ASL DETECTION & EMOTION RECOGNITION SYSTEM DESIGN AND DEVELOPMENT USING RNN ALGORITHM

[1]Ramesh Chandra Goud, [2]K.Ramesh, [3]K.Jagadiswar Reddy, [4]Loka Keerthi Reddy
[1,2,3]Assistant Professor, [4]UG Student, [1,2,3,4]Department of Computer Science Engineering, Brilliant Grammar School Educational Society Group of Institutions Integrated Campus, Hyderabad, India

ABSTRACT

This study suggests a Windows programmed that tries to create a system that identifies speech of patients with vocal cord diseases, recognizes sign language at the word level, and recognizes the patient's emotion from the audio or EEG data. Combining these algorithms aims to reduce barriers to communication for those who are unable to communicate with others due to a variety of speech or hearing issues. Those with speech issues or impairments find it exceedingly challenging to communicate with those around them while trying to articulate a word since many of their sounds are missing. Examples include dysarthria, stuttering, apraxia, cluttering, and other issues. Because even the receiver must learn and grasp the language to understand the words, sign language will make social interactions challenging. So we aim to provide an interface that increases the chances of healthy interaction for the people with the above mentioned disabilities. It takes in different types of words being represented in sign language as input and predicts the right word. Another functionality of givingout the normal speech will be added which helps in keeping it more interactive.

INTRODUCTION

Issue Statement and Purpose Sign language is used by those who are deaf or deaf-blind to communicate. A sign language is composed of a range of facial emotions, different hand shapes, movements, and orientations, as well as a number of different hand gestures. Several sign languages are used across the world by various cultures. As compared to spoken languages, they are extremely rare. Emotions have a significant role in how individuals interact with one another. So, communication will be strengthened by being able to identify a person's underlying emotion. Emotions can be utilized to reveal a person's emotional condition so that the right replies can be given. Due to various medical issues, a lot of people struggle to engage with their family and friends. Considering a few of those disabilities, our team developed this interface to help those individuals in communicating more efficiently.

Objectives

In order to improve the state of communication of people with dysarthric speech disorders and other speaking disabilities, as described in the previous section, the developed system will assist these people by detecting mispronounced words and providing the correct word as an output, as well as extracting the person's emotion. The text and audio will be given out to give them the confidence to communicate easily.
The following are the objectives of the project:

- To provide an interface that can convert the impaired speech into a normal one for a better communication for people with speech disorders.
- To ensure that the system is able to output a readable text from the images of differentwords in sign language.
- Make sure that right emotion is predicted from the audio files given as input.
- Able to provide an efficient and easy to use interface.

Scope

The primary intention of our project is to reduce the problems faced by people who are vocally handicapped or completely disabled to speak and communicate through sign language. We aim to do this by creating an interface that makes their intentions evident tothose around them by sensing their emotions anytime they wish to say or talk something. As well as making it easier to detect the word that is said in sign language even to the people who are not aware of it.

The following are our project scopes:

- The target groups of this system are people with speech disorders who cannot produce a fluent utterance of a word and patients with EEG signals for communication.
- The input file for sign language detection can be an image file that may represent adigit, alphabet, or word.
- The input file for emotion detection can be the audio or EEG signal of a patient.
- The application is made as a Windows application, which anyone with a PC withwindows can access.
- The system uses convolutional neural networks for the processing of the data anddetection of the word and emotion. Impact, significance and contribution

Usually, the systems nowadays cannot give efficient results when impaired speech is given as input. Therefore in this project, the following are the advancements and limitations that were overcome:

- The user can give in the impaired speech as input and gets the right emotion as theoutput.
- The system will not give text only as the output, but also the normal speech will beproduced and given as output when an image of a word in sign language is provided.
- It gives a better way of communication for the user with vocal disabilities.
- This system also performs rechecking the right word during the prediction process notto get any incorrect output.
- It has different options to choose from to give as input according to ease of the userand get the output in a way that people around them can understand in a better way.

Related Work:

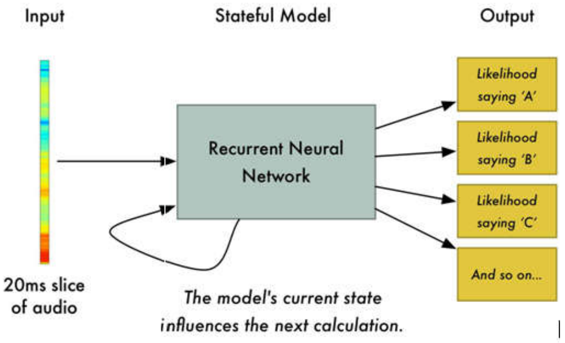
There were no speech recognition technology available in previous years to make conversions more comfortable. Many ASR (Automatic Speech Recognition) technologies have recently been developed to turn a person's speech into readable text. These systems are primarily designed to convert fluent speech. Automatic speech recognition systems (ASR) do a good job with fluent speech, but they struggle with dysfluent or stuttering speech. The development of ASRs, on the other hand, has never concentrated on speech with missing utterances. Furthermore, there are systems that can identify sign language.

Proposed work:

RNN algorithm

Firstly, the audio given as input will be broken into audio chunks of 20ms and then fed into a deep neural network.After feeding in, it will figure out the letter which matches the spoken sound.

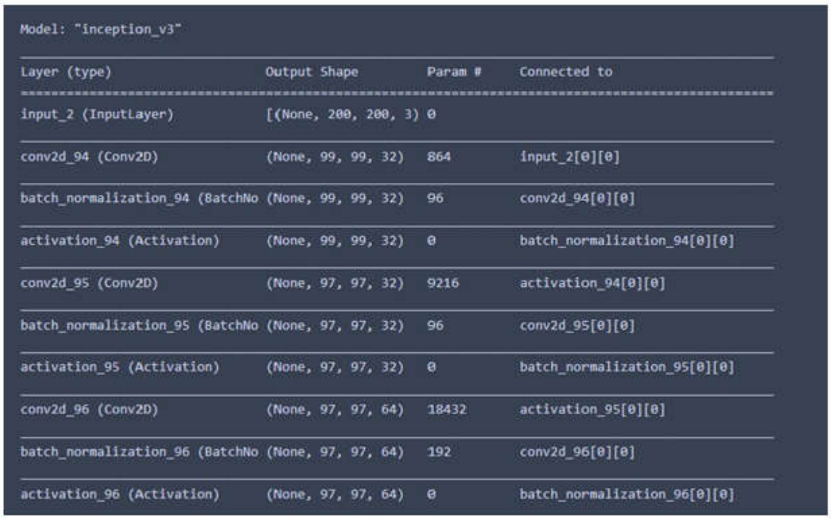
As this is a Dysfluent speech, some utterances might be missing eventually leaving gaps in the extracted word.RNN is a network that has a memory that decides future predictions. This is because as it predicts one letter it will affect the likelihood of the upcoming letter which it will predict too. This will help in giving out the perfect word from the dysfluent speech taken as input.



Results & Analysis

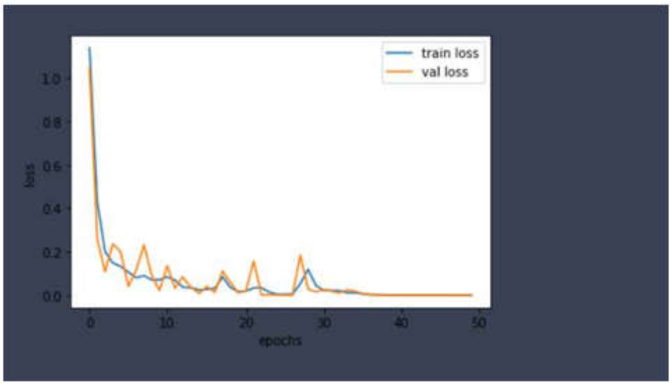
It is a model specifically and widely used to recognize the patterns and pixel categories inan image. It always ensures a decent amount of accuracy while training the model. It runs efficiently on cloud TPU. It tries to match the accuracy of the curves produced from the features extracted from the files containing images and gives out a result that is as accurate aspossible.

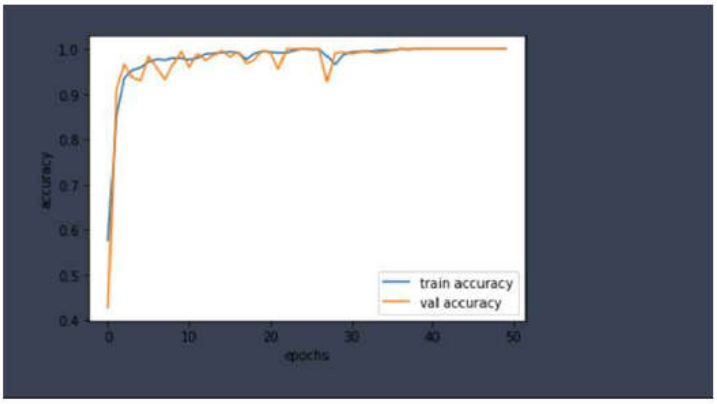
We have used the Inception V3 model to predict the word or number shown in sign languagein the image.



```
Model: "inception_v3"
__________________________________________________________________________________________________
Layer (type)                    Output Shape         Param #     Connected to
==================================================================================================
input_2 (InputLayer)            [(None, 200, 200, 3) 0
__________________________________________________________________________________________________
conv2d_94 (Conv2D)              (None, 99, 99, 32)   864         input_2[0][0]
__________________________________________________________________________________________________
batch_normalization_94 (BatchNo (None, 99, 99, 32)   96          conv2d_94[0][0]
__________________________________________________________________________________________________
activation_94 (Activation)      (None, 99, 99, 32)   0           batch_normalization_94[0][0]
__________________________________________________________________________________________________
conv2d_95 (Conv2D)              (None, 97, 97, 32)   9216        activation_94[0][0]
__________________________________________________________________________________________________
batch_normalization_95 (BatchNo (None, 97, 97, 32)   96          conv2d_95[0][0]
__________________________________________________________________________________________________
activation_95 (Activation)      (None, 97, 97, 32)   0           batch_normalization_95[0][0]
__________________________________________________________________________________________________
conv2d_96 (Conv2D)              (None, 97, 97, 64)   18432       activation_95[0][0]
__________________________________________________________________________________________________
batch_normalization_96 (BatchNo (None, 97, 97, 64)   192         conv2d_96[0][0]
__________________________________________________________________________________________________
activation_96 (Activation)      (None, 97, 97, 64)   0           batch_normalization_96[0][0]
```

We have acquired the train accuracy and loss graphs to get an idea about the correctness inprediction of our model.





The datasets we considered are taken from Kaggle where for ASL, we have considered a dataset that contains all 26 alphabets and 1-9 numbers for training the data. And for emotion recognition the dataset considered contains audio samples from 24 actors expressing different emotions for training the model. Using the above mentioned models and methodology, wehave achieved an overall accuracy of 73.33%.

CONCLUSIONS

Our system helps in improving the communication of people with speaking impairments and disorders. The recordings focus on a small vocabulary, including basic smartphone commands, such as “open contacts”, “start call”, “end call”. It predicts the word from the given images of sign language and generates the text from them. A normal speech will also be given out as output for better communication.

Future Scope

We look forward to add more features to the application and improve the performance of the models used without effecting the user experience. One way to achieve this is to find better datasets so that models can learn quicker than earlier and produce more accurate results. We are also planning to add audio output feature for emotion as well.

REFERENCES

1. Shakeel A. Sheikh ., Md Sahidulla ., Fabrice Hirsch ., Slim Ouni (2021). Machine Learning for Stuttering Identification: Review, Challenges & Future Directions.
2. Colin Lea., Vikramjit Mitra., Aparna Joshi., SachinKajarekar., Jeffrey P. Bigham (2021). SEP-28K: A Dataset for Stuttering event detection from Podcasts with people who stutter.
3. Rosanna Turrisi1., Arianna Braccia., Marco Emanuele., Simone Giulietti., Maura Pugliatti., Mariachiara Sensi., Luciano Fadiga1., Leonardo Badino (2021). EasyCallcorpus: a dysarthric speech dataset.
4. Sanjita. B. R., Nipunika. A., Rohita Desai., Department of ECM Sreenidhi Institute of Technology and Science, Telangana, India (2020). Speech Emotion Recognition using MLP Classifier.
5. Dr.R.L.K.Venkateswarlu., Dr. R. Vasantha Kumari., G.VaniJayaSri., International Journal of Scientific & Engineering Research Volume 2 (2011). Speech Recognition By Using Recurrent Neural Networks.
6. Aditya Amberkar., Gaurav Deshmukh., Parikshit Awasarmol., Piyush Dave., MCT’s Rajiv Gandhi Institute of Technology, Mumbai, Maharashtra (2018). Speech Recognition using Recurrent Neural Networks.
7. Suharjito., Ricky Anderson., Fanny Wiryana., Meita Chandra Ariesta., Gedeputra Kusuma., Sign Language Recognition Application Systems for Deaf-Mute People: A Review Based on Input-Process-

Output.

8. Sukhpreet Kaur., Nilima Kulkarni., CSE,MIT SOE, MIT ADT University, Pune, India (2021). Emotion Recognition – A review.
9. Analytics Insight., Speech Emotion Recognition through Machine Learning(2020).Article.
10. Amal Nair., A Beginners guide to Scikit-Learn’sMLPClassifier (2020) Article.,Analyticsindiamag.