e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:04/April-2024 Impact Factor- 7.868 www.irjmets.com

SIGN LANGUAGE TRANSLATOR INTO TEXT & AUDIO BY HAND GESTURES USING OPEN CV AND CNN

Mrs. Rajeshwari. P[*1], Saradhambal. R[*2], Vivek Kumaar. V.S[*3]

[*1]Assistant Professor Of Software Systems, Sri Krishna Arts And Science College, Coimbatore, Tamil Nadu, India.

[*2,3]PG Student Of Computer Science, Sri Krishna Arts And Science College, Coimbatore, Tamil Nadu, India.

ABSTRACT

The primary challenge in disability often stems from the inability to speak and hear. Those affected usually adopt alternative communication methods like sign language. Utilizing computer vision, a machine learning-driven application aims to bridge the communication gap between signers and non-signers. Specifically, the focus is on a desktop application ,uses a computer's webcam to capture real-time sign language gestures. It employs a Convolutional Neural Network (CNN) to interpret signs from video frames as hand gestures, which is recognized by background subtraction, motion detection, contour extraction. The webcam captures sign language gestures, translates them into text, and subsequently converts the text into audible speech for mute individuals.

Keywords: Open CV, Convolutional Neural Network, Text, Speech, Hand Gesture Recognition.

I. INTRODUCTION

The means of communication between those with mute people primarily relies on sign language and assuming mutual comprehension. This study aims to address this communication challenge by proposing a potential solution. The development of a live sign language interpreter stands as a pioneering leap in accessibility, particularly for individuals who are deaf or hard of hearing. The process involves capturing video input from a camera, followed by hand detection and tracking using OpenCV's image processing capabilities. The hand gestures are then analyzed and classified into corresponding sign language symbols using algorithms, allowing the system to interpret the user's intended message.

To enhance the accuracy and robustness of the system, a Convolutional Neural Network (CNN) is trained on a comprehensive dataset of sign language gestures. The trained model is integrated into the real-time processing pipeline, enabling the system to recognize a wide range of signs with high accuracy.

Sign language translator was used for both dumb and non-dumb individuals, using Convolutional Neural Networks (CNNs) to interpret and convert signs into text and speech, is a significant technological endeavor. CNNs are adept at processing visual data, making them valuable for analyzing gestures and movements in sign language.

The translation process involves converting the identified gestures into both textual and audio representations. The hand gestures are classified by some features like skin color, motion of hand, edge of background[1]. By these features background subtraction is applied to the hand gesture and sign language was recognized. This integrated approach not only facilitates effective communication between sign language users and non-signers but also addresses the diverse communication preferences within the dumb and hard of hearing community. By offering translations in both audio and text, the system ensures accessibility for individuals and satisfy the vary needs and requirements.

The development of a sign language translator featuring real-time hand recognition, audio translation, and text generation represents the significant advancement in breaking down the communication barriers and promoting universal accessibility[2]. This technology has the potential to empower individuals in their daily interactions, fostering a more inclusive and understanding society.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:04/April-2024 Impact Factor- 7.868 www.irjmets.com

II. METHODOLOGY

Algorithm:

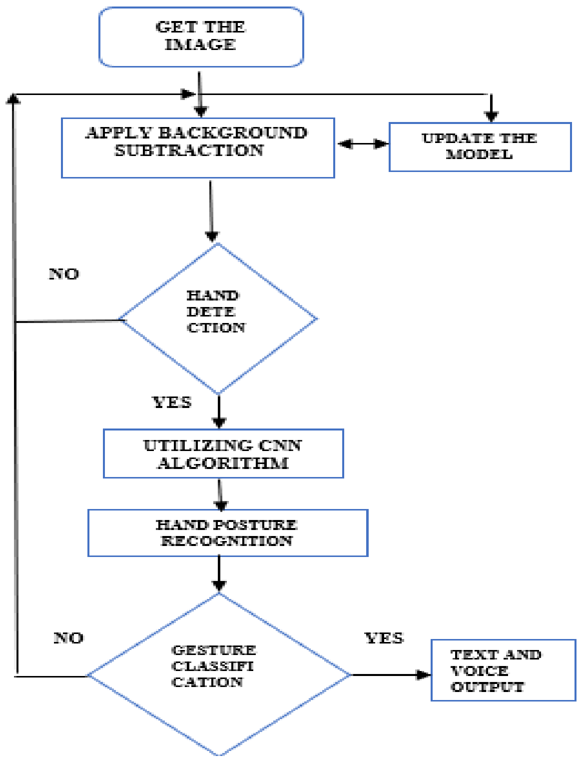
Step 1: Data collection and preprocessing collects a diverse dataset of sign language gestures, ensuring uniformity in size and background. Normalize images to mitigate variations and augment the dataset.

Step 2: Background subtraction and motion detection utilize the OpenCV's background subtraction algorithms to extract moving hands or gestures from the background. Implement motion detection algorithms to identify significant movement in the frame. Threshold motion detection to trigger gesture recognition appropriately.

Step 3: Gesture recognition with CNN designs a CNN architecture suitable for recognizing sign language gestures. Split the dataset into training, validation, and testing sets. Train the CNN model using the training dataset and fine-tune it based on validation performance. Evaluate the trained model using the testing dataset.

Step 4: Integration and output integrate used for background subtraction and motion detection modules with the gesture recognition model. Map recognized gestures to corresponding text representations. Convert recognized text into speech output using text-to-speech (TTS) libraries or services.

Step 5: Testing, evaluation, and deployment are final steps. Test the integrated system with sign language gestures under different conditions. Evaluate system performance with accuracy, speed. Gather feedback from users and testers for further improvements. Optimize the system for real-time performance and efficiency if necessary.



2.1 Flowchart

In this flow chart:

Each block represents a stage or process in the system. Arrows indicate the flow of data from one stage to another. The input is processed sequentially through various stages, leading to the final output. The system takes input from a video feed, processes it through hand detection, pose estimation, and gesture classification stages, translates the recognized gestures into text or speech, and finally presents the output to the user.

III. PROBLEM DESCRIPTION

Designing an effective sign language translator that adeptly converts gestures into written text and spoken language presents a formidable technical challenge. This endeavor hinges on deploying sophisticated technology that comprehends the subtleties inherent in diverse sign languages. It necessitates an intricate

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:04/April-2024 Impact Factor- 7.868 www.irjmets.com

understanding of hand movements and accurately transcribing these gestures into coherent text and subsequently vocalizing them in real time.

The complexities are manifold, encompassing variations among sign languages, regional dialects, and the need for a comprehensive database facilitating precise interpretation. These intricacies highlight the necessity for advanced machine learning models adept at recognizing and translating the diverse linguistic elements of sign languages. For this task, an extensive dataset representing various regional sign languages, alongside their corresponding spoken and written languages, would be essential. Training a CNN model would involve teaching it to recognize and understand the nuances of sign language gestures, encompassing distinct hand movements. Achieving seamless integration of text-to speech technology is crucial to ensure a fluid transition from text to spoken language.

The complexity lies in training the CNN to recognize a wide array of gestures and variations while ensuring real-time translation for effective communication[3]. Additionally, integrating speech synthesis technology would enable the translation of sign language into spoken language for accessibility to non-signing individuals.

Overcoming these challenges mandates not only technological advancements but also a deep understanding of the diverse needs within the deaf and hearing-impaired communities. The ultimate goal is to craft an inclusive and accessible communication tool that bridges the gap between sign language and spoken language, empowering individuals with diverse communication abilities.

IV. EXISTING SYSTEM

Existing sign language translation systems cooperated with Support Vector Machines (SVM) algorithm that resulted weak for gesture recognition and translation. SVMs excel in classifying and distinguishing between various sign language gestures. They are adept at establishing optimal decision boundaries in the feature space, allowing for accurate gesture differentiation, especially in datasets with clear feature distinctions between signs[4]. Existing systems often employ depth cameras to capture hand movements and gestures. These inputs are then processed using algorithms that analyze the spatial and temporal features of the gestures. Feature extraction methods like Deep Learning-based approaches are applied to recognize and classify these gestures. Additionally, some systems utilize pose estimation algorithms that track key points or landmarks on the user's hands or body to know the sign language gestures more accurately. For translation, once the gestures are recognized, Support Vector Machine algorithm or rule-based systems are employed to convert these gestures into textual representations. These systems often undergo extensive training using datasets that contain annotated sign language gestures. They aim to improve accuracy and inclusivity across different sign languages and variations within them[5]. While these existing systems showcase significant advancements, ongoing research continues to refine algorithms, improve recognition accuracy, expand vocabulary coverage, and enhance real-time performance for more seamless and effective communication between sign language users and non-signing individuals.

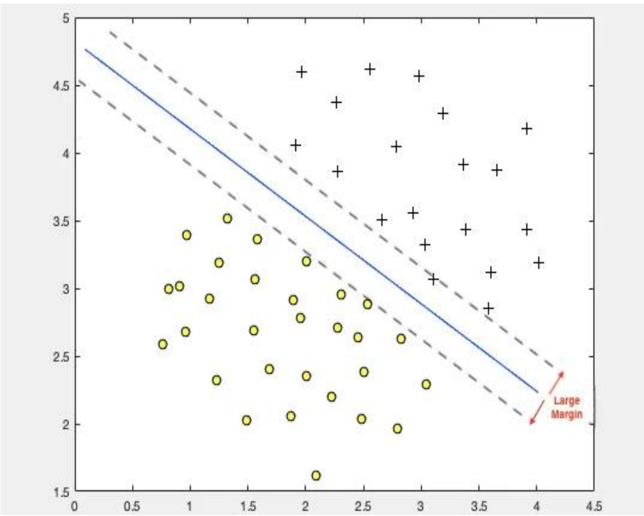


Fig 1: Linear classification of data using SVM model

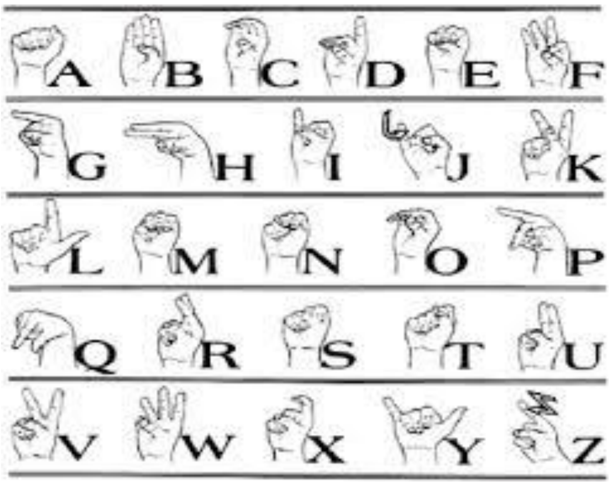


Fig 2: Representation of alphabets in signs

V. PROPOSED SYSTEM

The intended system strives to narrow the communication barrier faced by the deaf and mute community. Its primary goal is to furnish a tool enabling expression through sign language across diverse social and professional environments. The design of this implementation emphasizes adaptability, empowering users to personalize and broaden the sign language lexicon to suit their individual requirements.

A sign language translator system using Convolutional Neural Networks (CNNs) used to convert gestures into text and speech requires a comprehensive approach. Firstly, assembling a diverse dataset encompassing various sign language gestures[6]. This dataset should encapsulate the intricate hand movements inherent in sign languages. Real-time gesture recognition forms the system's backbone, necessitating algorithms that process live video inputs, extract pertinent features from background subtraction, motion detection, contour extraction and feed them into the trained CNN model for interpretation[7]. Once gestures are recognized, the system translates them into textual representations, maintaining grammatical and syntactical accuracy.

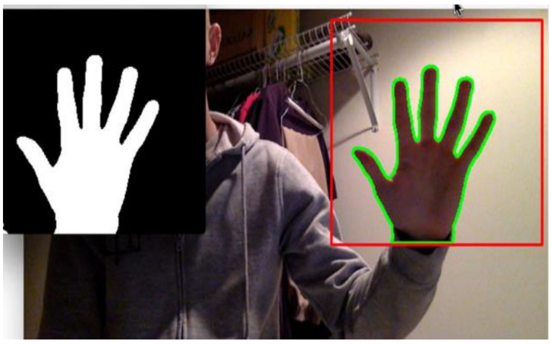


Fig 3: Representation of number five in sign

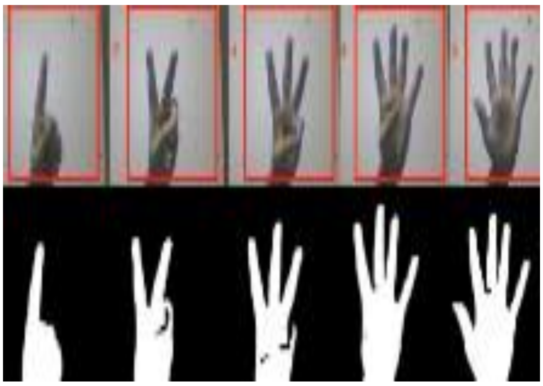


Fig 4: Representation of numbers in signs

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:04/April-2024 Impact Factor- 7.868 www.irjmets.com

The system places a strong emphasis on its ability to adapt to regional sign languages and corresponding spoken languages, ensuring inclusivity across diverse linguistic backgrounds[8]. Thorough testing across various sign languages and linguistic nuances is crucial to validate accuracy, speed, and usability. Its primary objective is continuous evolution through learning from user interactions to improve accuracy over time. Key considerations providing accessibility features, integrating emerging technologies, and implementing privacy measures. Community engagement and collaboration with sign language educators and experts are essential to ensure the system's cultural relevance and effectiveness. The next phase involves incorporating text-to-speech (TTS) technology to convert translated text into spoken language, enabling communication for non-signing individuals.

Ultimately, the system aims to break down communication barriers between deaf and hearing individuals by leveraging CNNs' capabilities to interpret sign language gestures accurately and facilitate seamless translation into accessible text and speech[9].

The sign language translation system operates through a series of distinct stages. Initially, it receives input from a video feed capturing sign language gestures, which then undergoes preprocessing to enhance its quality. Subsequently, the system detects and localizes hand regions within the video frames using techniques like Haar cascades or deep learning-based object detection methods. These estimated hand poses are then fed into a Convolutional Neural Network (CNN) model trained for gesture classification, which extracts features and categorizes them into corresponding sign language gestures. Once the gestures are recognized, they undergo translation into text representations using predefined mappings between gestures and textual equivalents. Optionally, the translated text can be further converted into synthesized speech using text-to-speech (TTS) synthesis for auditory output. Finally, the translated text and/or synthesized speech are presented to the user via an output interface, such as a screen or speaker, thereby facilitating communication accessibility for individuals with hearing or speech impairments.

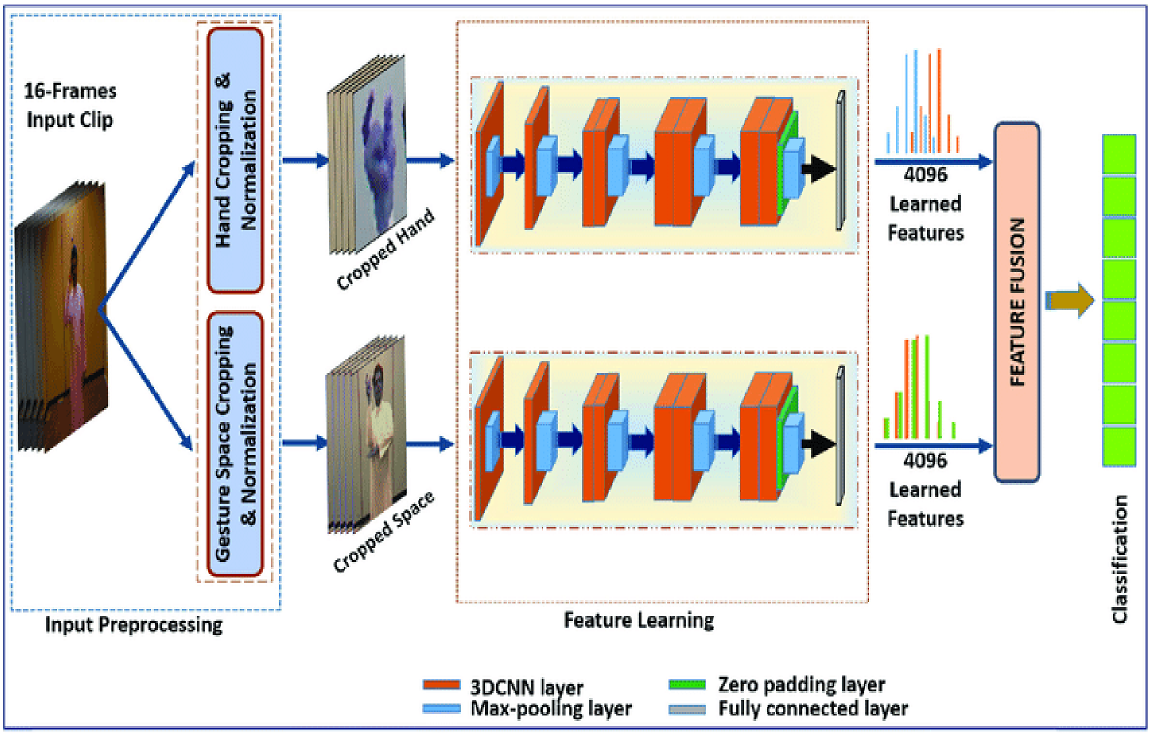


Fig 5: Convolutional Neural Networks Layers

VI. RESULT AND DISCUSSION

The utilization of Convolutional Neural Networks (CNNs) stands as a cornerstone in the advancement of sign language translation systems, primarily due to their exceptional ability to process visual data such as hand gestures. Through the adept use of CNNs, these systems can accurately interpret and categorize intricate gestures, essential for effective translation. CNN architectures are uniquely suited for learning complex patterns within images, enabling precise recognition of the diverse range of gestures present in sign language. However, despite the high accuracy achieved during testing phases, the real-world deployment of these systems has been met with challenges. Issues such as difficulty in recognizing gestures under low-light conditions and sensitivity to varying camera angles and backgrounds have been identified as significant hurdles. To address these challenges, extensive optimizations have been undertaken, focusing on refining hand pose tracking and segmentation algorithms, alongside fine-tuning the CNN model's hyperparameters to bolster robustness. Additionally, experiments involving different image preprocessing techniques and CNN architectures have been conducted, aimed at enhancing system performance. Encouragingly, the results of real-time testing indicate promising strides towards the practical application of sign language translation systems in real-world scenarios.

The successful development of a hand gesture recognition system holds immense promise for improving communication accessibility for individuals with hearing or speech impairments[10]. By overcoming the challenges posed by real-world scenarios through the integration of CNNs and rigorous optimization efforts, these systems can bridge the gap between sign language users and the broader community. Through continuous refinement and innovation, sign language translation systems can achieve higher levels of accuracy and efficiency, thereby empowering individuals with hearing or speech impairments to communicate effectively in diverse settings. Hand gestures, universal language, enhance communication by conveying meaning and emotions while advancements in technology, particularly in sign language recognition and human-computer interaction, harness the power of gestures to improve accessibility and facilitate intuitive interaction with digital devices and virtual environments. OpenCV, a versatile computer vision library, empowers developers to create powerful applications for image and video processing tasks, including object detection, facial recognition, and augmented reality, through its comprehensive suite of functions and algorithms. Ultimately, the advancement of these systems not only enhances accessibility but also promotes inclusivity, fostering a more inclusive society where communication barriers are minimized, and all individuals can participate fully and equally.

Fig 6: Finger spelled Alphabet (Top row [A, B, C, D, G], Bottom row [H, I, N, O, P])

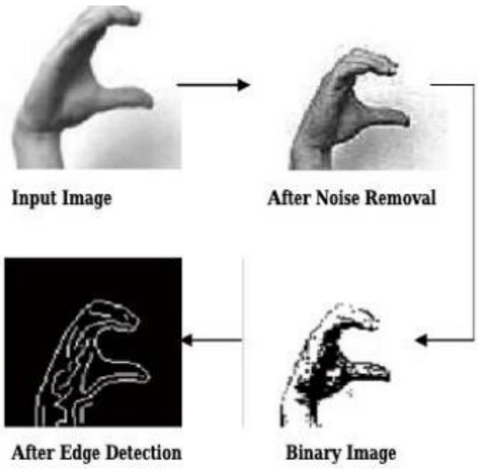


Fig 7: The process of detecting edges using the Edge Detection algorithm from an input image.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:04/April-2024 Impact Factor- 7.868 www.irjmets.com

Fig 8: Gesture Classification

Fig 9: Excepted Output (The conversion of sign into text and audio)

VII. CONCLUSION

Using the CNN algorithm, computer vision problems can be effectively addressed, as demonstrated by the creation of a 95% accurate finger spelling translator for sign language. CNNs are highly favored in sign language recognition for several reasons. Firstly, they possess the ability to automatically learn complex patterns and features from raw data without requiring manual feature engineering. This capability is particularly advantageous in sign language recognition, where hand gestures can vary significantly in appearance and context. Additionally, CNNs exhibit spatial invariance, meaning they can identify patterns irrespective of their position or orientation within an image, aligning well with the diverse spatial configurations of hand gestures in sign language. Furthermore, CNNs are designed with parameter sharing and local connectivity, making them computationally efficient and scalable, which is beneficial when working with large datasets or complex images. Despite these strengths, it's essential to consider the specific requirements and constraints of each application when selecting an algorithm for sign language recognition. This achievement opens the door to expanding the project to encompass additional sign languages through dataset creation and CNN training. Sign languages, often used in contextual settings rather than solely finger typing, can benefit from this initiative, tackling a portion of the translation challenges. The primary objective of eliminating interpreter has been met successfully. By integrating OpenCV with machine learning techniques, this endeavor fosters inclusive communication and fosters a more supportive environment for individuals in the deaf and mute community.

Despite challenges, optimizations and experimentation enhance system performance, paving the way for improved communication accessibility for individuals with hearing or speech impairments.

VIII. FUTURE SCOPE

The future of hand gesture recognition, propelled by technologies like OpenCV, promises groundbreaking advancements across numerous domains. As research and development efforts continue to evolve, we anticipate significant strides in accuracy, robustness, and versatility in hand gesture recognition systems. These advancements will enable more precise interpretation of intricate hand movements and gestures, facilitating seamless communication and interaction between humans and machines.

In fields such as human-computer interaction (HCI), the integration of hand gesture recognition holds tremendous potential for revolutionizing user interfaces and experiences. By enabling intuitive and natural interaction with digital devices, such as smartphones, tablets, and wearable technology, hand gesture recognition can enhance productivity, efficiency, and accessibility. Future applications may include touchless control interfaces, where users can navigate menus, manipulate objects, and execute commands through simple hand gestures, reducing reliance on traditional input methods like keyboards and touchscreens. Hand gesture recognition has promising implications in healthcare, where it can be utilized for various purposes ranging from assisting individuals with physical disabilities to enhancing surgical procedures and medical diagnostics. For instance, gesture-controlled prosthetic devices can offer greater mobility and autonomy to amputees, while hand gesture recognition systems integrated into medical imaging equipment can streamline diagnostic workflows and improve precision in surgical interventions.

In the automotive industry, hand gesture recognition holds the potential to enhance driver safety and convenience by enabling gesture-based controls for in-vehicle infotainment systems, navigation, and driver assistance features. By allowing drivers to perform tasks without taking their hands off the steering wheel or eyes off the road, gesture recognition contributes to safer and more intuitive vehicle interactions. The integration of hand gesture recognition into virtual reality (VR) and augmented reality (AR) applications opens up new dimensions of immersive experiences and interactive storytelling. From gaming and entertainment to training simulations and architectural visualization, hand gesture recognition enriches the user experience by enabling natural interaction and manipulation of virtual objects and environments.

The future of hand gesture recognition powered by OpenCV holds promise for transformative innovations that enhance communication, interaction, and accessibility across a wide range of domains, ultimately enriching lives and driving progress in society.

IX. REFERENCES

[1] Vaidya, O., Gandhe, S., Sharma, A., Bhate, A., Bhosale, V., Mahale, R. "Development of a hand gesture-based communication device for the deaf and mute" presented at the IEEE Bombay Section Signature Conference (IBSSC), Mumbai, India, December 4–6, 2020.

[2] Marin, G., Dominio, F., Zanuttigh, P. "Hand gesture recognition using Leap Motion and Kinect devices" presented at the IEEE International Conference on Image Processing (ICIP), Paris, France, October 27–30, 2014.

[3] Saleem, M.I., Otero, P., Noor, S., Aftab, R. "Full duplex smart system for the Deaf & Dumb and normal people" presented at the Global Conference on Wireless and Optical Technologies (GCWOT), Mlaga, Spain, October 6–8, 2020.

[4] Deb, S., Suraksha, Bhattacharya, P. "Augmented Sign Language Modeling (ASLM) on smartphone—an assistive learning and communication tool for inclusive classrooms" published in Procedia Computer Science, 2018.

[5] Howard, A.G. et al. "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications" presented on arXiv, 2017.

[6] Rishi, K. et al. "Two-way sign language conversion using neural networks" presented at the 8th International Conference on Advanced Computing and Communication Systems (ICACCS2022), Coimbatore, India, March 25–26, 2022.

[7] Anupama, H.S. et al. "Automated sign language interpreter using data gloves" presented at the International Conference on Artificial Intelligence and Smart Systems (ICAIS), Coimbatore, India, March 25–27, 2021.

[8] Kawai, H., Tamura, S. "Deaf-and-mute sign language generation system" published in Pattern Recognition, 1985.

[9] Bhadauria, R.S., Nair, S., Pal, D.K. "A Survey of Deaf Mutes" published in the Medical Journal of Armed Forces India, 2007.

[10] Raheja, J.L., Singhal, A., Chaudhary, A. "Android-Based Portable Hand Sign Recognition System" presented on arXiv, 2015.