IoT Based Sign Language Recognition System

Punsara K.K.T
Department Of Computer Systems Engineering
Sri Lanka Institute of Information Technology
Malabe, Sri Lanka
1.thilanpunsara@gmail.com

Premachandra H.H.R.C
Department Of Computer Systems Engineering
Sri Lanka Institute of Information Technology
Malabe, Sri Lanka
ravipremachandra@gmail.com

Chanaka A.W.A.D
Department Of Computer Systems Engineering
Sri Lanka Institute of Information Technology
Malabe, Sri Lanka
achinthachanaka97@gmail.com

Wijayawickrama R.V
Department Of Computer Systems Engineering
Sri Lanka Institute of Information Technology
Malabe, Sri Lanka
wijayawick@gmail.com

Nimsiri Abhayasinghe
Department Of Electrical and Electronic Engineering
Sri Lanka Institute of Information Technology
Malabe, Sri Lanka
nimsiri.a@sliit.lk

Rajitha de Silva
Department Of Computer Systems Engineering
Sri Lanka Institute of Information Technology
Malabe, Sri Lanka
rajitha.d@sliit.lk

2020 2nd International Conference on Advancements in Computing (ICAC) | 978-1-7281-8412-8/20/$31.00 ©2020 IEEE | DOI: 10.1109/ICAC51239.2020.9357267

***Abstract*—Sign language is the key communication medium, which deaf and mute people use in their day-to-day life. Talking to disabled people will cause a difficult s ituation s ince a non-mute person cannot understand their hand gestures and in many instances, mute people are hearing impaired. Same as Sinhala, Tamil, English, or any other language, sign language also tend to have differences according to the region. This paper is an attempt to assist deaf and mute people to develop an effective communication mechanism with non-mute people. The end product of this project is a combination of a mobile application that can translate the sign language into digital voice and IoT-enabled, light-weighted wearable glove, which capable of recognizing twenty-six English alphabet, digits, and words. Better user experience provides with voice-to-text feature in mobile application to reduce the communication gap within mute and non-mute communities. Research findings a nd r esults from the current system visualize the output of the product can be optimized up to 25%-35% with an enhanced pattern recognition mechanism.**

***Index Terms*—sign language, Internet of Things, Gesture recognition, Smart glove, Recurrent neural network**

I. Introduction

According to the latest statistics of the World Health Organization, 5% of the world population cannot hear a single word [1]. It is a tragedy, which leads them to mutism since they cannot hear or learn words to speak. In conditions such as Apraxia of Speech, Cerebral Palsy, and Aphasia people suffering from the inability to speak. To reduce this gap between mute and non-mute people, sign language will act as a bridge between them and act as the main communication method for those who cannot express their voice. In such situations, they will need a communication mechanism to express their ideas with each other. As a non-mute person, sharing ideas with a mute person will cause a difficult situation for both. Because one person cannot hear any sound and other one cannot understand the hand gestures. In such situations, the non-mute community may have to make sure that expressions of deaf & mute people may understand them very well. The more non-mute people do not understand them via sign language, the more they will avoid having human interaction with the community. There are 6909 distinct spoken languages in the world today [2]. Same as that, 125 sign languages use around the globe in different countries [3].These devices sense and record user activities, predict their future behavior, and prepare everything one step ahead according to the user's preference or needs, giving him/her the most convenience, comfort, efficiency, and security. [4]Currently, existing gesture recognition techniques can be divide In to 2 main categories based on the method;

1) - Computer vision-based method
2) - Sensor-based method

Computer vision-based gesture recognition [5]can be less accurate and less comfortable to the end user since, it involves many aspects such as motion modeling, motion analysis, pattern recognition, and color segmentation [6].Considering the sensor-based hand gesture recognition mechanisms, different sensors provide a set of data according to the joints and finger separation that characterizes a hand gesture.

IoT Based Sign Language Recognition System is an attempt to reduce the communication barrier between the mute and non-mute community and assist the non-mute community to understand the hand gestures. This research uses a combination of concepts related to Machine learning, Natural language processing, and IoT (Internet of Things). The final product of this research is expected to come in the form of a mobile application and a wearable hand glove. This hand glove architecture is designed with considering the better user experience and mobile application is consists of features such as new user training, voice-to-text with iOS and Android platforms.

The remainder of the paper is arranged as follows: Section II elaborates the background survey including a theoretical comparison of existing sign language interpreters. Section III explains the approach of implementing the Glove and whole environment in detail. Section IV explains the Test results from the process, Section V explains discussion and finally Section VI concludes the paper conclusion.

II. Background review

Researches in the sign language recognition systems are mainly based on computer vision and sensor based recognition mechanisms. The image processing techniques [7] using the camera to capture the image/video. Examinations of the data with static images and identify the image using various algorithms and create sentences for that into the display [7] Camera place to direct the place that captures highest available hand movements, higher resolution camera take up more calculation time and hold more memory space. A deaf person always need a high performance camera permanently and cannot use in a public place.

Another research determines sign language recognition system using a hand glove. [8] [9] In this design, mute person need to wear a glove consist of 5 flex sensor for each finger and motion tracker. Data are directly coming from every 5 sensors and process sensor data with static data to produce sentences. It's using a neural network to increase the completion of the system. The main advantage of this system is the fast response in real-time applications. Its movable device and cost of the complete device is higher since the hardware used are expensive [9].

In another research, researches developed a sign language recognition system using a portable Accelerometer (ACC) and Surface Electro Myogram (sEMG) [10]these sensors are used to detect the hand gesture. ACC used to capture movement information of hand and arms. These Sensor output signals are input to the computer and process to identify the hand gesture and provide speech and text with both [10]. But none of the above methods provides users with two-way communication and as well as a graphical picture of each sign. Our proposed system will be capable of delivering the two-way conversation with visualizing pictures of relevant signs in the app with a user-friendly manner.

Other than using flex sensors, a team of researchers used Potentiometer to extract the data from fingers [11]. Above system design was created to work with virtual reality applications like replacing the conventional input devices like joysticks in video games with the data glove. Also, the Robot control system to regulate machine activity at remote sensitive sites.

This proposed outcome will be capable of delivering the two-way conversation with graphical picture app with a user-friendly manner. This is all so will help both deaf people and normal people. Based on the arrangement done by this device, the user is capable of understanding the communication between mute people and the non-mute people vice verse and deciding if the new user came up with the device provide the instruction of how to do the signs and how success user is which should be done to manage the device or glove. This whole idea is not tested in yet. So using two-way communication is very new. Mobile App already exists. But broadcasting the output with graphical interfaces is also a new attempt.

As it could be clearly understood through the previous survey about the current systems, several tries have been taken in the current past towards two-way communication using various concepts such as Image Processing, Neutral networks and also various sensors. The outcomes of those research studies have been implemented, tested and published while some have been deployed as commercial level applications like android App. Throughout those analyses of the available research none of the existing products is capable of identifying the differences between how a new user can use the device in a better manner. Here we are supposing a method to train the signs for a user and after that period, user can test how successful it would be.

Taking these reasons into consideration, the system implemented under this research study effectively brings about a solution to this issue by analyzing the device for a specific sign language type (American Sign Language) and gives a time slot to practice the system and test the results also.

III. Methodology

Proposed system is designed to identify and translate the hand gestures into a digital voice as final outcome. In implementation, system is consist of a software module and a hardware module. Accelerometer, flex sensors and printed circuit board includes in hardware module. As the initial step, system will capture flex sensor and accelerometer readings.Record it one by one for each sign by using a push-button remote (push button).

A push button is used to input the boundaries of a single data frame corresponding to a certain sign in the data stream. Number of data sets for one sign can be obtained and save them in a CSV file format. In the same manner different data sets for different signs (100) can be collected. These signs referred as mean data set. As the next step get overall CSV file set and calculate the mean value for each sign.

After getting the mean values, collection of single CSV files set for each different sign. For example, if there are 10 signs,there will be x axis data for each 10 signs and saved in one CSV file. Similarly, now there are 8 CSV files for for X,Y,Z z-axis and 5 flex sensors. In addition to that, recorded data set also included in the system. As previous this also has 8 CSV files for X,Y,Z axis and also for the 5 flex. The only difference here is it is not necessary to calculate the

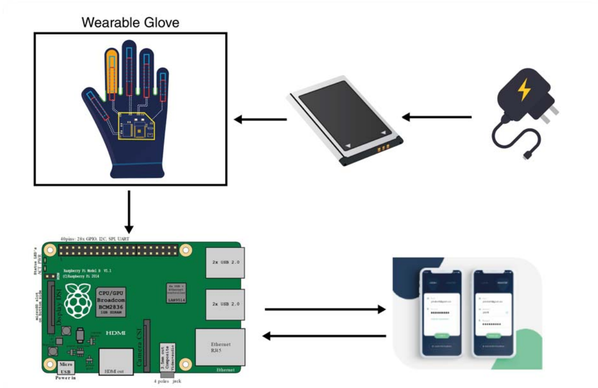

Fig. 1: System overview

mean value. The next step is to identify the mean squared error (MSE). Using recorded data and the actual data set can calculate the MSE value.

Up to this moment there are two different data sets available in the system(actual sign data set and recorded data set). According to the graph here in figure 2 it visualize our data set which consists of the actual gesture capture with their mean values. In addition to that, data set which consists of the recorded data set also available. Now according to the figure 3 graph,comparison of 2 graphs can be made by shifting them. As above, shifting the frame one by one will lead to calculate the MSE value. As a assumption, assume recorded data set and it has 120 data points. Initially it's necessary to fix the actual sign size. It will range 0 to 80 and then check it. Then shift by one value and next 1 to 81 ranges and check it. Likewise shifting the actual sign graph on our recorded sign graph and calculate the MSE value. As a last step of first phase,data can be stored in an array. After comparing the graphs by shifting method MSE value array can be generated. After checking the complete graph, point of lowest value in graph can be identified as the point where consist the lowest error.

Identification of fix value is essential since it used to compare the MSE values. To calculate this fix value, same two graphs will be used.Actual one (for a one sign) and calculate the MSE value of it. In a perfect error-less scenario outcome should be 0.00. But obviously there is are range differences between same sign graphs. Because of that calculating this in several times we have to select a fix one. As the final results for the fix value we got the answer as 0.05. After getting fix value now it is easy to scale out our MSE values as following.

1) Ex- maximum error value = 0.05
2) If our minimum MSE value is (E) $\leq$ 0.05

This result can be accepted, since mean of the recorded data set is equal to the exact sign and very similar to the sign that we used to shift (Actual data set).

We have eight CSV file up to now (both actual and recorded). This need to done to the same e CSV files also and we have to collect the results in several time for same graphs also to get a idea of the pattern that will take(ex R1, R1.1 R1.2). After getting results it can be visualize as below.

After getting data from the sensors and calculating the MSE values, cost function will provide the processed data in to a Long short-term memory network (LSTM). Once the data received, LSTM will try to recognize a pattern with input data and give an output. If we consider 1,2,3,4& 5 are five different signs,according to the below graph, output is identified as 1,2,3 &4. This is a perfect, error less situation since none of the other signs identified by accelerometer axis or flex sensors other than the original sign. It comes only if accelerometer axis and flex sensors identify the correct sign without any interference with other signs. If we check the output 5, it clearly visible all 8 sensors identify the output as "5" yet Y,Z axis and F3 sensor identified "5" with 3,4 signs as well. This is an error, yet it can be normalized, since once the same pattern identified by the LSTM network, it will keep the pattern in the memory. Whenever a similar pattern or pattern with minor changes identified by LSTM network it will give the output as pattern with most similarities.

Once a pattern is recognized through LSTM, processed data will feed in to a another LSTM network for smooth the outcome of the sentence. Smoothing the outcome is essential because from the first LSTM network we only get identified word series. Processing is essential for give a user-friendly outcome to the end user. After getting a complete, meaningful sentence from second LSTM network it will transmitted over WiFi to a mobile application. From mobile application identified sentence will be expose to the end user in voice format.

The designed Sign language recognition system has the capability of training an inexperienced user to the system with inbuilt training mode.Once a new user registered through the mobile application first time,user will be directed to the training mode.According to the given instructions user may complete the training in predefined time. Once the training session is completed, accuracy of new users hand gestures will be calculated and provide with percentage.

A. *PCB Design*

Customized PCB was designed to obtain the signals from 5 flex sensors and accelerometer with optimized space usage to reduce the weight of the device.Other than using whole modules, this PCB is designed with separate ICs, sensors and SMD components to reduce the space usage.

1) Esp wroom 32D IC
2) MPU 6050 Module
3) CP2102 (serial communication)

To build the serial communication we connect the CP2102 through a micro USB port. In such cases like, WiFi failure or battery power decrease we can directly connect the board and the Raspberry Pi via a micro USB cable. Also, we provide power, through the micro USB port.Data is directly coming from 5 flex sensors and MPU 6050 sensor.Through WiFi connection, data will transfer to the Raspberry Pi for processing.

When we take the data in live data will store in an array that is coming from module. This array size is depending on the

X					Y					Z					F1					F2					F3					F4					F5					Output
1	0	0	0	0	1	0	0	0	0	1	0	0	0	0	1	0	0	0	0	1	0	0	0	0	1	0	0	0	0	1	0	0	0	0	1	0	0	0	0	1
0	2	0	0	0	0	2	0	0	0	0	2	0	0	0	0	2	0	0	0	0	2	0	0	0	0	2	0	0	0	0	2	0	0	0	0	2	0	0	0	2
0	0	3	0	0	0	0	3	0	0	0	0	3	0	0	0	0	3	0	0	0	0	3	0	0	0	0	3	0	0	0	0	3	0	0	0	0	3	0	0	3
0	0	0	4	0	0	0	0	4	0	0	0	0	4	0	0	0	0	4	0	0	0	0	4	0	0	0	0	4	0	0	0	0	4	0	0	0	0	4	0	4
0	0	0	0	5	0	0	0	4	5	0	0	3	0	5	0	0	0	0	5	0	0	0	0	5	1	0	0	0	5	0	0	0	0	5	0	0	0	0	5	5
0	0	0	0	5	0	0	0	4	5	0	0	3	0	5	0	0	0	0	5	0	0	0	0	5	1	0	0	0	5	0	0	0	0	5	0	0	0	0	5	5
0	0	0	0	5	0	0	0	4	5	0	0	3	0	5	0	0	0	0	5	0	0	0	0	5	1	0	0	0	5	0	0	0	0	5	0	0	0	0	5	5

Fig. 2: LSTM Result

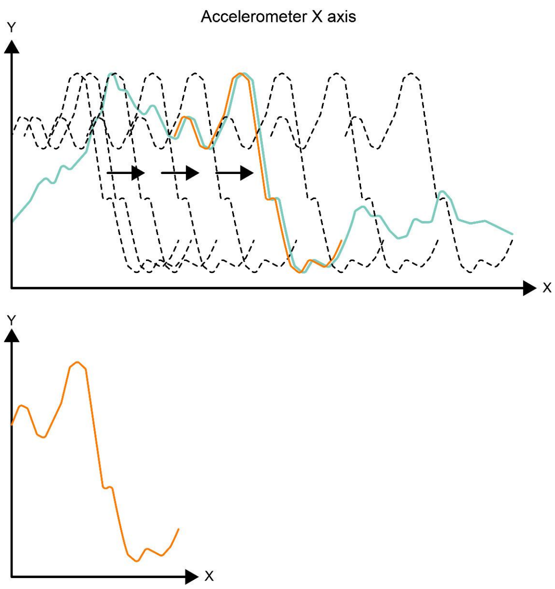

Fig. 3: shifting the graph

Fig. 4: PCB Design

largest size count on the mean data set. When the array count is fully then it will send for the process. Again for the same process before it storing 50% of data elements from the array will be deleted. Then the data will shift to the first elements and also adding. Then as in previous when the array count is fully then it will send for the process. This whole process will be evaluated when we catch live data.

IV. Test Results

Experiments were mainly conducted with the test graph results. For example, observe the sign ’HELLO’. For the ’HELLO’ sign, we should obtain the mean data set and recorded data set. It will show from order in figure 5 and 6 here (For a one Axis).

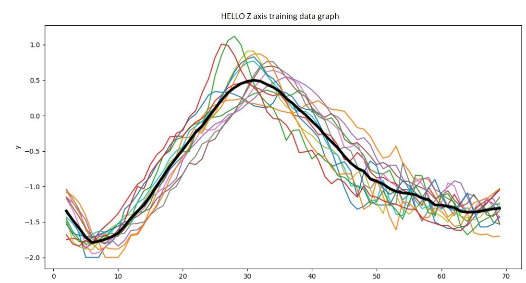

Fig. 5: Mean data set graph

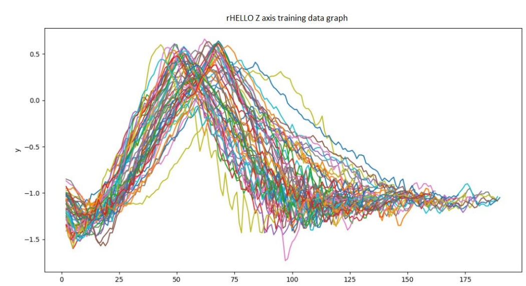

Fig. 6: Recorded data set graph

In the initial step when we record the sign first we have do the same sign again and again to get the most suitable one.After getting the mean values, recorded data set in the ’HELLO’ shift the frame, one by one and calculate the MSE. Before that, we calculate the fix MSE value previous (mentioned in the methodology).Then take those output to the LSTM as a input and predict the final result. After that trained period we can predict our specific sign as it is. For this example we get the ‘Hello’ sign and finally we can get the output result as ‘Hello’

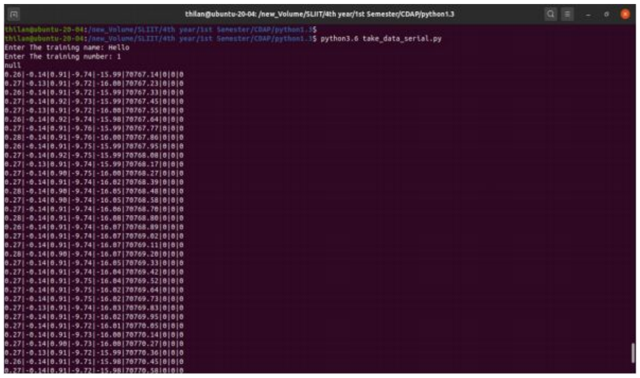

Fig. 7: Take data from sensors

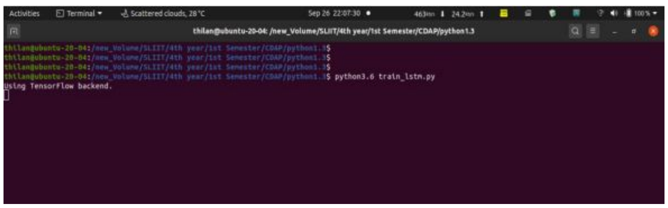

Fig. 8: Train LSTM model

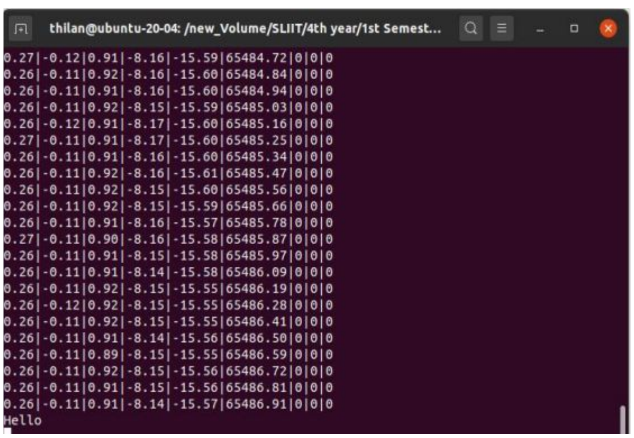

Fig. 9: Final Output

1) Ex- maximum error value = 0.05
2) If our minimum MSE value is (E) $\leq$ 0.05

If the results are agreed to above criteria, we can accept the results. It visualize, recorded data set the exact sign is very similar to the sign that we used to shift ('HELLO' sign).

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i)^2 \tag{1}$$

- N - Total number of Data points
- yi - Actual output value
- ŷi - Predicted output value
- yi-ŷi - The absolute value of the residual

Training session from this system has been done considering the new user (new to sign language) want to practice the sign with the glove. The initial version of this product only supports to predict the new user to assist, obtain a motion about the moving the sign language correctly. Several strategies could be followed to improve the accuracy of these predictions. In order to improve the accuracy of the model, there will be time period to all signs to do the same time. If time slot is not allocated, user practice the sign with exceeding the time limit. It will donate, provide more error value and the accuracy will be low.

Let's assume 'Hello' sign is practiced. In the program number of the sign need to be specifically mentioned. Because in the training mode first we have to select a specific sign and all these sign have a sign number. If we want to train the 'Hello' one it would be no 1. So we have to change the no in our code according to the requirements that needed.

Then in the given period after we finished the doing sign the results will be shown after shifting process in the inside. If the new user does the sign correctly it gives a good result as a percentage mark. Below figure shows us a good result that is done in a given period. In any case the given period if we try out the difference sign (not the selected one) in that case it will show us also an error value.

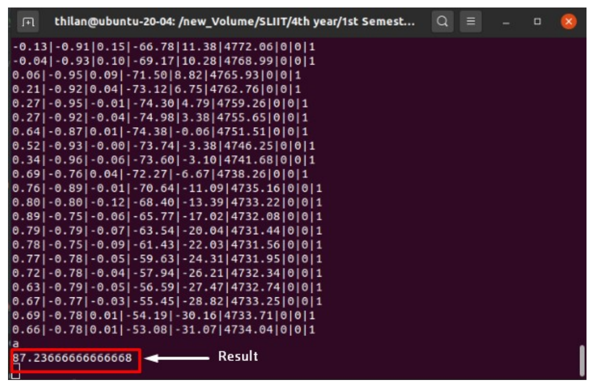

Fig. 10: Training mode results

V. Discussion

Study referencing shows the World Health Organization, Over 5% of the world's population or 466 million people have disabling hearing loss (432 million adults and 34 million children)[1]. In children under 15 years of age, 60% of hearing loss is attributable to preventable causes [1]. This figure is greater in low- and middle-income countries (75%) as matched to high-income countries (49%) [1]. Overall, preventable causes of childhood hearing loss include:

Taking these matters of the impact of hearing problems in the world, it has been discovered that a solution to identify these sign language communicate and predict how two-way communication is done and how the new user familiar with the sign language using the training mode. Hence this research study is based on using machine learning to predict the variation of the signs and neural network to identify the specific signs by using the data. Two types of sensors are used to capture the data of a hand gesture. The flex sensor and accelerometer are used to capture the readings from a hand gesture in a multidimensional way. Five flex sensors are used to capture the finger movements using the resistance of the flex sensors located on every finger of the glove. Five flex sensors will be used for a single hand since the project focuses on hand gesturing of American Sign Language. The American Sign Language has been selected since the gesture of the language is only based mostly on a single hand. Furthermore, the accelerometer (GY-521) positioned on the top of the glove will be used to measure the acceleration force of the hand gesture. The data is taken by the sensor will be sent ESP wroom 32D sensor for further processing as well.

Captured data transmitted via MPU 6050 Module to Raspberry-Pi and it will be processed in artificial neural network [12]. In this process ANN's output will be a collection of words, letters or numbers which will not give a proper sense to the end user. To overcome this obstacle, Natural Language Processing mechanism can be used in the proposed system.

The overall application is designed from using Flutter and Adobe XD. When considering the structure of mobile application, it has different interfaces to illustrate information to the user which is included different features;

1) User Mode
2) Training Mode
3) Battery Level

After the connection, the mobile application with the main control system user can direct to the home interface and it has user-mode tab and training mode tab if the user selected user mode user can communicate with other persons. Then user-selected training mode a new user can clearly get understanding of what is sign language and user can test whether how success the attempt of doing the signs in given time.Provide text and animation for illustrate sign language Normal person is talking with the deaf person that voices are converting to the hand gestures and that gestures are display through the mobile application with text

VI. Conclusion

End product of this project is useful for handicapped mute community, which will develop a bridge between those who comprehend sign language and those who do not. Initial version of this sign language recognition system supports American sign language and consist of two main components i.e wearable glove and mobile application.

Hand glove is consist of several sensors including flex sensors, gyroscope and battery monitoring sensors which can be used to obtain hand gestures. Raspberry Pi device establish a wireless communication between glove to obtain the gathered data for processing.

Mobile application is capable of identifying the voices of non-mute person, express the glove generated signal's electronic voice and training mode for new users. Mobile application and Raspberry pi is connected via WiFi and remaining battery percentage of the glove can also be monitored. Flutter has been used for develop the mobile application and it is compatible with iOS, Android and Web platforms.

For further developments, performance wise, sign recognition algorithm needs to fine tuned for better performance. Range of different hand signs can be trained and stored in the application, therefore accurate and smooth outcome can obtain for the end user. User experience can be improved further by optimizing the current size and weight of the device with high performance,less weight hardware components.

References

[1] Solvang Iselin Ertzgaard, Naalsund Kristin, Tønder Sofie, Hansen Giske Sindberg, Hagan Tobias Bang, Mnyanyi Cosmas, Tronstad Tron Vedul, Nkya Aslam, and Øygarden Jon. Prevalence of hearing impairment among primary school children in the kilimanjaro region within tanzania. *International Journal of Pediatric Otorhinolaryngology*, 130:109797, 2020.

[2] Mizanur Rahman, Md Salah Uddin, Md Samaun Hasan, Apurba Ghosh, Sadia Afrin Boby, Arif Ahmed, Shah Muhammad Sadiur Rahman, and Shaikh Muhammad Allayear. Anatomical analysis between two languages alphabets: Visually typographic test transformation in morphological approaches. In *International Conference on Advances in Computing and Data Sciences*, pages 330–339. Springer, 2020.

[3] Dolors Rodríguez-Martín, Catalina Rodríguez-García, and Anna Falcó-Pegueroles. Ethnographic analysis of communication and the deaf community's rights in the clinical context. *Contemporary nurse*, 54(2):126–138, 2018.

[4] Xu Li, Rongxing Lu, Xiaohui Liang, Xuemin Shen, Jiming Chen, and Xiaodong Lin. Smart community: an internet of things application. *IEEE Communications magazine*, 49(11):68–75, 2011.

[5] Ying Wu and Thomas S Huang. Vision-based gesture recognition: A review. In *International Gesture Workshop*, pages 103–115. Springer, 1999.

[6] Amy Cyphert. Tinker-ing with machine learning: The legality and consequences of online surveillance of students. *Nevada Law Journal*, 20, 2020.

[7] Rini Akmeliawati, Melanie Po-Leen Ooi, and Ye Chow Kuang. Real-time malaysian sign language translation using colour segmentation and neural network. In *2007 IEEE Instrumentation & Measurement Technology Conference IMTC 2007*, pages 1–6. IEEE, 2007.

[8] Priyanka Lokhande, Riya Prajapati, and Sandeep Pansare. Data gloves for sign language recognition system. *International Journal of Computer Applications*, 975:8887, 2015.

[9] Kirsten Ellis and Jan Carlo Barca. Exploring sensor gloves for teaching children sign language. *Advances in Human-Computer Interaction*, 2012, 2012.

[10] Yun Li, Xiang Chen, Xu Zhang, Kongqiao Wang, and Z Jane Wang. A sign-component-based framework for chinese sign language recognition using accelerometer and semg data. *IEEE transactions on biomedical engineering*, 59(10):2695–2704, 2012.

[11] Aaisha Parveen and Rohitha UM. Microcontroller based hand gesture recognition system using flex sensor for disabled people. *International Journal of Computer Applications*, 975:8887.

[12] A Cochocki and Rolf Unbehauen. *Neural networks for optimization and signal processing*. John Wiley & Sons, Inc., 1993.