

SIGN LANGUAGE RECOGNITION FOR SENTENCE LEVEL CONTINUOUS SIGNINGS

Ishika Godage, Ruvan Weerasignhe and Damitha Sandaruwan

University of Colombo School of Computing, Colombo 07, Sri Lanka

ABSTRACT

It is no doubt that communication plays a vital role in human life. There is, however, a significant population of hearing-impaired people who use non-verbal techniques for communication, which a majority of the people cannot understand. The predominant of these techniques is based on sign language, the main communication protocol among hearing impaired people. In this research, we propose a method to bridge the communication gap between hearing impaired people and others, which translates signed gestures into text. Most existing solutions, based on technologies such as Kinect, Leap Motion, Computer vision, EMG and IMU try to recognize and translate individual signs of hearing impaired people. The few approaches to sentence-level sign language recognition suffer from not being user-friendly or even practical owing to the devices they use. The proposed system is designed to provide full freedom to the user to sign an uninterrupted full sentence at a time. For this purpose, we employ two Myo armbands for gesture-capturing. Using signal processing and supervised learning based on a vocabulary of 49 words and 346 sentences for training with a single signer, we were able to achieve 75-80% word-level accuracy and 45-50% sentence level accuracy using gestural (EMG) and spatial (IMU) features for our signer-dependent experiment.

KEYWORDS

Sign Language, Word-Level Recognition, Sentence-Level Recognition, Myo Armband, EMG, IMU, Supervised Learning.

1. INTRODUCTION

According to Wikipedia, communication is the act of conveying meanings from one entity or group to another through the use of mutually understood signs and semiotic rules. There are many approaches for communication. Such as voice and speech, writing, manual signs, and gestures etc.

These communication methods can be divided into two different forms. The first is verbal communication methods and the second, non-verbal communication methods. Verbal communication describe the processes of communicating with words, whether written or spoken. Non-verbal communication is defined as the process of using the wordless message to generate meaning. Examples of nonverbal communication include haptic communication, chronemic communication, gestures, body language, facial expressions and eye contact.

There is no doubt that communication plays a vital role in human life. Communication helps to share information and knowledge. It also helps humans make new relationships, and express ideas, feelings, emotions and thoughts.

There are two conditions to be satisfied for a successful communication, namely,

- 1) There must be at least two parties who involve in the communication
- 2) Both parties must use a common communication platform

Most ordinary people (without any hearing/speaking disability) use verbal communication methods (e.g. voice and speech) for their communication. However, deaf and speaking-impaired people use non-verbal communication methods (mostly signs and gestures) for their communication. These two groups (ordinary people and deaf and speaking impaired people) therefore use different platforms for their communication. Because of this problem, there is a communication barrier between these two people groups when they need to communicate with each other.

Figure 1 demonstrates the communication barrier between a deaf person and an ordinary person. A deaf person uses sign language and the ordinary person uses voice or text. As mentioned previously, there are two conditions to be satisfied for a successful communication. Ordinary person to ordinary person and deaf person to deaf person communications satisfies those conditions. However, deaf person to ordinary person communication does not satisfy the second condition of using a common communication platform. In the figure, they attempt to use sign language as their communication platform which deaf person can understand but the ordinary person cannot. Since they do not use a common communication platform, their communication fails.

In the proposed solution, we create a sign language translator which can recognize sentence level continuous signings and translates them into a natural language. While it is translating signs into text/voice, it improves the practical usability of the system by employing a simple wearable device to capture the signs.

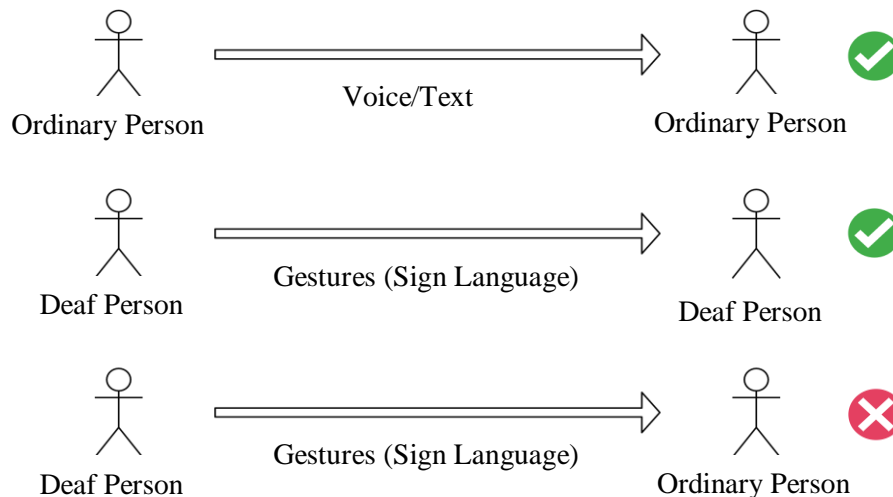


Figure 1. Communication methods

Figure 2 elaborates how our solution simulates a common platform. It captures the signs of a sign language and translates them into text/voice. As a result, the ordinary person would be able to understand the sign that the deaf person has performed. In this research, we do not concern ourselves with the communication in the other direction: i.e. from the hearing person to the hearing-impaired person. In our solution, we capture sentence level continuous signings in Sri Lankan Sign Language and translate them into Sinhala natural language.

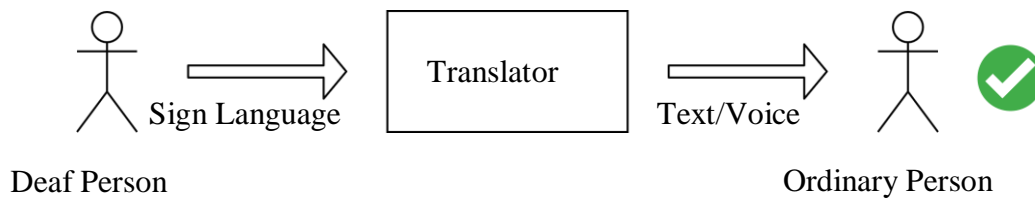


Figure 2. The communication method of the proposed solution

2. BACKGROUND

2.1. Sign Language

There are around 300 sign languages in the world [1]. These languages are different from each other. Word order in sentences can differ between these languages as well as from written text. Sign language is a visual language that incorporates gestures, facial expressions, head movements, body language and even the space around the speaker. Hand signs are the foundation of the sign language. Many signs are iconic, meaning the sign uses a visual image that resembles the concept it represents. Actions are often expressed through hand signals that mimic the action being communicated.

2.2. Sri Lankan Sign Language (SLSL)

Sri Lankan Sign Language is a visual-gestural language based on hand movements and the body (including facial expressions, lip moments, head movement). In Sri Lankan Sign Language, it can represent alphabets of normal languages (Sinhala, English) and it can represent other signs for each word. Currently, Sri Lankan Sign Language contains around 2000 signs [18, 19]. It also has regional signs across Sri Lanka.

British introduced the sign language to Sri Lanka. Hence, Sri Lankan Sign Language has been developed for years with the influence of British Sign Language (BSL). Because of that, there are some similarities in between Sri Lankan Sign Language and British Sign Language.

2.3. Electromyography (EMG) and Initial Measurement Units (IMU).

2.3.1. Electromyography (EMG)

Electromyography (EMG) is the detection and recording of the electrical signal produced by muscle tissue as it contracts. EMG depends on several factors such as the thickness and temperature of the skin, the thickness of the fat between the muscle and the skin, the velocity of the blood flow, and location of the sensors.

2.3.2. Initial Measurement Unit (IMU)

According to the Wikipedia, an IMU is an electronic device that measures and reports a body's specific force, angular rate, and sometimes the magnetic field surrounding the body, using a combination of accelerometers and gyroscopes, sometimes also magnetometers.

3. RELATED WORK

3.1. Word Level Sign Language Recognition Systems

3.1.1. Kinect Device Based Solution [2]

Kalin Stefanov and Jonas Beskow proposed a method for automatic recognition of isolated Swedish Sign Language (SSL) signs for the purpose of educational signing-based games. Two datasets consisting of 51 signs have been recorded from a total of 7 (experienced) and 10 (inexperienced) adult signers. Signer-dependent recognition rate is 95.3% for the most consistent signer. Hidden Markov Model (HMM) have been used as the model. Signer-independent recognition rate is on average 57.9% for the experienced signers and 68.9% for the inexperienced.

3.1.2. Data Glove Device Based Solution [3]

Wu jiangqin et al proposed a Chinese Sign Language recognition system based on data glove. In this paper, a simple word-level sign language recognition system is presented. 26 sign language words were used for this experiment. There are primarily 3 methods were used for sign language recognition. Such as template matching, neural networks and Hidden Markov Model HMM. The Recognition rate of testing samples is over 90%.

3.1.3. Leap Motion Device Based Solution [4]

Deepali Naglot and Milind Kulkarni proposed a system for recognition of 26 different alphabets of American Sign Language (ASL) using leap motion controller (LMC). LMC is 3D non-contact motion sensor which can track and detects hands, fingers, bones and finger-like objects. Multi-Layer Perceptron (MLP) is executed on a dataset of total 520 samples and Recognition rate of the proposed system is 96.15%.

3.1.4. Image/Video Based Solution (Vision Based) [5]

Manar et al introduce the use of different types of neural networks in human hand gesture recognition for static images as well as for dynamic gestures. A static gesture is a particular hand movement represented by a single image, while a dynamic gesture is a moving gesture represented by a sequence of images. This work focuses on the ability of neural networks to assist in Arabic Sign Language (ArSL) hand gesture recognition. This work focuses on the 28 letters of the Arabic alphabet. Fully recurrent architecture has had a performance with an accuracy rate of 95% for static gesture recognition.

3.1.5. EMG and IMU Based Solution [6]

Jian Wu et al proposed a real-time American SLR system leveraging fusion of surface electromyography (sEMG) and a wrist-worn inertial sensor at the feature level. A feature selection is provided for 40 most commonly used words and for four subjects. SVM was used as the classifier model. Their system achieves 95.94% recognition rate.

3.2. Sentence Level Sign Language Recognition Systems

3.2.1. Kinect Device Based Solution [7]

Edon Mustafa and Konstantinos Dimopoulos developed a system which uses SigmaNIL framework to recognize alphabet, number, word, and sentence of Kosova Sign Language. The recognition rate for one sentence from three testers is 73%.

3.2.2. Data Glove Device Based Solution [8]

Noor Tubaiz et al proposed a glove-based Arabic Sign Language recognition system using a novel technique for sequential data classification. The dataset contains 40 sentences using an 80-word lexicon. Data labelling is performed using a camera to synchronize hand movements with their corresponding sign language words. Modified k-Nearest Neighbor (MKNN) approach is used for classification. The proposed solution achieved a sentence recognition rate of 98.9%.

3.2.3. Image/Video Based Solution (Vision Based) [9]

Daniel Kelly et al presented a multimodal system for the recognition of manual signs and non-manual signals within continuous Irish sign language sentences. In this paper, they proposed a multichannel HMM-based system to recognize manual signs (hand gestures) and non-manual signals (E.g. facial expressions, head movements, body postures, and torso movements). Signer has to make pauses between words, to segment the words in a sentence. They have considered about 8 words. Using 4 words at a time they have created sentences. Their system achieved a detection ratio of 95.7%.

3.2.4. EMG and IMU Based Solution [10]

Xu Zhang et al presented a framework for hand gesture recognition based on the information fusion of a three-axis accelerometer (ACC) and multichannel electromyography (EMG) sensors. In this framework, the start and end points of meaningful gesture segments are detected automatically by the intensity of the EMG signals. 72 Chinese Sign Language (CSL) words and 40 CSL sentences are classified using a decision tree and multi-stream hidden Markov models. Overall word recognition accuracy is 93.1% and a sentence recognition accuracy is 72.5%.

We observed that EMG and IMU based solutions have sufficient accuracy, they can be enhanced as mobile solutions and they improve the practical usability of the system. Therefore, we planned to use EMG and IMU based device for this research. Instead of using electrodes, we chose Myo gesture control armband which is a commercial-off-the-shelf device for this research as the data capturing device [20]. The following 3.3 and 3.4 sections show the existing works which use the Myo gesture control armband as the data capturing device.

3.3. Word Level Sign Language Recognition Systems Using Myo Gesture Control Armband

Celal Savur and Ferat Sahin proposed a system [11] to identify recognize the American Sign Language alphabet letters (26) and a one for the home position. As a classification method, Support Vector Machine (SVM) and Ensemble Learning algorithm were used. Accuracies are 80% and 60.85% respectively. Only one hand use to perform gestures.

Prajwal Paudyal et al proposed SCEPTRE [12] which utilizes two non-invasive wrist-worn devices (Both arms were used) to decipher gesture-based communication. The system uses a

multitiered template-based comparison system for classification on input data from accelerometer, gyroscope, and electromyography (EMG) sensors. They tried to identify 20 signs of American Sign Language and the system was able to achieve an accuracy of 97.72 % for ASL gestures.

3.4. Sentence Level Sign Language Recognition Systems Using Myo Gesture Control Armband

Best of our knowledge, we were unable to find literature which tries to recognize sentence level continuous signing using Myo gesture control armband. However, the proposed system used Myo gesture control armband to recognize sentence level continuous signings.

3.5. Related Research Projects about Sri Lankan Sign Language Interpretation

Dulan Manujith and Nihal Kodikara invented a Sinhala figure spelling interpretation system [13]. Herath et al proposed an image-based sign language recognition system for Sinhala sign language [14]. Kulaveerasingam et al invented a system which is gesture based intercommunication platform for hearing-impaired people [15]. Pumudu Fernando and Prasad Wimalaratne proposed sign language translation approach to Sinhalese language [16]. Madushanka et al introduced a framework for Sinhala sign language recognition and translation using a wearable armband [17].

4. DESIGN

Initially, we created a framework for recognizing sentence level continuous signings of Sri Lankan Sign Language and for translating them into a natural language (Sinhala). Figure 3 shows the flow of the study.

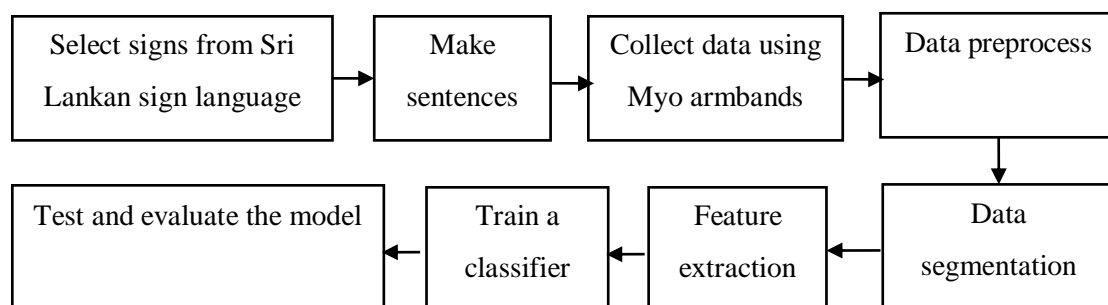


Figure 3. The flow of the study

4.1. Creating a Sri Lankan Sign Language Corpus

Sri Lankan Sign Language was selected as the sign language for this research project. There are around 2000 signs in Sri Lankan Sign Language [18, 19]. For this study, 49 of them were selected. The selected signs are common and useful signs in our day to day life. These 49 signs include nouns, pronouns nouns, and verbs only.

In this sentences creation process, we used SOV (Subject + Object + Verb) structure as the structure of the sentences. Each sentence consists of three words. Those are subject, object, and verb. 346 sentences were created using those 49 selected signs. Table 1 shows that 49 words which were selected as subjects, objects, and verbs. Figure 4 shows the frequencies of each sign which are selected for this study.

Table 1. Selected signs.

Class	Subject	Class	Object	Class	Verb
1	මම (<i>mama</i>) / I	19	මේසය (<i>mēsaya</i>) / Table	36	අදිනවා (<i>adinavā</i>) / Pull
2	අපි (<i>api</i>) / We	20	පුටුව (<i>puṭuva</i>) / Chair	37	අඳිනවා (<i>añḍinavā</i>) / Draw
3	ඔහු (<i>ohu</i>) / He	21	බර (<i>bara</i>) / Weight	38	අරිනවා (<i>arinavā</i>) / Open
4	ඇය (<i>āya</i>) / She	22	චිත්‍රය (<i>citraya</i>) / Painting	39	ඉගෙනගන්නවා (<i>igenagannavā</i>) / Study
5	ඔවුන් (<i>ovun</i>) / They	23	දොර (<i>dora</i>) / Door	40	ඉරනවා (<i>iranavā</i>) / Tear
6	අම්මා (<i>ammā</i>) / Mother	24	ජනේලය (<i>janēlaya</i>) / Window	41	උණුකරනවා (<i>uṇukaranavā</i>) / Boil
7	තාත්තා (<i>tāttā</i>) / Father	25	ඉංග්‍රීසි (<i>iṅgrīsi</i>) / English	42	උයනවා (<i>uyanavā</i>) / Cook
8	අක්කා (<i>akkā</i>) / Elder Sister	26	පොත (<i>pota</i>) / Book	43	එල්ලනවා (<i>ellanavā</i>) / Hang
9	අයියා (<i>ayiyā</i>) / Elder Brother	27	පත්තරය (<i>pattaraya</i>) / News Paper	44	පලනවා (<i>palanavā</i>) / Split
10	නංගී (<i>naṅgī</i>) / Younger Sister	28	කඩදාසිය (<i>kaḍadāsiya</i>) / Paper	45	යනවා (<i>yanavā</i>) / Go
11	මල්ලී (<i>mallī</i>) / Younger Brother	29	වතුර (<i>vatura</i>) / Water	46	බොනවා (<i>bonavā</i>) / Drink
12	දුව (<i>duva</i>) / Daughter	30	මාලු (<i>mālu</i>) / Fish	47	ලියනවා (<i>liyanavā</i>) / Write
13	පුතා (<i>putā</i>) / Son	31	එළවළු (<i>eḷavalu</i>) / Vegetable	48	වීසිකරනවා (<i>visikaranavā</i>) / Throw
14	නැන්දා (<i>nændā</i>) / Aunt	32	රෙදි (<i>redi</i>) / Clothes	49	සිටුවනවා (<i>siṭuvanavā</i>) / Plant
15	මාමා (<i>māmā</i>) / Uncle	33	දර (<i>dara</i>) / Firewood		
16	ආච්චී (<i>āccī</i>) / Grand Mother	34	ගෙදර (<i>gedara</i>) / Home		
17	සීයා (<i>sīyā</i>) / Grand Father	35	ගස (<i>gasa</i>) / Tree		
18	යාලුවා (<i>yāluvā</i>) / Friend				

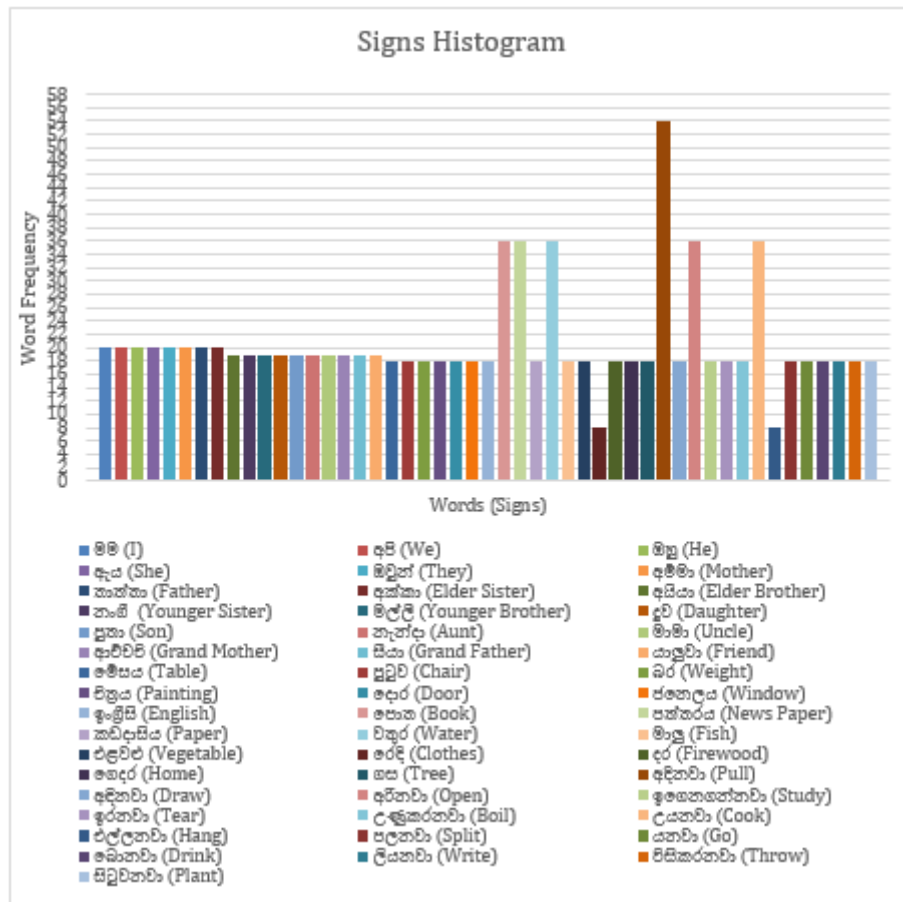


Figure 4. Signs histogram

4.2. Device Selection

In this research, Myo gesture recognition armband was selected as our data capturing device. This device was developed and introduced by Thalmic Labs Inc as a new way of using hand gestures to interact with computers and mobile devices (especially as an input/controlling device). Before Myo armband was selected as the data capturing device we had to consider three things.

- 1) Sign recognition accuracies of EMG and IMU based techniques
- 2) The mobility of the device
- 3) User convenience of the device

A Myo armband gives EMG and IMU data. There are 8 EMG signals and 10 IMU signals. Myo armband has 8 EMG sensors, 1 accelerometer sensor, 1 gyroscope sensor, and 1 magnetometer sensor.

4.3. Data Collection

After creating the sentences, data were collected using a sign language interpreter. Myo armband was used as our data collection device. Since we use both arms to perform signs, two Myo armbands were used. The Myo armbands were connected to two different computers using two Bluetooth adapters. After connecting armbands with the computers via Bluetooth, by running a

C++ program with the help of Myo SDK, EMG and IMU data were captured and stored in CSV files separately.

To avoid the speed variations when performing signs, a metronome was used as a supporting tool. A metronome is a device that produces an audible click or another sound at a regular interval that can be set by the user, typically in beats per minute. Thus, then the signer performs all the sentences (346) in the same rhythm. In the metronome, 5seconds were considered. In each second, the signer performed the particular sign according to the sentence. Rest sign was performed in 1st and 5th seconds. The first sign, second sign and third sign in the sentence were performed in 2nd, 3rd, and 4th seconds respectively. The metronome was screened in a separate display while performing signs. Figure 5 depicts the data collection design. Moreover, it's necessary to have a common starting and ending point for all the sentence to recognize a particular sentence when it gets started or ended.

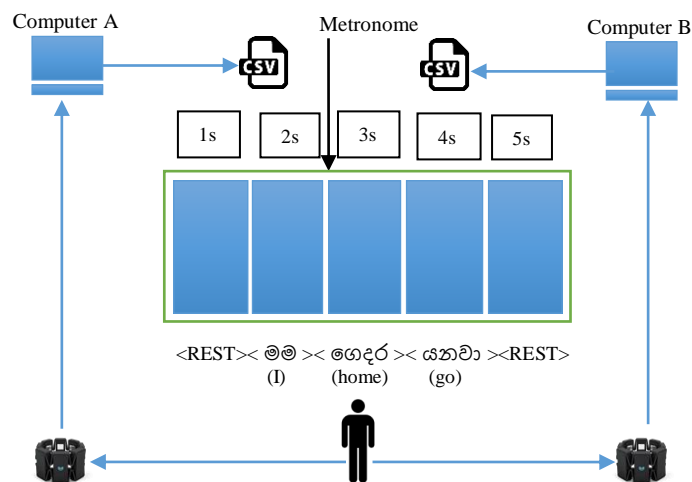


Figure 5. Data collection design

4.4. Data Pre-process

It is not a better idea to use raw data as it is for the classification process. Because there are many issues with the raw data. Such as unwanted data (e.g. noise), incomplete data, inconsistent data etc. Our collected data contain EMG data and IMU data. Therefore, we had to use digital signal processing (DSP) techniques to pre-process the collected raw data.

1) Pre-processing methods of EMG data

- Resampled the signals
- Removed the DC offset
- Applied full wave rectification
- Used low pass Butterworth filter
- Conducted zero-phase digital filtering

2) Pre-processing methods of IMU data

- Used the moving average filter

4.5. Data Segmentation

Each sentence has 3 words and each signal contains 3 signs. The aim of this step is to segment each sign separately. As a result of segmentation, there will be 3 segments per sentence. For a single sentence, one Myo armband gives 18 signals (8 EMG and 10 IMU). Since two armbands were used for the data collection, we had to segment 36 signals and saved the segmented signs separately.

We carried out a manual segmentation method. Since we used a metronome as a supporting tool, we knew that the length of a sentence which is 5 seconds and the rest sign was performed in 1st and 5th second. First, second and third signs in the sentence were performed in 2nd, 3rd, and 4th seconds respectively. Since 2nd, 3rd, and 4th seconds contain the valid signs of a particular sentence, all signals were segmented within that each time period.

4.6. Feature Extraction

Features are the unique attributes of a particular data. Features are the input to the machine learning models. According to the existing work [17], we selected the following features.

- 1) Mean Absolute Value

$$MAV = \frac{1}{N} \sum_{n=1}^N |x_n|$$

- 2) Variance

$$VAR = \frac{1}{N-1} \sum_{n=1}^N x_n^2$$

- 3) Standard Deviation

$$SD (\sigma) = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2}$$

Since we are interested in 3 features, there are 108 (=36*3) features for an each sign.

4.7. Feature Reduction and Feature Selection Methods

Feature reduction and selection are two techniques of feature engineering. Basically, what it does is, identifying the most important features. In this study, we have used below feature reduction and feature selection methods.

- 1) PCA - Principal Component Analysis
- 2) US - Univariate Selection
- 3) SVD - Singular Value Decomposition
- 4) RFE - Recursive Feature Elimination
- 5) RF - Random Forest

4.8. Machine Learning Model Training

In this research project, we applied supervised learning techniques. Because this is a classification problem. Therefore, we had to train a classifier. We selected 5 classifiers and trained all them using the training data. The training data set was composed of 241 sentences

(723 words) and the test data set included 105 sentences (315 words). We got the 10-fold cross-validation accuracy of all the models and selected the highest accuracy given classifier as the final classifier for this study.

- 1) NB - Gaussian NB
- 2) LDA - Linear Discriminant Analysis
- 3) RFC - Random Forest
- 4) LR - Logistic Regression
- 5) RC - Ridge Classifier

Then we created a framework for carrying out the task covered in previously in real-time. Real-time hand gesture recognition is one of the most challenging research areas in the human computer interaction field. In the initial experiment, we conducted an offline training and offline testing. However, in this experiment, we conducted offline training and online testing.

Data capturing, pre-processing, segmentation and feature extraction techniques are the same as the initial experiment. However, we used a previously trained model to recognize and translate sentence level continuous signings in real-time. Figure 6 shows the flow diagram of the real-time classification experiment.

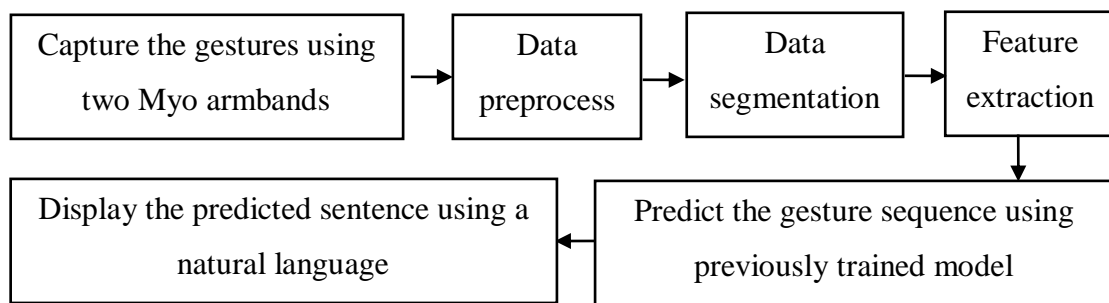


Figure 6. Flow diagram of real-time classification experiment

5. RESULTS AND EVALUATION

As the first experiment, we input all features to each model. As mentioned in the design section, 5 models were trained using 5 Machine Learning algorithms and feature vectors. We then compared the cross-validation accuracies of each model. The cross-validation results are given in Table 2. The Linear Discriminant Analysis classifier performed best for average 10-fold cross-validation accuracy.

Since Linear Discriminant Analysis (LDA) classifier showed the highest cross-validation accuracy, we selected LDA as the classifier for the final study. We then trained the LDA classifier using all the features (108). Finally, we got a word level testing accuracy of between 75% - 80%. The full sentence level accuracy varies from 45% to 50%. Figure 7 shows the confusion matrix of the LDA classifier while Table 3 shows the average precision, recall, and F1-score of the LDA model.

Table 2. Average 10-fold cross validation accuracy.

Model	Average 10-Fold cross-validation score	standard deviation
Logistic Regression (LR)	0.597774	0.047929
Linear Discriminant Analysis (LDA)	0.761796	0.064298
Ridge Classifier (RC)	0.675114	0.067444
Random Forest Classifier (RFC)	0.603387	0.052306
Gaussian Naïve Bayes (NB)	0.560731	0.067124

Table 3. Average precision, recall and F1-score of the LDA model.

Precision	Recall	F1-score
0.81	0.79	0.79

Precision - Precision is the ratio of correctly predicted positive observations to the total predicted positive observations. The question that this metric answer is of all signs that labeled as a correct sign, how many actually correct signs? High precision relates to the low false positive rate. We have got 0.81 average precision value which is pretty good.

Recall (Sensitivity) - Recall is the ratio of correctly predicted positive observations to all observations in actual class. The question recall answer is: Of all the signs that have true class, how many did we label? We have got an average recall of 0.79 which is good for this model as it's above 0.5.

F1-score - F1-score is the weighted average of Precision and Recall. Therefore, this score takes both false positives and false negatives into account. Intuitively it is not as easy to understand as accuracy, but F1-score is usually more useful than accuracy, especially if you have an uneven class distribution. Accuracy works best if false positives and false negatives have a similar cost. If the cost of false positives and false negatives are very different, it's better to look at both Precision and Recall. In our case, the average F1-score is 0.79. Table 4, 5 and 6 show the class, precision, recall and F1-score of the LDA model in different F1-score ranges.

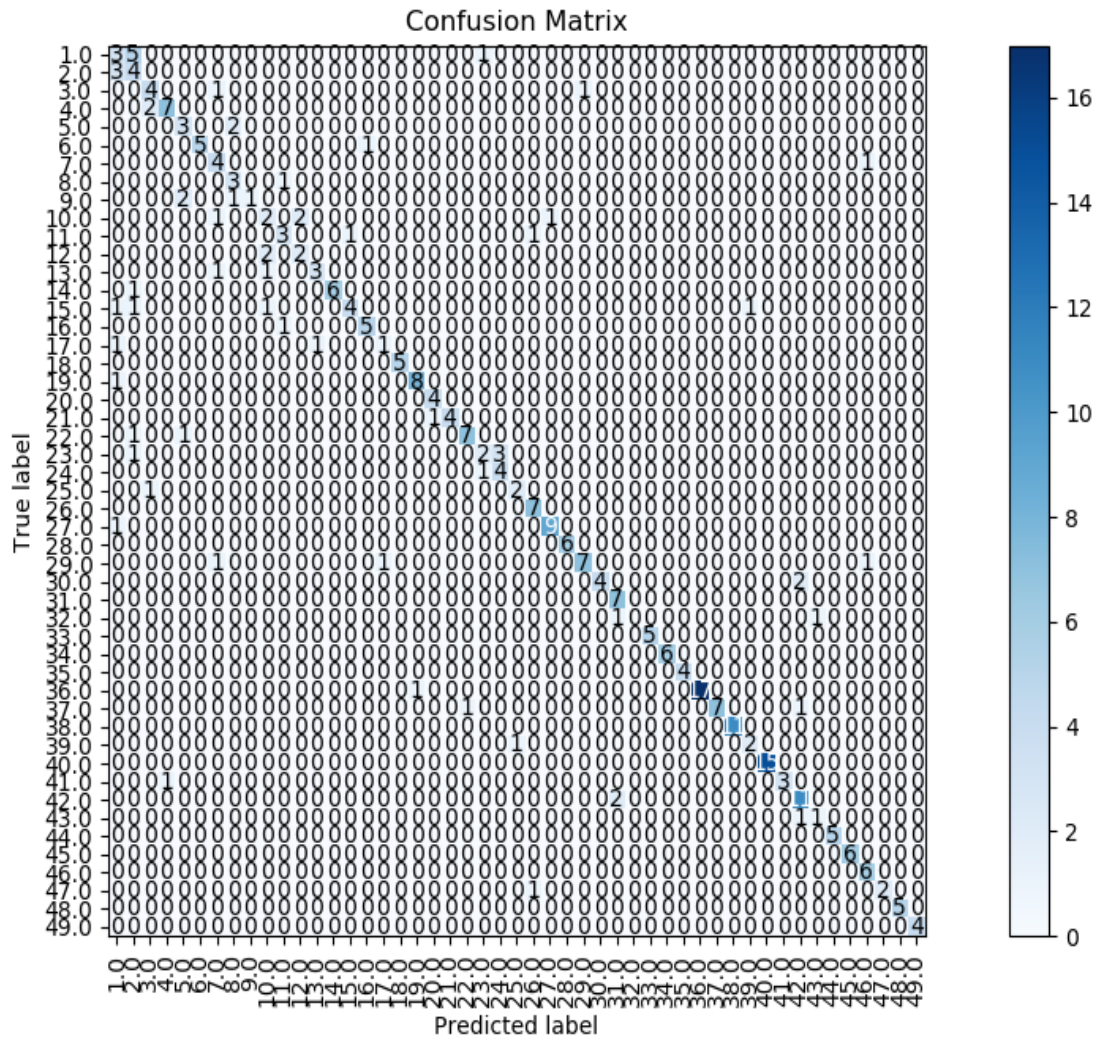


Figure 7. Confusion matrix

Table 4. Precision, recall and F1-score < 0.6.

Class	Precision	Recall	F1-score	Support
1	0.3	0.33	0.32	9
2	0.31	0.57	0.4	7
5	0.5	0.6	0.55	5
9	1	0.25	0.4	4
10	0.33	0.33	0.33	6
12	0.5	0.5	0.5	4
17	0.5	0.33	0.4	3
23	0.5	0.33	0.4	6
32	0	0	0	2
43	0.5	0.5	0.5	2

Table 5. Precision, recall and $0.6 \leq F1\text{-score} \leq 0.8$.

Class	Precision	Recall	F1-score	Support
3	0.57	0.67	0.62	6
7	0.5	0.8	0.62	5
8	0.5	0.75	0.6	4
11	0.6	0.6	0.6	5
13	0.75	0.6	0.67	5
15	0.8	0.5	0.62	8
24	0.57	0.8	0.67	5
25	0.67	0.67	0.67	3
29	0.88	0.7	0.78	10
30	1	0.67	0.8	6
39	0.67	0.67	0.67	3
42	0.73	0.85	0.79	13
47	1	0.67	0.8	3

Table 6. Precision, recall and $F1\text{-score} > 0.8$.

Class	Precision	Recall	F1-score	Support
4	0.88	0.78	0.82	9
6	1	0.83	0.91	6
14	1	0.86	0.92	7
16	0.83	0.83	0.83	6
18	1	1	1	5
19	0.89	0.89	0.89	9
20	0.8	1	0.89	4
21	1	0.8	0.89	5
22	0.88	0.78	0.82	9
26	0.78	1	0.88	7
27	0.9	0.9	0.9	10
28	1	1	1	6
31	0.7	1	0.82	7
33	1	1	1	5
34	1	1	1	6
35	1	1	1	4
36	1	0.94	0.97	18
37	1	0.78	0.88	9
38	1	1	1	11
40	1	1	1	15
41	1	0.75	0.86	4
44	1	1	1	5
45	1	1	1	6
46	0.75	1	0.86	6
48	1	1	1	5
49	1	1	1	4

However, F1-scores of classes 1, 2, 5, 9, 10, 12, 17, 23, 32 and 43 are less than 0.60 (Table 4). Especially, F1-score of class 32 is 0.00. The reason would be there are not enough examples (There are only 2 examples). Table 7 shows the category of each class which has the F1-score less than 0.6.

Table 7. Category of each class which has the F1-score less than 0.6.

Classes	Category (subject, verb, object)
1, 2, 5, 9, 10, 12, 17	Subject
23, 32	Object
43	Verb

According to Table 7 above, we can observe that most of the classes which have F1-score less than 0.60 belong to the subject category. Therefore, we can come to the decision that our model is unable to recognize signs which belong to the subject category. By the way, it is not possible to observe a clear diagonal and values are spread in the class range 1 – 18 in the confusion matrix (Figure 7). Therefore, we can confirm that our model is unable to recognize signs which appear in the subject category of the sentence more than those in other categories by looking at the confusion matrix further.

Figure 8 shows the ROC curve of the LDA classifier. It shows that all classes (49 signs) have high percentages of the area under the ROC curve. A rough guide for classifying the accuracy of a diagnostic test is the traditional academic point system. Table 8 shows AUC-ROC results according to the academic point system. According to the traditional academic point system also we can confirm that all the signs belong to the excellent category. Finally, we can conclude that model performed well at word level classification in the continuous sentence signing task.

As mentioned previously, the sentence level accuracy of the recognizer varies between 45% and 50%. A sentence consists of 3 words (subject + object + verb). If at least one of the signs is predicted wrongly, the entire sentence will be classified as incorrect. Table 9 shows, how the position of the sign contributes to sentence level errors.

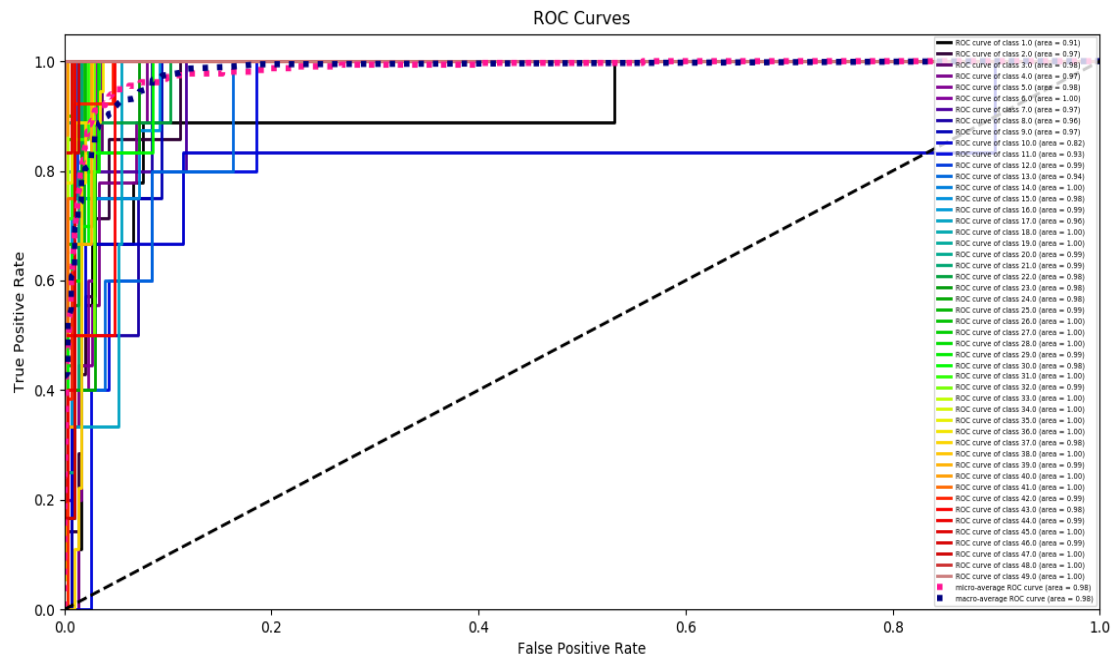


Figure 8. ROC curve

Table 8. AUC-ROC results according to the academic point system.

Points	No. of classes (Signs)
0.90 - 1.00 = excellent (A)	49 (all signs)
0.80 - 0.90 = good (B)	0
0.70 - 0.80 = fair (C)	0
0.60 - 0.70 = poor (D)	0
0.50 - 0.60 = fail (F)	0

Table 9. The contribution of positions of sign for sentences misclassification.

1st Sign (Subject)	2nd Sign (Object)	3rd Sign (Verb)	No. of Misclassified Sentences (n)	Percentage (n/104) *100%
Misclassified	Correctly Classified	Correctly Classified	31	29.8%
Misclassified	Misclassified	Correctly Classified	5	4.8 %
Misclassified	Misclassified	Misclassified	1	0.9 %
Misclassified	Correctly Classified	Misclassified	2	1.9 %
Correctly Classified	Misclassified	Correctly Classified	9	8.7 %
Correctly Classified	Misclassified	Misclassified	3	2.9 %
Correctly Classified	Correctly Classified	Misclassified	3	2.9 %
Number of misclassified sentences (total)			54	51.9 %

According to Table 9 above, we can observe that predicting the 1st sign of a sentence incorrectly accounts for 57% of all errors. Therefore, we can conclude that our model is weak at recognizing the 1st sign (Subject of the sentence) correctly relative to other positions. We have already discussed this issue by looking at the confusion matrix (Figure 7) and the F1-score (Table 4, 5 and 6). Misclassification of one sign directly affects the overall sentence level accuracy. We observed that there are two main reasons for such misclassification.

- 1) Similarities of signs
- 2) Incorrect sign segmentation

Then we wanted to experiment, how feature reduction and feature selection techniques affect the accuracy of the model. We observed the 10-fold cross-validation accuracy of each classifier with 5 feature reduction and selection techniques.

We reduced the original number of features (108) to 20, 40, 60, 80 and 100 by using the above mentioned 5 methods. However, we were unable to observe any significant improvement of the model, when we trained the model using 20 and 60 features. The model did show significant improvement in the 40, 80 and 100 feature cases.

Figure 9 shows how the cross-validation accuracy varies in each model after training them with the reduced set of 40 features. The LDA model shows the highest cross-validation accuracy (0.757831) after reducing the features using the random forest method. The LDA model showed the highest cross-validation accuracy (0.780723) after reducing the features to 80 using the principal component analysis (PCA) method. The LDA model showed the highest cross-

validation accuracy (0.791566) after reducing the features to 100 using the singular value decomposition (SVD) method.

Initially, we had 108 features and the accuracy varied between 75% and 80% (Baseline). However, we observed that the LDA model has the highest cross-validation accuracy even with reduced features. The accuracies and numbers of features can be summarized as follows.

- 1) 40 features - 0.757831
- 2) 80 features - 0.780723
- 3) 100 features - 0.791566

Even though 80 feature and 100 features instances have high accuracy than 40 feature instances, the number of features is closer to the initial number of features (108). In practical deployments, we need to consider both the model's accuracy and the number of features in order to select the best feature reduction method. Hence, we selected the 40 feature (random forest) model as the final model.

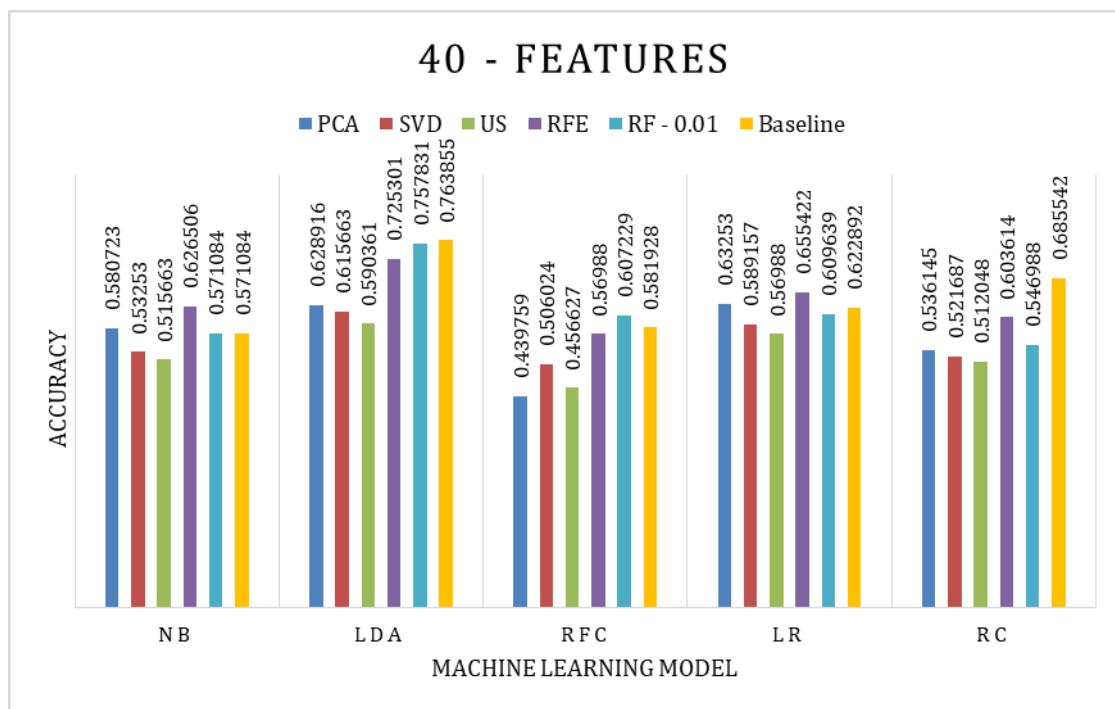


Figure 9. Cross-validation accuracies of each model using the reduced set of 40 features

We have conducted the above mentioned experiments as offline experiments which all the training data and testing data collected previously. Then pre-process, segment, extract the features from the raw data and finally trained the classifier. As the next experiment, we wanted to conduct gesture recognition in a real-time manner. Because the final goal is to use this system in real-world scenarios. Here, we used two previously trained classifiers to predict the signs.

- 1) The classifier which has been trained using all the features (108 features) (Classifier-1)
- 2) The same classifier which has been trained after feature reduction (40 features). (Classifier-2)

Average prediction time of a sentence using classifier-1 (LDA and 108 features) in real-time is 17.4 seconds. Average prediction time of a sentence using classifier-2 (LDA and 40 features) in real-time is 13.6 seconds.

We can observe that the classifier-2 which was trained using 40 features for the sentence prediction shows less prediction time than classifier-1 in the real-time scenario. Even though, that 13.6 seconds of time is not suitable for the real-time scenario, we can observe that the prediction time of a sentence is reduced when we reduced the number of features. Therefore, we can state that there is an effect to the prediction time when we reduced the features.

We translated the gestures in a real-time manner. In order to do that we have created a python application. Example of a real-time classification output is shown in Figure 10. It shows the output as “යාලුවා චිත්‍ර අඳිනවා” (*yāluvā citra añdinavā*). (Friend draws paintings).

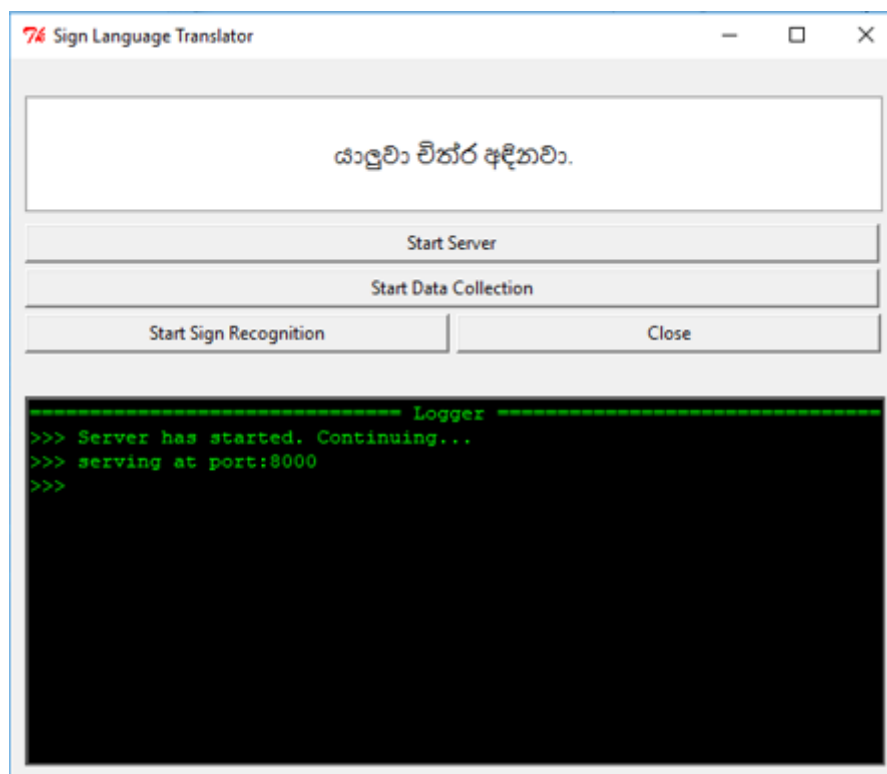


Figure 10. Example output of a real-time gesture classification

5.1. Comparison of the Results of the Proposed Solution and Related Work

This research project is an extended version of previous research which has been conducted at University of Colombo School of Computing and title of that publication is “Framework for Sinhala Sign Language Recognition and Translation Using a Wearable Armband” [17] 2016. Prajwal Paudyal et al proposed another work which is SCEPTRE [12] 2016. Table 10 shows the comparison between the proposed solution and the above mentioned two research projects ([12], [17]).

Table 10. The comparison of the results of the proposed solution and two main references.

	Myo Armband [17] (EMG and IMU based solution)	Myo Armband [12] (EMG and IMU based solution)	Proposed Solution
Sign Language	Sri Lankan Sign Language	American Sign Language	Sri Lanka Sign Language
Word Level	Yes	Yes	Yes
Sentence Level	No	No	Yes
User Dependent	Yes	Yes	Yes
Accuracy (around)	100% (Word Level)	97.72% (Word level)	75%-80% (Word Level) 45%-50% (Sentence Level)
Number of Signs	3	20	49 (Words) 346 (Sentences)
Method	ANN	Multitiered template-based comparison system	Linear Discriminant Analysis
Real-time	No	Yes	Yes
Real-time recognition time	-	0.552 S (Word Level)	13.6 S (Sentence Level)

6. CONCLUSION

The aim of this research was to bridge the communication gap between hearing/speaking impaired and ordinary people by proposing a framework for recognize sentence level continuous signings of Sri Lankan Sing Language and translate them into a natural language (Sinhala).

In order to conduct the research, we created a dataset using a single sign language interpreter and Sri Lankan Sign Language was selected as the sign language. After completing this research project, that dataset will be publicly available. Then we trained models and got promising results for sign language recognition for both word level and sentence level continuous signings which are 75%-80% accuracy for the word level and 45% - 50% accuracy for the sentence level continuous signing.

The accuracy of the recognition of sentence level continuous signings in real time manner vary between 45% - 50%. However, we used the model which is trained using only 40 features after feature reduction for this scenario, because after feature reduction, we got less prediction time (13.6s) than prediction time when we used all features (17.4s).

Finally, the proposed solution improves the usability and mobility, because we use the wireless, lightweight wearable device and we did not use any unnatural method to identify the moment epenthesis.

7. FUTURE WORK

Our proposed solution showcased a proper outcome based on the scope of the research study which can be extended in several ways.

- 1) Increase the number of signs
- 2) Increase the words per sentence
- 3) Identify an automatic way to segment the signs
- 4) Reduce real-time classification time

REFERENCES

- [1] "List of sign languages", En.wikipedia.org. [Online]. Available: https://en.wikipedia.org/wiki/List_of_sign_languages. [Accessed: 01- Jan- 2019]
- [2] K. Stefanov and J. Beskow, "A Real-time Gesture Recognition System for Isolated Swedish Sign Language Signs", Proceedings of the 4th European and 7th Nordic Symposium on Multimodal Communication (MMSYM 2016), Copenhagen, 29-30 September 2016, no. 141, pp. 18-27, 2017. [Accessed 1 January 2019]
- [3] Wu jiangqin, Gao wen, Song yibo, Liu wei and Pang bo, "A simple sign language recognition system based on data glove", ICSP '98. 1998 Fourth International Conference on Signal Processing (Cat. No.98TH8344). Available: 10.1109/icosp.1998.770847 [Accessed 1 January 2019]
- [4] D. Naglot and M. Kulkarni, "Real time sign language recognition using the leap motion controller", 2016 International Conference on Inventive Computation Technologies (ICICT), 2016. Available: 10.1109/inventive.2016.7830097 [Accessed 1 January 2019]
- [5] M. Maraqa, F. Al-Zboun, M. Dhyabat and R. Zitar, "Recognition of Arabic Sign Language (ArSL) Using Recurrent Neural Networks", Journal of Intelligent Learning Systems and Applications, vol. 04, no. 01, pp. 41-52, 2012. Available: 10.4236/jilsa.2012.41004 [Accessed 1 January 2019]
- [6] J. Wu, Z. Tian, L. Sun, L. Estevez and R. Jafari, "Real-time American Sign Language Recognition using wrist-worn motion and surface EMG sensors", 2015 IEEE 12th International Conference on Wearable and Implantable Body Sensor Networks (BSN), 2015. Available: 10.1109/bsn.2015.7299393 [Accessed 1 January 2019]
- [7] E. Mustafa and K. Dimopoulos, "Sign Language Recognition using Kinect", Conference: Conference: 9th South East European Doctoral Student Conference, pp. 271-285, 2014. [Accessed 1 January 2019]
- [8] N. Tubaiz, T. Shanableh and K. Assaleh, "Glove-Based Continuous Arabic Sign Language Recognition in User-Dependent Mode", IEEE Transactions on Human-Machine Systems, vol. 45, no. 4, pp. 526-533, 2015. Available: 10.1109/thms.2015.2406692 [Accessed 1 January 2019]
- [9] D. Kelly, J. Reilly Delannoy, J. Mc Donald and C. Markham, "A framework for continuous multimodal sign language recognition", Proceedings of the 2009 international conference on Multimodal interfaces - ICMI-MLMI '09, 2009. Available: 10.1145/1647314.1647387 [Accessed 1 January 2019]
- [10] Xu Zhang, Xiang Chen, Yun Li, V. Lantz, Kongqiao Wang and Jihai Yang, "A Framework for Hand Gesture Recognition Based on Accelerometer and EMG Sensors", IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans, vol. 41, no. 6, pp. 1064-1076, 2011. Available: 10.1109/tsmca.2011.2116004 [Accessed 1 January 2019]
- [11] C. Savur and F. Sahin, "American Sign Language Recognition system by using surface EMG signal", 2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC), 2016. Available: 10.1109/smc.2016.7844675 [Accessed 1 January 2019]
- [12] P. Paudyal, A. Banerjee and S. Gupta, "SCEPTRE", Proceedings of the 21st International Conference on Intelligent User Interfaces - IUI '16, 2016. Available: 10.1145/2856767.2856794 [Accessed 1 January 2019]
- [13] D. M. Wathugala and N. D. Kodikara, "A Sinhala finger spelling interpretation system using nearest neighbor classification", proc of 4th International information technology conference Envisioning an eNation, Colombo, Sri Lanka, 2002. [Accessed 1 January 2019]
- [14] H. C. M. Herath, W. A. L. V. Kumari, W. A. P. B. Senevirathne and M. B. Dissanayake, "Image based sign language recognition system for Sinhala sign language", Proceedings of SAIM Research Symposium on Engineering Advancements 2013, pp. 107-110, 2013. [Accessed 1 January 2019]
- [15] N. Kulaveerasingam, S. Wellage, H. M. P. Samarawickrama, W. M. C. Perera and J. Yasas, "The Rhythm of Silence" - Gesture Based Intercommunication Platform for Hearing-impaired People (Nihanda Ridma)", 2014. [Accessed 1 January 2019]
- [16] P. Fernando and P. Wimalaratne, "Sign Language Translation Approach to Sinhalese Language", GSTF Journal on Computing (JoC), vol. 5, no. 1, 2016. Available: 10.7603/s40601-016-0009-8 [Accessed 1 January 2019]
- [17] A. Madushanka, R. Senevirathne, L. Wijesekara, S. Arunatilake and K. Sandaruwan, "Framework for Sinhala Sign Language recognition and translation using a wearable armband", 2016 Sixteenth International Conference on Advances in ICT for Emerging Regions (ICTer), 2016. Available: 10.1109/ictcr.2016.7829898 [Accessed 1 January 2019]

- [18] "Sri Lankan Sign Language", Lankasign.lk. [Online]. Available: <http://www.lankasign.lk/index.html>. [Accessed: 01- Jan- 2019]
- [19] Sri Lanka Sign Dictionary, 2nd ed. SRI LANKA CENTRAL FEDERATION OF THE DEAF, 2012
- [20] "Myo Gesture Control Armband tech specs", Welcome to Myo Support. [Online]. Available: <https://support.getmyo.com/hc/en-us/articles/202648103-Myo-Gesture-Control-Armband-tech-specs>. [Accessed: 01- Jan- 2019]

© 2021 By AIRCC Publishing Corporation. This article is published under the Creative Commons Attribution (CC BY) license.