. MATERIALS AND METHODS

3.1. Data

In this work, well studied publicly available sEMG-based hand gesture recognition dataset [20] is used. The details data acquisition procedure is provided below. Eighteen healthy able-bodied subjects have been recruited for data collection. Their ages range between 23 to 26 years. They performed eight different types of gestures with the use of a non-invasive wearable acquisition device. Each electrode's (array of size 8×2) diameter is 3mm, and the horizontal distance between each electrode is 7.5mm, and the vertical distance is 10.05mm, shown in Figure 2(a). The electrode is wet and disposable, and the contact impedance is less than 3kΩ. The first acquisition module is located at the height of the radio-humeral joints on the extensor digitorum communis muscle, shown in Figure 2(b). The other modules are located clockwise, which formed an array of an 8 × 16 electrode. A band-passed filter is applied at 20–380Hz, and the sampling frequency is 1000Hz and sampled with a 16-bit A/C conversion. The signal is normalized to the range [-1 to 1], which corresponds to the actual voltage range -2.5mV to 2.5mV. An ARM controller is used to pack the 8 acquisition modules. The 8 performed hand gestures are 'Thumb up', 'Extension of index and middle, flexion of the others', 'Flexion of ring and little finger, extension of the others', 'Thumb opposing base of little finger', 'Abduction of all fingers',' Fingers flexed together in fist', 'Pointing index', and 'Adduction of extended fingers' and the gesture class is defined as the g1, g2, g3, g4, g5, g6, g7, and g8, respectively. This also provides us the opportunities to study high-density and sparse multi-channel sEMG data in gesture recognition. The duration of each gesture is 3-10 s and the same trial is repeated ten times. A resting posture of 7 s is used for gesture alteration in order to avoid fatigue. During the performance of gesture, subjects are not enforced a pre-defined contraction.

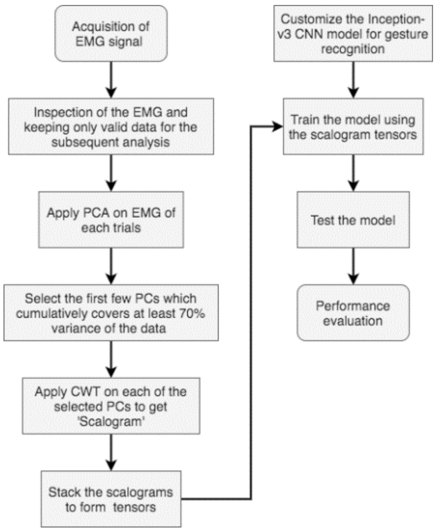


Fig. 1. The basic flow diagram of the proposed gesture recognition task.

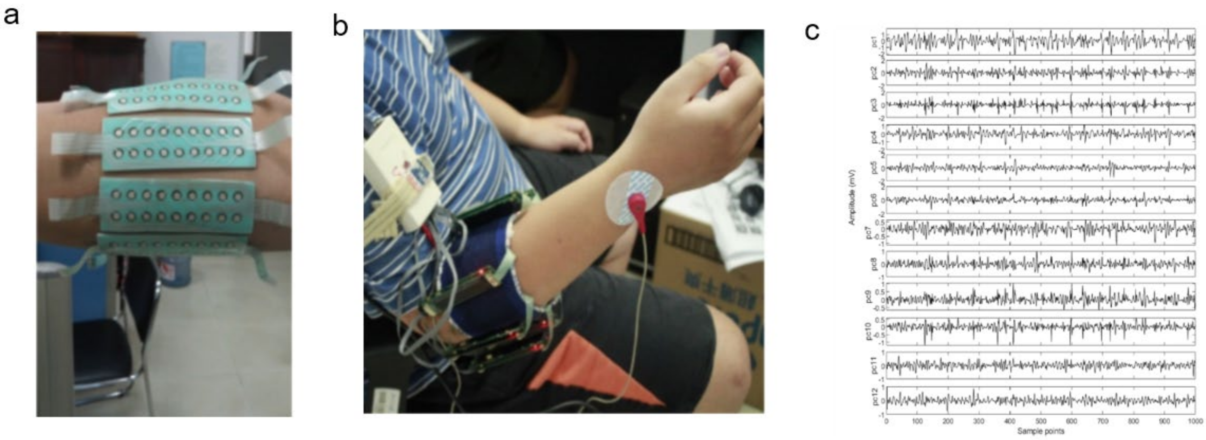


Fig. 2. Working methods:
a) The arrangement of the sEMG electrode arrays on the hand [20]; b) The electrode arrays with acquisition device wearing on right hand [20]; c) First 12 principal components of the sEMG data of a single trial.

3.2. Data reduction using principal component analysis

The Principal Component Analysis (PCA) [25] is a widely used tool for data reduction. The technique allows a smaller number of Principal Components (PCs) to capture a greater amount of the variance than the original variables, thus becoming a popular technique for data reduction. The technique helps to evoke structure for exploratory study without any a priori conceptions of the underlying structure. To minimize the dimensionality, PCA will be used to resolve the first objective of this study.

A covariance matrix is computed from EMG dataset, which consists of samples with variables that represent the channels. For extracting the number of PCs expressed corresponding to the significant eigenvalues are calculated from the elbow-plot of eigenvalues chosen. How many principal components account for the total variation in the data should be examined before selecting the PCs for the downstream analysis. One of the main objectives of using PCA is to replace the original features of the data with a very smaller number of principal components, however, it discards very little information. A major concern in selecting the number of principal components is how small the subset can be without serious information loss. Selecting a (cumulative) percentage of total variation to which one desires that the selected PCs contribute, say 80 % or 90 % is possibly the well-adopted criterion for selecting the number of principal components. The selected number of principal components is then determined by the smallest number of PCs for which the chosen threshold percentage is exceeded. This study considered at least 70 % percent of variation accounted for the first few principal components.

3.3. Scalogram using CWT

EMG signal being analyzed in this study is time-variant and non-stationary, thus only obtaining the frequency information by Fourier analysis is insufficient. Moreover, Continuous wavelet transform solves the window length-related problem of STFT by using the variable length window function. Thus, time-frequency analysis of the EMG signal wavelet becomes an indispensable technique. A wavelet is defined as a small wave. The energy of the wavelet is concentrated in time thus providing a useful tool for the nonstationary, transient, or time-variant EMG signal. The wavelet has oscillating wave-like properties but still has the ability to time and frequency analysis of the signal simultaneously with a robust mathematical foundation. The general idea of the wavelet transformation is about shifting and scaling a "mother wavelet" $\psi(t) \epsilon L_2(R)$, considering the family of functions as:

$$\psi_{a,b}(t) = \frac{1}{\sqrt{|a|}} \psi\left(\frac{t-b}{a}\right) dt \tag{1}$$

were $a, b \; \epsilon \; R(a \neq 0)$ and the equality, $\| \psi_{a,b}(t) \| = \| \psi(t) \|$ is ensured by the normalization (we assume for now that the parameter of the function a can be positive as well as negative).

3.4. Deep CNN for gesture recognition

CNNs are deep artificial network that automatically learns hierarchies of invariant features. CNNs are invariant, which means that at different locations the same transformation is applied and features are treated as distinctive attributes. A Convolutional neural network is widely used to process data that has a known spatial relationship or topology like grids, in other words, it has a very regular structure. It consists of multiple hidden layers in between the input and output layers, where the hidden layers can be convolutional, pooled or completely connected. Convolution operations are performed in CNNs, especially in the convolutional layers, and there is a kernel for each convolution operation. The kernel or filter is a lower dimensional matrix than the original image, usually 3x3 or 5x5 dimensional. The filter is gone over the image during the convolution operation to detect a specific character from the image. The filter is moved over the image with a given stride size. How many pixels the filter moves horizontally or vertically to the input image is defined by the stride. An output is generated in the form of a feature map, after the convolution is applied to the input image using a convolution filter. Because the filter is much smaller than the image, the size of the feature map is smaller than the input. One may use padding to maintain the dimensionality, that is, surround the input with padding. Two values are mostly used for padding, either with zeros or with edge values can be the padding. The size of the feature maps is preserved by using padding in the CNNs.

We have found Inception [26] more suitable for our problem of gesture recognition. First, the Inception v3 model shows very good performance in image classification on the ImageNet dataset. Second, it outperformed other very deep learning networks like VGG16, VGG19, Resnet-50 etc. Finally, our gesture recognition task has turned into an image classification problem after getting the CWT of the time domain PC of the EMG and finding Inception v3 applicable. For this reason, in this study, we use a very special CNN, Inception v3.

4. RESULTS AND DISCUSSION

4.1. Experimental environment details

Hand gesture recognition is performed using a surface EMG signal collected from a highly cited and well-studied public EMG database. The dataset is chosen so that the experimental results obtained in this study could be compared with the existing studies. After collecting data, it is very carefully examined to check if there are any distorted signals in the dataset. But we have not found any distorted data in the dataset, which made us easy to use the full volume of the dataset for the model training and testing for gesture recognition. Matlab R2018b is used for partial data analysis, calculating the time-frequency image that is the final input of the deep learning-based model.

Later on, Python is used to create, train, and test of deep learning algorithms. We use tensorflow library to build the deep learning model. The model training platform is built on a gaming laptop computer with Intel(R) Core (TM) i7-7700HQ (2.80GHz, 2808MHz, 4 Cores and 8 Logical processors) CPU, and NVIDIA GPU GeForce GTX 1060 6GB GPU and 16GB of RAM, running on a Windows 10 64-bit system. For the deep learning model building, training, and testing Python 3.5, CUDA 9.0 is used.

4.2. Principal components selection

The sEMG data are separated for every trial. Principal Component Analysis (PCA) is applied to each of the trails of the whole data set. There are 128 channels in the array electrode grid and the sampling frequency of the signal is 1000 Hz. For this reason, in each trial, there are total 128 × 1000 data points. After applying PCA we get 128 PCs. Figure 2 c) shows the plot of the first 12 PCs. However, all of the PCs do not carry very useful information about the original data. This fact encourages us to discard some of the principal components for the subsequent analysis. But now the question is how to discard the PCs that carry very insignificant information. From Figure 3 it is observed that the first PC covers 29.3 % of the total variance of the EMG data collected from a single trial. The second PC covers 23.3 % of the total variance of the EMG data collected from the same trial. The third PC covers 19.8 % of the total variance of the EMG data. As the indices of the principal components rise the variance of the EMG covered by each PC decreases. From the 9 PC, the variances are about zero. This indicates these PCs carry very insignificant information. Figure 4 shows the cumulative variances of the EMG covered by the principal components. The first three principal components covered more than 70% variance of the original data. To increase the cumulative variance even a little amount more than 70%, it needs to add many more PCs. For this reason, we set the threshold 70% for selecting the number of PCs for the use of the input of the deep learning model. As a result, the first three PCs have been chosen for the subsequent analysis of gesture recognition.

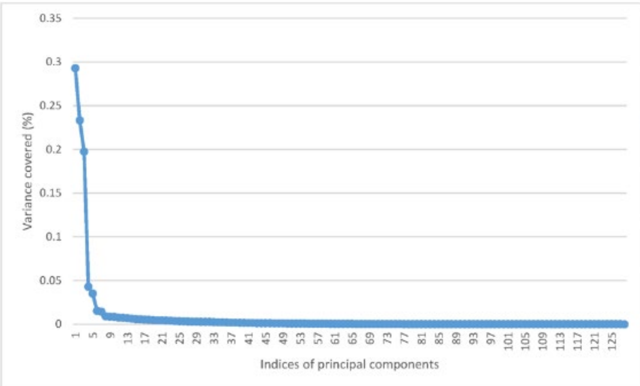


Fig. 3. Plotting of the variance covered (in percentage) by each of the principal components of the EMG signal.

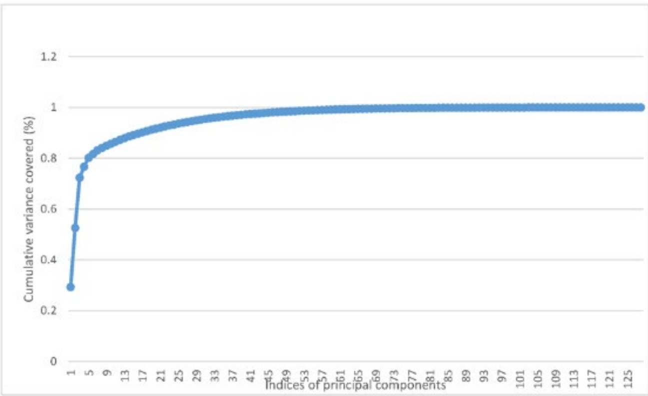


Fig. 4. Plotting of the cumulative variance covered (in percentage) by the principal components of the EMG signal.

4.3. Preparation of the input for the model

After the selection of the principal components, the signal-to-image conversion method has been applied to them and the obtained scalograms are used to create the input tensor of the deep learning model. We apply CWT to each principal component individually. The time-frequency represented images of the CWT, scalogram, are stacked to make the tensor. Figure 5 shows the overall procedure of the tensor creation.

The tensor consists of the scalogram of selected principal components. Three principal components have been selected for the subsequent experiments. Thus all three principal components are to the subject of the continuous transform. From Figure 5 it can be seen that scalograms of different principal components have different frequency patterns. These scalograms are used to make the multidimensional tensor, the input of the deep learning model.

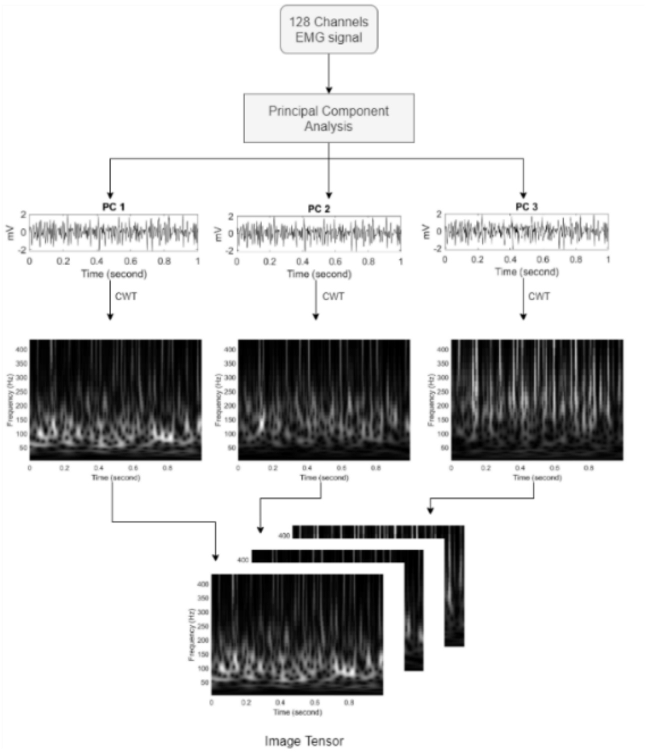


Fig. 5. The over procedure of the input tensor preparation for the deep learning model.

From Figure 5 it can be seen that the CWT is applied to the first three principal components individually. The obtained scalograms are stacked to create the tensor input.

There are 18 subjects, each subject performs eight different gestures and 10 trials for each gesture. We prepare the input tensor deep learning model for each trial, so we have total 1440 tensors. After splitting the total tensors with 10-fold cross-validation, we have total 1296 tensors in the training set and 144 tensors for validation and testing for each fold.

4.4. Performance measurement metrics

The accuracy of a classifier determines how accurately the model classifies the samples and in machine learning studies performance measurement is a crucial task. In quantitative measurement, we usually use accuracy, sensitivity, specificity, and f1 score. The accuracy of a classifier measures how correct the predicted values are. The sensitivity pattern classifier indicates how many of the correct results are produced.

The f1-score provides a balanced average result. In the case of the visualization of the performance of the classification problem AUC (Area Under the Curve), ROC (Receiver Operating Characteristics) curve is used in this study. It is a very crucial model evaluation metric. In this study, a confusion matrix is also used to evaluate the classifier. It is a table to illustrate the performance of the classifier on the test dataset.

The first step of the training of the model is to prepare the dataset for the model. All the data are split into training and testing/validation set by a ten-fold cross-validation method. The cross-validation technique is used to evaluate the classification model by splitting the database into a training set and a test set. In 10-fold cross-validation, the original dataset is randomly split into 10 equal-size subsamples. Among the 10 subsample sets, 9 subsamples sets are used for the training purpose and the remaining single subsample set is used for the model evaluation.

In order, to make sure that each of the 10 subsamples is used exactly once as the validation set, the cross-validation procedure is repeated 10 times. The results for each fold are then averaged to obtain a single measurement of the experiments.

4.5. Experimental results and gesture classification accuracies

The Receiver Operating Characteristics (ROC) curves in Figure 6 investigate and employs the relationship between the sensitivity and specificity of the classifier. The curves of various classes are shown in individual colors. The area under curve (AUC) statistic gives an empirical measure of the classification performance of the classifier based on the area under a ROC curve. It evaluates the performance of the gesture recognition model on the trial-1' test set. Here AUC is expressed on a scale of 0 to 1.

The micro-average AUC is reported as 0.98 and the macro-average AUC is also reported as 0.98. The AUCs are expressed as 0.94, 0.97, 0.97, 1.00, 1.00, 1.00, 1.00, and 1.00 for the gesture classes g1, g2, g3, g4, g5, g6, g7, and g8, respectively.

The confusion matrix shown in Figure 7(a) summarizes the classification performance of the inception v3 classifier with respect to the fold 1 test data. It is represented by a matrix of two dimensions and indexed in one dimension by the true label of the gesture and in the other dimension by the label of that the classifier assigns. Here is this confusion matrix for an eight-class gesture recognition task, with the classes g1, g2, g3, g4, g5, g6, g7, and g8.

The first row of the confusion matrix shows that 18 gestures belong to the gesture class 'g1' and that 16 are correctly classified as belonging to 'g1', and two are misclassified as belonging to other classes. One gesture that is truly belonging to 'g1' class is misclassified as belonging to 'g3' class, and another gesture that is truly belonging to 'g1' class is also misclassified as belonging to 'g6' class.

The second row of the confusion matrix shows that 18 gestures belong to the gesture class 'g2' and that 17 are correctly classified as belonging to 'g2'. One gesture that is truly belonging to 'g2' class is misclassified as belonging to 'g1' class. The third row of the confusion matrix shows that 18 gestures belong to the gesture class 'g3' and that 17 are correctly classified as belonging to 'g3'. One gesture that is truly belonging to 'g3' class are misclassified as belonging to 'g1' class.

For all the remaining rows in the confusion matrix, the true gesture labels are classified correctly. The bar plot shown in Figure 7(b) represents the overall gesture classification performance of the inception v3 model in the 0 to 1 scale with fold-1 test set. The classification accuracy is 0.99306, sensitivity is 0.97222, and specificity is 0.99603.

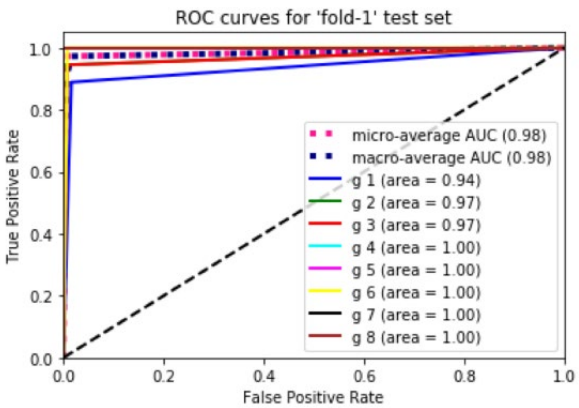
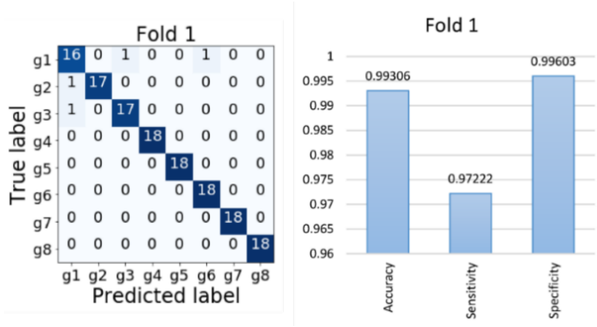


Fig. 6. Model performance evaluation of Inception v3 model for gesture recognition for 'fold-1' test dataset. The Receiver Operating Characteristics (ROC) curve of various classes is shown in individual colors.



a) b)

Fig. 7. a) Confusion matrix of the model performance evaluation of Inception v3 model for gesture recognition b) Bar plot of accuracy, sensitivity, and specificity for 'fold-1' test dataset.

The average performance of the proposed inception-v3 based classifier for all the test folds is shown in Figure 8. The average is calculated across all the test results of the model. The average classification accuracy of 99.44 %, the sensitivity of 97.78 %, and the specificity of 99.68 % are obtained.

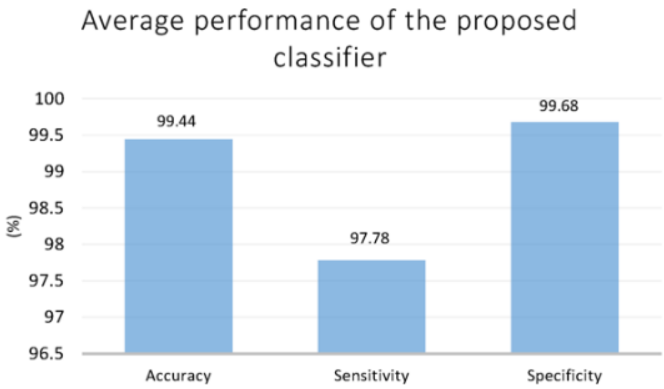


Fig. 8 The average performance of the proposed inception-v3 based classifier.

4.6. Compare performances of Inception v3 and sequential CNNs

Four sequential CNN networks have also been implemented to compare the results of the inception v3 based model and the small sequential CNN. The first sequential CNN consists of 2 convolutional layers, the second sequential CNN consists of 3 convolutional layers, the third sequential CNN consists of 4 convolutional layers, and finally, the fourth sequential CNN consists of 5 convolutional layers. The gesture recognition accuracies are shown in Figure 9. It is observed that the classification accuracy is increased with the increase of the convolution layer. The Inception v3 outperformed all the sequential CNN. The first sequential CNN consists of 2 convolutional layers (2 CONV LAYER), the second sequential CNN consists of 3 convolutional layers (3 CONV LAYER), the third sequential CNN consists of 4 convolutional layers (4 CONV LAYER), the fourth sequential CNN consists of 5 convolutional layers (2 CONV LAYER), and the Inception v3 (INCEPTION V3) obtained classification accuracy of 17.71 %, 17.71 %, 19.79 %, 32.81%, and 99.44 %, respectively.

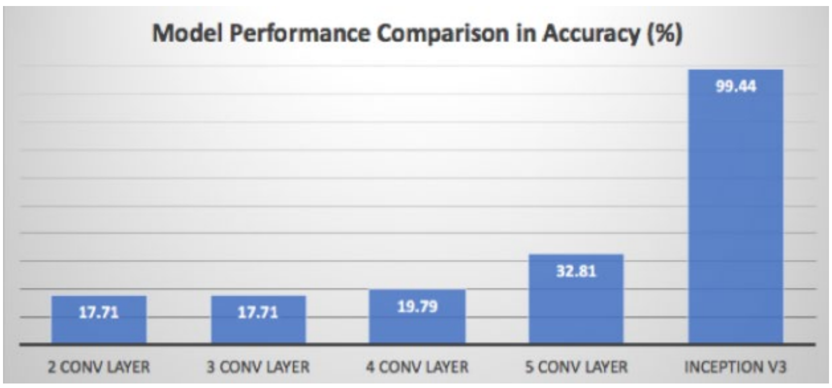


Fig. 9. Performance comparison of the CNN networks for gesture classification.

4.7. Discussion

In this paper, gesture recognition is performed using EMG signal, and deep network is used for analysis. Since an array of electrodes are used for the EMG acquisition, there are 128 channels and therefore, they produce a high volume of data. We apply PCA for the feature extraction and dimension reduction of the dataset. We have selected three PCs for the subsequent analysis. The CWT has been applied on these PCs to get the time-frequency representation of the signals. These images, also called scalogram, are then fed to the deep learning-based gesture recognition model. We use inception v3 architecture for this purpose and utilize the transfer learning technique to initialize the weight the network and fine-tuned it to train. We have got very good accuracy with the inception v3-based gesture recognition model. Here we split the dataset in ten-folds and performed ten-fold cross validation. Lastly, we compare our results obtained with the results of the existing studies. The details comparisons are shown in Figure 10. Our proposed method for hand gesture recognition has reported the best performance and outperformed the existing studies. This system could be used in many applications where hand gestures are control commands and in many rehabilitation systems.

Sl. No.	Existing Studies	Methods	Accuracy
1	An EMG Gesture Recognition System with Flexible High-Density Sensors and Brain-Inspired High-Dimensional Classifier [27]	Brain-inspired high-dimensional (HD)	96.64%
2	Gesture recognition by instantaneous surface EMG images [11]	Deep convolutional network	89.3%
3	EMG-based hand gesture recognition for real-time biosignal interfacing [28]	k-NN and Bayes	94%
4	sEMG-Based Gesture Recognition with Convolution Neural Networks [16]	Parallel multiple-scale convolution architecture	78.71%
5	Gesture recognition method based on a single-channel sEMG envelope signal [29]	KNN algorithm and the soft margin SVM	75.8% and 79.4%
6	Hand Gesture Recognition Using Machine Learning and the Myo Armband [30]	k-nearest neighbor and dynamic time warping algorithms	86%
7	Robust hand gesture recognition with a double channel surface EMG wearable armband and SVM classifier [31]	SVM	95% to 100%
8	Surface EMG-Based Hand Gesture Recognition via Hybrid and Dilated Deep Neural Network Architectures for Neurorobotic Prostheses [32]	Hybrid Recognition Model (HRM) and Temporal Convolutional Network Model (TCNM)	HRM 98.01% and TCNM 92.5%
9	Surface EMG hand gesture recognition system based on PCA and GRNN [33]	PCA and GRNN	95.1%
10	Gesture recognition with EMG signals based on ensemble RNN [34]	RNN network	90.78%
11	A novel attention-based hybrid CNN-RNN architecture for sEMG-based gesture recognition [35]	CNN-RNN	94.5%
12	Hand gesture recognition based on surface electromyography [36]	Hidden Markov model	79%
13	sEMG-Based Hand-Gesture Classification Using a Generative Flow Model [37]	generative flow model (GFM)	63.86 ± 5.12%
14	Transradial Amputee Gesture Classification Using an Optimal Number of sEMG Sensors: An Approach Using ICA Clustering [38]	independent component analysis (ICA) and Icasso clustering	> 95%
15	A Novel Surface Electromyographic Signal-Based Hand Gesture Prediction Using a Recurrent Neural Network [39]	recurrent neural network (RNN)	89.6%
16	A novel hand gesture recognition method based on 2-channel sEMG [40]	back propagation (BP) neural network	96.41% to 99.70%
17	Hand Gesture Recognition based on Surface Electromyography using Convolutional Neural Network with Transfer Learning Method [41]	CNN + Transfer Learning	Around 90%
18	R-CNN and wavelet feature extraction for hand gesture recognition with EMG signals [42]	R-CNN	96.48%
Proposed	Deep Learning for EMG-based Hand Gesture Recognition using Principal Component Analysis and Continuous Wavelet Transform	Deep CNN, principal component analysis (PCA) and CWT	99.44%

Fig. 10. The classification performance comparison of the proposed methodology with some very recently published deep learning-based methods.

So we can conclude that our proposed method for hand gesture recognition has reported the best performance and outperformed the existing studies. This system could be used in many applications where hand gestures are control commands and in many rehabilitation systems.

5. CONCLUSIONS

The main focus of this work is electromyography-based hand gesture recognition using a deep learning algorithm. Hand gesture recognition has become a particular interest in research. Electromyography (EMG) has already been used in hand gesture recognition since EMG acquisition is easy and can be recorded both in a non-invasive or an invasive way. For hand gesture recognition many applications use noninvasive EMG signals. However, the classification of hand gestures using EMG is not easy because the time domain EMG signal has almost similar characteristics and the result obtained in many existing studies are not very perfect. In this paper, for obtain perfect classification accuracy deep Inception v3 model has been employed. In this study, our contribution is to generate the image input of the deep learning idea from the time-domain EMG signal that can capture both temporal and spatial information. In addition, to make the image of the time domain signal, it is tried to reduce the high-volume data coming from the array electrode. Principal component analysis (PCA) has been employed for dimension reduction purposes. Continuous wavelet analysis (CWT) has been employed for the signal-to-image conversion of the time domain principal components of the EMG signals. A classification accuracy of 99.44 % is achieved by using the proposed inception v3-based gesture recognition model. This result is much better than the results obtained using the sequential CNN consisting of a few convolutional layers. The results obtained from the proposed architecture have also been compared with the results of the existing studies. It has been observed that the gesture recognition rate obtained in this work is much better than the results of the existing studies. Thus, it can be concluded that the proposed hand gesture classification technique can recognize hand gestures very well and can be used in many EMG-based applications.

REFERENCES

[1] Saponas, T.S., Tan, D.S., Morris, D., Balakrishnan, R., Turner, J., Lan-Day, J.A., Enabling always available input with muscle-computer interfaces, Proceedings of the 22nd annual ACM symposium on User interface software and technology, 2009, p. 167-176.
[2] Phinyomark, A., Phukpattaranont, P., Limsakul, C., A review of control methods for electric power wheelchairs based on electromyography signals with special emphasis on pattern recognition, IETE Technical Review, no. 28, 2011, p. 316-326.
[3] Scheme, E., Englehart, K., Electromyogram pattern recognition for control of powered upper-limb prostheses: state of the art and challenges for clinical use, Journal of Rehabilitation Research and Development, vol. 48, no. 6, 2011, p. 643-659.
[4] Yousefi, J., Hamilton-Wright, A., Characterizing EMG data using machine-learning tools, Computers in Biology and Medicine, no. 51, 2014, p. 1-13.
[5] Padmanabhan, P., Puthusserypady, S., Nonlinear analysis of EMG signals a chaotic approach, The 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, vol. 1, 2004, p. 608-611.
[6] Scheme, E., Englehart, K., Training strategies for mitigating the effect of proportional control on classification in pattern recognition based myoelectric control, Journal of Prosthetics and Orthotics, vol. 25, no. 2, 2013, p. 76-83.
[7] Amma, C., Krings, T., Boer, J., Schultz, T., Advancing muscle computer interfaces with high-density electromyography, Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, 2015, p. 929-938.
[8] Rojas-Martınez, M., Mananas, M., Alonso, J., Merletti, R., Identification of isometric contractions based on high density EMG maps, Journal of Electromyography and Kinesiology, no. 23, 2013, p. 33-42.
[9] Hwang, H.J., Hahne, J.M., Muller, K.R., Channel selection for simultaneous and proportional myoelectric prosthesis control of multiple degrees-of-freedom, Journal of Neural Engineering, no. 11, 2014, art. no. 056008.
[10] Xia, P., Hu, J., Peng, Y., EMG based estimation of limb movement using deep learning with recurrent convolutional neural networks, Artificial Organs, no. 42, 2018, p. 67-77.
[11] Geng, W., Du, Y., Jin, W., Wei, W., Hu, Y., Li, J., Gesture recognition by instantaneous surface EMG images, Scientific Reports, no. 6, 2016, art. no. 36571.
[12] Boostani, R., Moradi, M.H., Evaluation of the forearm EMG signal features for the control of a prosthetic hand, Physiological Measurement, no. 24, 2003, p. 309.

[13] Zardoshti-Kermani, M., Wheeler, B.C., Badie, K., Hashemi, R.M., EMG feature evaluation for movement control of upper extremity prostheses, IEEE Transactions on Rehabilitation Engineering, no. 3, 1995, p. 324-333.
[14] Phinyomark, A., Phukpattaranont, P., Limsakul, C., Feature reduction and selection for EMG signal classification, Expert Systems with Applications, no. 39, 2012, p. 7420-7431.
[15] Xia, P., Hu, J., Peng, Y., EMG based estimation of limb movement using deep learning with recurrent convolutional neural networks, Artificial Organs, no. 42, 2018, p. 67-77.
[16] Ding, Z., Yang, C., Tian, Z., Yi, C., Fu, Y., Jiang, F., sEMG based gesture recognition with convolution neural networks, Sustainability, no. 10, 2018, art. no. 1865.
[17] Khushaba, R.N., Correlation analysis of electromyogram signals for multiuser myoelectric interfaces, IEEE Transactions on Neural Systems and Rehabilitation Engineering, no. 22, 2014, p. 745-755.
[18] Gijsberts, A., Atzori, M., Castellini, C., Muller, H., Caputo, B., Movement error rate for evaluation of machine learning methods for SEMG based hand movement classification, IEEE Transactions on Neural Systems and Rehabilitation Engineering, no. 22, 2014, p. 735-744.
[19] Allard, U.C., Nougarou, F., Fall, C.L., Giguere, P., Gosselin, C., Laviolette, F., Gosselin, B., A convolutional neural network for robotic arm guidance using SEMG based frequency-features, IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2016, p. 2464-2470.
[20] Du, Y., Jin, W., Wei, W., Hu, Y., Geng, W., Surface EMG based inter-session gesture recognition enhanced by deep domain adaptation, Sensors, no. 17, 2017, art. no. 458.
[21] Englehart, K., Hudgins, B., Parker, P.A., Stevenson, M., Classification of the myoelectric signal using time-frequency based representations, Medical Engineering and Physics, no. 21, 1999, p. 431-438.
[22] Addison, P.S., Wavelet transforms and the ECG: a review, Physiological Measurement, no. 26, 2005, p. 155-199.
[23] Karlsson, S., Gerdle, B., Mean frequency and signal amplitude of the surface EMG of the quadriceps muscles increase with increasing torque—a study using the continuous wavelet transform, Journal of Electromyography and Kinesiology, no. 11, 2001, p. 131-140.
[24] Toledo, C., Munoz, R., Leija, L., sEMG signal detector using discrete wavelet transform, Pan American Health Care Exchanges, 2012, p. 62-65.
[25] Hotelling, H., Analysis of a complex of statistical variables into principal components, Journal of Educational Psychology, no. 24, 1933, p. 417-441.
[26] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A., Going deeper with convolutions, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, p. 1-9.
[27] Moin, A., Zhou, A., Rahimi, A., Benatti, S., Menon, A., Tamakloe, S., Ting, J., Yamamoto, N., Khan, Y., Burghardt, F., An EMG gesture recognition system with flexible high-density sensors and brain inspired high dimensional classifier, IEEE International Symposium on Circuits and Systems (ISCAS), 2018, p. 1-5.
[28] Kim, J., Mastnik, S., Andre, E., EMG based hand gesture recognition for realtime biosignal interfacing, Proceedings of the 13th International Conference on Intelligent User Interfaces, 2008, p. 30-39.
[29] Wu, Y., Liang, S., Zhang, L., Chai, Z., Cao, C., Wang, S., Gesture recognition method based on a single-channel sEMG envelope signal, EURASIP Journal on Wireless Communications and Networking, 2018, p. 1-8.
[30] Benalcazar, M.E., Jaramillo, A.G., Zea, A., Paez, A., Andaluz, V.H., Hand gesture recognition using machine learning and the myo armband, 25th European Signal Processing Conference (EUSIPCO), 2017, p. 1040-1044.
[31] Tavakoli, M., Benussi, C., Lopes, P.A., Osorio, L.B., de Almeida, A.T., Robust hand gesture recognition with a double channel surface EMG wearable armband and SVM classifier, Biomedical Signal Processing and Control, no. 46, 2018, p. 121-130.
[32] Rahimian, E., Zabihi, S., Atashzar, S.F., Asif, A., Mohammadi, A., Surface EMG-based hand gesture recognition via hybrid and dilated deep neural network architectures for neurorobotic prostheses, Journal of Medical Robotics Research, no. 5, 2020.
[33] Qi, J., Jiang, G., Li, G., Sun, Y., Tao, B., Surface EMG hand gesture recognition system based on PCA and GRNN, Neural Computing and Applications, no. 32, 2020, p. 6343-6351.
[34] Zhou, X.F., Wang, X.F., Wu, Z.K., Korkhov, V., Gaspary, L.P., Gesture recognition with EMG signals based on ensemble RNN, Optics and Precision Engineering, no. 28, 2020, p. 424-442.
[35] Hu, Y., Wong, Y., Wei, W., Du, Y., Kankanhalli, M., Geng, W., A novel attention-based hybrid CNN-RNN architecture for sEMG-based gesture recognition, PloS one, vol. 13, no. 10, 2018, art. no. e0206049.
[36] Samadani, A.A., Kulic, D., Hand gesture recognition based on surface electromyography, 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2014, p. 4196-4199.
[37] Sun, W., Liu, H., Tang, R., Lang, Y., He, J., Huang, Q., sEMG-based hand-gesture classification using a generative flow model, Sensors, vol. 19, no. 8, 2019, art. no. 1952.

[38] Naik, G.R., Al-Timemy, A.H., Nguyen, H.T., Transradial amputee gesture classification using an optimal number of semg sensors: an approach using ICA clustering, IEEE Transactions on Neural Systems and Rehabilitation Engineering, no. 24, 2015, p. 837-846.
[39] Zhang, Z., He, C., Yang, K., A novel surface electromyographic signal-based hand gesture prediction using a recurrent neural network, Sensors, vol. 20, no. 14, 2020, art. no. 3994.
[40] Yu, H., Fan, X., Zhao, L., Guo, X., A novel hand gesture recognition method based on 2-channel sEMG, Technology and Health Care, no. 26, 2018, p. 205-214.
[41] Chen, X., Li, Y., Hu, R., Zhang, X., Chen, X., Hand gesture recognition based on surface electromyography using convolutional neural network with transfer learning method, IEEE Journal of Biomedical and Health Informatics, vol. 25, no. 4, 2020, p. 1292-1304.
[42] Shanmuganathan, V., Yesudhas, H.R., Khan, M.S., Khari, M., Gandomi, A.H., R-CNN and wavelet feature extraction for hand gesture recognition with EMG signals, Neural Computing and Applications, no. 32, 2020, p. 16723-16736.