Sign Language to Speech Translation

Aishwarya Sharma
M. Tech. Artificial Intelligence
Mukesh Patel School of Technology Management and Engineering, NMIMS
Mumbai, India
aishwaryansharma@gmail.com

Dr. Siba Panda
Mukesh Patel School of Technology Management and Engineering, NMIMS
Mumbai, India
siba.panda@nmims.edu

Prof. Saurav Verma
Mukesh Patel School of Technology Management and Engineering, NMIMS
Mumbai, India
saurav.verma@nmims.edu

***Abstract*— Although a certain fraction of the world suffers from speech and hearing disabilities, sign language is not a widespread language across the world. In today's oral world, a gesture-based language is not particularly popular with the general masses. However, sign language itself is a fully developed language with multiple regional dialects across the globe. Therefore, to aid with a smoother communication between the speaking and non-speaking world, technical developments can be introduced. A considerable amount of work has been done in this direction. The basic need of the hour is for an application that can function in real-time and can facilitate real-time conversations between a person who can sign in sign language and one that cannot. This paper proposes an application that works on this problem statement. In order to construct such an application designed to respond in real-time and aid a live conversation between a speaking and non-speaking individual, it is necessary to allow for live video inputs to be made to the app which would then be translated to speech. This paper proposes the use of convolutional neural networks (CNN) alongside the use of Text-to-Speech translator. By use of the CNN algorithm, the gestures can be identified by the proposed application and converted to text, which can then be converted into speech.**

***Keywords*— Indian Sign Language; real-time; gesture recognition; neural networks; machine learning; text-to-speech algorithms**

I. Introduction

Sign languages, while popular with the speech and hearing disabled communities, do not have high popularity with a majority of the speaking world, thereby posing a barrier in the communication between the two masses. These languages are recognized as natural languages with their own grammar and lexicon, making them fairly complicated to be understood by a layman.

Sign Languages vary based on location, with one of the most popular ones being American Sign Language (ASL) and British Sign Language (BSL). Stands to reason that in India, the Indian Sign Language (ISL) is the version of the sign language in use. While this dialect of signing is highly popular in the country, the speaking masses, understandably, have little to no comprehension of this. The learning resources for this version of the sign language are however massive, with the Indian government being supportive of creating an online dictionary.

The first edition of the ISL Dictionary comprised of 3000 terms and was launched in March, 2018, while the second edition was launched in February, 2019 [1]. This initiative by the Indian Sign Language Research and Training Centre (ISLRTC) under the Department of Empowerment of Persons with Disabilities (DEPwD) of Ministry of Social Justice & Empowerment, [2] is mindful of the fact that discrimination against such minorities do occur even in modern times [3]. As a step towards spreading awareness, the resources for the learning of sign language need to be made available for public consumption.

In India alone, over 50 lakh people suffer from hearing disabilities as per the 2011 Census [4] which made for 19 percent [5] of the total population. Close to 20 lakh people suffer from speech impairments [4], making for roughly 7 percent of the population [5]. In order to bridge this communication gap and aid with the ease of life for these minorities, the use of artificial intelligence aided technology is introduced.

Several technologies have been proposed and created to help with the translation of sign languages to spoken languages and vice versa. Although efforts are being made in the direction, technologies for everyday use are yet to be developed and popularized. The text or speech to sign translators that exist today are relatively basic in design, while the sign to text or sign to speech translators are not enhanced sufficiently to benefit the target audiences or to adapt to a daily routine of individuals.

Therefore, it becomes the need of the hour to look into the development of tools that can adapt to the day-to-day lives of civilians with speech and hearing disabilities. What is required is a solution that allows for the non-hearing/non-speaking communities to be able to interact seamlessly in the real world with hearing/speaking individuals. In other words, a device or an interface needs to be developed that can aid any individual to be able to interpret sign language as it is being spoken to them.

This paper proposes such a system: an interface that can translate sign to speech in real time. For the purposes of this implementation, the approach taken would be as follows:

1. Capturing of the gestures as input for the interface
2. Gesture recognition
3. Translation of signs to text
4. Conversion of the text to speech and presenting an audio translation as output

A central requirement here that this implementation depends on, is the need for the system to provide feedback in real time. In ideal situations where communication between someone who does and does not understand sign language is to occur, the requirement would immediately be to have the translation done as soon as the signer finishes signing. This would require for:

a. The distinction to be made about the start and end of a sentence, and
b. For there to be zero to minimal lag in the translation

Signing can be broken down into three major components: stationary gestures, gestures involving movement and facial expressions. However, a majority of the signs can be identified by the hand-based gestures. Facial expression and body language are generally more related to the expression of the sentiment underlying the sign. Although, it is worth noticing that the differences in body language from one individual to another, results in a difference in signing from one person to another. Meaning, two people could show variance in signing the same message due to differences in their generic body languages. This is akin to people having different handwritings and for the same word, different people will write differently and there would be variances observed in the final written text.

With regards to the implementation of this design, the primary focus would be the recognition of the hand gestures. Later stages of development would also involve the facial expressions and possibly the body language of the signer. The implementation would be divided into phases basis this.

The initial phase of implementation of this design would relate to the recognition of the stationary gestures focusing primarily on hand gestures. For this, the approach would be as follows:

1. Constructing a dataset of stationary signs
2. Using image processing and computer vision, identification of the hand gesture in the input image
3. Creation of a gesture recognition model that would recognize and translate the sign to text
4. Integration of machine learning algorithms that would help the system identify signs with a higher accuracy, the greater the number of times that the system is run and for higher the number and variety of inputs that it is run for

The next phase of the implementation would work to incorporate the recognition and identification of gestures with motion involved as well as the identification, while further phases would look into the translation of signs dependent on facial expressions and body language.

II. Literature Review

A number of technologies have been developed for the translation of sign languages across the world. Some of these technologies make use of hardware while others are purely algorithmic software applications. Systems have been proposed and developed wherein glove-like [6] extensions can be worn by the signer so that the gestures may be captured, identified and consequently translated. These gloves help capture the hand movements of the signer with heightened accuracy and post which these gestures are translated through appropriate algorithms.

The shortcoming with such devices is that they only capture the gestures made via hand and not the signs exhibited via expression and body language. Therefore, as an alternate to resolve this, pictures or videos are taken as inputs.

In such applications, the user input is taken via cameras on either mobile devices or specialized devices with in-built cameras are utilized. For instance, a Tamil Nadu based team worked on a mobile based application for which video inputs are utilized [7]. One of the methods of capturing inputs for this application is to take video inputs by attaching a camera to a cap. The signer would wear this cap and her/his actions would be captured from the top-view by this modified cap.

As previously discussed, all sign languages comprise of hand gestures, facial expressions and body language. Models of these nature focus on the hand-based gestures and do not widen the scope to the other aesthetics of signing. However, these models do provide results with varying accuracy within the scope of their work.

If body language were to be taken into account then the approach should be changed to accepting video inputs which capture the facial expressions as well as hand gestures of the signers. This would enable the capturing of the other factors in signing and make it possible to work on the data and recognize the signs more effectively.

This is one of the primary challenges faced in sign language translation, that is, the variety of factors involved in signing. Besides the basic aesthetics of signing in any of the sign languages, there are other varying factors to be considered. There exist individualistic variations in the signing of the same message. That is to say that if two people were to sign the same word or sentence, there would be variations in the signs exhibited purely out of the differences in their individual behaviors and body languages. There also exist physical and structural differences in the hands of individuals. These factors matter when feeding signs in as an input to an algorithm or an application. An algorithm is only as good as the training database. Hence, it must be seen to it that the training database comprises of a variety of signs from a variety of signers to capture maximal possible variations.

The input images or videos, then undergo processing through image processing and computer vision techniques such as Skin detection and Canny Edge detection. These techniques can be explicitly deployed or integrated with the gesture recognition algorithms implemented.

Based on the nature of the model proposed, a variety of algorithms have historically been made use for development of such translators. Depending on the requirements in each case, algorithms and concepts such as SURF (Speeded-Up Robust Features), SIFT (Scale Invariant Feature Transformation) [8], Neural Networks have been made use of.

Neural Networks or Artificial Neural Networks (ANN) are a biomimicry-based computing networks system, meaning they are based on the naturally occurring connection between neurons in animal brains. As a result, these are connectionist systems. ANN can be adapted to aid in deep learning giving rise to the entire category of Deep Neural Networks.

Deep neural networks can be implemented via a variety of mathematical methods. Convolutional Neural Networks (CNN) are one such implementation. CNN, also known as Space Invariant Artificial Neural Networks (SIANN), refer to a fully connected system where the networks are based on convolutions in one or more layers of neuron connectivity.

CNNs or ConvNets find high utility in image recognition and classification and are popularly applied in the fields of facial recognition, object detection, amongst others. [9]

The results of the implementation of the above algorithm should serve to provide a text output of the signs presented in the input. This output would then need to be converted to audio to serve the purposes of this system. Converting the text to speech would therefore be the last building block of this equation.

Several APIs and libraries exist which aid in the conversion of text to speech. These packages are generally open source and exist for a plethora of languages, thereby, providing flexibility of implementation by removing the constraint of the language used.

In Python, several Test-to-Speech translating libraries, such as the Google Test-to-Speech (gTTS) [10], exist. This is an online library, meaning, the program utilizing this would require active internet connection for this library to function correctly in real time. While this library, cannot function in offline mode, the library is highly programmable and customizable and can be made use of to store the outputs generated in separate audio files.

The primary motivation behind the work here is to develop technology to enable the translation of Indian Sign Language (ISL). While the model is scalable and given the appropriate training dataset, can be utilized for different dialects of Sign Language, the idea is to work with ISL due to lack of work done for the same. The ideal is to develop an easy-to-use and widely available application for the deaf and mute community in India. Undeniably, work has been done in this direction, however, there is still scope for attaining higher translation accuracies and especially with real-time translation.

III. Methodology

A heavy amount work has been done for the translation and transcribing of the American Sign Language (ASL). By comparison, the work done on the Indian Sign

Language has been limited and grants greater scope for innovations.

Although the proposed model here could be designed to adapt to any sign language by training the model on a database for that language, here, the Indian Sign Language (ISL) is specifically taken into consideration with the aim of translating the signs to spoken English.

The approach taken for the implementation of this model involves the following steps:

1. Creation/Compiling of a database with signs and gestures in chosen sign language
2. Gesture Recognition from input feed via application and implementation of a Neural Network based Algorithm
3. Processing using Classification and Machine Learning Techniques to train the model and increase the accuracy of translation
4. Generation of a text translation for input sign/gesture
5. Conversion of Text to Speech

The implementation of the above would be followed by a training and testing phase to train and enhance the model.

The proposed model is to run in real-time, that is to say that the ultimate design would accept video inputs and work on that. However, in the initial phases of development, the focus would be on capturing and translating stationary gestures. Therefore, the inputs would be taken as still images of signs that do not involve any motion. Specifically with regards to ISL, stationary signs are usually utilized for a majority of the alphabets and for simple words. Complex grammatic structures such as sentence formation require gestures with motion.

The dataset should ideally be comprised of a range of pre-determined set of gestures as made by a token number of individuals. The intention behind using such a dataset is to identify the same gestures as made by different individuals. The tool constructed should be able to identify the gestures accurately irrespective of the person signing. This would also mean that the variances in the styles and body languages of individuals would need to be covered in order for the tool to accurately recognize and translate signs.

The variations in the dataset would also serve to cover the differences in the physical attributes of people – including differences in structure of hands (length of fingers, size of palm, et al) and skin texture – as well as variations in image quality. Furthermore, the dataset should involve variations in other graphic factors, such as the attire of the signer or the color of their clothes, which might impact the capturing and identification of the signs by the algorithm during image processing.

The wider the varieties in the training dataset, the more accurate and stable the resultant model would be. The variety included in the dataset would help avoid overfitting of the model and thereby, provide relatively better and apt outcomes for the images or signs processed through the system. Increasing the variation in the dataset would lead to reduced bias in the resulting, trained model.

Once compiled, the dataset would be required to be split into three categories – training, validation and testing data. The training and validation data would have to be labelled for the system to be able to understand the classifications to be made. That would mean, that every sign or gesture in the training data would need to be labelled with the equivalent English translation.

The testing data would be a randomized mix of signs and gestures with no classification provided, for which the algorithm of the model will run and test accuracy.

With the dataset compiled and segregated, the model can be run to train and verify its algorithm. For this system, convolutional neural networks (CNN) are used. The visual inputs are fed into the CNN algorithm which has three primary functions:

a. Processing input graphic data for algorithmic use
b. Classification of signs and gestures and training the model by running a number of iterations on the training dataset
c. Gesture recognition by providing equivalent text output for input signs

The visual inputs can be processed using an assortment of image processing and computer vision techniques. This is integrated in the neural network algorithm employed. The algorithm then works to identify and classify gestures. Moreover, machine learning techniques are incorporated in the algorithm in order to allow for more accurate classifications and for the resultant system to adapt and improvise with use. This serves to train the system better with every iteration of data processed by it.

The output of this algorithm is a text message equivalent to the sign. This resulting text output is then run through a text-to-speech conversion model, thereby providing the final, requisite output.

For this, a number of libraries and APIs exist that can function as text-to-speech translators and even be customized to suit requirements. By making use of these, the derived text can be converted to provide an audio output. Basis the library used, the audio can either be played in real time and/or stored as per utilization. For the purposes of this system, the preference would obviously be to play the audio output in real time.

This would complete the system requirement and would therefore end the process, thereby translating sign to speech.

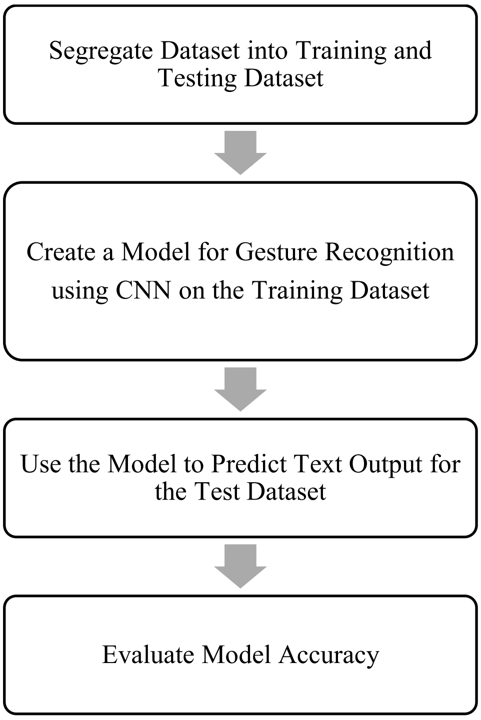

Fig. 1: Flowchart of the proposed sign to speech translator

IV. Algorithm

Although the algorithm could be developed on various platforms, the implementation for this system would be done in Python. This is because Python is a high-level, interpreted language that finds application in systems relating to artificial intelligence. It is also one of the open source languages with a plethora of libraries which allow for enhanced developments and implementations.

Keeping in mind the requirements for the initial phase of implementation, a dataset was compiled of stationary signs from a pre-existing, open source GitHub repository [11]. This dataset comprises of ISL signs as signed by different users. For each sign, images have been taken as signed by different users and with varying degree of image clarity. This would help strengthen the model during training. This composite dataset was then divided into training and test images.

The CNN algorithm makes use of Keras Deep Learning library. Keras offers pre-trained networks that can aid with the identification and classification of object categories with high accuracy [12]. It can classify visual inputs in up to thousand object categories. In order to customize these classifications, Keras is run on top of TensorFlow. This would allow for the inputs to be classified into relevant and requisite categories, thereby allowing for the construction of a specialized model.

For the model to be able to attain a high accuracy in gesture recognition, it is necessary for it to run multiple iterations on the training dataset. These iterations can be defined by determining the number of Epoch cycles to be run. Keras is built as a machine learning algorithm. With a vast and varied training dataset, and a sufficient number of runs, the model can adapt and offer greater the chances of producing accurate results.

```
Model: "sequential_1"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
resnet50 (Model)             (None, 2048)              23587712
_________________________________________________________________
dense_1 (Dense)              (None, 256)               524544
_________________________________________________________________
dropout_1 (Dropout)          (None, 256)               0
_________________________________________________________________
dense_2 (Dense)              (None, 15)                3855
=================================================================
Total params: 24,116,111
Trainable params: 528,399
Non-trainable params: 23,587,712
_________________________________________________________________
```

Fig. 2: Model Architecture of the Convolutional Neural Network

The results in this case, would be the text output for the sign input. This output is then processed by running it through a text to speech translator.

Once run through this translator the outputs can either be played or stored and kept in audio file format. This resulting output is treated as the final outcome of this system which would therefore, in conclusion, be able to convert a sign to speech.

V. Implementation

For the Phase 1 of implementations, a dataset of stationary gestures was utilized for model creation and prediction. Once designed, a number of tests were run on the algorithm to determine its accuracy. Basis different combinations of the input data and the number of iterations the neural network was run for, accuracy of the model was determined.

Multiple instances of the input dataset were generated and processed through the algorithm for varying number of Epoch cycles to verify the accuracy of the model generated.

For a given input dataset comprising of a set of gestures, the model was trained for 25, 50, 75 and 100 Epoch cycles. The Log-loss function was observed to decrease with every cycle with an in increase in accuracy. It was found that higher the number of Epoch cycle, greater the stability of the accuracy.

The model accuracy grew up to the range of 85-95% for the training dataset and in the range of 75-85% for the validation dataset. The accuracy for every Epoch cycle for both, the training and the validation dataset, reflect that the model has been suitably trained and is neither underfitted nor overfitted.

As the accuracies increase with every cycle, the loss decreases. The graphical representation of the loss function and accuracies are provided in the following images.

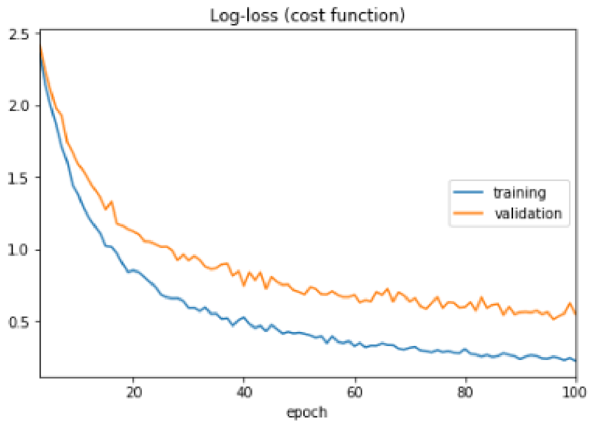

Fig. 3: Loss Function for a dataset of 15 signs and 100 Epoch cycles

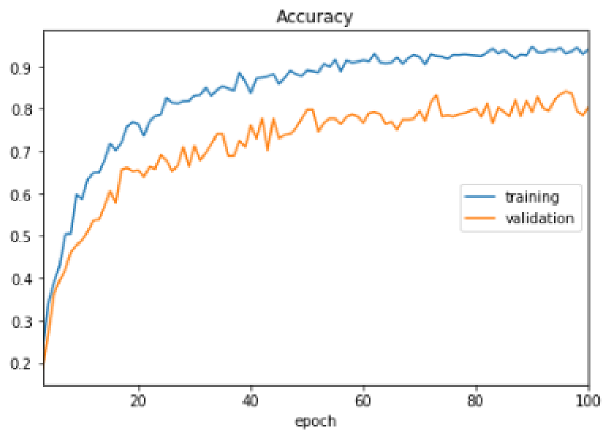

Fig. 4: Accuracy for a dataset of 15 signs and 100 Epoch cycles

Generally, the model appeared to reach stability after approximately 50 Epoch cycles. This was true for a variety of input datasets taken.

The trained model was then used to identify test signs. The model generated a confidence interval for a test image by generating a confidence interval for every sign it was trained on and then opting for the highest interval. Accordingly, the model labels a sign and translates text to speech.

In certain instances, the model may not be able to differentiate between similar signs. This would be improved upon by further iterations and implementation of deep learning models to the algorithm.

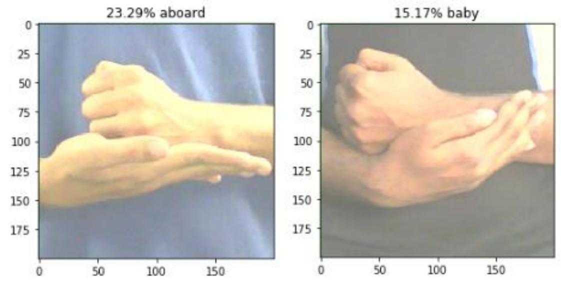

Fig 5: Correctly identified test signs along with their confidence intervals as identified by the trained model, developed using a training dataset of 15 signs and 75 Epoch cycles

VI. Conclusion

In India, where reports of bias against the speech and/or hearing-impaired communities are not unheard of [3], having a tool that aids in the communication of the speaking community with these minorities, helps level the field by opening a dialogue between the two masses. In doing so, it reducing the barriers for the aforementioned minorities.

A primary advantage of this model is that it is designed to be an interface that functions in real time and would be available to the masses. Should this be developed into a mobile application, it would ensure that the model would be easily accessible and effectively distributed amongst the target audiences.

There do exist limitations with this proposed system. Like any system, this system can also not promise a hundred percent accuracy of the translation of signs. Since this system is proposed to function in real-time, the inaccuracies in translation would be harder to identify and develop on. The reason behind this is that for the inaccuracies to be pointed out on the fully developed system, the user would have to cite the errors. The likelihood of a user flagging an error in real-time use would considerably be lower since, presumably, the user would be in a midst of a conversation.

However, this does not rule out complete the probability of real-time users flagging errors in translation. In-built machine learning algorithms can then utilize the flagged errors to improve upon the system.

Furthermore, the algorithm may yield inaccurate results for similar signs which could be deciphered by a person based on context or moods or other related factors, however, with this algorithm, context recognition is not a feature. It works primarily on translating clean, distinctly made hand gestures or signs.

Lastly, certain libraries used would depend on an active internet connection. This would mean that the final developed system would function only with an internet access.

VII. FUTURE SCOPE

The proposed system currently describes the translation from a particular variant of the sign language to spoken English. This could be built upon to create a more scalable model which could translate from a plethora of dialects of sign languages – be it Indian, American or British sign language – to any spoken language as chosen by the user. This could potentially allow for someone signing in Pakistani Sign Language to communicate with someone who understands Italian.

Another aspect of development would be sentiment analysis of the signer. As mentioned before, sign language is also heavily dependent on the facial expressions and body language of the signer. Therefore, encapsulating the intent or the sentiment of the signer, alongside the gesture s/he is making would lead to a more accurate system which would be considerate of the sentiment or emotion of the signer. This could also potentially help enhance the accuracy of the translation as certain signs are dependent on these factors. Incorporation of these factors and expansion of the model on the same would ideally grant a smoother conversation in the real world.

REFERENCES

[1] ISL Dictionary Launch, 2019 [Online]
Available:http://www.islrtc.nic.in/isl-dictionary-launch

[2] 2nd Edition of Indian Sign Language Launched, 2019 [Online]
Available:https://currentaffairs.gktoday.in/2nd-edition-indian-sign-language-dictionary-launched-02201966449.html

[3] NPR, A Mom Fights To Get An Education For Her Deaf Daughters, 2018 [Online]
Available:https://www.npr.org/sections/goatsandsoda/2018/01/14/575921716/a-mom-fights-to-get-an-education-for-her-deaf-daughters

[4] Department of Empowered Persons With Disabilities (Divyangjan), State UT Wise Persons with Disability: Number of Disabled Persons Disability Wise as per Census 2011, 2015 [Online]
Available:http://disabilityaffairs.gov.in/content/page/state-ut-wise-persons.php

[5] Disabled Persons in India: A Statistical Profile 2016, 2016 [Online]
Available:http://mospi.nic.in/sites/default/files/publication_reports/Disabled_persons_in_India_2016.pdf

[6] D. Sharma, D. Verma, P. Khetarpal, 'Labview Based Sign Language Trainer Cum Portable Display Unit For The Speech Impaired', in 2015 Annual IEEE India Conference (INDICON), 2015

[7] Y. Madhuri, G. Anitha, M. Anburanjan, 'Vision-Based Sign Language Translation Device', in 2013 International Conference on Information Communication and Embedded Systems (ICICES), 2013

[8] SIFT:Theory and Practice [Online]
Available:http://aishack.in/tutorials/sift-scale-invariant-feature-transform-introduction/

[9] Understanding of Convolutional Neural Network (CNN) — Deep Learning, 2018 [Online]
Available:https://medium.com/@RaghavPrabhu/understanding-of-convolutional-neural-network-cnn-deep-learning-99760835f148

[10] gTTS.PyPI, gTTS 2.0.4 [Online]
Available: https://pypi.org/project/gTTS/

[11] Indian Sign Language Recognition System [Online]
Available: https://github.com/abhishekdudhal/Indian-Sign-Language-Recognition-System

[12] ImageNet: VGGNet, ResNet, Inception, and Xception with Keras, 2017 [Online]
Available:https://www.pyimagesearch.com/2017/03/20/imagenet-vggnet-resnet-inception-xception-keras/

[13] Convert Text to Speech in Python [Online]
Available:https://www.geeksforgeeks.org/convert-text-speech-python/

[14] Text to Speech [Online]
Available:https://pythonprogramminglanguage.com/text-to-speech/

[15] J. Singha, K. Das, 'Recognition of Indian Sign Language in Live Video', in International Journal of Computer Applications (0975 – 8887) Volume 70– No.19, May 2013

[16] P. V. V. Kishore, P. Rajesh Kumar, 'A Video Based Indian Sign Language Recognition System (INSLR) Using Wavelet Transform and Fuzzy Logic', in IACSIT International Journal of Engineering and Technology, Vol. 4, No. 5, October 2012

[17] D. Vishal, H. M. Aishwarya, K. Nishkala, B. T. Royan, T. K. Ramesh, 'Sign Language to Speech Conversion: Using Smart Band and Wearable Computer Technology', in IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), 2017

[18] V. N. T. Truong, C. Yang and Q. Tran, "A translator for American sign language to text and speech," 2016 IEEE 5th Global Conference on Consumer Electronics, Kyoto, 2016, pp. 1-2, doi: 10.1109/GCCE.2016.7800427.

[19] R. Shahriar, A. G. M. Zaman, T. Ahmed, S. M. Khan and H. M. Maruf, "A communication platform between bangla and sign language," 2017 IEEE Region 10 Humanitarian Technology Conference (R10-HTC), Dhaka, 2017, pp. 1-4, doi: 10.1109/R10-HTC.2017.8288892.

[20] A. Jarndal, A. Al-Maflehi, 'On Design and Implementation of A Sign-toSpeech/Text System' in 2017 International Conference on Electrical, Electronics, Communication, Computer and Optimization Techniques (ICEECCOT) , 2017

[21] S. Y. Heera, M. K. Murthy, Sravanti V. S., S. Salvi, 'Talking Hands – An Indian Sign Language to Speech Translating Gloves', in International Conference on Innovative Mechanisms for Industry Applications (ICIMIA 2017), 2017

[22] K. K. Dutta, S. K. Raju K., A. Kumar G.S. and S. A. Swamy B., "Double handed Indian Sign Language to speech and text," 2015 Third International Conference on Image Information Processing (ICIIP), Waknaghat, 2015, pp. 374-377, doi: 10.1109/ICIIP.2015.7414799.

[23] N. C. Camgoz, S. Hadfield, O. Koller, H. Ney and R. Bowden, "Neural Sign Language Translation," *2018 IEEE/CVF Conference*

on Computer Vision and Pattern Recognition, Salt Lake City, UT, 2018, pp. 7784-7793, doi: 10.1109/CVPR.2018.00812.

[24] J. Kunjumon and R. K. Megalingam, "Hand Gesture Recognition System For Translating Indian Sign Language Into Text And Speech," 2019 International Conference on Smart Systems and Inventive Technology (ICSSIT), Tirunelveli, India, 2019, pp. 14-18, doi: 10.1109/ICSSIT46314.2019.8987762.

[25] E. Abraham, A. Nayak and A. Iqbal, "Real-Time Translation of Indian Sign Language using LSTM," 2019 Global Conference for Advancement in Technology (GCAT), BANGALURU, India, 2019, pp. 1-5, doi: 10.1109/GCAT47503.2019.8978343.