Enhancing Collaborative Interaction with the Augmentation of Sign Language for the Vocally Challenged

Sukruth G L, Dr. Vijaya Kumar B P, Tejas M R, Rithvik K, Trisha Ann Tharakan
Department of Information Science and Engineering, M S Ramaiah Institute of Technology
Bangalore, Karnataka, India

***Abstract*—As per Census 2011, in India, there were 26.8 million differently abled people, out of which more than 25% of the people faced difficulty in vocal communication. They use Indian Sign Language (ISL) to communicate with others. The proposed solution is developing a sensor-based Hand Gesture Recognition (HGR) wearable device capable of translating and conveying messages from the vocally challenged community. The proposed method involves designing the hand glove by integrating flex and Inertial Measurement Unit (IMU) sensors within the HGR wearable device, wherein the hand and finger movements are captured as gestures. They are mapped to the ISL dictionary using machine learning techniques that learn the spatio-temporal variations in the gestures for classification. The novelty of the work is to enhance the capacity of HGR by extracting the spatio-temporal variations of the individual's gestures and adapt it to their dynamics with aging and context factors by proposing Dynamic Spatio-temporal Warping (DSTW) technique along with long short term memory based learning model. Using the sequence of identified gestures along with their ISL mapping, grammatically correct sentences are constructed using transformer-based Natural Language Processing (NLP) models. Later, the sentences are conveyed to the user through a suitable communicable media, such as, text-to-voice, text-image, etc. Implementation of the proposed HGR device along with the Bidirectional Long-Short Memory (BiLSTM) and DSTW techniques is carried out to evaluate the performance with respect to accuracy, precision and reliability for gesture recognition. Experiments were carried out to capture the varied gestures and their recognition, and an accuracy of 98.91% was observed.**

Keywords*—Hand Gesture Recognition (HGR); wearable sensors; Long-Short Term Memory (LSTM); Natural Language Processing (NLP); Dynamic Spatio-Temporal Warping (DSTW); Indian Sign Language (ISL)

I. Introduction

The need for communication stems from our natural desire to express ourselves. In a world that is becoming more inclusive, no one should feel left behind. The differently-abled community has been struggling historically to assimilate with the society. There exist natural communication barriers between the differently abled and abled people. This is because, the differently abled community, generally the vocally challenged community, uses sign language to communicate with others. Around 99% of the world do not understand sign language [1]. This creates a massive communication gap between the people in such communities who only use sign language for communication and those who do not understand sign language. The traditional solution for this has been to use an interpreter, who specializes in both sign language and spoken language. With the advent of technology, especially in the fields of artificial intelligence and machine learning, a hand gesture recognition system for sign language can be developed which can replace the interpreter, making the people of the differently abled people independent.

Sign languages (also known as signed languages) are languages that use the visual-manual modality to convey meaning. In India, the primary sign language used is the Indian Sign Language (ISL). ISL is used in the deaf and/or dumb community all over India.

ISL interpreters are an urgent requirement at institutes and places where communication between deaf and hearing people takes place, but India has around 300 certified interpreters, which is minuscule to the number of interpreters that are required [2]. The needs of the deaf and/or dumb community have long been ignored, and the problems have been documented by various organizations working for them. The aim is to bridge this gap with the help of a sign language interpreter by using HGR.

This paper proposes the use of sensor-based gesture recognition where sensors are placed on each finger and on the hand, through which the sensor data is collected based on the hand and finger movements. Such raw data is transformed to gesture data and classified using machine learning algorithms. The obtained results on classification of sequence of gestures is fed to the NLP models involving segmentation and grammar correction. Finally, the sentences are conveyed suitably through communicable media devices like speakers, displays, braille tablets, etc.

The remaining paper has been structured as follows. Section II discusses some related works done and the key points from different research papers. Section III elaborates on the proposed model for sign language gesture recognition. Section IV explains the implementation methodologies. Section V describes along with the inferences for the proposed model and results with respect to performance parameters. Section VI provides the concluding remarks and future scope.

II. Related Work

Some of the related works in the areas of hand gesture recognition and NLP related to differently abled communication are abstracted and discussed in the section.

Flex sensors are used to develop a HGR based system by capturing the finger movements by making use of the

Flex ADC values, voltage, resistance, and the ratio between the flex voltage and source voltage values. These inputs are then processed through a GRU and classification is done by implementing maximum a posteriori [3].

Another methodology wherein a yarn based stretchable sensor array (YSSAs) was made use of, which is based on the change in voltage as and when a gesture is being performed [4]. It is noticed here that the YSSAs, is not a commercially available product. The Flex or the IMU sensors are available commercially and at a low cost, which would provide ease of access to the materials of the data glove.

In a system of sensor-based HGR by making use of accelerators and gyroscopes, the most important part is the determination of the starting and stopping points whilst performing the continuous hand gesture and it was solved by implementing an LSTM model in [5]. Based on a many-to-many interface scheme, where output is produced by the LSTM at each time step, a collection of the output sequence is viewed as an output path.

To tackle long-distance dependencies, a frame stream density compression (FSDC) algorithm is introduced in [6] for detecting and reducing redundant similar frames. The traditional encoder is replaced in a neural machine translation (NMT) module with an improved architecture, which incorporates a temporal convolution (T-Conv) unit and a dynamic hierarchical bidirectional GRU (DH-BiGRU) unit sequentially.

It is often necessary in practical situations to attempt parsing an incorrect or incomplete input. Since there are no signs for articles (i.e., the, is, an, etc.,) most sentences performed using ISL will be structurally and grammatically incorrect. To tackle this issue certain NLP solutions were explored, one of them was parsing incomplete sentences. A chart parser was used in [7] to try all possible parses on all possible inputs. The result is a parse forest for all the grammatically acceptable sentences that can be generated by the (non-necessarily deterministic) finite state automaton.

Recent work on Grammar Error Correction (GEC) has highlighted the importance of language modeling to improve the performance and also compared the probabilities of the proposed edits. The approach in [8] involves using Google's BERT and OpenAI's Generative Pre-Trained Transformer (GPT) and GPT-2. Using simple heuristics, the language model generates a score such that it reduces the error and confusion set of sentences.

A more efficient algorithm using the distant measurement scheme of Dynamic Time Warping (DTW) and learning method of Restricted Coulomb Energy (RCE) neural network is proposed in [9]. A test platform is constructed using an IMU sensor to verify that real-time learning and recognition was possible using the proposed algorithm. DTW-based acknowledgment calculations have shown great execution for HGR.

Inferencing from the above literature review, using sensors and a hand glove design is provided with a sensor-based data collection system to capture finger and hand movements, a gesture classification unit along with dynamic spatio-temporal warping (DSTW) algorithm, grammar correction unit and a text to communicable media unit.

The data collected from the glove is in the form of time-series data, which is a type of Sequential Data. Sequential neural networks are designed to classify sequential data. Recurrent neural networks have the vanishing and exploding gradient problems, which is handled by LSTM by introducing cell state to decide what information to retain and forget long term. BiLSTM is used because it has shown better accuracy than LSTM models.

III. Proposed Model

In this work, the novelty involves in designing a suitable sensor-based glove that consists of flex and IMU sensors to collect the hand and finger movements as signs and gestures. Here, the glove is calibrated for different dynamics of hand and finger movements collected as time-series data. Such time-series data is analyzed for spatio-temporal variations using machine learning algorithms that use neuro-computing models. The different spatio-temporal variations are mapped to the respective gestures taken from ISL dictionary [10]. However, to capture similar gestures performed by individuals that vary w.r.t time and space, the dynamic spatio-temportal warping technique is adopted. To improve the accuracy and generalization, different gestures with time and space variations with continuous learning for classification and repeated mapping to the ISL dictionary can be performed to avoid over-fitting and under-fitting. From the trained model, a given series of gestures is converted to a meaningful sentence using a transformer-based Natural Language Processing (NLP) model. Such sentences can be converted into the text/speech or any other media that an abled human can understand.

Here, the learning methods include feed forward neural network model with memory-based learning integrated with DSTW as similarity metric along with a BiLSTM neural network model with Online Learning. Gestures belonging to the same category which are partially and semantically similar can be classified accurately even if there are time and space variations in any given gesture, as the experience increases with learning. The basic idea of training using memory-based learning is that it classifies the different gestures and also their similarities with the previously seen gesture data.

The ability of the proposed system is to compute and abstract the spatio-temporal variations in the data items derived from hand and finger movements, as time-series data items and their correlations using feed-forward neuro-computing models.

In time-series data analysis, dynamic spatio-temporal warping is an algorithm for measuring similarity between two temporal sequences, which may vary in time and space. For instance, similarities in walking could be detected using DSTW, even if one person is walking faster than the other, or if there is acceleration and/or deceleration during the course of an action. DSTW is illustrated in Section IVB with mathematical formulation.

The complete proposed model and its functional modules is shown in Fig. 1. The sensors data corresponding to a gesture taken from ISL video dictionary [10], is collected from the hand glove and it is pre-processed to extract time and spatial features for its respective meaning. Such mapping is continuously fed into neuro-computing learning modules until the gesture meanings are mapped correspondingly. However,

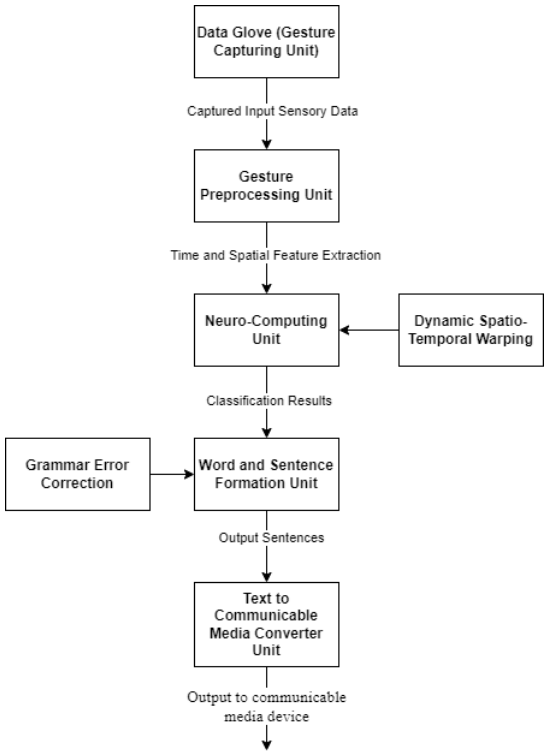

Fig. 1. Functional modules of the proposed hand gesture recognition (HGR) system.

the speed and area of occupancy in performing the same gestures by different individuals will vary, hence, DSTW algorithm is integrated to classify and map the meaning to the ISL dictionary. The sequence of gestures performed are integrated to generate meanings by using transformer-based NLP model with grammar correction to obtain a grammatically correct sentence. Such generated sentences are converted into communicable media by using text-to-speech, text-to-images, etc. as per the user's requirement.

A. Moving Average Smoothing

The captured data for the performed gestures will be in the form of time-series data and is bound to be noisy. To eliminate the fine-grained fluctuation between time steps, smoothing is used over the time-series data. Here, the smoothing is intended to reduce noise and thus clearly reveal the signal of underlying causal processes. For time-series data, moving averages are a straightforward and widely used method of smoothing. Moving Average Smoothing, represented in Fig. 2, involves creating a new time-series where the values are the average of raw observations in the original time-series.

A window size, referred to as the window width, is necessary when using a moving average. This specifies how many unprocessed observations were utilized to determine the moving average value. In order to calculate the average values in the new series, the window defined by the window width is moved along the time-series, hence giving the name "moving average".

After the noise has been removed through the moving average smoothing process, the time and spatial features (i.e. the bend angle of flex sensors and the Roll, Pitch and Yaw of the IMU Sensor) is fed to the neuro-computing unit, wherein the preprocessed data is fed to various machine learning and neural network models for the classification of the gestures performed.

B. Dynamic Spatio-Temporal Warping (DSTW)

This subsection discusses the dynamic spatio-temporal warping algorithm used for comparing the ISL gesture data collected through the glove, consisting of flex and IMU sensors placed suitably to abstract the clear movement of fingers and hand.

DSTW uses time and space variations as that of time in DTW [11], [12]. The DSTW algorithm and how it can be applied to handle time and space variations of two similar gestures and their analysis is explained below.

Considering two data sequences belonging to gestures S and T of varying time lengths N and M samples, which are sampled at the same rate (i.e. the time taken to perform the two gestures S and T may not be the same). Let:

$$S = (S_1, S_2, ..., S_N) \; and \; T = (T_1, T_2, ..., T_M) \tag{1}$$

be the measured values at the sampling times $t_1, t_2, ... \; t_N$ & $t_1, t_2, ... \; t_M$ for S and T, respectively.

In order to find the similarities between the two sequences, a cost matrix C of $N \times M$ dimension is defined and is formulated using the following equations:

$$C(0,0) = 0 \tag{2}$$

$$C(i,0) = \infty \; \forall \; i \in [1, N] \tag{3}$$

$$C(0,j) = \infty \; \forall \; j \in [1, M] \tag{4}$$

$$C(i,j) = |S_i - T_j| + \; min(C(i-1,j), \; C(i,j-1), \; C(i-1,j-1)) \; \forall \; i \in [1,N] \; and \; j \in [1,M] \tag{5}$$

With this cost matrix, the goal is to find the optimal path, which has the minimal overall cost that leads to gestures S and T having similarities to the maximum extent as compared to dissimilar gestures, as per the gestures given in the ISL video dictionary.

In order to have accuracy in the similarity measure is to follow certain necessary conditions i.e. to find the optimal path by traversing the minimum value in the cost matrix involving the time series gesture data formulated in eq. (5) are: (i) To travel the cost matrix from top left corner to bottom right corner, (ii) the path should be incremental in steps, (iii) the path should move from one cell (i, j) at a time, either to the right $(i+1, j)$ or bottom $(i, j+1)$ or bottom-right $(i+1, j+1)$ cell by choosing the cell with minimum of the costs $C(i+1, j)$, $C(i, j+1)$, $C(i+1, j+1)$ respectively as represented by equations 8, 9, 10, respectively.

Here, the optimal path O with length L can be obtained by traversing the minimum value of the cost matrix, is as follows:

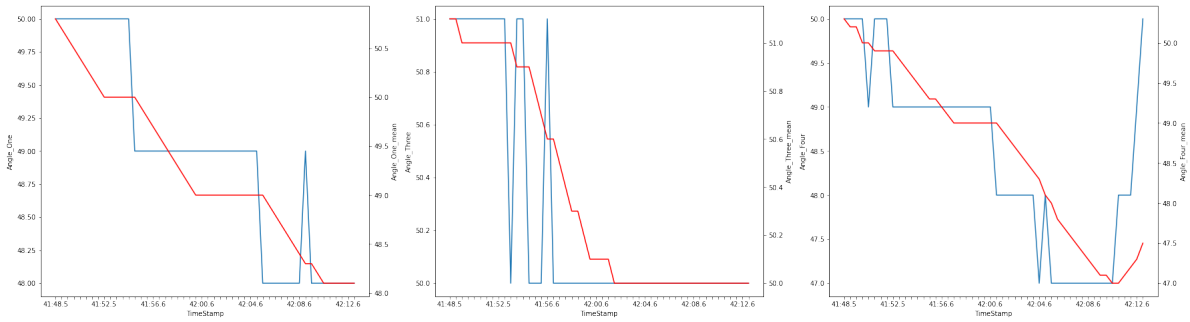

Fig. 2. Moving average implementation on gestures captured by data glove.

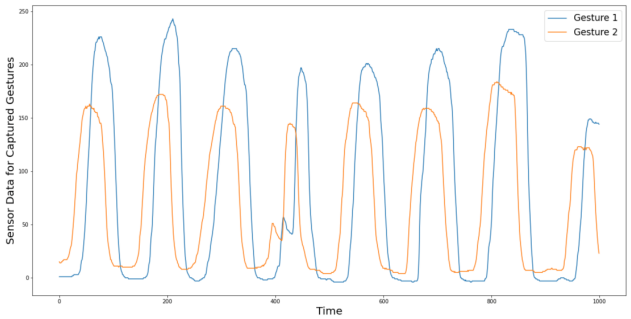

Fig. 3. Comparing two similar gestures of varying lengths with DSTW.

$$O = (O_1, \ldots, O_l, \ldots, O_L) \tag{6}$$

$$O_l = (i, j) \ for \ some \ i \in [1, N] \ and \ j \in [1, M] \tag{7}$$

$$O_1 = (1, 1) \ and \ O_L = (N, M) \tag{8}$$

$$O_1 \leq O_2 \leq \ldots \leq O_{L-1} \leq O_L \tag{9}$$

$$(O_{l+1} - O_l) \in (1, 0), (0, 1), (1, 1) \ for \ l \in [1, L-1] \tag{10}$$

The time series data discussed above and their respective mapping in space and time domain along with their sampled values is illustrated with examples below.

Two gestures of the same kind having different time and space variations can be compared using DSTW. Consider two gestures of the same kind, $G = [1, 2, 3, 4, 5]$ and $G^{'} = [1, 1, 2, 2, 3, 3, 4, 4, 5, 5]$. Both the gestures are the same but the first gesture is being performed twice as fast as the second one. The same is the case with space variations. Consider two gestures of the same kind $G = [2, 2, 3, 3, 3]$ and $G^{'} = [3, 3, 4, 4, 4]$. These gestures are varying with respect to space, which means the magnitude of the gestures being performed is different, but the gesture remains the same.

In order to handle the above situation, DSTW Algorithm 1 is proposed to integrate time and space for comparing the gestures carried out with different time durations to handle both the intra-sampling and inter-sampling variations. Comparison of two gestures using DSTW algorithm is shown in Fig. 3. The output of Algorithm 1 is the distance between the two gestures G and $G^{'}$.

C. BiLSTM Architecture

Similarly, the spatio-temporal variations with similarity principles are considered intrinsically in BiLSTM architecture and is explained below. BiLSTM architecture for classifying the gestures, as shown in Fig. 4, consists of two LSTM cells; forward LSTM and backward LSTM [13]. The forward LSTM processes the time-series gesture information from left to right and its hidden state $\overrightarrow{h}$ can be shown as

$$\overrightarrow{h} = LSTM(x_t, \overrightarrow{h_{t-1}}) \tag{11}$$

where x_t is the time series gesture data point at time t and $\overrightarrow{h_i}$ is the hidden state of the forward LSTM at time $t = i$.

The backward LSTM processes the time-series gesture information from right to left and its hidden state $\overleftarrow{h}$ can be expressed as

$$\overleftarrow{h} = LSTM(x_t, \overleftarrow{h_{t+1}}) \tag{12}$$

where $\overrightarrow{h_i}$ is the hidden state of the backward LSTM at time $t = i$.

Finally, the output of BiLSTM can be summarized by concatenating the forward and backward states as

$$h_t = [\overrightarrow{h_t}, \overleftarrow{h_t}] \tag{13}$$

Correlation of $\overrightarrow{h_t}$ and $\overleftarrow{h_t}$ at time t is performed to get better classification and feature extraction, since gestures will have spatio-temporal variations with redundant and varying features in different time spans.

In order to bring better accuracy of feature extraction in gestures, online learning is proposed. Online learning represented in Fig. 5 is done by performing the training process with one data point at a time. Fig. 5 shows the training process of the model where G_i is processed at time $t = i$. This approach is common when working with sequential data.

Once these gestures are classified and their meanings are mapped using ISL, then the obtained meanings are fed sequentially into the word and sentence formation unit with transformer-based NLP model to form grammatically correct sentences.

Algorithm 1: Dynamic Spatio-Temporal Warping

Input: Collected data for two gestures $G = [g_1, \ldots, g_N]$,and $G^{'} = [g_1^{'}, \ldots, g_M^{'}]$ as arrays of varying lengths N and M
Output: Similarity measure between G and $G^{'}$

$SpaceTimeGap \leftarrow matrix[0 \ldots N, 0 \ldots M]$
$for\ i \leftarrow 0\ to\ N\ do$
  $for\ j \leftarrow 0\ to\ M\ do$
    $SpaceTimeGap[i, j] \leftarrow \infty$
$SpaceTimeGap[0, 0] \leftarrow 0$
$for\ i \leftarrow 1\ to\ N\ do$
  $for\ j \leftarrow 1\ to\ M\ do$
    $cost \leftarrow distanceMeasure(g_i, g_j^{'})$
    $SpaceTimeGap[i, j] \leftarrow$
$cost + min(SpaceTimeGap[i-1, j],\ SpaceTimeGap[i, j-1],\ SpaceTimeGap[i-1, j-1])$
$return\ SpaceTimeGap[N, M]$

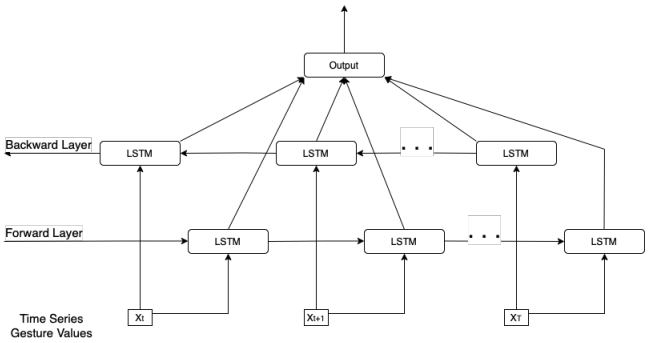

Fig. 4. BiLSTM architecture

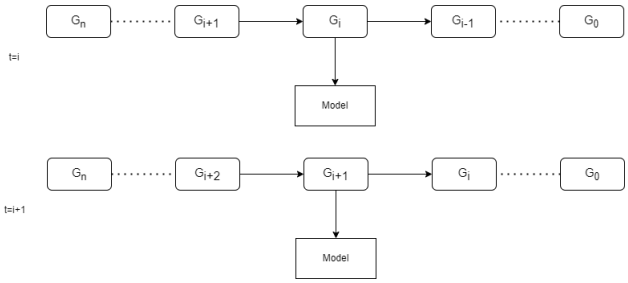

Fig. 5. Training process of BiLSTM model using online learning

IV. Implementation

Using the proposed model discussed in Section IV, a sensor-based HGR data glove was developed, is shown in Fig. 6 and it's circuit diagram in Fig. 7, that is capable of capturing hand and finger movements.

The data glove was developed from scratch by making use of five 4.5" Flex Sensors, one for each finger and an MPU6050 GY-521 IMU sensor on the dorsal (back) side of the hand and an Arduino Mega 2560 microcontroller. The flex sensor is used to calculate the finger bend angle for each of the fingers and the IMU sensor is used to calculate the Euler Yaw, Pitch and Roll values. A second glove was developed in the same fashion to replicate the process and validate the feasibility of building of the glove. Both the gloves are shown in Fig. 6.

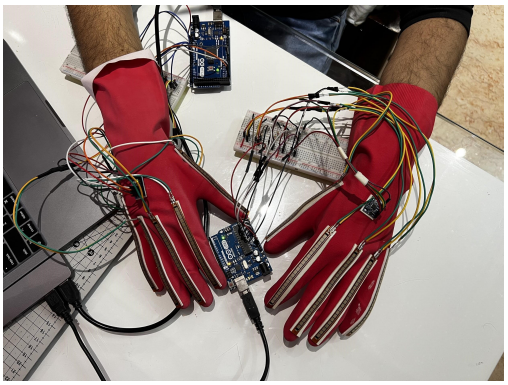

Fig. 6. Developed hand glove with embedded and sensing device to test the proposed model.

The raw data collected by the data glove is then processed by amplifying the signal and passed through low-pass filters to get rid of interferences and noises by using Moving Average Smoothing. It is then passed to the A/D converter of the microcontroller, capable of processing the time series data for machine learning algorithms to extract the features relating to the hand and finger movements carried out for the gestures intended by the users. This data is then processed through a neuro-computing system capable of classifying the gesture being performed in real time. The various hand gestures are mapped to their respective meanings and stored in a dictionary. These act as classes to create a labeled dataset for further training process.

Different gestures are considered for training the feed-forward neural network model integrated with DSTW and the BiLSTM model. For each gesture, 100 data frames were collected. After collecting the data, having taken 5 flex sensor angles shown in Fig. 8 captured at intervals of 20ms. A feed-forward neural network architecture is used with memory-based learning as the learning method to classify the gestures into their meanings. Dynamic spatio-temporal warping is used to compare time series data with spatial and temporal variations. The algorithm shows time complexity of $O(m * n)$. During the experimentation, numba python library for multi-processing was used to bring down the computation time from 1 hour to around 30 seconds. This resulted in increasing the

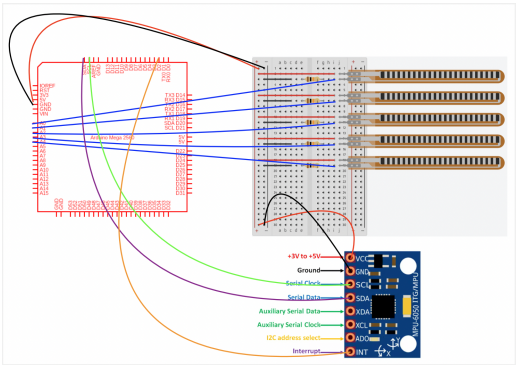

Fig. 7. Circuit diagram of the proposed HGR model.

TABLE I. GRAMMAR ERROR CORRECTION (GEC) OUTPUTS FOR SAMPLE INPUTS

Input	Restructured Sentence	GingerIt	Gramformer	Actual Sentence
Your phone number what	What your phone number	What's Your phone number	'What is your phone number ?', 'What is Your phone number.'	'What is your number?'
I go theatre	I go theatre	I go theatre	'I go to the-atre.', 'I went to theatre'	I go to a the-atre

number of times the gestures were trained in a given period of time.

Once the gestures are classified, they act as the input to the NLP model that aims to determining the start and stop of a sentence and further segmentation is performed. They are then processed to correct the grammar and form meaningful sentences. DeepSegment [14], is a GloVe + BiLSTM Conditional Random Field (CRF) sequence model, to segment sentences with no punctuations. A trend was noticed in the ISL structure and grammar, where the interrogative sentences ended with the wh-pronoun, and the sentence was restructured by making use of the spaCy [15] library. However, the sentences were still incomplete, as there are no signs in ISL for articles. To combat this, grammar correcting libraries such as GingerIt [16], an LSTM based architecture, and Gramformer [17], a GPT-2 based model were used. Once this is done, the grammatically correct sentence then acts as an input to the text-to-speech generative algorithm which aims at conversion of the sentence to audio which can be played over a sound system / mobile device / speaker which helps the person at the receiving end understand what the user is trying to communicate. Table I illustrates the comparison between some of the grammar correction technologies like GingerIt and Gramformer. It was observed that Gramformer gave more accurate results with respect to grammar correction.

The data is fed into a BiLSTM model represented in Fig. 9 with 64 neurons in its input layer, which are BiLSTM in nature, another BiLSTM layer with 64 neurons as the hidden layer, and one neuron in the output layer, which gives us the class of the gesture. Adam Optimizer is used with Mean Square Error as the loss function.

The simulation of the collected Flex Sensor angles is shown in Fig. 10. It represents the angles collected for 10 fingers on both hands using 10 graphs. To simulate the angle at which the finger is bent, two line segments are used having the same angle between them. One line segment is the line segment joining $(0.0, 1.0)$ and $(1.0, 1.0)$ and the other line segment is plotted using the Circle equation with center at (h, k), $(x-h)^2+(y-k)^2 = a^2$ for which the points on the circle are $x = h + a\cos(\theta)$ and $y = k + a\sin(\theta)$. Therefore, for the circle with center at $(1, 1)$ and radius $a = 1$, the circle equation becomes $(x-1)^2+(y-1)^2 = 1$ and for which the points on the circle becomes $x = 1 + \cos(\theta)$ and $y = 1 + \sin(\theta)$. Since the angles being plotted are in the fourth quadrant, $\theta = 360° - \theta_f$ where θ_f is the flex angle.

V. RESULTS AND DISCUSSION

Performance of the model was carried for different gestures in Indian sign language involving hand and finger movements. Feed-forward neural network with memory-based learning integrated with Dynamic Spatio-temporal Warping and BiLSTM models were compared to classify the time series data collected from the data gloves. BiLSTM was found to have better performance during the experimentation. Trasformer-based NLP model for grammar correction modules was found to work accurately for the contexts considered during the experimentation. Different performance parameters were considered in the experimentation, some of them are discussed below.

With the Feed forward neural network model integrated with DSTW, the data is split into train data - 70% and test data - 30% resulting in an accuracy of 96.74% as shown in Fig. 11.

With the BiLSTM model, the data is split into train data - 70% and test data - 30% resulting in an accuracy of 98.91% as shown in Fig. 12.

VI. CONCLUSION

The development of wearable devices for sign language translation enhances the collaborative interaction with the augmentation of reality. In comparison to the vast number of existing technologies present to aid the deaf population, such specific wearable devices will be inexpensive and easy to use. Flex and IMU sensors calibrated and integrated by using hardware and software filters to collect accurate data of hand and finger movements with low amount of noise are proposed. Machine learning methodologies like Feed-forward neural network with memory-based learning and BiLSTM integrated with DSTW has enhanced the performance of gesture recognition as per the ISL video gestures. From the experimentation, BiLSTM model was found to have better performance with Feed-forward neural network with DSTW with accuracy of 98.91%. Multiprocessing features were used to improve the computation time during the classification. The gesture sequences with meanings used by NLP models for grammar correction were compared and it was found that transfomer-based models outperformed other types of models and parsers. Obtained sentences were conveyed to the user through communicable media as per the user's requirement.

Gesture recognition devices can be used in multiple ways and in various fields of science, like in the medical field for

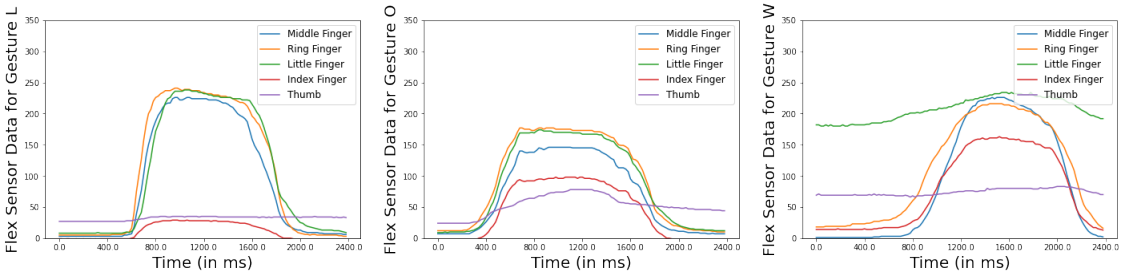

Fig. 8. Flex sensor data for gestures L, O and W.

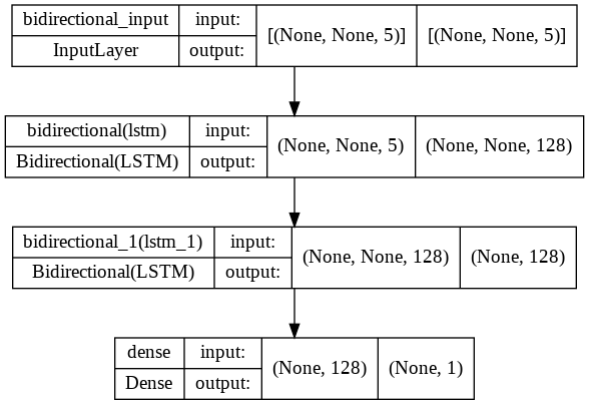

Fig. 9. BiLSTM model

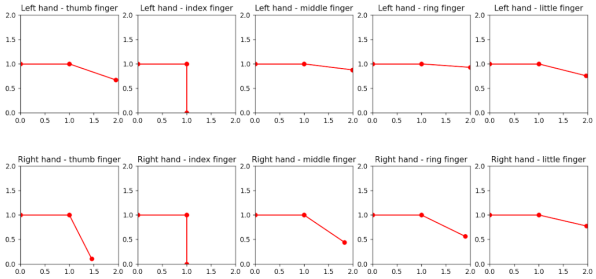

Fig. 10. Simulation of collected flex sensor data

	precision	recall	f1-score	support
Gesture L	1.0000	0.9706	0.9851	34
Gesture O	0.9655	0.9655	0.9655	29
Gesture W	0.9333	0.9655	0.9492	29
accuracy			0.9674	92
macro avg	0.9663	0.9672	0.9666	92
weighted avg	0.9681	0.9674	0.9676	92

Fig. 11. Classification report of the feed-forward neural network with DSTW model

	precision	recall	f1-score	support
Gesture L	0.9688	1.0000	0.9841	31
Gesture O	1.0000	0.9688	0.9841	32
Gesture W	1.0000	1.0000	1.0000	29
accuracy			0.9891	92
macro avg	0.9896	0.9896	0.9894	92
weighted avg	0.9895	0.9891	0.9891	92

Fig. 12. Classification report of the BiLSTM model

remote operations/surgeries using HGR, as well as during the recovery of the patients either through physical rehabilitation or physical therapy that can be done remotely with gesture recognition. In other examples it can be used in interacting with a computer through HGR, virtual and augmented reality applications using gesture recognition, gesture controlled IoT home appliances and much more.

A. Future Scope

This proposed model can be built on and adapted to enable the communication between varied disabled communities, like a person from the deaf community and a person from the blind community. The placement of the sensors on the glove could be optimized to ensure that the data collected is representative of the hand gesture being performed. To make the glove more responsive and usable in real-world scenarios, the data processing and classification can be improved to work in real-time. The number of classes of gestures that the glove is able to recognize should include all the gestures in the ISL.

ACKNOWLEDGMENT

The authors thank the Artificial Intelligence and Robotics Technology Park (ARTPARK) division of the Indian Institute of Science (IISc), Bangalore for providing resources and technical support. We also acknowledge the Information Science and Engineering department of M S Ramaiah Institute of Technology for their support.

REFERENCES

[1] United Nations, International day of sign languages, https://www.un.org/en/observances/sign-languages-day, 2022

[2] Faculty of Disability Management and Special Education, "Indian Sign Language Portal", https:// indiansignlanguage.org/, 2022

[3] Chuang, W.C., Hwang, W.J., Tai, T.M., Huang, D.R., Jhang, Y.J. Continuous finger gesture recognition based on flex sensors, 2019

[4] Zhou, Z., Chen, K., Li, X., Zhang, S., Wu, Y., Zhou, Y., Meng, K., Sun, C., He, Q., Fan, W., Fan, E., Lin, Z., Tan, X., Deng, W., Yang, J., Chen, J. Sign-to-speech translation using machine-learning-assisted stretchable sensor array, 2020

[5] Tai, T.M., Jhang, Y.J., Liao, Z.W., Teng, K.C., Hwang, W.J. Sensor-based continuous hand gesture recognition by long short-term memory, 2018

[6] Zheng, J., Zhao, Z., Chen, M., Chen, J., Wu, C., Chen, Y., Shi, X., Tong, Y. An improved sign language translation model with explainable adaptations for processing long sign sentences, 2020

[7] Lang, B. Parsing incomplete sentences, 1988

[8] Alikaniotis, Dimitrios & Raheja, Vipul. The Unreasonable Effectiveness of Transformer Language Models in Grammatical Error Correction, 2019

[9] Patil, S., Bidari, I., Sunag, B., Gulahosur, S.V., Shettar, P. Application of hmi technology in automotive sector, 2016

[10] Government of India, Indian Sign Language Research and Training Center (ISLRTC), http://www.islrtc.nic.in./, 2022

[11] Jangyodsuk, P., Conly, C., Athitsos, V. Sign language recognition using dynamic time warping and hand shape distance based on histogram of oriented gradient feature, 2014

[12] Barth, J., Oberndorfer, C., Pasluosta, C., Schülein, S., Gassner, H., Reinfelder, S., Kugler, P., Schuldhaus, D., Winkler, J., Klucken, J., Eskofier, B.M. Stride segmentation during free walk movements using multi-dimensional subsequence dynamic time warping on inertial sensor data, 2015

[13] Hameed, Z., Garcia-Zapirain, B. Sentiment classification using a single-layered bilstm model, 2020

[14] notAI tech, "Deepsegment", https://github.com/notAI-tech/deepsegment, 2020

[15] Explosion, "spacy", https://github.com/explosion/spaCy, 2022

[16] Azd325, "Gingerit" https://github.com/Azd325/gingerit, 2022

[17] Damodaran, P. "Gramformer", https://github.com/PrithivirajDamodaran/, 2022