MYOELECTRIC HUMAN COMPUTER INTERACTION USING LSTM-CNN NEURAL NETWORK FOR DYNAMIC HAND GESTURES RECOGNITION

A Thesis

by

QIYU LI

Submitted to the Graduate and Professional School of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Chair of Committee,	Reza Langari
Committee Members,	Shankar P. Bhattacharyya
	Kiju Lee
Head of Department,	Guillermo Aguilar

December 2021

Major Subject: Mechanical Engineering

ABSTRACT

Human-computer interaction(HCI) has become a trendy research field recently. Many HCI systems are based on bio-signal analysis and classification. EMG signal which is formed due to muscle activation, is used in this thesis. sEMG signals play a central role in many applications, including clinical diagnostics, control of prosthetic devices, and some human-machine interactions. These applications are commonly referred to as myoelectric control. Many factors would influence the classification in myoelectric control, and limb positions are focused on in this thesis.

Two research goals in this thesis are:

1. Decease the effect of arm positions when recognizing a gestures. To tackle this issue, a CNN-LSTM neural network is introduced. Compared to Dr. Shin's work, the new model is able to classify more gestures with more positions.

2. Apply the new model to a human-computer interaction system. A 7-DoF Kinova robot arm is used here. And a go and grasp task is designed to test the system. For most cases, the myoelectric control system finished the task successfully in an acceptable time longer than a joystick control. In addition, this control method is easier to master compared to a joystick control.

In conclusion, this research focuses on EMG-based dynamic gestures recognition with multiple limb positions. First, the CNN-LSTM neural network, which combined the advantages of CNN and LSTM is proposed in this thesis. Then this model is used for a myoelectric control system. A 7-DoF robot arm is controlled by human gestures via the system.

DEDICATION

To my supportive mother, father and my grandfather

ACKNOWLEDGMENTS

I would like to thank my parents and grandfather(Shidian Li), who gives me unlimited support and love. Also, thank you to my friends who encourage me every time I need it.

Thank you to all my committee members: Prof. Shankar P. Bhattacharyya, Prof Kiju Lee, and mainly my advisor Prof. Reza Langari who offer the most significant help during my master period. He is always kind and supportive to me. I really enjoy the discussion with him where I am inspired, encouraged and guided.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supported by a thesis committee consisting of Professor Reza Langari who is my advisor, Professor Kiju Lee of the Department of Mechanical Engineering and Professor Shankar P. Bhattacharyya of the Department of Electrical Computer Engineering.

All other work conducted for the thesis was completed by the student independently.

Funding Sources

No other outside source of funding was provided.

NOMENCLATURE

HCI	Human-computer interaction
EMG	Electromyography
LDA	Linear discriminant analysis
SVM	Super vector machine
KNN	K-nearest neighbors
FS	Forward Selection
GA	Genetic Algorithm
PCA	Principal component analysis
ICA	Independent component analysis
DNN	Deep neural network
CNN	Convolutional neural network
RNN	Recurrent neural network
TKEO	Teager-Kaiser energy operator
DWT	Dynamic time wraping
LSTM	Long short-term memory

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION AND LITERATURE REVIEW

1.1 Literature Review

Human-computer interaction(HCI) has become a very popular research field recently. Many HCI systems are based on bio-signals analysis and classification. One of the essential bioelectrical signals is the Surface ElectroMyoGraphic (sEMG) signal, which is formed due to muscle activities, including muscle excitation and contraction[1]. EMG signals play a central role in many applications, including clinical diagnostics, control for prosthetic devices and for human-machine interactions[2]. These applications are commonly referred to as myoelectric control.

1.1.1 ElectroMyoGraphic Signal

ElectroMyoGraphic(EMG) signals are electrical signals 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[32]. In invasive methods, the prosthetic device needs to be connected directly to the targeted muscle tissues surgically whereas, in non-invasive methods, signals are recorded from the surface of the muscle regions[35]. Also, multiple-channel electrodes could cover more areas, which is more effective than single-channel electrodes.

Pattern recognition for EMG classification usually consist of data pre-processing, data segmentation, feature extraction, dimensionality reduction, and classification stages[4].

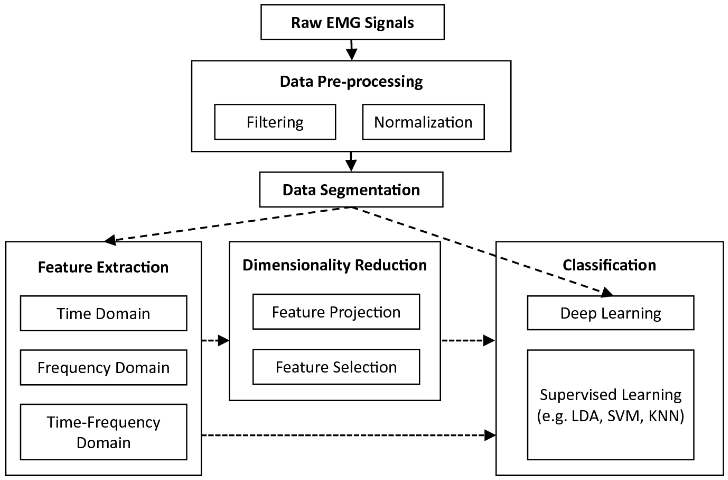

Figure 1.1: EMG Pattern Recognition Structure[4]

In the pre-processing stage, raw EMG signals' noise like motion artifacts would be reduced. Then for the segmentation part, the two main techniques for data segmentation include adjacent windowing and overlapping windowing. Also, for real-time systems, the length of these segments plus any computation must be less than 300 ms to avoid noticeable delays[5]. A lot of studies focus on feature extraction because different features could influence the accuracy of classification. And EMG features have been commonly divided into three categories: frequency-domain, time-domain, and time-frequency domain[6][11][8]. The frequency-domain features are obtained using Fourier transform having relatively poor real-time performance. On the contrary, the time-domain features are direct and easily extracted from the time series of the original sEMG signal without any conversion process[9]. Dr.Shin[3] use eight time domain features for his classification, and the features are in table 1.1.

Feature Extraction Name	Definition
Mean absolute value	$\mathrm{MAV} = \frac{1}{N}\sum_{i=1}^{N}\lvert x_i\rvert$
Root mean square	$\mathrm{RMS} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} x_i^2}$
Integrated EMG	$\mathrm{IEMG} = \sum_{i=1}^{N}\lvert x_i\rvert$
Waveform length	$\mathrm{WL} = \sum_{i=1}^{N-1}\lvert x_{i+1} - x_i\rvert$
Zero crossing	$\mathrm{ZC} = \sum_{i=1}^{N-1}[sgn(x_i \times x_{i+1}) \cap \lvert x_i - x_{i+1}\rvert \geq th]$
Slope sign change	$\mathrm{SSC} = \sum_{i=2}^{N-1}[sgn((x_i - x_{i+1}) \times (x_i - x_{i+1})) \cap (\lvert x_i - x_{i+1}\rvert \geq th \cup \lvert x_i - x_{i-1}\rvert \geq th)]$
Skewness	$\mathrm{SKW} = \frac{\frac{1}{N}\sum_{i=1}^{n}(x_i - \bar{x})^3}{\left(\frac{1}{N}\sum_{i=1}^{n}(x_i - \bar{x}_i)^2\right)^{\frac{3}{2}}}$ ($\bar{x}$ is sample mean)
Auto-regressive coefficients (a_i)	$x_k = \sum_{i=1}^{P} a_i x_{k-i} + e_k$ (a_i: AR coefficients, e_k: an error term, P: the order of AR)
N: the size of samples; $th =$ threshold	$sgn(x) = \begin{cases} 1, if\ x \geq th \\ 0, othewise \end{cases}$

Table 1.1: Time Domain Features (reprinted with permission from [3])

Dimensionality reduction is for selecting an optimal subset of good performance features or combining features and their projection. Sequential forward selection (SFS), genetic algorithms (GA), principal component analysis (PCA), and independent component analysis (ICA) are common techniques for this process[10]. For the final part, supervised learning like linear discriminant analysis(LDA) is used for years. With the development of deep learning recent years, more and more deep neural networks have been proposed. Mukhopadhyay et al. [11]introduced a deep neural network(DNN) which includes an input layer, several feed-forward layer and softmax layer could achieve 98.88% accuracy. Atzori et al.[16] have used a convolutional neural network(CNN) to classify 50 hand movement.

With the view of Mitra and Acharya[13], gestures can be divided as static motion and dynamic motion. Static motion was represented by repeated and constant multiple-dimensional EMG features. On the contrary, dynamic motion is expressed by a temporal sequence of multiple EMG features that are changed during the motion. According to Dr. Shin's work, the accuracy would be smaller when apply the before pattern recognition process to a dynamic motion with different limb positions than apply pattern recognition process to a static motion. Dr. Shin proposed a sequence-based classification in Fig 1.2. The three different steps from previous pattern recognition in Fig 1.1 are vital for the classification. TKEO is used to detect start and end points of dynamic motions. DTW is employed to align different time-length dynamic motions like Fig 1.3 shows. And Pearson product-moment correlation coefficient is used for template matching[3].

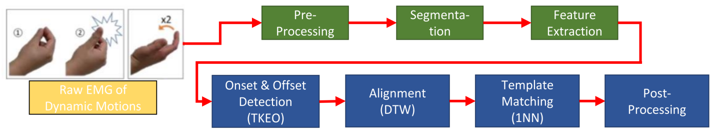

Figure 1.2: Sequence-based Classification (modified from [3])

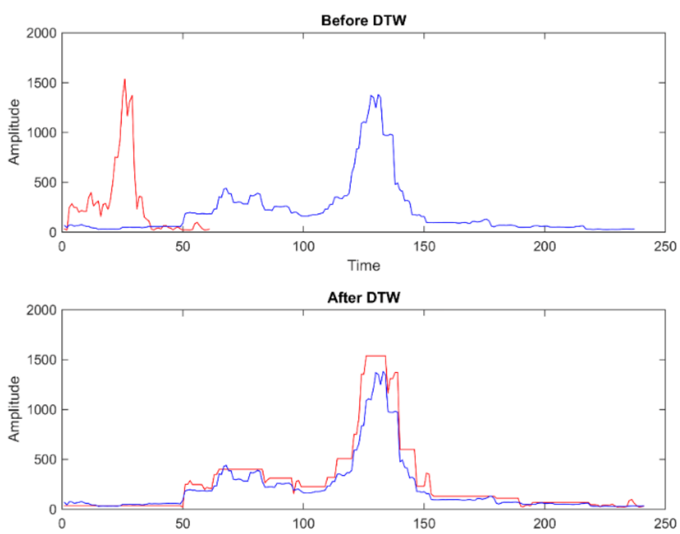

Figure 1.3: Motions with Different Time Length Before and After Alignment by DTW(reprinted with permission from [3])

1.1.2 Myoelectric Control

Myoelectric control is an advanced technique concerned with the detection, processing, classification, and application of myoelectric signals to control human-assisting robots and other devices[22]. Myoelectric control of neuroprostheses has been proposed since the introduction of motor neuroprostheses (Vodovnik et al., 1965). The advantages of myoelectric control include a simple sensor structure (two metal conductors separated by a fixed distance), a sensor that can be applied equally well to many different muscles and the possibility of providing both on/off and proportional control (Popovic et al. , 2001)[21]. With the development of machine learning and deep learning, more and more researcher apply advanced algorithms for classification, which promotes myoelectric control into real world problem significantly.

1.2 Observation

Gestures include dynamic and static motions recognition is a time-series problem. Therefore, recurrent neural network(RNN) is proper for this. They take all previous steps of the sequence into consideration and do not require a fixed size for the input sequences. Especially, long short term memory network should be used which are able to learn and capture long term temporal dependencies[33].

1.3 Thesis Objective

This project aims to design a system that enables people to use gestures with an MYO armband to manipulate a robot arm. This system has two main parts: 1) recognition part: using neural networks to recognize dynamic gestures. 2) control part: implementing a relative functions on a robot arm based on different gestures. In general, this system collects raw EMG signals from people when they make dynamic motions, then recognizes the gestures mapped to a specific function to control a 6-DOF robot arm with a 1-DOF gripper. This system should also ensure:

- Position Independent: Ensure that gestures could be recognized while limb positions change.
- Respond Promptly: Ensure that the time of completing a task by using myoelectric control

is not much longer than other control methods.

Two devices are used in this research. First is the Myo armband, which was employed in the study for collecting EMG signals. It includes eight EMG sensors and one IMU sensor. The sampling rate of the Myo armband is 200Hz. The Myo armband is put on the right forearm of the participant.

Figure 1.4: MYO Armband

Seocnd one is a 7 DoF Kinova robot arm and one of the degree is for the gripper.

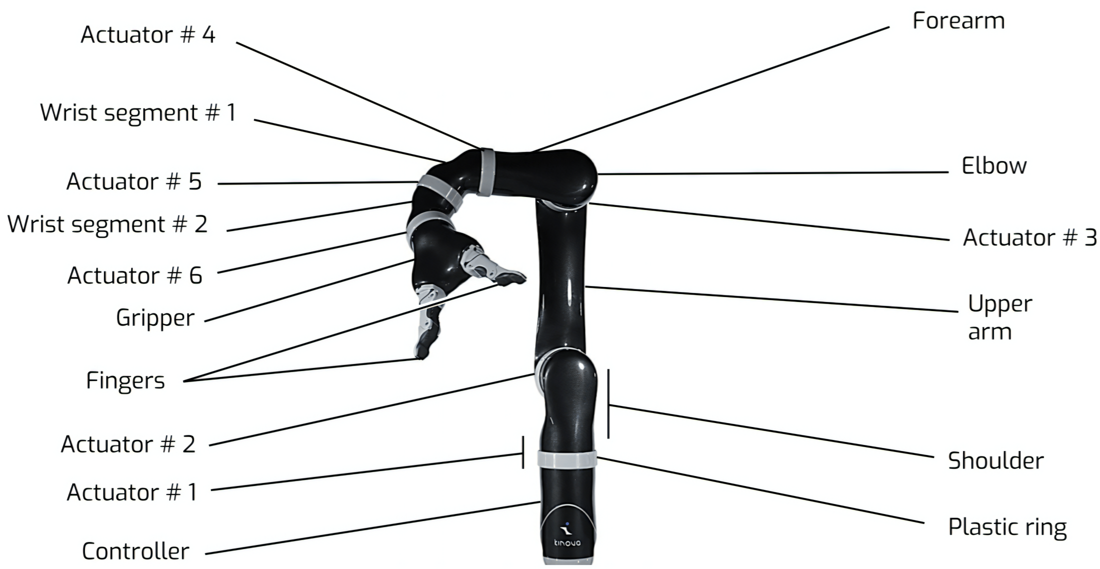

Figure 1.5: Kinova Robot Arm

2. PROPOSED METHOD FOR GESTURE RECOGNITION

2.1 Introduction

But we still face some challenges commonly associated with this lack of reliability in practical conditions that can be roughly categorized into four confounding factors: limb position factor, constrain intensity factor, electrode shift factor and within/between day factor[16]. Anand Kumar-Mukhopadhyay and SumanSamui[11] explored a deep neural network (DNN) based classification system for the upper limb position invariant myoelectric signal. The results turned out that the average accuracy of test data based on the DNN model trained by one limb positions same as test data is over 95% and the average accuracy of test data could only reach about 70% accuracy when the limb position of the training data is not same as test data. Also, the DNN model trained by all positions gained 75% accuracy approximately on test data for every position. The limb positions made a significant impact on gestures recognition.

(Evan Campbell et al., 2020) explained how the limb positions influence in detail: the muscular activity that maintains limb positions against gravitational force is dependent on the position of the limb[16]. When the limb positions are different, the supplemental muscle activities are different. Additionally, while in different positions the underlying topography of the muscle fibers may shift relative to the electrodes[16].

Dr. Shin's work[3] shows that dynamic gestures are in principle more reliable indicators of intent. From his sequence based pattern recognition model, accuracies of dynamic gestures which has temporal information are higher than static gestures. Based on Dr. Shin's work and the challenges I mentioned before, a CNN-LSTM neural network is introduced to classify five dynamic gestures with five arm positions in this section.

2.2 Proposed Method

2.2.1 EMG Gestures Dataset

There are five dynamic gestures with five arm positions in the dataset. The gestures are shown below.

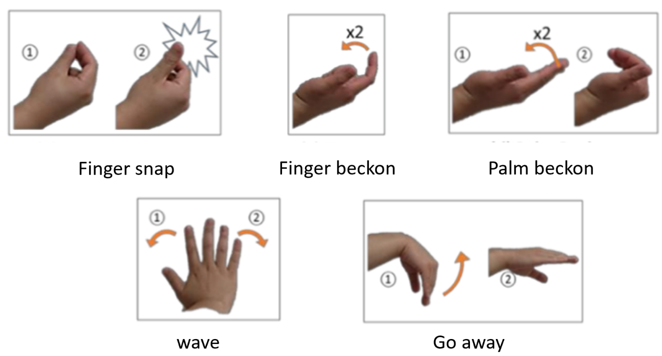

Figure 2.1: Dynamic Gestures
Note: this figure is reprinted with permission from
Sungtae Shin, 2016[3]

For each gestures, there are five common arm positions selected when performing the gesture.

Figure 2.2: Five Limb Positions of Finger Snap

Figure 2.3: Five Limb Positions of Finger Beckon

Figure 2.4: Five Limb Positions of Palm Beckon

Figure 2.5: Five Limb Positions of Wave

Figure 2.6: Five Limb Positions of Go Away

The training data is from seven human subjects. Every gesture at each position would be repeated for 10 times. After finishing seven experiments, there are 7*5*5*10 = 1750 samples. For each sample, there is a 8*680 matrix where 8 is the the amount of channels and 680 is the time steps (The matrix would be pad zeros to make every matrix same length). Tn original dataset is in below.

gesture	label	number
finger snap	0	350
finger beckon	1	350
palm beckon	2	350
wave	3	350
go away	4	350

Table 2.1: Original Dataset

Since the dataset is not very large based on the amount of gestures and arm positions. In image recognition, the data augmentation methods like figure flipping are usually used to add more data to the dataset and the accuracy always be improved after the augmentation. According to (Wei yang et all., 2018) research results[18], two kinds of data-augmentation operations:

1. Electrodes shift. The Myo armband may shift when human subjects performs gestures and changes arm positions. Also, it is hard to guarantee that every human subjects put the armband on exact same place of the forearm.

2. Wrong arrangement (on the other forearm). The arrangements of electrodes on the two forearms should be mirrored.

helped in EMG-based gesture recognition. Therefore, the following augmentation operations are implemented. Basically, the order of 8 channels is changed. The left part is the original order, the first four rows on the right part are electrodes shift condition and the last row is wrong arrangement condition.

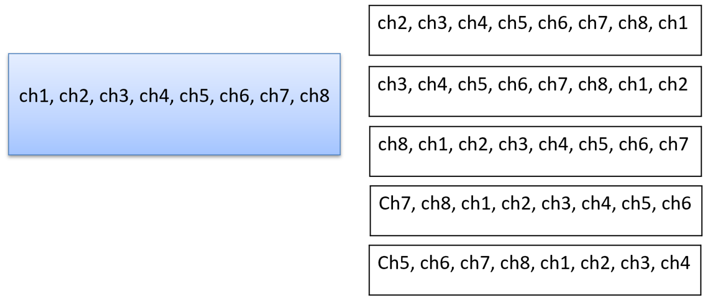

Figure 2.7: Augmentation Operations

After the augmentation, the new dataset was in the below table.

gesture	label	number
finger snap	0	1750
finger beckon	1	1750
palm beckon	2	1750
wave	3	1750
go away	4	1750

Table 2.2: Original Dataset

To get to know the function of the augmentation, in the following section, two models are trained by using the two datasets respectively.

2.2.2 CNN-LSTM network Structure

Convolutional Neural Network known as CNN has a strong feature extraction function. In a convolutional layer, each channel contains a set of parameters known as convolution kernel, which connects to a small part of input data of fixed size from the previous layer. That small part is usually called a patch and moves along all dimensions with a predefined increment size at a time. Every channel's convolution kernel (i.e. connection weights) evolves with the training process[33].Therefore, the convolutional layer is able to capture the special local information of the input data. There are 1D, 2D and 3D convolutional layers. 1D convolutional layer is commonly used in time series data since the kernel only moves along the time axis which preserve the important temporal information.

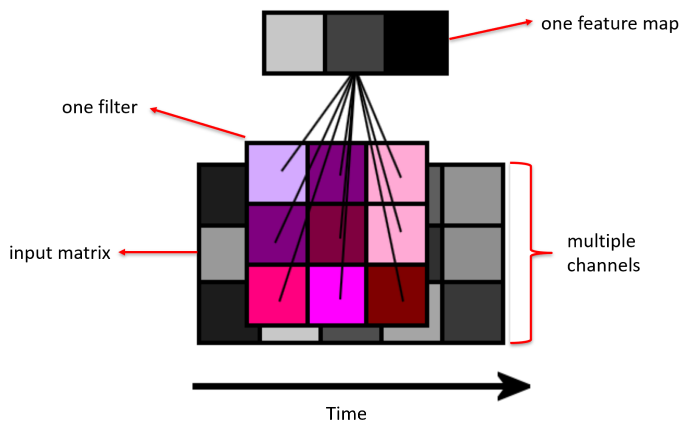

Figure 2.8: 1D Convolution with Multiple Channels

In this research, each gesture's signal is a 8*680 matrix as a input for convolutional layer where 8 is 8 channels from 8 sensors and 680 is time steps. 1D convolutional layer with 8 channels is used here.

For a complete CNN part, after convolutional layer, it follows batch normalization layer and drop out layer.

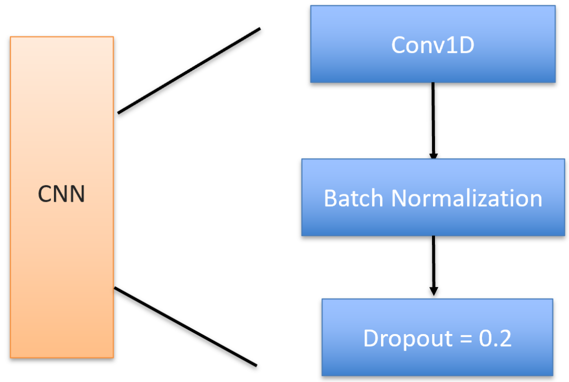

Figure 2.9: CNN Architecture

Broadly speaking, batch normalization is a mechanism that aims to stabilize the distribution (over a minibatch) of inputs to a given network layer during training. This is achieved by augmenting the network with additional layers that set the first two moments (mean and variance) of the distribution of each activation to be zero and one respectively. Then, the batch normalized inputs are also typically scaled and shifted based on trainable parameters to preserve model expressivity[19]. Dropout prevents overfitting and provides a way of approximately combining exponentially many different neural network architectures efficiently. The term "dropout" refers to dropping out units (hidden and visible) in a neural network. By dropping a unit out, we mean temporarily removing it from the network, along with all its incoming and outgoing connections as in the below figure. The choice of which units to drop is random[20].

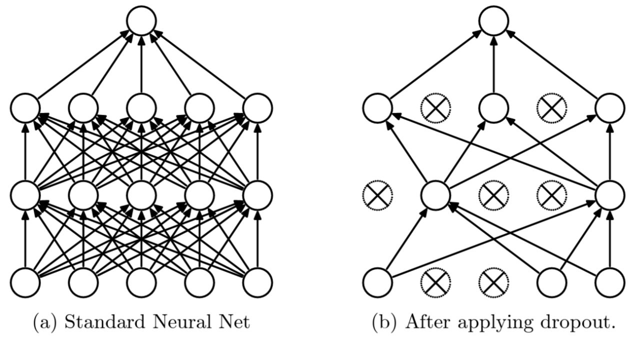

Figure 2.10: Dropout[20]

Long Short-Term Memory Network known as LSTM network is a type of recurrent neural networks.RNN is wildly adopted in research areas concerned with sequential data like text, audio and video[34]. Its internal state could store information which enables it to learn the relevant information of the input data when the gap is not large.

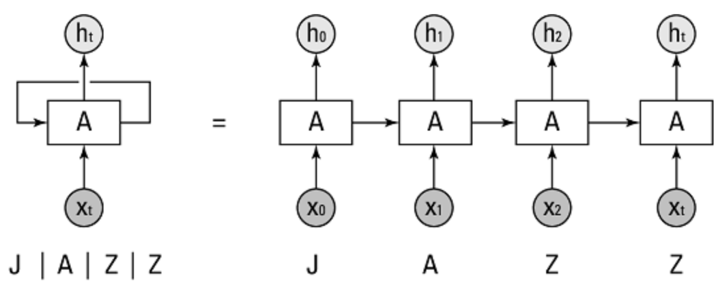

Figure 2.11: RNN Example

. But due to the limited information storage, RNN are unable to learn when there is a big gap. LSTM was proposed for the long-term memory, whose unit includes three gates which enable the network handle long-term dependencies well.

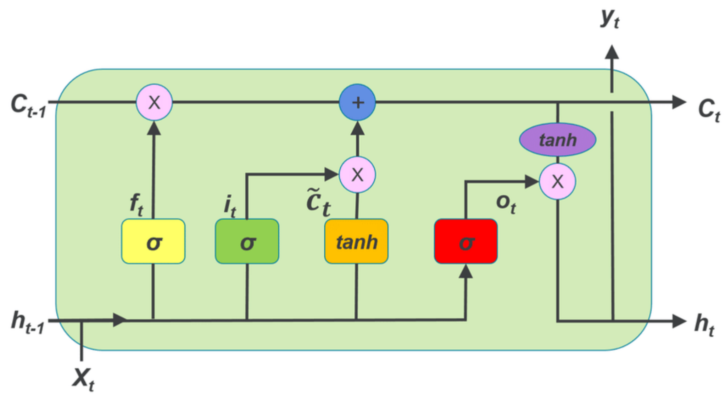

Figure 2.12: A LSTM Unit

X= X(1), X(2) ... ... X(t) is an input sequence, where x(t) is a sequential data of length T. In the LSTM cell, the hidden unit h_t and each parameter at the time t are given by the Eqs. (1), (2), (3), (4), (5), (6) as follows:

$$C_t = f_t * C_{t-1} + i_t * \widetilde{C}_t \tag{2.1}$$

$$h_t = O_t * tanh(C_t) \tag{2.2}$$

$$f_t = \sigma((X_t + h_{t-1})W^f + b^f) \tag{2.3}$$

$$I_t = \sigma((X_t + h_{t-1})W^i + b^i) \tag{2.4}$$

$$O_t = \sigma((X_t + h_{t-1})W^o + b^o) \tag{2.5}$$

$$\widetilde{C}_t = \sigma((X_t + h_{t-1})W^c + b^c) \tag{2.6}$$

*: element-wise multiplication

+: element-wise addition

W^f, W^i, W^o, W^c are weight matrixes which will be calculated when the model are trained. b^f, b^i, b^o, b^c are biases. $I_t$is an input gate determines what part of X_t should be stored in the cell C_t and f_t is a forget gate decide what part of the previous cell C_{t-1} would be forget. The output gate O_t controls what part of the cell state to export as short-term memory h_t. In addition, the cell unit C_t is the recurrent unit, having the activation function $tanh$ is computed from the input of the current frame and the state of the previous frame h_{t-1}. The hidden state h_t is obtained through the $tanh$ activation and memory cell I_t[14].

In a trained network, the values of different gates vary by the information. When the information is relevant, forget gate would be 0 and input gate would be 1, then the new useful information would be stored in the cell and the old information would be forgotten. In contrast, once the information is irrelevant, forget gate would be 1 and input gate would be 0 so that new information would not be stored and the old information would be saved and passed to next cell. Consequently, the entire network can learn easily the long term dependencies between the sequences[9].

Using CNN to generate features which still maintain temporal information, then LSTM is used for the temporal information. A CNN+LSTM neural network is shown in below.

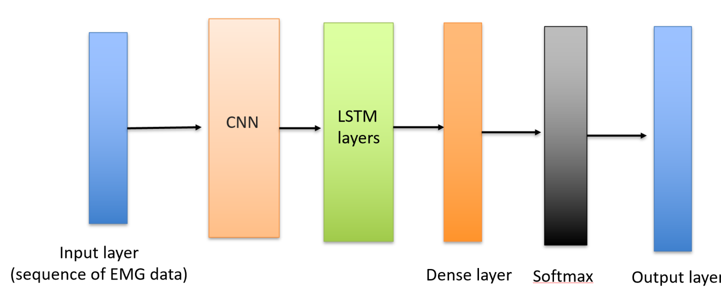

Figure 2.13: CNN-LSTM Neural Network

2.2.3 Training Results

The dataset is split into training dataset and validation dataset, which are 80% and 20% respectively. Therefore, for the original dataset, there are 1750*0.8 = 1400 samples for training and 350 samples for validation. And for dataset after the augmentation, there are 10500*0.8 = 8400 samples for training and 2100 smaples for validation. The validation accuracy of original dataset is appproximate 90% which is lower than the model trained by new dataset. The training accuracy is 95.8% and validation accuracy is 92.7%. Therefore, the model trained by new dataset (after augmentation) is chosen.

2.3 Real Time Gestures Recognition

2.3.1 Real Time Recognition System

Based on the last section which shows a CNN-LSTM model has a good performance on validation data. a real-time recognition system is proposed here.

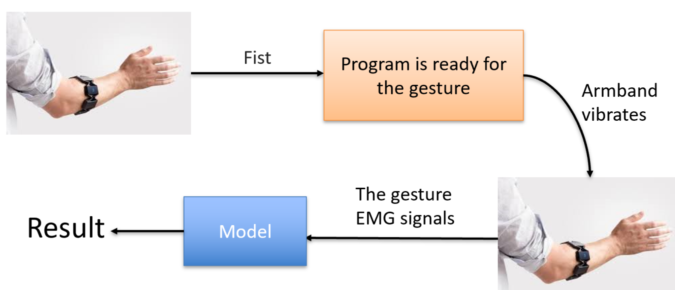

Figure 2.14: Real-time Recognition System

In this system, the fist is a trigger to notify the armband to collect the following gestures data. For the trigger mechanism, a specific value as a threshold is set to identify if the sum of the absolute values of each channel is higher than the threshold, which indicate a muscle activity.Figure below shows the signals for no hand motion and fist. After the fist, the armband vibrates. Then gesture's signals would be recorded in the following 3 seconds. Compared to non-fist system, this system has a clear start for one gesture and less noise of the signals, which increase the classification accuracy.

2.3.2 Data Collection

The test data is from 12 human subjects. The armband is put on the right forearm of every participant. Each gesture at each arm position is performed by the participant 10 times. There are 12*5*5*10 = 3000 samples for test.

2.4 Results

2.4.1 Real-time Classification Accuracy

The overall accuracy is 84.2%. The accuracy for each positions is in the table.

1	2	3	4	5
82%	82.14%	81.68%	84.5%	85.66%

Table 2.3: Accuracy in Each Limb Position

2.4.2 Results Analysis

From the table 2.3, the accuracies vary in a small range among different arm positions. To get more details of the the accuracies, the distribution of accuracies on different positions for different gesture is in the fig 2.15.

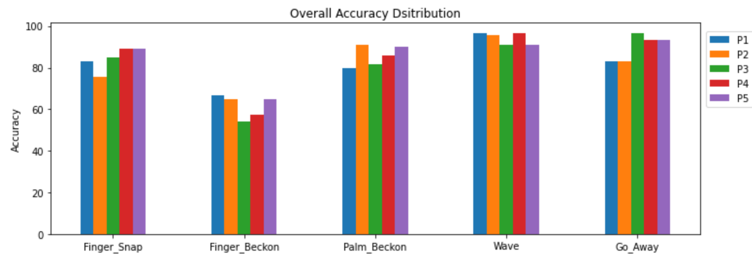

Figure 2.15: Accuracy Distribution

As we can see, for each gesture, the accuracies on different limb positions are similar. In addition, some gestures are easier to classify like wave, and some are not like finger beckon. The muscle activities of different dynamic motions are different may cause different recognition difficulties.

To further explore the recognition condition, the detailed best and worst accuracy of subjects are shown below. The five confusion matrixes represent the accuracy condition at five arm positions.

The best accuracy on one subject is 90.4%.

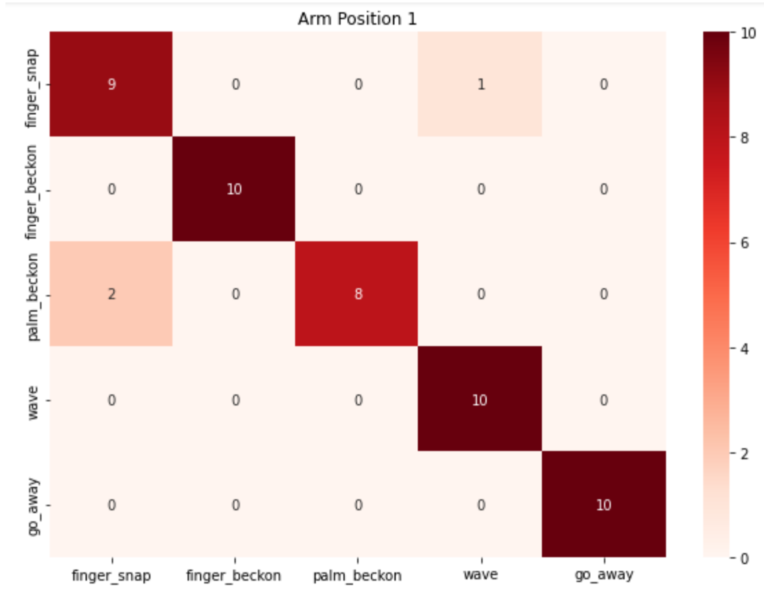

Figure 2.16: Confusion Matrix on P1 position from Best Condition

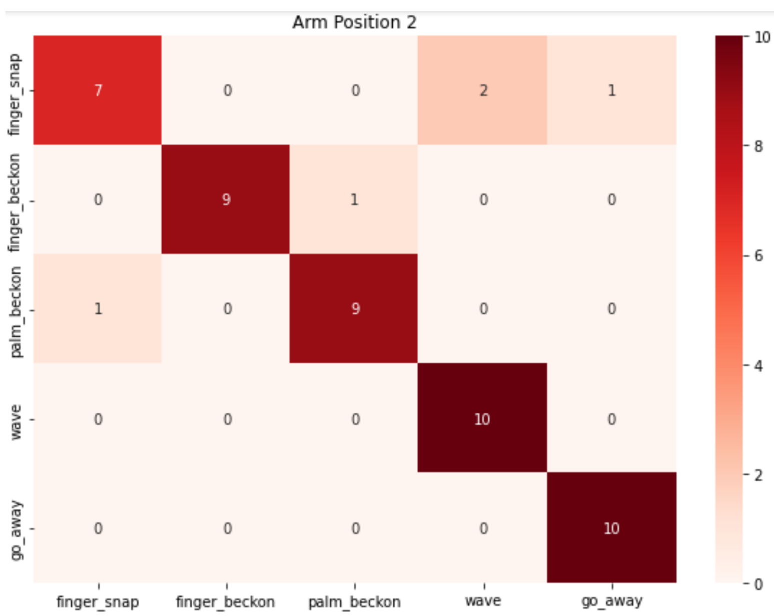

Figure 2.17: Confusion Matrix on P2 position from Best Condition

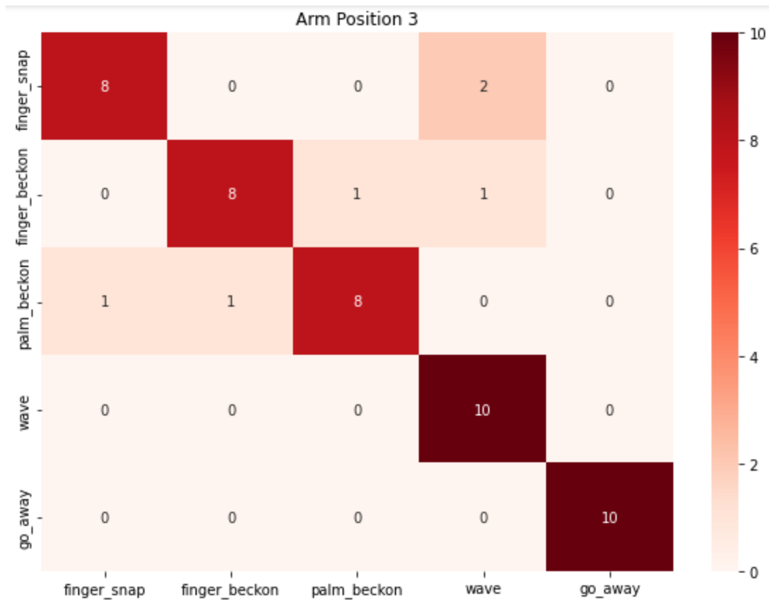

Figure 2.18: Confusion Matrix on P3 position from Best Condition

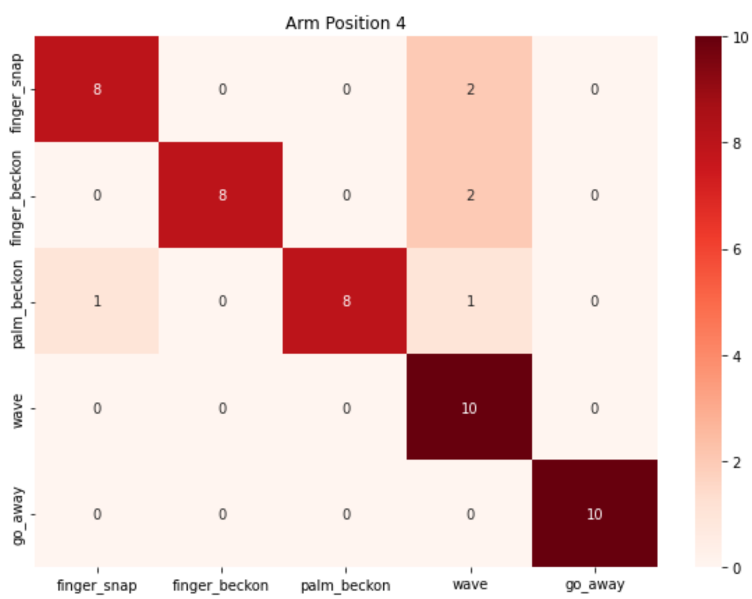

Figure 2.19: Confusion Matrix on P4 position from Best Condition

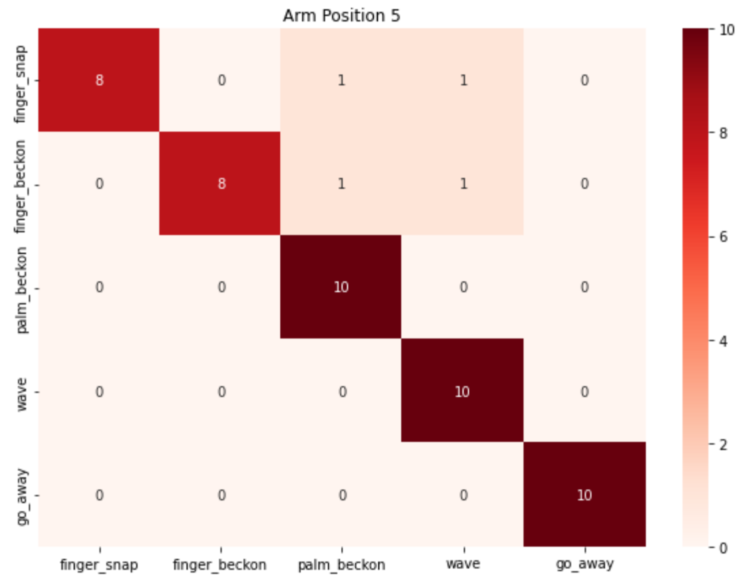

Figure 2.20: Confusion Matrix on P5 position from Best Condition

The worst accuracy on one subject is 78%.

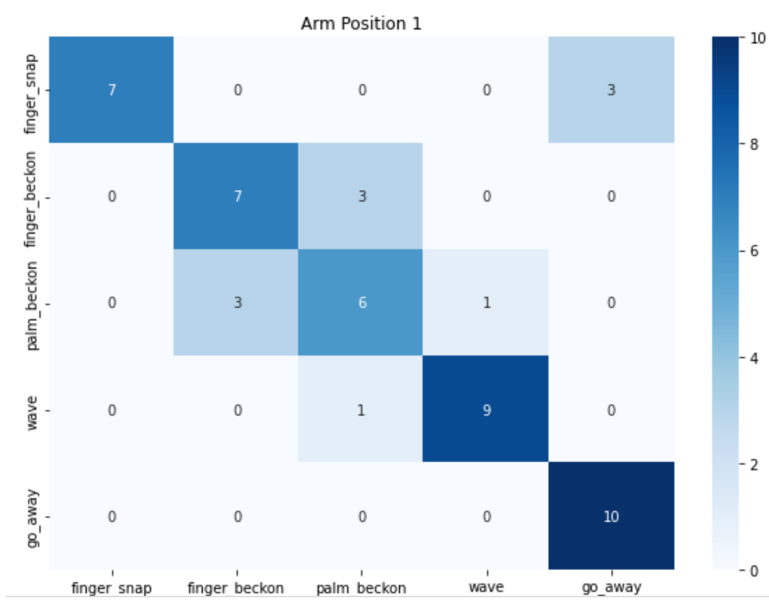

Figure 2.21: Confusion Matrix on P1 position from Worst Condition

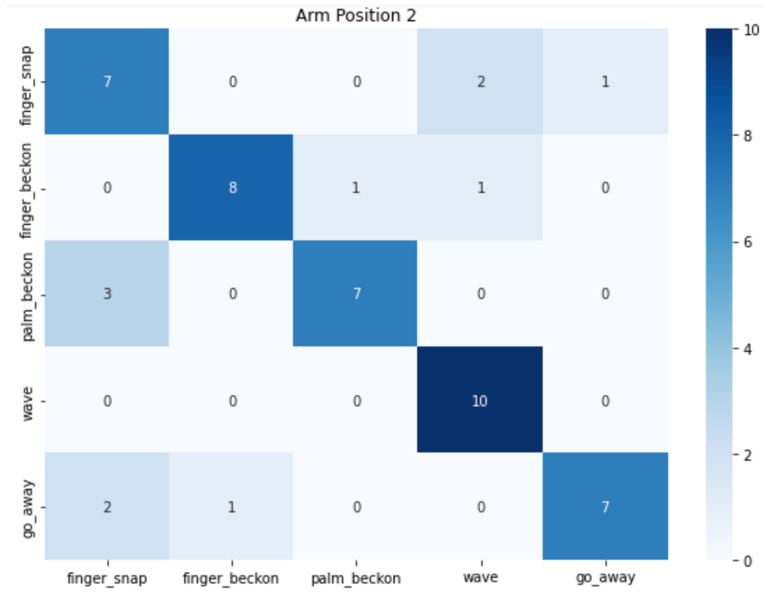

Figure 2.22: Confusion Matrix on P2 position from Worst Condition

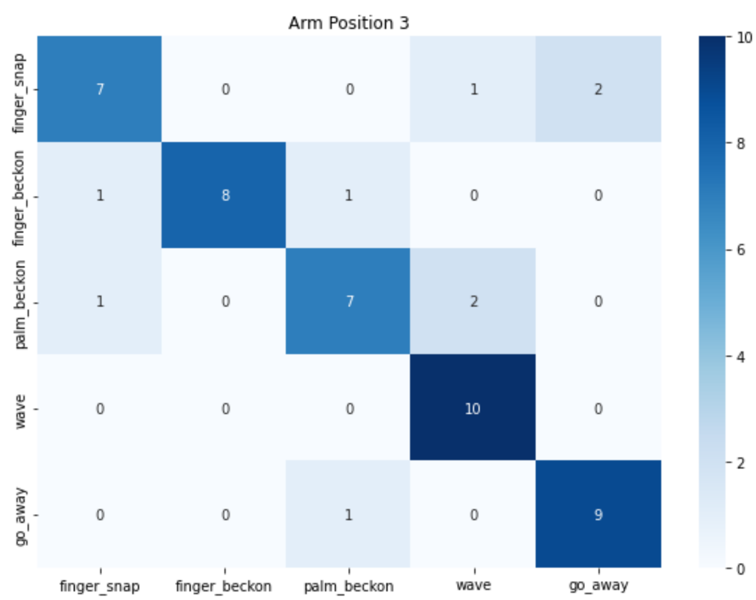

Figure 2.23: Confusion Matrix on P3 position from Worst Condition

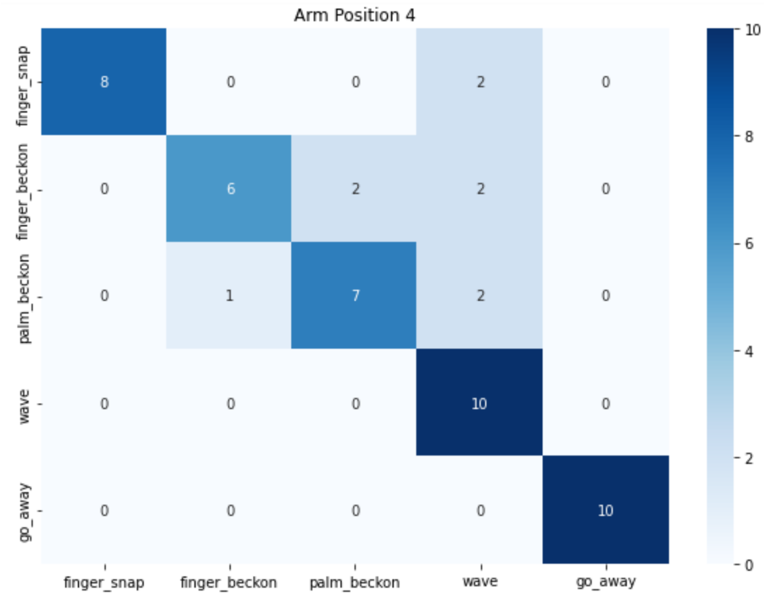

Figure 2.24: Confusion Matrix on P4 position from Worst Condition

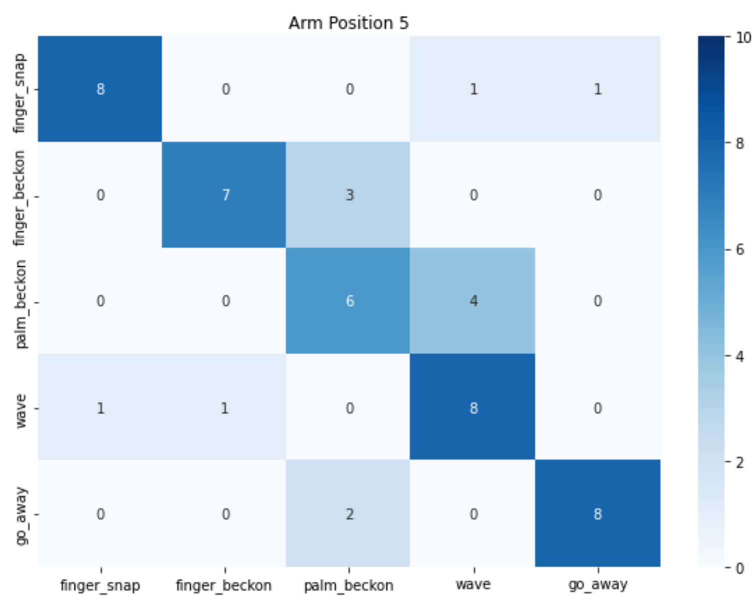

Figure 2.25: Confusion Matrix on P5 position from Worst Condition

From the matrixes, for one subject, performing the gestures in different arm positions would produce different accuracies, but the difference is not very large in this model. The model is relatively arm position-invariant.

For different subjects, their accuracies are different; the model is not entirely subject independently. The accuracy seems to be slightly influenced by subjects' proficiency in gestures. To show this clearly, the accuracies of finger snap from two subjects are shown below.

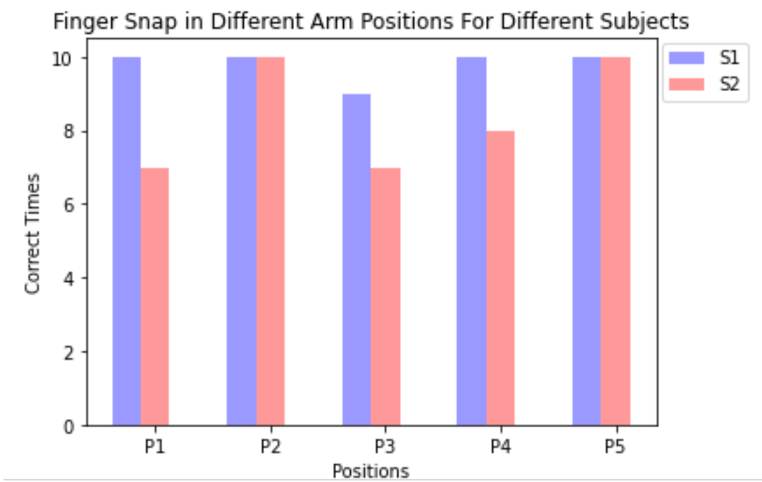

Figure 2.26: Finger Snap Recognition of Two Subjects

In the figure 2.26, finger snap has obvious different accuracies for different subjects. The overall accuracy of S1 is higher than it of S2. And at each arm positions the accuracy of S1 is higher or equal to the accuracy of S2.

2.4.3 Discussion

From the previous analysis, there are three observations:

1. The model is relatively position independent. The effect of arm positions still exists but is small.

2. Different gestures impact recognition accuracy. In this model, influence on gestures may be bigger than positions for classification accuracy. The gestures share less common would get

better results.

3. The model is affected by subject dependency. The best and worst results show it clearly. Due to the different performing habits for different human subjects, more data from various people may help.

2.5 Conclusion

This chapter aims to develop a model that is able to classify gestures performing at multiple arm positions. A CNN-LSTM model is introduced here, trained by data from 7 human subjects with five categories of dynamic motions and each type of gesture includes five arm positions. The training and validation accuracy are both over 90%. Then a real-time recognition based on this model for five dynamic gestures is designed and tested. The overall accuracy is 84.2%, which is lower than the validation accuracy, but it is acceptable. The detailed accuracy distribution shows that the model is relatively position-independent.

3. REAL-TIME HCI SYSTEM FOR A 7-DOF ROBOT ARM CONTROL

3.1 Introduction

Nowadays, human-computer interaction has become more and more popular. Researchers are trying to find a way for humans to communicate intuitively with a device. One approach for this is myoelectric control, which sends human intention to a device like a robot arm using EMG signals generated by muscle activities. Based on the need for myoelectric control, we proposed a real-time control system to control a 6-DoF robot manipulator with 1-DoF gripper. The Myo armband is used during the control process, which collects data of dynamic gestures. After getting the gestures information, the computer would send a particular control command to the robot arm based on the control algorithm.

3.2 Method

In this myoelectric control system, the gestures recognition part is introduced in chapter 2. To estimate the performance of the system, the operation time of the system and joy stick control are recorded.

3.2.1 Proposed Myoelectric Control

The control loop is in the below figure. This is a feed forward control. Once the specific gestures are performed, their signals would be classified by the CNN-LSTM neural network. The model here is trained by the data collected in chapter 2. The output of the model would be an input to the control part, which sends control commands to the robot arm.

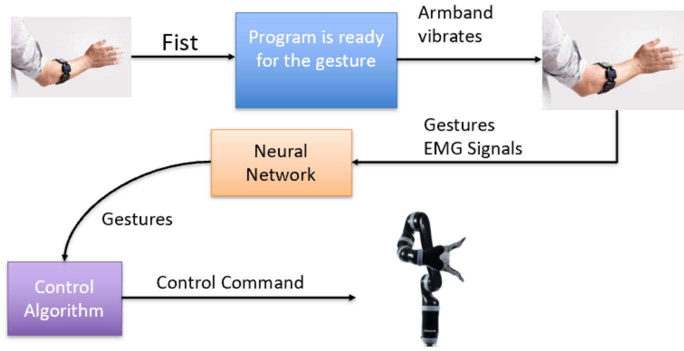

Figure 3.1: Myoelectric Control System for the Kinova Robot Arm

Four gestures are used to complete four pieces of the path to finish a task in the control algorithm. Each motion corresponds to a control command which enables the robot to move along a specific route. Since the classification accuracy is not 100%, there will be wrong recognition results, leading to wrong commands and paths. The mechanism to identify gestures' order is involved in the control algorithm. The i in the algorithm enables the commands to be always sent in the correct order. In other words, the robot would move in a specific trajectory of the task every time. If there is a wrong classification because the output and i do not match, no command would be sent. Human subjects need to repeat the gestures in the order list until the robot move and then perform the next motions. The control loop is in the below figure. This is a feed forward control. Once the specific gestures are performed, their signals would be classified by the CNN-LSTM neural network. The model here is trained by the data collected in chapter 2. The output of the model would be an input to the control part, which sends control commands to the robot arm.

```
i = 1
if fist:
{
if motion == gestures 1 and i == 1:
   control command 1
   i = i + 1
if motion == gestures 2 and i == 2:
   control command 2
   i = i + 1
if motion == gestures 3 and i == 3:
   control command 3
   i = i + 1
if motion == gestures 1 and i == 4:
   control command 4
   i = i + 1
 if i == 5：
     print("task finished")
}
```

Figure 3.2: Control Algorithm

3.2.2 Protocol of Experiment

In order to estimate the performance of the proposed myoelectric control system. A go and grasp task is designed for it. First, the bottled water is put at a certain point where it is reachable for the robot arm. The robot arm starts from its home position, approaches the bottled water and closes its gripper to grab the water. Then move back to the home position with the water. The experiment setup is as below.

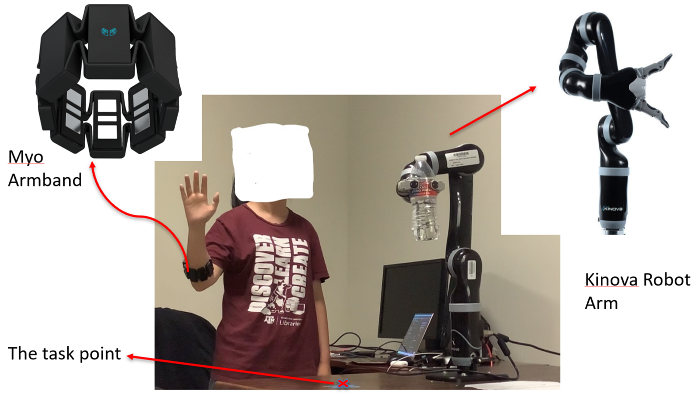

Figure 3.3: Experiment Setup

The trajectory of the myoelectric control is learned from the joystick control mode. So we use the joystick to finish the task first. Then the path of the robot arm is recorded and written as the control commands assigned to pre-defined gestures.

3.2.3 Performance Measurement

To measure the performance of the robot arm, the operation time of the task is chosen. Since the joystick sends the command directly and myoelectric control needs time to perform gestures, data collection, recognition and code processing, it assumed that joystick would use less time than myoelectric control.

The participant would be asked to use a joystick to complete the task and repeat it three times. The participant would be given about 10 minutes to get familiar with the joystick. Then they need to use the myoelectric control system to repeat the task three times. The time of every task would be recorded. If the manipulator knocks down the bottled water during one task, the task will keep going, and the time to reset the bottled water would be in the task time.

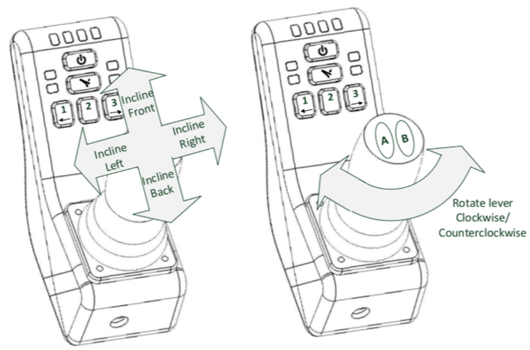

Figure 3.4: Joystick

3.3 Results and Discussion

3.3.1 Experiment Results

The average time of task of 12 human subjects for two control modes is in the below.

Subject	Average Time of Myoelectric Control(s)	Average Time of Joystick Control(s)
1	41.6	37.0
2	36.2	35.6
3	36.6	52.7
4	44.1	26.6
5	57.3	30.06
6	43.2	43.4
7	48.6	43.0
8	40.6	23.0
9	50.6	31.62
10	36.87	41.9
11	28.4	55.8
12	36.4	37.91

Table 3.1: Average Time for Joystick Control and Myoelectric Control

For most subjects, the average time of joystick control is less than myoelectric control. But for some subjects like subject 5, the time for the two control mode is similar. And for subject 6, the myoelectric control time is much smaller than joystick control. The times of three joystick control and three myoelectric control of subject 1, subject 5 and subject 6 are shown as follows, representing three conditions from table 3.1

Control Mode	1	2	3
Myoelectric Control	46.65	48.85	29.4
Joystick Control	41.43	32.49	37.12

Table 3.2: Task Time of Subject 1

Control Mode	1	2	3
Myoelectric Control	60	60.78	51
Joystick Control	44.45	23.98	21.75

Table 3.3: Task Time of Subject 5

Control Mode	1	2	3
Myoelectric Control	51.61	44.72	33.35
Joystick Control	47.53	37.88	46.78

Table 3.4: Task Time of Subject 6

3.3.2 Discussion

There are several observations from the results.

- The majority of results of the human subjects fit the assumption that myoelectric control would use a longer time than the joystick control. But time differences are not huge. Considering the extra processing time of executing the motions, catching the movements, wrong classification and recognition, the differences are acceptable.

- For two control modes, the third task time is always shorter than the first task time. Practice could improve the performance for both control mode.

- From subject 1 and subject 6, the time of joystick control do not reduce monotonically. Because even human subjects would learn from the previous task, they cannot find the best path every time.

Moreover, the fixed trajectory makes it impossible for the myoelectric control algorithm here to complete other tasks unless the commands are modified according to the specific requirements. But this helps the control system perform stably in every task. During the experiments, myoelectric control is able to grab the bottled water successfully without knocking it down. But it happened when using a joystick control because the trajectory of the myoelectric control method is fixed. And the path by using joystick changes every time. In addition, the myoelectric control mode is easier to start than the joystick control. Human subjects do not need to learn the rules and practice for the task. Based on the discussion before, myoelectric control establishes a potential in human-computer interaction.

4. SUMMARY AND CONCLUSIONS

4.1 Summary of the Work

In this research a human computer interaction system is developed based a dynamic gestures recognition system. Compared to other gestures recognition, the proposed method has the following advantages:

- Using CNN to generate the features instead of calculating time or frequency domain features manually before input
- Compared to many researches which focus on gestures recognition at the same limb position and have bad performance when the limb position changes, the model in this research is relatively position-invariant.
- Compared to Dr. Shin's work[3] which involves 3 gestures and 4 limb positions in the real time recognition, this research includes 5 gestures and 5 limb positions. The model is more robust on gestures and limb positions for real-time classification.

Additionally, proposed myoelectric control system works well on go and grasp task. This system could be used in other conditions to translate human intentions to a device.

4.2 Further Work

In this research, there exist two main parts: EMG-based dynamic hand gestures recognition and HCI system for a robot arm control. Based on the results described in the previous chapters, the possible future works are as below:

1. Removing the trigger(fist) in the real-time system.
2. Speeding up processing time. This could be done by using a better GPU.
3. Improving accuracy for the classification. Other advanced methods could be tried such as transfer learning. Also more data from various types of people my help since the amount

of gestures and arm positions is not small and there are considerable differences among different human beings.

4. Increasing recognizable gestures
5. Reducing the arm position's effect to gestures recognition. In other words, the gestures performing at the random arm position could be classified successfully.
6. Decreasing the effect of human subject dependency.
7. Applying the proposed myoelectric control system on other applications like a drone control.
8. Adding feedback in the myoelectric control system.

REFERENCES

[1] Guo, Shuxiang; Pang, Muye; Gao, Baofeng; Hirata, Hideyuki; Ishihara, Hidenori. 2015. "Comparison of sEMG-Based Feature Extraction and Motion Classification Methods for Upper-Limb Movement" Sensors 15, no. 4: 9022-9038

[2] Neta Rabin, Maayan Kahlon, Sarit Malayev, Anat Ratnovsky. 2020. "Classification of human hand movements based on EMG signals using nonlinear dimensionality reduction and data fusion techniques" Expert Systems with Applications, Volume 149

[3] Shin, Sungtae (2016). Myoelectric Human Computer Interaction Using Reliable Temporal Sequencebased Myoelectric Classification for Dynamic Hand Gestures. Doctoral dissertation, Texas A M University. Available electronically from http : / /hdl .handle .net /1969 .1 /174271.

[4] Phinyomark A., Campbell E., Scheme E. 2020. "Surface Electromyography (EMG) Signal Processing, Classification, and Practical Considerations" Biomedical Signal Processing. Series in BioEngineering. Springer, Singapore. https://doi.org/10.1007/978-981-13-9097-51

[5] Englehart, K., Hudgins, B. 2003. "A robust, real-time control scheme for multifunction myoelectric control". IEEE Trans. Biomed. Eng. 50(7), 848–854.

[6] C.L. Ng, M.B.I. Reaz. "Evolution of a capacitive electromyography contactless biosensor: design and modelling techniques". Measurement, 145 (2019), pp. 460-471

[7] M. Simão, N. Mendes, O. Gibaru, P. Neto. "A review on electromyography decoding and pattern recognition for human-machine interaction". IEEE Access, 7 (2019), pp. 39564-39582

[8] M.S. Hazam Majid, W. Khairunizam, A.B. Shahriman, I. Zunaidi, B.N. Sahyudi, M.R. Zuradzman. "EMG feature extractions for upper-limb functional movement during rehabilitation". International Conference on Intelligent Informatics and Biomedical Sciences (ICIIBMS) (2018), pp. 314-320

[9] Ying Wang, Qun Wu, Nilanjan Dey, Simon Fong, Amira S. Ashour. "Deep back propagation–long short-term memory network based upper-limb sEMG signal classification for automated rehabilitation". Biocybernetics and Biomedical Engineering, Volume 40, Issue 3, 2020, Pages 987-1001, ISSN 0208-5216

[10] Oskoei, M.A., Hu, H. "Myoelectric control systems—a survey". Biomed. Signal Process. Control 2(4), 275–294 (2007). https://doi.org/10.1016/j.bspc.2007.07.009

[11] A. K. Mukhopadhyay and S. Samui, "An experimental study on upper limb position invariant EMG signal classification based on deep neural network," Biomed. Signal Process. Control, vol. 55, p. 101669, Jan. 2020, doi: 10.1016/j.bspc.2019.101669.

[12] M. Atzori, M. Cognolato, and H. Müller, "Deep learning with convolutional neural networks applied to electromyography data: a resource for the classification of movements for prosthetic hands," Front. Neurorobotics, vol. 10, pp. 9–9, Sep. 2016, doi: 10.3389/fnbot.2016.00009.

[13] S. Mitra, T. Acharya, Gesture recognition: A survey, IEEE Transactions300on Systems, Man, and Cybernetics, Part C (Applications and Reviews)37 (3) (2007) 311–324.

[14] Safa Ameur, Anouar Ben Khalifa, Med Salim Bouhlel. "A novel hybrid bidirectional unidirectional LSTM network for dynamic hand gesture recognition with Leap Motion". Entertainment Computing, Volume 35, 2020,100373, ISSN 1875-9521

[15] M. Jabbari, R. N. Khushaba and K. Nazarpour, "EMG-Based Hand Gesture Classification with Long Short-Term Memory Deep Recurrent Neural Networks," 2020 42nd Annual International Conference of the IEEE Engineering in Medicine Biology Society (EMBC), Montreal, QC, Canada, 2020, pp. 3302-3305, doi: 10.1109/EMBC44109.2020.9175279.

[16] Campbell, Evan; Phinyomark, Angkoon; Scheme, Erik. 2020. "Current Trends and Confounding Factors in Myoelectric Control: Limb Position and Contraction Intensity" Sensors 20, no. 6: 1613. https://doi.org/10.3390/s20061613

[17] M. A. Ozdemir, D. H. Kisa, O. Guren, A. Onan and A. Akan, "EMG based Hand Gesture Recognition using Deep Learning," 2020 Medical Technologies Congress (TIPTEKNO), 2020, pp. 1-4, doi: 10.1109/TIPTEKNO50054.2020.9299264.

[18] W. Yang, D. Yang, J. Li, Y. Liu and H. Liu, "EMG dataset augmentation approaches for improving the multi-DOF wrist movement regression accuracy and robustness," 2018 IEEE International Conference on Robotics and Biomimetics (ROBIO), 2018, pp. 1268-1273, doi: 10.1109/ROBIO.2018.8664790.

[19]Shibani Santurkar, Dimitris Tsipras, Andrew Ilyas, Aleksander Madry, "How Does Batch Normalization Help Optimization?" arXiv:1805.11604

[20]Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, Ruslan Salakhutdinov, "Dropout: A Simple Way to Prevent Neural Networks from Overfitting" Journal of Machine Learning Research 15 (2014) 1929-1958

[21]K.L. Kilgore, 13 - Sensors for motor neuroprostheses, Editor(s): Andreas Inmann, Diana

Hodgins, In Woodhead Publishing Series in Biomaterials, Implantable Sensor Systems for Medical Applications, Woodhead Publishing, 2013, Pages 401-436, ISBN 9781845699871.

[22]Mohammadreza Asghari Oskoei, Huosheng Hu, Myoelectric control systems—A survey, Biomedical Signal Processing and Control, Volume 2, Issue 4, 2007, Pages 275-294, ISSN 1746-8094.

[23]D. Bai, T. Liu, X. Han, G. Chen, Y. Jiang and Y. Hiroshi, "Multi-Channel sEMG Signal Gesture Recognition Based on Improved CNN-LSTM Hybrid Models," 2021 IEEE International Conference on Intelligence and Safety for Robotics (ISR), 2021, pp. 111-116, doi: 10.1109/ISR50024.2021.9419532.

[24] Y. Wu, B. Zheng and Y. Zhao, "Dynamic Gesture Recognition Based on LSTM-CNN," 2018 Chinese Automation Congress (CAC), 2018, pp. 2446-2450, doi: 10.1109/CAC.2018.8623035.

[25]Juan Cheng, Fulin Wei, Chang Li, Yu Liu, Aiping Liu, Xun Chen, Position-independent gesture recognition using sEMG signals via canonical correlation analysis, Computers in Biology and Medicine, Volume 103, 2018, Pages 44-54, ISSN 0010-4825.

[26]M. T. Wolf et al., "Decoding static and dynamic arm and hand gestures from the JPL BioSleeve," 2013 IEEE Aerospace Conference, 2013, pp. 1-9, doi: 10.1109/AERO.2013.6497171.

[27]B. Milosevic, E. Farella and S. Benatti, "Exploring Arm Posture and Temporal Variability in Myoelectric Hand Gesture Recognition," 2018 7th IEEE International Conference on Biomedical Robotics and Biomechatronics (Biorob), 2018, pp. 1032-1037.

[28]Mads Jochumsen, Muhammad Asim Waris and Ernest Nlandu Kamavuako, "The effect of

arm position on classification of hand gestures with intramuscular EMG", May 2018 Biomedical Signal Processing and Control 43, DOI:10.1016/j.bspc.2018.02.013

[29] D. Xiong, D. Zhang, X. Zhao and Y. Zhao, "Deep Learning for EMG-based Human-Machine Interaction: A Review," in IEEE/CAA Journal of Automatica Sinica, vol. 8, no. 3, pp. 512-533, March 2021, doi: 10.1109/JAS.2021.1003865.

[30]Tsinganos, Panagiotis Cornelis, Bruno Jan, Cornelis Jansen, Bart Skodras, Athanassios. (2018). Deep Learning in EMG-based Gesture Recognition. 107-114. 10.5220/0006960201070114.

[31]Jaramillo-Yánez, Andrés, Marco E. Benalcázar, and Elisa Mena-Maldonado. 2020. "Real-Time Hand Gesture Recognition Using Surface Electromyography and Machine Learning: A Systematic Literature Review" Sensors 20, no. 9: 2467.

[32]Shanmuganathan, V., Yesudhas, H.R., Khan, M.S. et al. R-CNN and wavelet feature extraction for hand gesture recognition with EMG signals. Neural Comput Applic 32, 16723–16736 (2020).

[33]D. Huang and B. Chen, "Surface EMG Decoding for Hand Gestures Based on Spectrogram and CNN-LSTM," 2019 2nd China Symposium on Cognitive Computing and Hybrid Intelligence (CCHI), 2019, pp. 123-126, doi: 10.1109/CCHI.2019.8901936.

[34] Yong Yu, Xiaosheng Si, Changhua Hu, Jianxun Zhang; A Review of Recurrent Neural Networks: LSTM Cells and Network Architectures.Neural Comput 2019; 31 (7): 1235–1270. doi: https://doi.org/10.1162/necoa01199

[35]Tribikram Pradhan, Prashant Kumar, Sukomal Pal, CLAVER: An integrated framework of

convolutional layer, bidirectional LSTM with attention mechanism based scholarly venue recommendation, Information Sciences, Volume 559, 2021, Pages 212-235, ISSN 0020-0255.