Bridging the Gap: Real-Time Sign Language Translation

Prof. S. R. Karjol[1], Mohammed Mujtaba M Mulla[2], Abhishek Channabasavaraj Kinagi[3], Avaneesh N Devareddi[4], Prajwal Siddappa Hallur[5]

[1]Assitant Professor, Department of Computer Science and Engineering, B.V.V. Sangha's Basaveshwar Engineering College, Bagalkot

[2,3,4,5]Students, Department of Computer Science and Engineering, B.V.V. Sangha's Basaveshwar Engineering College, Bagalkot

Abstract* - This research explores the development of a real-time sign language translation mobile application designed to bridge the communication gap for deaf and mute individuals. Leveraging cutting-edge machine learning algorithms and computer vision techniques, the application captures and translates sign language gestures into text, enhancing accessibility and inclusivity. The study reviews existing systems like Microsoft Kinect and SignAloud gloves, and various academic efforts employing convolutional neural networks (CNNs) and long short-term memory networks (LSTMs) to improve sign language recognition accuracy. Key features of the proposed solution include real-time video processing, accurate text translation, and a user-friendly interface, with a strong focus on privacy and security. Through meticulous development and rigorous testing, the application aims to empower deaf individuals to communicate seamlessly with the broader population, fostering greater social inclusion and participation. This project not only addresses immediate communication barriers but also sets the stage for future advancements in assistive technologies, contributing to a more inclusive society.

Index Terms* - Deep Learning, Sign Language Recognition, Communication Accessibility, Real-Time Translation, Wearable Technology.

I. INTRODUCTION

The act of communication involves transferring information between individuals, typically through oral or written means. However, for hearing impaired and mute individuals, traditional communication methods are not feasible, leading to isolation. Sign language has provided a vital means for these individuals to express themselves, with over 1.5 billion people globally experiencing hearing loss that necessitates rehabilitation, according to the World Health Organization. Sign language, with its distinct grammar and syntax, is a visual system of communication used by deaf and mute individuals. There are 143 different sign languages worldwide, including American Sign Language, British Sign Language, and others.

Despite technological advancements improving daily life and facilitating global interaction, deaf individuals still face significant barriers in education, employment, and public life due to communication challenges. By 2050, it is projected that 2.5 billion people will have some degree of hearing loss, with 700 million requiring rehabilitation. The reliance on sign language interpreters highlights the ongoing need for better communication solutions. As technology continues to evolve, there is hope for a future where deaf individuals can communicate more easily with the broader population, reducing the need for human translators and fostering greater inclusivity. This research aims to explore the potential of technological advancements to bridge the communication gap and improve accessibility for the deaf and mute community.

II. MOTIVATION

The motivation for this project stems from the significant communication barriers faced by hearing impaired and mute individuals, which contribute to their isolation and difficulties in daily life. Despite the existence of sign language as a vital communication tool, the need for interpreters and the limited understanding of sign language among the general population hinder full integration and equal opportunities for deaf individuals. With over 1.5 billion people globally experiencing hearing loss and projections indicating that by 2050 nearly 2.5 billion people will have some degree of hearing loss, there is an urgent need for innovative solutions to address these challenges.

Technological advancements have transformed many aspects of modern life, yet deaf individuals continue to encounter obstacles in education, employment, and public interactions. The reliance on sign language interpreters underscores the communication gap that persists despite these advancements. This project is driven by the potential of emerging technologies to bridge this gap, enhancing communication accessibility and fostering a more inclusive society.

By leveraging advancements in technology, we aim to develop solutions that enable seamless communication between deaf individuals and the broader population without the need for human translators. This project aspires to dismantle the barriers that isolate the deaf and mute community, ensuring they have equal opportunities to participate fully in all aspects of life. Through this research, we seek to contribute to a future where improved communication accessibility is a reality for all.

III. LITERATURE OVERVIEW

- Microsoft Kinect and wearable gloves represent innovative solutions in the field of sign language recognition, utilizing depth-sensing cameras and sensors to capture hand movements and gestures for translation into text or speech. Deep learning models, powered by computer vision and machine learning techniques, are trained on extensive datasets of sign language gestures to accurately interpret hand movements, facial expressions, and body postures, paving the way for real-time communication solutions.

- Microsoft Kinect offers robust sign language recognition capabilities by capturing the 3D structure of hand and body movements, enabling accurate interpretation of sign language gestures. SignAloud gloves provide a wearable technology solution for interpreting sign language gestures and converting them into text or spoken language, enhancing communication between sign language users and non-signers.

- Research papers delve into various aspects of sign language recognition, from real-time Indian Sign Language recognition systems to deep learning methods for converting sign language into text format. Proposed systems utilize computer vision, deep learning, and image processing techniques to recognize hand gestures, track facial expressions, and bridge the communication gap between deaf and mute individuals and the rest of society.

- Efforts focus on addressing the challenges faced by the deaf community in communication, highlighting the importance of recognizing different sign languages and gestures for effective interpretation. Advanced methodologies, such as convolutional neural networks (CNNs), long short-term memory networks (LSTMs), and transfer learning, demonstrate significant improvements in accuracy and real-time sign language recognition.

The literature underscores the ongoing advancements in sign language recognition technology, with a focus on minimizing communication barriers and enhancing accessibility for the hearing impaired and mute population. Future research aims to further refine existing systems, improve accuracy, and develop more inclusive communication solutions to empower deaf and mute individuals in all aspects of life

IV. OBJECTIVES

To develop a real-time sign language translation mobile application, empowering individuals who are deaf or voiceless to communicate effectively with hearing individuals by:

- Real-Time Video Processing: Utilizing machine learning algorithms to capture and analyze sign language gestures from video input on Android devices.
- Accurate Text Translation: Converting recognized sign language gestures into real-time textual output, enabling clear and unambiguous communication.
- Accessibility Focus: Designing an intuitive and user-friendly interface specifically catering to the needs of deaf and voiceless users.
- Privacy and Security: Ensuring user privacy and data security by employing robust and ethical data handling practices.
- Promote Inclusion: Building bridges between individuals with and without hearing disabilities, fostering greater social inclusion and participation.

This project aims to break down communication barriers and empower individuals with disabilities to engage in everyday interactions with confidence and autonomy.

V. METHODOLOGY

The working methodology of our project is,

a. Once the app is started and user starting interacting through hand gestures
b. The system begins by capturing video input from a webcam.
c. Each frame from the video is extracted and converted into a single image.
d. The image is subjected to hand detection. This step aims to identify and isolate the hand region within the frame.
e. Features are extracted from the detected hand region. These features could be characteristics like the hand’s shape, orientation, and finger positions.
f. The extracted features are then matched against a reference database of features corresponding to known signs.
g. The matching process generates a result that indicates how well the extracted features correspond to a particular sign in the database.
h. Based on the matching outcome, the system recognizes the sign language gesture and potentially translates it to text and speech.

The working modules of our project are,

- Camera: The Camera class serves as the entry point for capturing video of sign language gestures. It accesses the device's camera hardware and continuously streams video input to the application.
- Video Preprocessing: Once the video is captured, it undergoes preprocessing using segmentation and transformation techniques to enhance the quality and clarity of the sign language gestures. This step helps improve the accuracy of the subsequent prediction process.
- Sign Language Prediction Model: The preprocessed video data is fed into the Sign Language Prediction ML Model, which is a neural network trained to recognize and interpret sign language gestures. This model analyzes the video frames in real-time and predicts the corresponding signs being performed.
- Display of Prediction: The Display class is responsible for visualizing the predictions made by the ML model. It renders the recognized sign language gestures and optionally displays the text translation output alongside the gestures.
- Text Translation: If the user opts for text translation, the Text Input class is activated. It converts the recognized sign language gestures into text format, which is then passed to the Language Translation ML Model.
- Text Preprocessing: Before translation, the raw text undergoes preprocessing in the Preprocess Text class. This includes tokenization, vocabulary identification, and grammar normalization to prepare the text for accurate translation.
- Language Translation Model: The preprocessed text is then fed into the Language Translation ML Model, which translates the text into the desired language selected by the user. The translated text is then sent back to the Display class for visualization.
- Data Management: The Store Data class manages all data-related operations, including storing captured videos, accessing stored data for training or testing purposes, and maintaining user preferences and settings. It ensures efficient storage and retrieval of data while adhering to privacy and security protocols.

Throughout the development process, collaboration between these classes ensures seamless integration of functionalities, robust performance, and a user-friendly experience for individuals using the sign language translation mobile application. Detailed diagrams and flowcharts depicting the interaction between these components facilitate clear communication and effective implementation of the project.

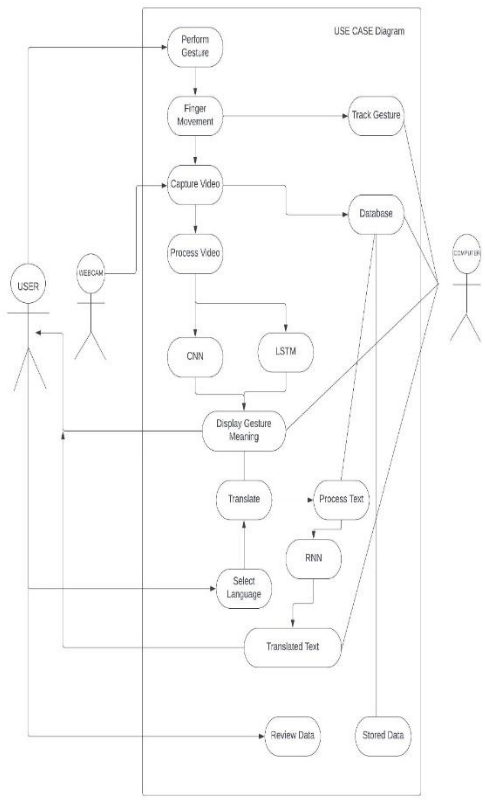

Fig. 5.1. Use Case Diagram

VI. ARCHITECTURE

1. Camera: this class will load the camera and capture the video of the gesture/sign.

2. Preprocess of video: this class will preprocess the video/image using segmentation and transformation.

3. Sign language prediction ML model: Here, the preprocessed video/image will be loaded, and the prediction results will be displayed using a neural network.

4. Display: this class has the function to load and display the prediction of gesture/sign and text translation output.

5. Text Input: If the user selects to translate the text, this class will trigger, and it will be used to convert the text into the desired language.

6. Preprocess Text: in this class, we will be preprocessing the text input, i.e., raw text. preprocessing will include tokenization, vocab finding, and grammar and pass the preprocessed text to the ML model.

7. Language Translation ML Model: Here, the preprocessed text will be converted into the selected language, and the results will be sent to the display class.

8. Store Data: all the data-related things will be covered in this class, from storing the data to accessing the data.

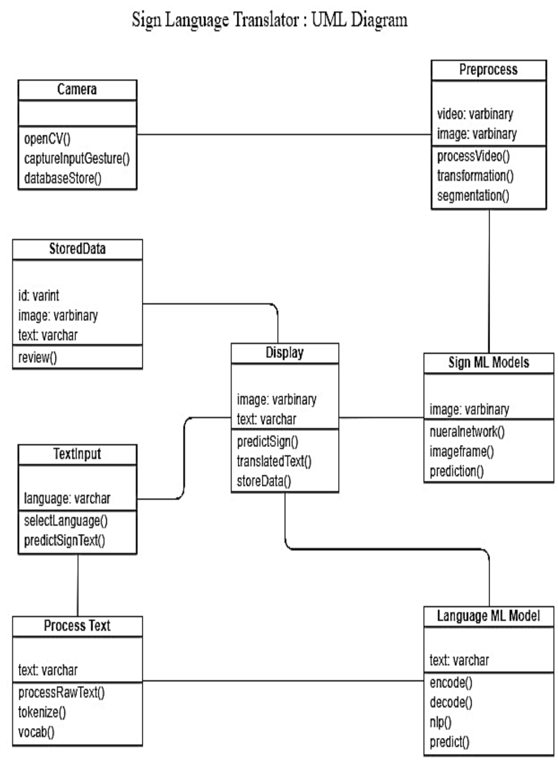

Fig. 6.1. Data Architecture.

VII. WORKING SYSTEM

Real-time sign language to text and speech translation is a complex process with several phases necessary for precise and effective communication. The important steps in this method are listed below:

- Understanding individual sign gestures.
- Train the Machine Learning Model for Image to Text Translation.
- Forming words
- Forming sentences
- Creating content
- Obtaining audio output A.

A.Flow Diagram

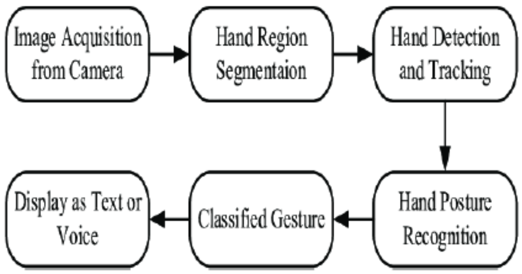

Fig. 7.1. Flow Chart

Figure 7.1 shows a flow chart that visualizes the project's sequential processes and technique. Each stage is described in full below:

1. Image Acquisition

The movements are collected using a web camera and an OpenCV video feed. This stream catches the whole signing period and extracts frames. Frames are processed into grayscale pictures with a 50x50 pixel resolution for uniformity across the collection.

2. Hand Region Segmentation & Hand Detectionand Tracking

The collected photos are examined for hand motions. The preprocessing phase improves the visibility of gesture parts, leading to higher prediction accuracy. The procedure includes separating hand areas from the rest of the picture.

3. Hand Posture Recognition

Preprocessed photos are put into a trained Keras Convolutional Neural Network (CNN) model. The model assigns probabilities to each gesture, then selects the label with the highest likelihood.

4. Display as Text & Speech

Recognized motions are combined to make words. Python's pyttsx3 module is used to translate the words into speech. This text-to-speech technology enhances user experience by mimicking spoken communication.

B. Convolutional Neural Network for Detection

CNNs are particularly successful in computer vision, inspired by the visual cortex in the human brain. Filters or kernels analyze pixel data and add weights to identify certain traits. CNN architecture consists of many layers, including convolution, max pooling, flatten, dense, dropout, and fully connected neural networks. These layers identify characteristics from low to high levels in a picture.

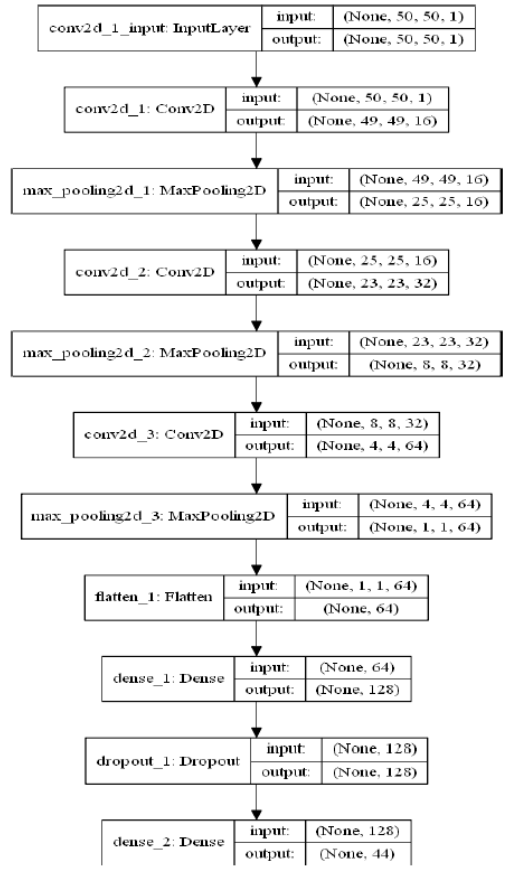

Fig. 7.2. CNN Architecture

C. The CNN Architecture Functioning

The CNN model for this project has 11 layers, including three convolutional layers: The first convolutional layer uses 16 2x2 filters to identify low-level elements in 50x50 grayscale pictures, such as lines. The result is a 49x49x16 activation map with a ReLU layer to remove negative values and a max pooling layer to reduce it to 25x25. The second convolutional layer identifies mid-level characteristics like angles and curves using 32 3x3 filters, resulting in a 23x23x32 activation map. A max pooling layer further reduces the image to 8x8x32. The third convolutional layer uses 64 5x5 filters to identify high-level elements such as gestures and forms, resulting in a 4x4x64 activation map. This map is then reduced to 1x1x64 by a max pooling layer.

D. Recognition of Alphabets and Numbers

To determine the bounding boxes of distinct objects, Gaussian background subtraction involves modeling each background pixel with a combination of K Gaussian distributions. Static colors are considered background, but bounding boxes are meant to accommodate shifting pixels. These photos are used to train a Convolutional Neural Network model that distinguishes gesture indications from background. The CNN detects common patterns in the training data, successfully distinguishing movements from the backdrop. This technology allows for real-time translation of sign language motions into text and voice, improving communication for the deaf and mute people.

E. Algorithm

Algorithm for real-time sign language to text conversion and start.

S1: Adjust the hand histogram to match the skin tone and lighting circumstances.
S2: Use data augmentation to increase the dataset and reduce overfitting.
S3: Separate the dataset into train, test, and validation sets.
S4: Train the CNN model to fit the dataset.
S5: Create a model report that contains accuracy, error, and confusion matrix.
S6: Run the prediction file, which predicts specific gestures, combines them into words, displays them as text, and sends the vocal output.
Stop

VIII. RESULTS

Our model achieved 95.8 percent accuracy with only layer 1 and 98.0 percent accuracy when layer 1 and layer 2 were combined, outperforming most previous research publications on American sign language. Most research articles focus on the use of technologies like Kinect for hand detection. While most of the above 21 applications involve Kinect devices, we aimed to create a project that can be used using readily accessible materials. Our method uses a normal camera instead of the expensive Kinect sensor, making it accessible to a wider audience.

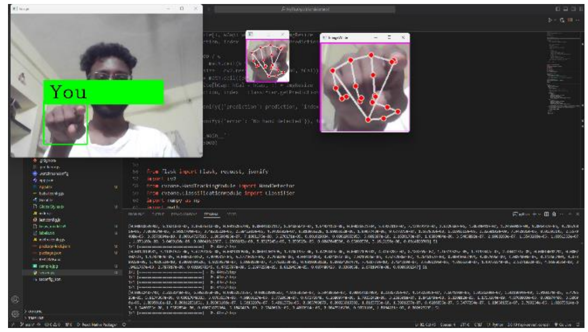

Fig. 7.3. Results

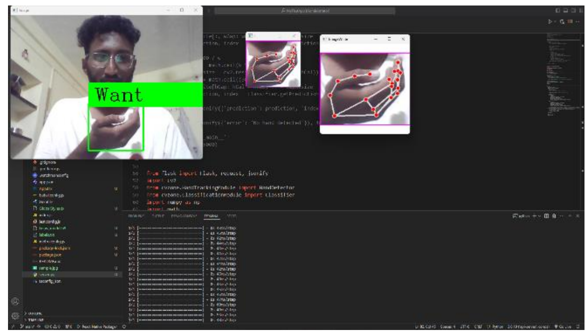

Fig. 7.4. Results

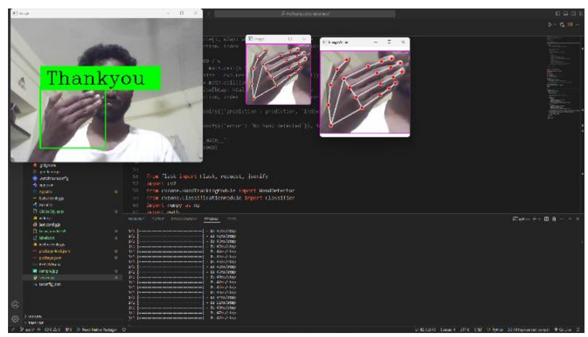

Fig. 7.5. Results

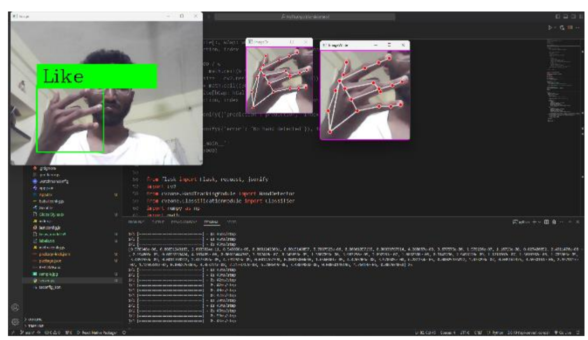

Fig. 7.6. Results

CONCLUSION

Through meticulous user-centered development, our team has successfully crafted a real-time sign language translation mobile application that stands as a beacon of accessibility and inclusivity. By deeply understanding the needs of our users, we have designed an intuitive interface that empowers individuals who are deaf or voiceless to communicate effortlessly with hearing individuals. Leveraging cutting-edge machine learning algorithms, we have achieved remarkable accuracy in capturing and translating sign language gestures into real-time textual output, thereby breaking down communication barriers like never before.

Our integration of screen reader compatibility, voice control, and customizable text size and color ensures a truly inclusive experience, catering to users across all abilities. Rigorous testing across platforms and adherence to accessibility standards have culminated in a seamless deployment on app stores, ready to make a transformative impact on the lives of individuals with disabilities and their broader communities. The success of this project not only lies in its immediate outcomes but also in its potential to stimulate further research and development in the realm of sign language technologies. By promoting social inclusion and participation, this endeavor paves the way for a more inclusive society where everyone, regardless of their abilities, can communicate with confidence and autonomy. As we look to the future, we envision a world where innovation continues to flourish, driven by the imperative to create solutions that empower and uplift all members of society.

REFERENCES

[1] Mahesh Kumar N B , “Conversion of Sign Language into Text”, International Journal of Applied Engineering Research ISSN 0973-4562 Volume 13, Number 9.

[2] Akash Kamble , Jitendra Musale , Rahul Chalavade , Rahul Dalvi , Shrikar Shriyal, “Conversion of Sign Language to Text”, International Journal for Research in Applied Science & Engineering Technology.

[3] Bikash K. Yadav, Dheeraj Jadhav, Hasan Bohra, Rahul Jain, “Sign Language to Text and Speech Conversion”, International Journal of Advance Research, Ideas and Innovations in Technology.

[4] Shubham Thakar, Samveg Shah, Bhavya Shah and Anant V. Nimkar, “Sign Language to Text Conversion in Real Time using Transfer Learning”, IEEE Explore.

[5] Shaheen Tabassum , Raghavendra R, “Sign Language Recognition and Converting into Text”, International Journal for Research in Applied Science & Engineering Technology (IJRASET) .

[6] Divya S, Kiruthika, S Nivin Anton A L and Padmavahi S, “Segmentation, Tracking And Feature Extraction For Indian SignLanguage Recognition” IJCSA, vol.4, no.2.

[7] Shi, 1., and C. Tomasi, "Good Features to Track," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.

[8] S. M Mahesh Kumar, “Conversion od Sign Language into Text” International Journal of Applied Engineering Research ISSN 0973-4562 Volume 13.

[9] Kohsheen Tiku, Jayshree Maloo, Aishwarya Ramesh, Indra R, “Real-time Conversion of Sign Language to Text and Speech,” Second International Conferenceon Inventive Research in Computing Applications, Coimbatore, India.

[10] Kusurnika Krori Dutta , Satheesh Kumar Raju K, Anil Kumar G S, Sunny Arokia Swarny, “Double Handed Indian Sign Language to Speech and Text”, Third International Conference on Image Information Processing.

[11] I. Krak, I. Kryvonos and W. Wojcik, "Interactive systems for sign language learning," 6th International Conference on Application of Information and Communication Technologies (AICT), Tbilisi, 3, doi: 10.1109/ICAICT.2012.6398523.

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

[13] Pigou L., Dieleman S., Kindermans PJ., Schrauwen B. Sign Language Recognition Using Convolutional Neural Networks. In: Agapito L., Bronstein M., Rother C. (eds) Computer Vision -

ECCV Workshops. ECCV Lecture Notes in Computer Science, vol 8925. Springer, Cham

[14] Bheda, Vivek and Dianna Radpour. “Using Deep Convolutional Networks for Gesture Recognition in American Sign Language.” ArXiv abs/1710.06836 : n. pag.

[15] M.V, Beena. “Automatic Sign Language Finger Spelling Using Convolution Neural Network : Analysis.”

[16] Tolentino, Lean Karlo S. et al. “Static Sign Language Recognition Using Deep Learning.” International Journal of Machine Learning and Computing 9: 821-827.

[17] Umang Patel and Aarti G. Ambedkar, “Moment Based Sign Language Recognition for Indian Language” International Conference on Computing, Communication, Control and Automation (ICCUBEA).

[18] Bhargav Hegde, Dayananda P, Mahesh Hegde, Chetan C, “Deep Learning Technique for Detecting NSCLC”, International Journal of Recent Technology and Engineering (IJRTE), Volume-8 Issue-3, pp. 7841-7843