Gesture Detection using Tensor flow lite Efficient Net Model for Communication and E-learning Module for Mute and Deaf

Snehal Patil, Yash Shah, Payal Narkhede, Abhinav Thakare, Rahul Pitale

Abstract*: Human communication plays a vital role; without communicating, day-to-day tasks seem difficult to complete. And the world has an almost 5% population that struggles with hearing or speaking disability, which contributes to 430 million people worldwide, and this will grow up to 900million just in the next 25 to 30 years. With the increasing noise pollution, hearing capacity degrades, leading to various hearing problems. The WHO statistics show that 32million kids are acoustically impaired. With disabilities, there are multiple issues these people face, such as lack of learning facilities, job opportunities, communication platforms, etc. These people need a cooperative environment to express, learn at their pace and level of understanding. This paper focuses on developing an application that bridges the gap between these acoustically disabled people and people unknown to their way of communication. The proposed research is an edge device application provides features like a gesture to text, speech to text, e-learning platform, and Alert mechanism. This paper majorly focuses on developing a friendly all in one platform for mute and deaf community for communication, learning and emergency alerts. The research was conducted with two approaches the traditional CNN and Tensorflow lite EfficientNet model to train the ASL (American Sign Language) dataset for the communication platform, where we obtained accuracy of 98.91% and 98.82% respectively. To overcome the computational barriers of traditional CNN approach, Tensorflow lite EfficientNet model was brought into the picture. The proposed methodology would help build a platform for the deaf and mute community to express themselves better and gain wider exposure to the world.

***Keywords:* ASL, android application, CNN, image classification, E-learning, alert mechanism, firebase.**

I. INTRODUCTION

Communication with others is needed for anything and everything while performing day-to-day activities. People who are blessed with the ability to speak and hear can communicate through words. Different countries have their language of communication, and one needs to learn the language to communicate. Similarly, while speaking with mute and deaf people, a person needs to know their language of communication. So far, from the ancient period, people have expressed themselves using various hand gestures and movements; This brings the concept of Sign language into the picture. So far, there are around 300 different Sign languages. Widely used is American Sign language(ASL), but every country has its sign language and lacks in providing one universally accepted Sign Language for communication. These sign languages help the deaf and mute to communicate with people around them. But statistically speaking, the World Federation of the Deaf (WFD) stated that 72million out of 430 million could only communicate using sign language; This shows the lack of education provided facilities for these differently-abled. According to the report by WHO, unaddressed issues faced by these differently-abled people included lack of education, communication medium, unemployment, facilities, etc.

In this paper, we have focused on four of the issues faced by these people, namely gesture to text, speech to text, e-learning platform, and the alert system. Gesture to text will be a live prediction of gestures performed; This will be provided in the android mobile application so that anyone can use it easily. Furthermore, the application catches the words spoken and types on screen for mute and deaf people to read. These two features will smooth out and shorten the drift between two parties unknown of their respective languages. The current system also provides an e-learning platform where users can find various courses created considering the need of the Mute and deaf community. Finally, a feature that helps alert the people in emergencies like a fire, natural disaster, etc., using vibration patterns

II. LITERATURE SURVEY

There are various applications available to provide the features of text to speech for communicating with each other. Of course, these applications can only be helpful for people who know the language spoken or typed. But most mute and deaf people use sign language to communicate. Play store offers several applications like Deaf communication, Easy to Talk, Talk to Deaf, Mute Helper, etc.; these applications only provide the functionality of text to speech and speech to text. Some other applications like Mute speaker, Indian Sign Language, etc.; give icons to text/speech and speech to text. These are limited resources to communicate on a daily basis. Various other approaches can be broadly classified into Gloves-based and Image-based.

Manuscript received on June 10, 2021.
Revised Manuscript received on June 15, 2021.
Manuscript published on June 30, 2021.
* Correspondence Author
Snehal Patil*, Pursuing, Department of Computer Engg., PCCOE, Pune (Maharashtra), India. Email: snehal.patil17@pccoepune.org
Yash Shah, Pursuing, Department of Computer Engg., PCCOE, Pune (Maharashtra), India. Email: yash.shah17@pccoepune.org
Payal Narkhede, Pursuing, Department of Computer Engg., PCCOE, Pune (Maharashtra), India. Email: payal.narkhede17@pccoepune.org
Abhinav Thakare, Pursuing, Department of Computer Engg., PCCOE, Pune (Maharashtra), India. Email: abhinav.thakare17@pccoepune.org
Rahul Pitale, Assistant Professor, Department of Computer Engg., PCCOE, Pune (Maharashtra), India. Email: rahul.pitale@pccoepune.org

Gloves based approach is a traditional approach that uses the pose of hand and motion for predicting hand gesture[3]. In addition, gloves based approach primarily uses proximity sensors to determine the bends of fingers[4]. There are approaches through which continuous moment of data is being captured with the help of flex sensors; these sensors then send data to a microcontroller like Arduino and, based on sensor data, predict the output[11]. Similarly, using flex sensors, the angle between fingers is calculated and based on the angle value of the resistor, these values will be used to determine the prediction[9]. The gloves-based approach gives accuracy but has the additional cost of hardware, which necessarily is not affordable for many people.

Another widely used approach was using Microsoft-Kinect; it is a hardware device specially designed as a low price device combining RGB camera and depth sensor. First, the application captures an image using an RGB camera; further, the depth sensor measures the depth and forms a skeleton which is then analyzed. Finally, the noise and background are removed from the analyzed skeleton and passed for prediction [13].

Other approaches include the software and hardware method together, where the motion measurements are taken using sensors and using these electrical signals, and fuzzy logic predictions are made[8]. The software approach generally uses a webcam/ mobile device camera for image capturing and uses OpenCV for pre-processing and TensorFlow for the training model for predicting the gesture [4,7,12]. However, these are few approaches for gesture-to-text applications.

While talking about e-learning applications, the google play store offers few applications like Sign Language ASL pocket Sign, Deaf-ISL, deaf Sign Language, etc. These applications provide basic day-to-day life sign gestures through video-based lessons. Looking at common educational approaches for these acoustically disabled people, bilingual-bicultural, and auditory/oral methods are widely efficient and used[19]. The bilingual method is about using text and signs for teaching using video lectures. High visualization helps students to grasp the content well[17]. Pre-quiz and post-quiz methods also help in analyzing how much progress a person has made. This method experimented on deaf and mute college students for teaching advanced engineering level courses.[6]

Therefore after analyzing various approaches for gesture-to-text and e-learning applications, developing an android application would be a handy and easy solution. An android application consuming minimal memory space and uses mobile CPU efficiently for computation would be the target achieved by the proposed system.

III. PROPOSED SYSTEM

The various currently available applications we studied have text-to-speech and speech-to-text options. These are pretty basic for people to communicate efficiently. As we saw, the issues faced by these people are expressing themselves and evolving into today's world. That is a lack of communication and education medium, thus in our system, we have proposed the following features:

1. Gesture to Text
2. Speech to Text
3. E-Learning Platform
4. Alert Mechanism

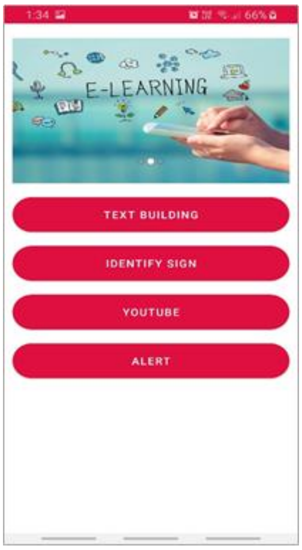

Fig -1: Home Page

A. Gesture to text

This feature focuses on live gesture-to-text prediction; this enables quick and fluent conversation without any latency in frame processing as compared to other mediums. This helps in bridging the communication gap by enabling people without any knowledge of sign language to understand the voice of acoustically disabled people. It predicts the English alphabets [a-z] and 3 additional gestures that are; delete, nothing and space.

The dataset we used here is a publicly available dataset, the ASL alphabet dataset; having around 87000 images. The dataset contains 29 classes with each comprising approx. 3000 images.

Further implementing this feature we followed two approaches:

1. Traditional CNN model
2. EfficientNet model

Concealed from the reviewers, and vice versa, throughout the review process. All submitted manuscripts are reviewed by three reviewer one from India and rest two from overseas. There should be proper comments of the reviewers for the purpose of acceptance/ rejection. There should be minimum 01 to 02 week time window for it.

1. Traditional CNN model:

As CNN is proven to be best as an image classification model, initiating this approach was an ideal start for the classification. Therefore proceeding with this approach, at first we resized all the images in the dataset into (64x64) and labelled them further.

The next step is splitting the dataset for test and train sets, where the ratio for test-train split is 0.2:0.8. Then applying One-Hot encoding onto the result set, converting it into categorical data so that our algorithm can understand and identify the relationship between labels. Then further on converting the train and test sets into float32 format and normalizing the pixel values by dividing them by 255, thus bringing them into the range of [0 - 1].

International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-10 Issue-8, June 2021

We added convolutional 2D layers and Maxpooling2D layers to our model. With "Relu" activation function for the convolution 2D layer and "sigmoid" activation function for the fully connected layer; the diagram of which is represented below:

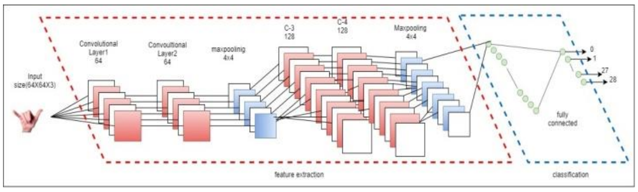

Fig -2: CNN model

The CNN model when integrated with edge device requires pre-processing such as rotation, resizing, normalization and reshaping for which we need to use libraries which are bulky and heavy for edge devices. Having those libraries and performing the steps as mentioned in the Fig -3 increases the computational space and time on edge devices. Thus, such bulky application would not be ideal when considering edge devices.

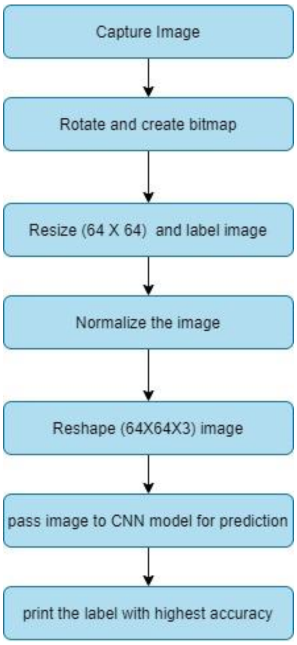

Fig -3: Traditional CNN Approach

The accuracy obtained through this traditional approach was 98.91%. But while implementing this model onto mobile/android devices, the pre-processing steps consumed time and extra space, and there was an additional requirement of the OpenCV library to be installed onto the device; which made it very heavy and bulky.

2. EfficientNet model

In order to compensate for the disadvantages of the traditional approach, the EfficientNet model was brought into the picture. EfficientNet lite model is specially designed for better performance on mobile, CPU, and GPU; thus overcoming our issue of computational resources constraints.

TensorFlow lite model maker library helps in simplifying the process of training our EfficientNet model using image_classifier.create() function. This model maker library makes use of transfer learning which results in better accuracy in fewer data and time. Hereafter using image_classifier.create() we get the EfficientNet lite model with the corresponding label file generated along with it. The model takes an input image of size 224x224 with three channels per pixel (RGB - Red Green Blue). Further making it of the size 224x224x3, this being the actual input to the first layer of the model. The label with the most probability as per the model is chosen to be our desired output.

Using this approach we obtained an accuracy of 98.28%. Also, the pre-processing of the input image was carried out without using any external or bulky library.

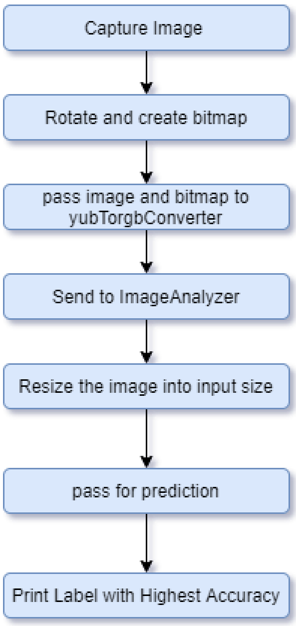

Fig -4: Gesture to Text Flow

In spite of the accuracy of the EfficientNet approach being 0.63% less than the traditional approach, it was taken into further implementation considering its advantages on mobile/android devices.

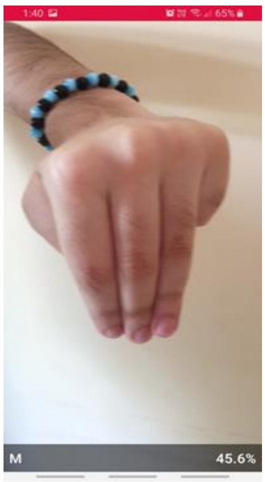

Fig -5: Gesture Prediction

Table- I: Comparison between the two approaches

Attributes	CNN	TF lite model
Accuracy	98.91%	98.82%
Pre-processing	Normalization, reshaping	reshaping
Memory required in edge device	>100 mb	<25 mb
Computation time required in edge device	Significantly more	Comparatively less

B. Speech to Text

Using the inbuilt features provided by the android allows conversion from speech to text. This is done with the help of Google Speech to Text API. Speech input is provided to the app or intent which is then streamed to a server, where the voice data will be converted into textual data and the response is sent back to our app. The Recognizer Intent is used in the Speech to text Android API, in which we basically trigger the Recognizer Intent and the speech input is fed to it, then the speech to text conversion outcome is returned back and displayed onto the label provided.

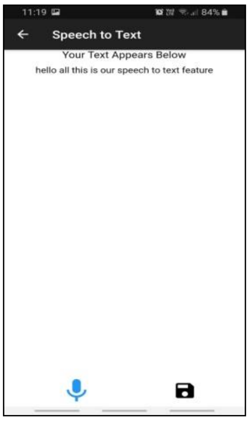

Fig - 6: Speech to text

C. E-Learning Platform

These mute and deaf people face economical challenges while incurring the high costs for learning and educational means due to the limited resources available for them to explore. Thus, to overcome this challenge of theirs, we have implemented this feature of E-Learning, where people or tutors from all over the world can upload videos through YouTube, where they just have to provide the unique id of their published videos. The videos and courses would be managed and moderated by a moderator before surfacing them onto our app. The videos' unique ids will be uploaded onto the Google Firebase, and these unique ids would be fetched on runtime from Firebase and listed onto our app. With the help of Youtube API, these videos can be played on the app. Thus providing an open-source and free-of-cost learning platform for these differently-abled people to enhance their skills and explore new opportunities around the world.

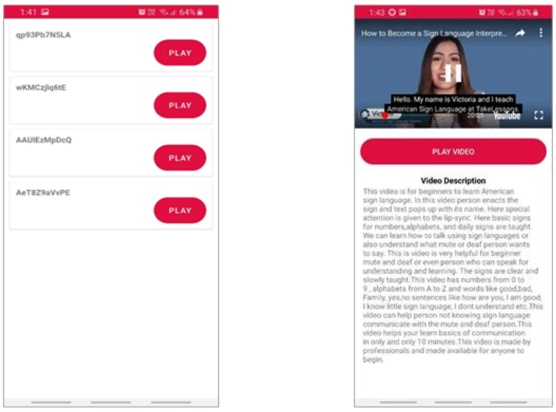

Fig - 7: E-Learning Platform

D. Alert Mechanism

These people are challenged with their hearing power, therefore it is difficult for them to figure out the alarming or chaotic situations around them, whether it may be a natural calamity or a man-made disaster. Thus, to overcome this challenge of theirs, we came up with this feature, where the location of the disaster or mishap can be uploaded onto the Firebase by any Rescue Department worker. And based upon these location changes in the Firebase, the users in the given perimeter of the location would be alerted through different and unique vibration patterns on their mobile phones, thus alerting them about their surroundings and urging them to reach out to a safe place.

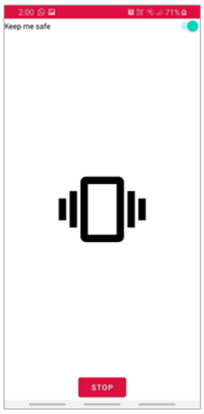

Fig -8: Alert Mechanism

IV. CONCLUSION

Mute and deaf people are quite dependent on human interpreter for communication and other needs. The proposed approach would help reducing the dependency of these people on human interpreter. Thus, the proposed approach of gesture detection using Tensorflow lite EfficientNet model is suitable for edge devices with less computational capabilities and making it compatible with minimum configuration devices. The e-learning approach proposed in this paper would abridge the gap to learn the industry skills for these people and also provide a platform for researchers and people to experiment with different learning techniques for these differently abled people.

Retrieval Number: 100.1/ijitee.H92040610821
DOI: 10.35940/ijitee.H9204.0610821
Journal Website: www.ijitee.org

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication

The proposed approach for emergency alerts would help in reducing the mishaps that include these people and help them save their own lives during emergencies.

Further work and scope of improving would be initialized with creating a dataset with varieties of images including varieties in the background, skin tones, light exposure, etc. Also, more gestures can be added to this predefined set which will increase its fluency. Text concatenation and sentence building, which will ease the person to understand the sign gestures fluently.

REFERENCES

1. International Conference on Communication and Signal Processing, July 28 - 30, 2020, India. A Brief Review of the Recent Trends in Sign Language Recognition Nimisha K P and Agnes Jacob
2. AAWAAZ: A Communication System for Deaf and Dumb Anchal Sood1 , Anju Mishra2 1 Student, Amity School of Engineering and Technology, Amity University, Noida, India 1 aanchalplp@gmail.com 2 Asst.Prof., Amity School of Engineering and Technology, Amity University, Noida, India 2 amishra1@amity.edu
3. Studies on Deaf Mobile Application Shelena Soosay Nathan1, a) Azham Hussain 1, b) and Nor Laily Hashim 1, c) 1 Universiti Utara Malaysia, 06010 Sintok, Kedah Darul Aman, a)Corresponding author: sn.shelena@gmail.com b)azham.h@uum.edu.my c)laily@uum.edu.my
4. Journal of Communications Technology, Electronics and Computer Science, Issue 8, 2016 ISSN 2457-905X MonVoix - An Android Application for hearing impaired people AishwaryaDanoji1 , Aishwarya Dhage1 , Rachana Kamat1 , PriyaPuranik1 , Prof Sharmila Sengupta2 BE Students, Department of Computer Engineering, Vivekanand Education Society's Institute of Technology, Mumbai-400074, India1 Professor, Department of Computer Engineering, Vivekanand Education Society's Institute of Technology, Mumbai-400074, India2.
5. Augmented Sign Language Modeling(ASLM) with interaction design on smartphone - an assistive learning and communication tool for inclusive classroomProcedia Computer Science 125 (ELSEVIER-2018)
6. Concha Batanero , Luis de-Marcos , Jaana Holvikivi, José Ramón Hilera, and Salvador OtónEffects of New Supportive Technologies for Blind and Deaf Engineering Students in Online Learning (IEEE-2019)
7. 2019 International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData).Assistive SIGN LANGUAGE Converter for DEAF AND DUMB B.Lakshmi, Rasheed Ahamed, Harshali Rane, and Ravi Kishore Kodali Dept. of ECE, NIT, Warangal
8. Sign Language to Speech Conversion Using Smart Band and Wearable Computer Technology Dasari Vishal, H M Aishwarya, K Nishkala, B Toshitha Royan, T K Ramesh Department of Electronics and Communication Engineering Amrita University Bengaluru, India.
9. Real-Time Translator for Sign Languages Zulfiqar A. Memon, M. Uzair Ahmed, S. Talha Hussain, Zahid Abbas Baig, Umer Aziz Department of Computer Science National University of Computer and Emerging Sciences (FAST-NUCES), Karachi {zulfiqar.memon, k132317, k132332, k132316, k132334}@nu.edu.pk
10. ICET2017, Antalya, Turkey 978-1-5386-1949-0/17/$31.00 ©2017 IEEE Design and implementation of deaf and mute people interaction system Assist. Lect. Fadia Noori Hummadi Al-Nuaimy Biomedical Engineering Department University of Baghdad Baghdad, Iraq fadia_noori@yahoo.com
11. Continuous Finger Gesture Recognition Based on Flex Sensors Wei-Chieh Chuang 1,†, Wen-Jyi Hwang 1,†* , Tsung-Ming Tai 2,†, De-Rong Huang 1,† and Yun-Jie Jhang 1,† 1 Department of Computer Science and Information Engineering, National Taiwan Normal University, Taipei 106, Taiwan; 60647012s@gapps.ntnu.edu.tw (W.-C.C.); 60447059s@ntnu.edu.tw (D.-R.H.); 60547051s@ntnu.edu.tw (Y.-J.J.) 2 NVIDIA Corp., 11001 Lakeline Blvd #100, Austin, TX 78717, USA; ntai@nvidia.com
12. 8th International Conference on Advances in Computing and Communication (ICACC-2018)Real time conversion of sign language to speech and prediction of gestures using Artificial Neural NetworkAbey Abraham , Rohini V CHRIST (DEEMED TO BE UNIVERSITY) ,Hosur Road, Bengaluru -560029, Karnataka , INDIA.
13. 2016 International Conference on Automatic Control and Dynamic Optimization Techniques (ICACDOT)Communication Interface for Deaf-Mute People using Microsoft Kinect,Ms vandana Baburao Malode Jawaharlal Nehru Engineering College · Electronics And Telecommunication Ph.D., MTECH(EDT), BE(ECT)
14. International Conference on Communication and Signal Processing, April 6-8, 2016, India Software Based Sign Language Converter Keerthi S Warrier, JyateenKumar Sahu, Himadri Halder, Rajkumar Koradiya, and Karthik Raj V
15. 2017 International Conference on Frontiers of Information TechnologyReal Time Translator for Sign Languages Zulfiqar A. Memon, M. Uzair Ahmed, S. Talha Hussain, Zahid Abbas Baig, Umer Aziz Department of Computer Science National University of Computer and Emerging Sciences (FAST-NUCES), Karachi
16. International Research Journal of Engineering and Technology (IRJET) Sign Language Interpreter using Image Processing and Machine Learning Omkar Vedak1, Prasad Zavre2, Abhijeet Todkar3, Manoj Patil4Student, Department of Computer Engineering, Datta Meghe College of Engineering, Mumbai University, Airoli, India 4Assistant Professor, Department of Computer Engineering, Datta Meghe College of Engineering, Mumbai University, Airoli, India.
17. ITHET 6th Annual International Conference, July 7 - 9, 2005, Juan Dolio, Dominican Republic, TC2-17Session T1A 0-7803-9141-1/05/$20.00 © 2005 IEEE July 7 – 9, 2005, Juan Dolio, Dominican Republic ITHET 6th Annual International Conference T1A-1 An e-Learning System for the Deaf people A.S.Drigas1 , D.Kouremenos2 , S. Kouremenos3 and J. Vrettaros4
18. International Journal of Human and Social Sciences 2:2 2007A Virtual Learning Environment for Deaf Children: Design and Evaluation Nicoletta Adamo-Villani
19. Development of Learning Software for Deaf: A Sample of Language Learning Material Shwan Hadi* and Oğuzhan Özdemir** *Software Engineering Department,College of Technology, ** Computer Education and Instructional Technology Department, College of Education, Firat University, Elazig, Turkey

AUTHORS PROFILE

Snehal Patil is currently pursuing her B.E. in Computer Engineering at Pimpri Chinchwad College of Engineering, Pune. Her area of interest are Data Science, Machine Learning and Deep Learning.

Yash Shah is currently pursuing his B.E. in Computer Engineering at Pimpri Chinchwad College of Engineering, Pune. His area of interest are Artificial Intelligence, Machine Learning and App development.

Payal Narkhede is currently pursuing her B.E. in Computer Engineering at Pimpri Chinchwad College of Engineering, Pune. Her area of interest are ML and Java developer.

Abhinav Thakare is currently pursuing his B.E. in Computer Engineering at Pimpri Chinchwad College of Engineering, Pune. His area of interest are ML, Java and Application Development.

Rahul Pitale has done his M.E. in Computer Engineering and is currently an Assistant Professor at Pimpri Chinchwad College of Engineering, Pune. He has an academic experience of 7 years. His area of interest are Data Science, Artificial Intelligence, Machine Learning and Deep Learning