

# Hand Gestures Recognition Using Convolutional Neural Networks

<sup>1</sup>A. Dhivya Bharathi, <sup>2</sup>R. Arun Prasath, <sup>3</sup>V. Kesavan

<sup>1</sup>Assistant Professor, K. Ramakrishnan College of Technology

<sup>2,3</sup>UG Student, K. Ramakrishnan College of Technology

Submitted: 05-04-2022

Revised: 16-04-2022

Accepted: 19-04-2022

## ABSTRACT:

sEMG is a widely used method of human-computer interaction that has been used in a number of scenarios. To enable sEMG classification, numerous methods of machine learning based approaches have also been developed. However, despite its popularity in the computer vision sector, In sEMG deciphering, the deep neural network has a very restricted use. In this research, we used a novel deep learning framework to classify hand gestures based on sEMG. In particular, we used a convolutional neural network (CNN) to classify sEMG with numerous sessions. This is more problematic because of the different time-varying biodynamics of the participants. As a result, we've looked at several CNN topologies in the hopes of finding an optimum design that can successfully discover hidden characteristics in signals. For surface EMG-based hand gesture recognition, the proposed CNN framework has a better classification accuracy, and the varied topologies have a substantial impact on CNN performance. This research lays a solid foundation for CNN to recognise multiple-session sEMG signal patterns.

**Keywords** : Deep learning, Convolutional neural network, Surface Electromyography, Feature extraction, Human computer interaction, Hand gesture.

## I. INTRODUCTION:

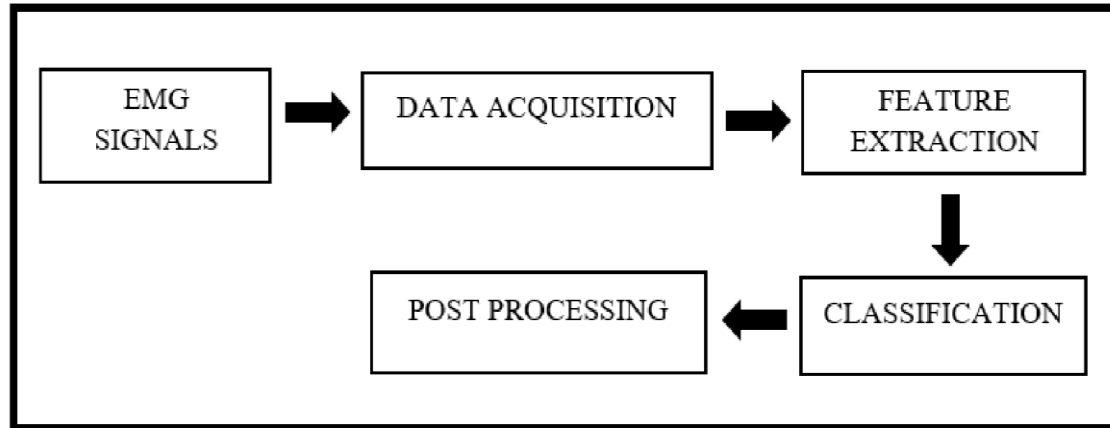
In recent years, human-computer interaction (HCI) has gotten a lot of attention and has become a popular tool for improving growth, particularly for handicapped and elderly people [1]. Human-computer interaction (HCI) refers to the human-designed framework that allows humans and computers to communicate and interpret human motives and motions into machine orders. For persons suffering from chronic motor sickness

due to a number of factors, such as Spinal Cord Injury (SCI), the quality of life will be considerably improved if they can easily and naturally interface with machines and carry out routine tasks using machine commands [2]. Different bioelectrical signals, such as Electro-Encephalogram (EEG) and Electromyogram (EMG), are used to link neurons to computers. These signals are created by bio systems from a variety of organs, tissues, and cells. Bio signal techniques, as compared to traditional input devices, offer a wealth of new opportunities for improving the human-computer interaction process and making it more intelligent and trustworthy. [4]

Among all of these bio signals, surface EMG (sEMG) is considered a promising source for developing HCI, since it is now employed in a wide range of applications such as robot arm control, prosthetic device control, fall detection, neurological rehabilitation, and so on. The electrical currents generated during the deflation and mitigation phases of muscles are referred to as sEMG. [5] Electrical activity of muscles generated by arm motions can therefore be converted into machine-control directives utilizing sEMG signals. Furthermore, using attachable electrodes, sEMG signals may be recorded on the skin's surface in a non-invasive manner.

We propose a CNN-based framework for hand gesture classification from the sEMG data in this paper. CNN reconstructs the original input, layer by layer, to the final class scores, using an architecture that mimics the neuron connection network of the human brain and visual cortex. The constant pattern in different portions of the sEMG signal may be evaluated by sharing equal weights between distinct neurons in the layer. In addition, investigate how the architecture of CNN affects the performance of sEMG pattern recognition. [7] Our tests reveal that the proposed system can

effectively detect hand motions from sEMG data. This effort will significantly improve the field of sEMG-based HCI.



## II. LITERATURE SURVEY:

Human-computer interaction has gained popularity and tremendous attention as it helped numerous differently-abled and elderly people. There are many who suffer from various hand injuries and survive without a hand.[1] The condition of their life could be enhanced if they could naturally move their hands. This is possible with devices that are developed using HCI. Various prosthetic arm controls have come into existence that utilizes HCI technology.

sEMG (Surface Electromyography) technique could be adopted that would recognize various hand gestures which will be more helpful to develop a prosthetic arm control.[3] When some movements occur in our hands, it is due to the electrical signals passed from our brain. All these are related to the field of bio signals. When the hands get into motion, the neurons from the human brain send various electrical signals to the muscles after which the muscle gets reacted to it. If these signals are captured using any technique, it could be converted into machine understandable commands and fed as input for the artificial arm controls. [5]

sEMG is a non-invasive technique commonly used to investigate muscle activation and fatigue, which can allow for continuous measurement. It is like the technique used to capture ECG signals. Several needle electrodes are inserted into the human muscles by applying some gel and the signals are acquired and recorded.[3] Due to various movements of the electrodes during the signal capturing, noises get accumulated with the original signal.

To nullify the noises involved, Short Term Fourier Transform technique is utilized. It's a

mathematical procedure that calculates the sinusoidal phase and frequency content of local sections as time passes.

There lies many traditional machine algorithms like Support Vector machine (SVM), Random Forest (RF), Linear Discriminant Analysis (LDA) to perform the hand gesture recognition. But these are not suitable when the tasks get complicated. Therefore, we move into the Deep learning terminology, by using various neural networks. There are many neural networks like ANN, RNN, KNN which could be used to train and test to recognize various hand gestures. But these are inappropriate for large datasets as the accuracy gets reduced.[1]

The best way for hand gesture recognition using deep learning technology would be using the Convolutional neural network (CNN). It can handle any kind of complicated situation and work with hundreds and thousands of datasets.[6] The resultant accuracy of CNN framework will be greater than all other neural networks.

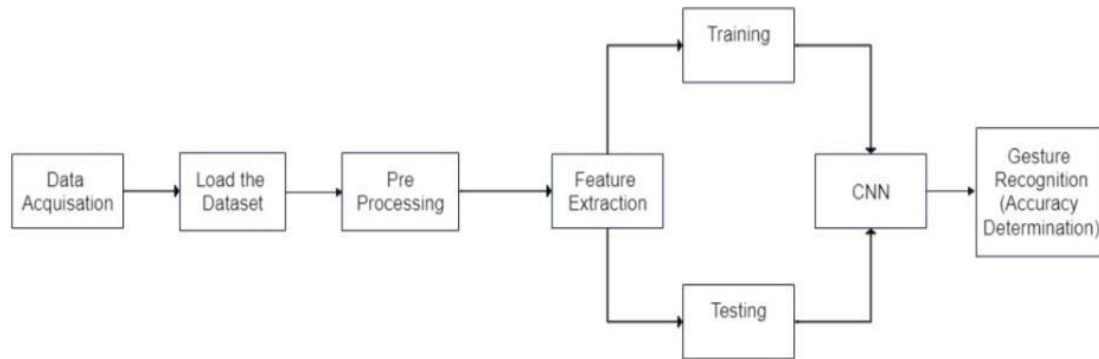
There exists various other gestures called deep gestures. For example, the gestures performed by a policeman are so difficult to recognize. Because, these gestures are so robust and it significantly changes over time. Both the gestures and also the background get changed frequently. To recognize these kinds of gestures, a robust CNN framework has to be administered.[7]

CNN reduces the process of extracting the essential features from the input data by enabling various hidden layers. When the input data is sent into the CNN framework, the essential features get extracted by a pooling layer that performs max pooling, min pooling and average pooling. After all

these features are extracted, it is then classified and the neural network gets trained and tested for hand gesture recognition. Therefore, CNN will be the

most probable neural network to perform Hand gesture recognition.

### III. SYSTEM ARCHITECTURE:



### IV. PROPOSED METHODOLOGY:

We offer a novel CNN framework for Hand Gesture Recognition in this paper. For huge datasets, traditional approaches are ineffective, and for difficult tasks, they become overly convoluted. Therefore CNN is found to be the most efficient one. When the EMG signals are decoded and converted as machine-understandable commands, these can be fed as input for the CNN framework. The main advantage of using the CNN network is, it doesn't require any specific feature extraction techniques to be performed. Instead, it automatically extracts all the essential features on its own.

CNN has its advantage in the field of self-driving cars which resulted in more than 95% accuracy, and most these days automated cars are being developed using the CNN framework. The input layer, hidden layers, and output layer are all layers that make up the CNN. The input data is processed at the hidden layer after it passes through the input layer. The pooling layer and fully connected neural networks are the hidden layers.

This is where the important characteristics are extracted. The whole matrix format input is split up into several small matrices and the dot product of the end matrix is calculated. With the result, Max - pooling is calculated (Maximum value among the matrix). Likewise, this operation is performed several times so that all the essential features get extracted and at last gets combined in the fully connected layer.

#### I). Dataset Collection:

To complete the required finger motions, eight individuals, six men and two women, varying in age from twenty to thirty-five years, were entered. The participants were completely healthy, with no neurological or muscle problems. Prior to participating in the study, everyone who took part in the activity gave their informed consent. The participants sat in a wheelchair with their arms supported and fastened in a constant position. To produce the datasets, eight EMG channels were securely positioned along the forearm's peripheral and evaluated by the Bagnoli desktop EMG system.

Each of the sensors was tested with a two-slot adhesive skin interface to ensure that they were securely attached to the skin. Each subject had a conductive gummy electrolyte solution (dermatrode reference terminal) put on their wrists during the research. A Delsys Bagnoli-8 speaker was used to amplify the EMG signals resulting in a total gain of thousand.

A 12-bit analog-to-digital converter (National Instruments, BNC-2090) was used to sample the signal at 4000 Hz, and the signal data were collected without difficulty using Professional EMGWorks Acquisition application. The EMG signals were then bandpass filtered between 20 and 450 Hz, with the 50 Hz line interference removed using a notch filter.

#### II). Data Pre-Processing:

CNN is inspired by the structure of the visual system and is capable of replicating how the human mind recognises and analyses complex input. CNN has achieved tremendous advancements in computer vision domains

including facial recognition software, object recognition, and automated vehicles on city streets. CNN is made up of a series of layers, including convolution, pooling, and wholly layers. While input information is changed across each layer, the hierarchical face features will be retrieved. In this work, we apply a STFT to analyze the sEMG signals before feeding into CNN.

Because the sEMG signal exhibits a lot of irregularity across time, we make use of the STFT. The Fourier transform cannot decipher the time-varying sEMG signal [8].

The Fourier transform's time resolution can be improved by examining and determining time frequencies. As a result, the short-term Fourier transform provides a method for determining time-localized frequency information.

First, a fixed-size window divides the signal. Second, The frequency content of each frame is determined using the Fourier transform. The frequency information variability can then be induced over time. To achieve short-term Fourier transform transformation, We used a window size of 512 samples and a step size of 62 samples. CNN is expected to learn more about the input data after translating the sEMG signal to the time-frequency domain and, as a consequence, improve its efficiency.

### III). Dataset Acquisition:

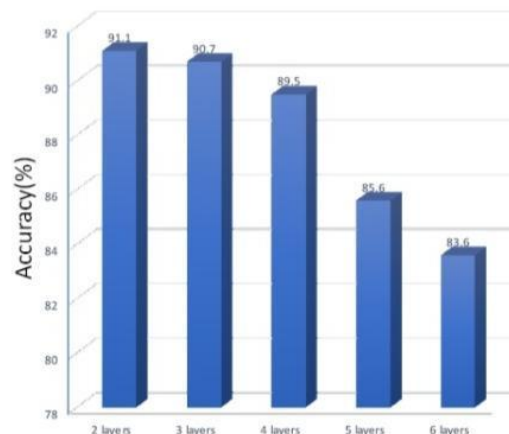
To process the acquired EMG signal, a STFT is applied. The noise caused by electrodes and hand motions is reduced as a result of this. This refined signal is converted to binary data so that computers can understand it.

### IV). Investigation of CNN Topologies:

Hidden layers are used in neural networks to learn and show non-linear movement, but designing neural networks hesitantly is difficult. The network architecture is one of the most important factors that influences CNN's efficiency. However, we cannot rely on the number of layers as a supposition. From the other hand, the number of levels is a model hyper parameter that we must specify. As a consequence, optimizing layer numbers can help us identify a good CNN design that supports representation. In this research, we look at several layer counts in order to find the best one for sEMG decoding. As there are too many layers, there is a risk of overfitting. We look at two-layer, three-layer, four-layer, five-layer, and six-layer topologies.

### V). Topology Comparision:

The topology is one of the most critical determinants of CNN's success [9]. Using two layers, three layers, four layers, five layers, and six layers, we perfect that topology of our projected CNN in this section. It's possible that as the number of layers grows, so does the efficiency. The topology with two layers has the highest precision or correctness (91.1%). There might be two explanations for this. To begin, we employ all three trails of the putEMG dataset, resulting in a massive database. Second, the two-layer structure is sufficient for detecting the sEMG signals hidden beneath the short-term Fourier transform. As a result, the optimal architecture for gesture recognition is two layers.



## V. RESULT AND DISCUSSION:

Thus, a new kind of CNN framework has been designed to recognize hand gestures. Other neural networks, such as the ANN, the KNN, and

the RNN, have lower accuracy. Whereas, this new CNN is estimated to provide more than 95% accuracy.

Neural Networks:	Accuracy:
ANN	80%
KNN	85%
CNN	>=95%

### Comparison of the accuracy of neural networks

With this CNN, we fabricate the network to get trained with the input dataset. For example, when we have 20 input samples, we divide them accordingly like 75% of input for training and 25% for testing. The most important phase is training and testing. The resultant accuracy is completely dependent on how well the network gets trained. The more the training is, the more will be the accuracy.

The confusion matrix produced after testing can be used to assess the accuracy. It's a table that's used to figure out a neural network's performance by calculating the difference between the actual and predicted values.

## VI. CONCLUSION:

Human computer interaction has considerably enhanced the standard of living for the crippled and elderly. Bio signal-based HCI approaches open up a slew of new opportunities for fine-tuning the HCI process and making it more simple and dependable in day-to-day farming. One of the most promising bio signals in HCI is sEMG. We thoroughly evaluate CNN's topology and choose an optimal topology with an outstanding precision or correctness of more than 90%. On difficult popular datasets, the suggested CNN model has been shown to detect help gestures with a high degree of precision or correctness. We don't need to segment the hand in the input figure or extract features deliberately, unlike other state-of-the-art approaches.

In reality, we may acquire higher confidence on the dataset in conjunction with a

simple design of the projected CNN model, and so perform better than the NUS-II dataset. We believe that the proposed technology can be used in fields such as sign language recognition and human-computer interaction. We intend to expand this work in the future to include strong hand movements as well.