

ML Algorithms For Deaf, Dumb , and Blind Assistive Device

P THANUSH

Dept. Electronics and Communication Engineering,
R.V. College Of Engineering,
Bangalore, India.
thanushreddy753@gmail.com

NAGARAJ BHAT

Dept. Electronics and Communication Engineering,
R.V. College Of Engineering,
Bangalore, India.
nbhat437@gmail.com

Abstract—Technology has become an integral part of our lives. This has a positive impact on society, and people benefit from it. This research is focused on helping people with impairments. Machine Learning and Artificial intelligence can support people with impairments to integrate into society effectively. The World Health Organization (WHO) reports that 5 of the world's population is deaf, blind and mute. All around the world, about 9.1 billion people are deaf and mute. In their daily life, they face plenty of problems with their communication. Sign language is a linguistic process that is employed for communication among normal people and handicapped people. Learning sign language is a challenge for those with no impairment. In this article, we have discussed the methods and the aspects of conversion from speech to text, Image to speech, text to speech, and sign language recognition which makes it easy for specially-abled people. The conclusion and future scope of the work are also examined.

Keywords— Machine Learning(ML), Artificial Intelligence(AI) Convolution neural network(CNN), Long Short-Term Memory (LSTM), Recurrent Neural Networks(RNN), Text-to-Speech Synthesis (TTS), Connectionist Temporal Classification (CTC) and, Hidden Markov Models(HMM).

I. INTRODUCTION

This Papers objective is to create a useful ML system for people who are dumb, deaf, or blind and treat them like normal people. The World Health Organization (WHO) estimates that 1 in 6 individuals globally, or more than 1.3 billion people, would have major disability by 2022. To enable people with physical, sensory, or cognitive disabilities to live and work more successfully and independently in all facets of their lives, a variety of assistive technologies (ATs) have been developed. Smart homes now combine artificial intelligence and machine learning techniques, leading to a plethora of systems, chatbots, augmentative communication devices, smart wheelchairs, and brain-computer interfaces. The most well-known area of artificial intelligence (AI) is likely machine learning (ML), which encompasses a wide range of innovative research projects and commercial developments that offer more effective, automated, and efficient algorithms to handle large amounts of data in a variety of fields (such as computer vision, neuroscience, speech recognition, language processing, human-computer interaction, health informatics, medical image analysis, recommender systems, fraud detection, etc.). The most effective AI for this technology is of the highest caliber. The accuracy of speech recognition and word predictability can be increased with the help of Natural Language Processing (NLP) and machine learning algorithms, which can reduce dictation errors and promote productive collaboration between students and teachers. Our plans for caring for our loved ones in the future may significantly rely on AI and ML due to the

increasing aging population, which is predicted to reach 2 billion people over the age of 60 by 2050. Physicians and businesspeople are already blazing a trail; just in the last ten years, investments in medical AI. Machine learning can be highly beneficial for assistive devices in several ways. Machine learning algorithms can analyse and learn from user data to personalize the assistive device's behaviour according to individual needs and preferences. The goal of this research is to aid those who have disabilities. ML/AI can help people with disabilities successfully integrate into society.

II. MOTIVATION

It's challenging to provide solutions for those with visual, hearing, and vocal impairments with just one assistive technology. Global estimates show that in the last 20 years, there have been fewer persons who are visually impaired due to eye-related disorders. The work is focused on developing a novel method that helps the visually impaired by enabling them to hear what is portrayed as text. This is accomplished using a method that takes an image and converts it into speech signals.

III. LITERATURE REVIEW

Literature reviews are crucial elements of academic research because they set the scene, identify the research need, direct methodological choice, and offer a critical evaluation of the body of knowledge. They advance scholarly discourse in a particular field of study and help produce new research, theoretical frameworks, and scholarly conversations. Comprehensive research and the compiling of pertinent sources are required for literature reviews. One useful tool for additional reading and citation in one's own research is the identification of essential references, seminal works, or influential scholars in the subject.

Different sign language recognition systems are presented in the first manuscript [6].A communication framework for DnM people was presented by the authors in [7]. They collected data using motion sensors attached to a glove while employing an automated sign language interpreter (SLI). Language Interpreter (SLI), where they collected information by attaching motion sensors to a glove. An Arduino board is used to collect the sensor data. A machine-learning technique is then used to process the gathered data. A 93% accuracy rate is achieved by the technology. [8] presents a speech recognition system that can recognize the voice of a registered speaker.

The term "AWAAZ" refers to a system that is presented in [10]. This technology allows Deaf and Mute persons to utilize it since it uses image processing to capture their motions. The image acquisition, segmentation, morphological erosion, and feature extraction processes are the major parts of the gesture

recognition system. A system that allows two-way communication is described in [11] as Using an artificial voice recognition method and a mobile application based on visualization, Deaf and Mute and Non-Deaf and Mute people are displayed. In [12], the authors created an Android-based application. The Video Relay Service (VRS) and the Principal Component Analysis (PCA) methodology are used in this. The VRS functions as a manual interpreter, translating hand gestures into voice and vice versa. The PCA algorithm detects the hand gesture.

The authors surveyed [14] and published the issues faced by Deaf and Mute children and developed a web cam-based application where the images are first acquired using a webcam, then PCA is applied to extract features, and the characters are recognized using training sets. Some solutions are presented in this paper. In order to analyze sequential time series data obtained by the Leap Motion device for gesture identification, the authors of [15] combined recurrent neural networks with Long Short-Term Memory (LSTM). Both the standard unidirectional LSTM and the bidirectional LSTM were employed by the authors. Based on the model mentioned here, together with other elements, a prediction network architecture known as the Hybrid Bidirectional Unidirectional LSTM is developed. Due to taking into account the spatial and temporal interactions between the network layers and the Leap Motion data, this model dramatically enhanced performance in both forward and reverse directions.

After going through all these papers there are still few things needed to a degree of grace we have the use of Braille system for blind we need to find a solution for it. Build a enhanced better performing tool for image to speech, speech to text and text to speech conversion with appropriate approach. So that it can be easily and duly implemented in everybody's life so that ML can make life easier and better in the future AI & ML.

IV. CONCEPTS

In this section we will have brief on all the topics and contents that we have used in the work so that it would become easier in the methodology and Results part to understand the topic present in it.

A. Machine Learning

In the realm of artificial intelligence (AI), machine learning (ML) is concerned with the creation of models and algorithms that allow computers to learn from experience and anticipate or decide based on data without being explicitly programmed. Making computer systems capable of autonomously learning from experience and improving their performance is the aim of machine learning. This will enable them to manage challenging jobs and arrive at reliable predictions or choices. Due to its capacity for data analysis and interpretation, pattern recognition, and prediction or decision-making, machine learning (ML) offers enormous potential for use in assistive technology.

B. Convolution Neural Networks(CNN)

Convolutional Neural Networks (CNNs) are used for a variety of tasks, notably in the field of computer vision, and are sometimes referred to as the "CNN algorithm". It has several layers, including fully linked layers, pooling layers, activation functions, and convolutional layers. Together, these layers extract and discover pertinent elements from the incoming data,

enabling the network to forecast or categorize the data. Labeled training data are frequently used to train the CNN algorithm. The input data and their related labels are fed into the network during training, and the network's parameters (weights and biases) are optimized such that the CNN algorithm can automatically recognize and extract useful features from the input data. The pooling layers down sample the feature maps, while the convolutional layers learn to recognize significant patterns and spatial hierarchies.

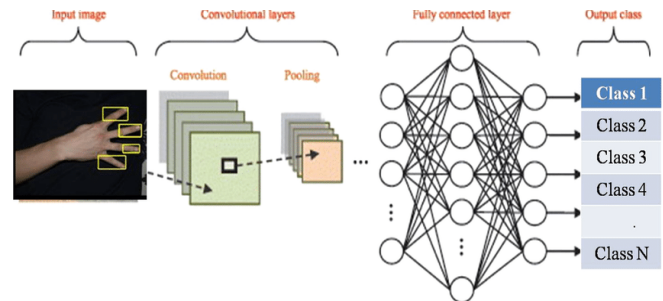


Fig 1 : Typical CNN network working

The CNN algorithm can be employed after training for a variety of tasks, including picture segmentation, object identification, and classification.

C. Recurrent Neural Networks (RNN)

An artificial neural network called a recurrent neural network (RNN) is made to handle sequential data where the sequence of the inputs is important. RNNs feature feedback connections, in contrast to conventional feedforward neural networks, which enable them to preserve an internal state or recollection of prior inputs. RNNs can recognise temporal relationships and generate predictions based on the context of prior inputs thanks to this memory. The capacity of RNNs to accommodate input sequences of different lengths is one of its fundamental characteristics. The prior concealed state serves as context for the present input as they process each input one at a time. RNNs are highly suited for applications like natural language processing, speech recognition, time series analysis, and machine translation due to their ability to model and produce sequences of data.

D. Text-to-speech (TTS)

The technique known as text-to-speech (TTS) transforms written text into spoken language. TTS is frequently applied in machine learning through the use of deep learning methods, notably neural networks. The objective is to train a model to produce excellent, natural-sounding speech from input text. The idea behind Text-to-voice (TTS) technologies is to turn written text into synthesized voice by fusing language analysis, acoustic modelling, and signal processing.

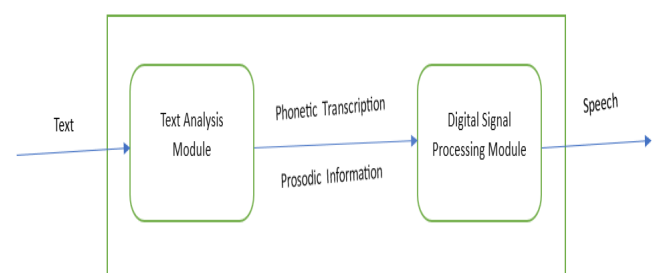


Fig 2 : Conversion from text to speech in TTS

In terms of quality, naturalness, and language support, TTS systems might differ. Deep learning and neural network approaches have significantly improved TTS systems, resulting in more expressive and human-like synthesised speech.

E. Hidden Markov Models (HMM)

In the areas of speech recognition, natural language processing, and bioinformatics, hidden markov models (HMMs) are a popular statistical modelling method for pattern identification and machine learning applications. Given that the underlying system comprises unobserved or hidden states, HMMs are generative probabilistic models that excel at modelling sequential data. The system being modelled in an HMM is thought to be a Markov process, in which a system's future state depends only on its present state and not on its past states. However, it is not possible to directly observe the present status. Instead, based on its hidden state, the system produces visible symbols or observations. Beyond speech recognition and natural language processing, HMMs have a variety of uses. They may be used for anomaly detection, bio sequence analysis, part-of-speech tagging, gesture recognition, and more. They offer an effective framework for modelling and analysing sequential data, particularly when the states of the underlying system are not easily accessible.

F. Connectionist Temporal Classification(CTC)

Machine learning applications of Connectionist Temporal Classification (CTC) include sequence modelling and sequence-to-sequence tasks. Automatic voice recognition (ASR), handwriting identification, and other sequence labelling issues are frequent jobs where CTC is used. The basic goal of CTC is to make it possible for models to be trained from beginning to end for tasks involving sequence prediction without the need for explicit input-output alignment. It enables the model to internalise learning the alignment, making it appropriate for jobs where the alignment is not known or clearly specified. In CTC, the input sequence and the output sequence can both have varying lengths. The objective is to learn a mapping between the input and output sequences without explicitly demanding correspondences between the input and output components.

V. METHODOLOGY FOR CONVERSION

The Overall ML methodology will be explained with each individual components methodology will be explained in the after subtopics in this section . We will get an all round idea about the flow of work in the session and then we will follow with the results section in the upcoming pages of the article.

A. Machine Learning Methodology

In order to train and test the ML model, collect relevant data. This might consist of human input, sensor data, or other pertinent data. Make that the data is accurate, varied, and reflective of the intended user base. To eliminate noise, manage missing values, normalise the data as necessary, and perform any necessary transformations, clean and pre-process the acquired data. Pre-processing procedures might change based on the type of data being utilised and the particular ML algorithms being applied. Determine or produce from the pre-

processed data relevant characteristics that will be utilised as inputs to the ML model.

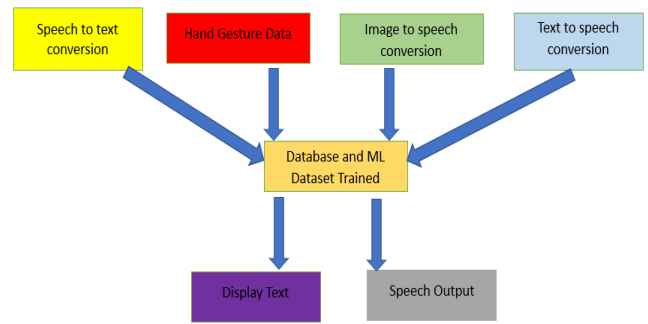


Fig 3 : ML Methodology

Implement the ML model in the software algorithms of the assistive device. Make sure the item is thoroughly tested to make sure it works as intended and meets the demands of the user. Think about things like accessibility, user interface, and real-time processing. Gather user feedback and make adjustments to the device's functioning and design depending on actual usage. ML model should be updated and improved continuously depending on fresh data or changing user needs.

B. Text to Speech Conversion Methodology

A person can use a process for creating a text-to-speech (TTS) conversion system that includes data gathering, model training, assessment, and deployment.

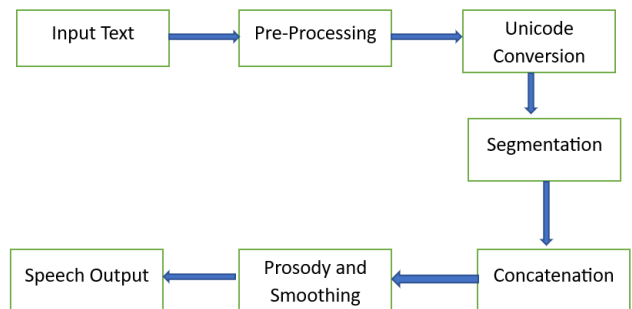


Fig 4 : TTS Methodology

Assemble a sizable, varied text sample with the related speech recordings. To manage any inconsistencies, unusual characters, or formatting problems, clean up and preprocess the text data. Transform the voice recordings into an appropriate acoustic representation, such as spectrograms or Mel-frequency cepstral coefficients (MFCCs). Pick a TTS model architecture that works. A common option is concatenative synthesis. Utilise the preprocessed text and audio information to train the chosen TTS model. Use a variety of measures to assess the effectiveness of the trained TTS model, including the mean opinion score (MOS), naturalness, intelligibility, and alignment with the input text.

C. Image-to-Speech Conversion Methodology

Compile a dataset with pairings of photos and the labels or descriptions that go with them in text. The input photos should be preprocessed to guarantee uniform size, resolution, and colour space. The written descriptions that go with the

photographs should be processed beforehand. Tokenization, punctuation removal, text conversion to lowercase, and vocabulary mapping for the terms used in the descriptions would be necessary for this. Identify visual elements in the image, and then provide the relevant speech.

D. Speech-to-Text Conversion Methodology

Accumulate a bunch of audio files with the appropriate text labels or transcriptions. To make the system more robust, make sure the dataset includes a variety of speakers, accents, and speech variants. The audio recordings should be preprocessed to guarantee consistency in format, sample rate, and length. Improve the quality by using methods like resampling, noise reduction, and normalization. From the preprocessed audio data, extract acoustic characteristics. Mel-frequency cepstral coefficients (MFCCs), delta MFCCs, and energy-based features are examples of frequently used features. The mapping between audio characteristics and their related textual labels should be taught to the model. A deep neural network is commonly used in the model, and its output layer predicts the probability of certain letters or units. Train the model to make its output even more precise. The model uses language models to increase transcription accuracy after taking the model's predictions as input. To create the final transcription, combine the outputs of the training models. Combining techniques like decoding algorithms, such as the Viterbi algorithm, can be used to identify the most probable grouping of words or units based on the probabilities from both models. Utilise relevant metrics to assess the performance of the trained models, such as the word error rate (WER) or character error rate (CER). To evaluate the precision and standard of the transcriptions produced by the system, use a separate assessment dataset.

E. Gesture-to-Text Conversion Methodology

A collection of flicks or photos in sign language and the related text descriptions. Make sure the dataset includes a variety of sign motions and is diverse in terms of backdrop colors, lighting, and hand gestures. The films or photographs of sign language should be preprocessed to guarantee uniformity in terms of size, quality, and colour. Create training, validation, and testing sets from the dataset. The validation set will be utilized for model selection and hyperparameter tweaking, while the training set will be used to train the model. Create a processing architecture that is appropriate for sign language films or pictures. Provide the architecture with the videos or pictures, then use a suitable loss function, such as categorical cross-entropy, to optimize the model's parameters. Measure relevant metrics such as accuracy, precision, recall, or F1 score to assess the model's performance in recognizing sign language gestures.

VI. RESULTS

In this section we will look in all the conversions and the algorithms used for the conversion with aspects chart plot and compare them so that best results can be seen.

A. Text to Speech Conversion

A popular method for producing voice from text input is text to speech synthesis utilising Hidden Markov Models (HMM)

and Text-to-voice Synthesis (TTS).The ability of TTS systems based on HMM to produce lifelike and understandable speech has been well-proven. Higher levels of naturalness, expressiveness, and prosodic quality still present obstacles, though. HMM-based TTS systems may be significantly enhanced by further advances in training data quality, voice modelling methods, and the application of deep learning techniques.

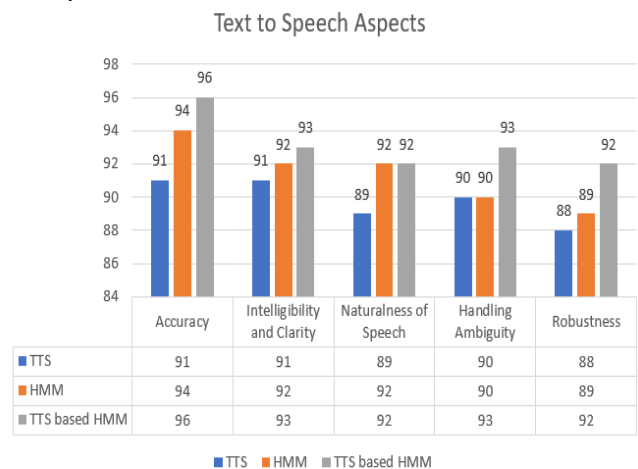


Fig 5 : Aspects of Text to Speech

As you can see that the TTs based HMM model has better performance compared to the individual based models of the algorithms this proves that the TTS based HMM model has better accuracy of prediction by greater percentage than others. The aspects chart plot is shown in figure 5.

B. Image to Speech Conversion

In order to convert pictures into voice, Long Short-Term Memory (LSTM) and Recurrent Neural Networks (RNN) algorithms have produced promising results. The LSTM and RNN models can detect temporal relationships and provide output that is coherent for speech. The capacity to produce understandable speech that matches the content of the input pictures has been established by the LSTM and RNN models. These algorithms have proven to be capable of producing precise, believable, and understandable voice output. To overcome obstacles to gaining increased naturalness, robustness, and computational efficiency, further study and development are required.

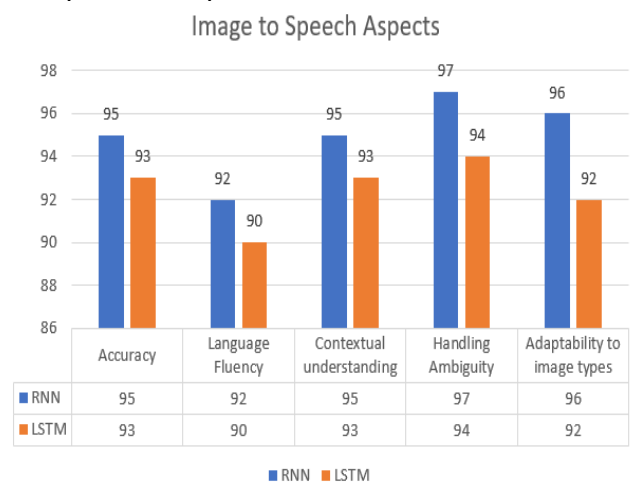


Fig 6 : Aspects of Image to Speech

C. Speech to Text Conversion

It has been extensively investigated to convert speech to text using Connectionist Temporal Classification (CTC) and Hidden Markov Models (HMM), each of which has advantages and disadvantages of its own. Systems for reliably transcribing spoken words and obtaining excellent recognition performance based on CTC and HMM have demonstrated encouraging results. CTC models may accurately capture the temporal relationships and patterns in speech, leading to accurate transcriptions. These models are often based on deep neural networks (DNNs) or recurrent neural networks (RNNs). HMM-based systems may successfully include linguistic context and boost transcription accuracy when used in conjunction with language models. When there are homophones or context-dependent changes, language models can improve recognition, aid with ambiguities, and represent the statistical regularities of language.

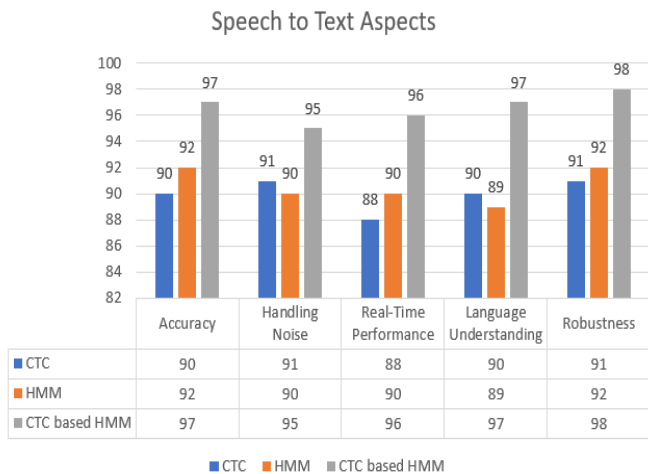


Fig 7 : Aspects of Speech to Text

The conversion of voice to text has showed promise using both CTC and HMM-based methods. While HMM-based systems excel at adding linguistic context through language models, CTC models enable end-to-end training and resilience to temporal misalignments. In order to further improve the accuracy, robustness, and efficiency of voice to text conversion systems, research is now focused on integrating the advantages of both methodologies and investigating cutting-edge deep learning techniques. The CTC based HMM system has validated the conversion in a better way so that the output can be so accurate.

D. Gesture to text Conversion

In order to comprehend hand motions, a CNN architecture is highly suited for examining spatial data in a picture or video frames. Multiple convolutional layers may be used in the design, followed by pooling layers to extract useful information. Depending on the project's unique goals, more fully linked layers are used for classification or regression tasks. There are training and validation sets created from the dataset. A method based on CNN that converts gestures or sign language to text shows potential for those who have speech impairments. The dataset's quality, the model's design, and the training procedure all affect the outcomes. The main benefit of utilising ML and CNN-based gesture or sign language detection is to help those who have trouble speaking interact with others.



Fig 8 : Gesture to Text Prediction

For this conversion, any other algorithm has not been considered other than CNN as we had time constraint while building the work the aspects of the CNN alone in shown in Figure 9.

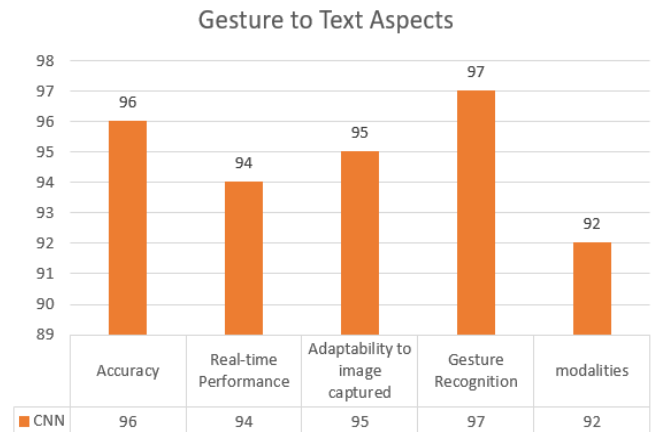


Fig 9 : Gesture to Text Aspects

The Accuracy and the aspects of the conversion are very good in numbers as you can see in the chart plot. The latency and the Real time performance of the CNN algorithm is good and useful for the prediction.

VII. CONCLUSION

CNN-based ML-based gesture or sign language to text conversion systems have enormous potential to change the lives of people with speech problems by enabling them to communicate effectively, participate actively, and be accepted in different facets of society. For the benefit of those with visual impairments, LSTM and RNN algorithms provide useful capabilities for transforming images into audio descriptions, giving them access to visual information via aural methods. Using TTS synthesis with HMM, it is possible to effectively translate written text into spoken speech. The system offers linguistic adaptability, natural-sounding voices that may be customised, and adaptation to various text inputs. For speech-to-text conversion, the combination of CTC and HMM algorithms offers a number of substantial benefits. Their resistance to temporal fluctuations, capacity to represent languages, flexibility in using them, noise resistance, scalability, and continuing are now useful tools for correctly

converting spoken language to printed form thanks to developments.

VIII. FUTURE SCOPE

The development of other cutting-edge technologies, such as computer vision, natural language processing, robotics, and wearables, may be taken advantage of by ML-based assistive devices. With the help of this integration, it will be possible for people with disabilities to interact in a variety of contexts in a more smooth and natural way. The future of assistive technology based on ML algorithms looks bright. For people with impairments, ML algorithms offer the potential to create highly personalised, context-aware, and adaptable solutions that improve engagement, communication, and independence. Continued innovation-spurring research, development, and interdisciplinary cooperation will increase the capabilities of assistive technology, enhancing accessibility and fostering inclusion for everyone.

REFERENCES

- [1] Vaidya, O.; Gandhe, S.; Sharma, A.; Bhate, A.; Bhosale, V.; Mahale, R. Design and development of hand gesture based communication device for deaf and mute people. In Proceedings of the IEEE Bombay Section Signature Conference (IBSSC), Mumbai, India, 4–6 December 2020; pp. 102–106.
- [2] Marin, G.; Dominio, F.; Zanuttigh, P. Hand gesture recognition with Leap Motion and Kinect devices. In Proceedings of the IEEE International Conference on Image Processing (ICIP), Paris, France, 27–30 October 2014; pp. 1565–1569.
- [3] Saleem, M.I.; Otero, P.; Noor, S.; Aftab, R. Full duplex smart system for Deaf & Dumb and normal people. In Proceedings of the Global Conference on Wireless and Optical Technologies (GCWOT), Mlaga, Spain, 6–8 October 2020; pp. 1–7.
- [4] Deb, S.; Suraksha; Bhattacharya, P. Augmented Sign Language Modeling (ASLM) with interaction design on smartphone—An assistive learning and communication tool for inclusive classroom. *Procedia Comput. Sci.* 2018, 125, 492–500.
- [5] Howard, A.G.; Zhu, M.; Chen, B.; Kalenichenko, D.; Wang, W.; Weyand, T.; Andreetto, M.; Adam, H. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. *arXiv* 2017, *arXiv:1704.04861*.
- [6] Rishi, K.; Prarthana, A.; Pravena, K.S.; Sasikala, S.; Arunkumar, S. Two-way sign language conversion for assisting deaf-mutes using neural network. In Proceedings of the 8th International Conference on Advanced Computing and Communication Systems, ICACCS 2022, Coimbatore, India, 25–26 March 2022; pp. 642–646.
- [7] Anupama, H.S.; Usha, B.A.; Madhushankar, S.; Vivek, V.; Kulkarni, Y. Automated sign language interpreter using data gloves. In Proceedings of the International Conference on Artificial Intelligence and Smart Systems (ICAIS), Coimbatore, India, 25–27 March 2021; pp. 472–476.
- [8] Kawai, H.; Tamura, S. Deaf-and-mute sign language generation system. *Pattern Recognit.* 1985, 18, 199–205.
- [9] Bhadauria, R.S.; Nair, S.; Pal, D.K. A Survey of Deaf Mutes. *Med. J. Armed Forces India* 2007, 63, 29–32.
- [10] Sood, A.; Mishra, A. AAWAAZ: A communication system for deaf and dumb. In Proceedings of the 5th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), Noida, India, 7–9 September 2016; pp. 620–624.
- [11] Yousaf, K.; Mehmood, Z.; Saba, T.; Rehman, R.; Rashid, M.; Altaf, M.; Shuguang, Z. A Novel Technique for Speech Recognition and Visualization Based Mobile Application to Support Two-Way Communication between Deaf-Mute and Normal Peoples. *Wirel. Commun. Mob. Comput.* 2018, 2018, 1013234.
- [12] Raheja, J.L.; Singhal, A.; Chaudhary, A. Android Based Portable Hand Sign Recognition System. *arXiv* 2015, *arXiv:1503.03614*.
- [13] Soni, N.S.; Nagmode, M.S.; Komati, R.D. Online hand gesture recognition & classification for deaf & dumb. In Proceedings of the International Conference on Inventive Computation Technologies (ICICT), Coimbatore, India, 26–27 August 2016; pp. 1–4.
- [14] Chakrabarti, S. State of deaf children in West Bengal, India: What can be done to improve outcome. *Int. J. Pediatr. Otorhinolaryngol.* 2018, 110, 3742.
- [15] Ameer, S.; Khalifa, A.B.; Bouhleb, M.S. Chronological pattern indexing: An efficient feature extraction method for hand gesture recognition with Leap Motion. *J. Vis. Commun. Image Represent.* 2020, 70, 102842.
- [16] Ameer, S.; Khalifa, A.B.; Bouhleb, M.S. A novel hybrid bidirectional unidirectional LSTM network for dynamic hand gesture recognition with Leap Motion. *Entertain. Comput.* 2020, 35, 100373.
- [17] Boppana, L.; Ahamed, R.; Rane, H.; Kodali, R.K. Assistive sign language converter for deaf and dumb. In Proceedings of the 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), Atlanta, GA, USA, 14–17 July 2019; pp. 302–307.
- [18] Suhajito; Anderson, R.; Wiryana, F.; Ariesta, M.C.; Kusuma, G.P. Sign Language Recognition Application Systems for Deaf-Mute People: A Review Based on Input-Process-Output. *Procedia Comput. Sci.* 2017, 116, 441–448.
- [19] Patwary, A.S.; Zaohar, Z.; Sornaly, A.A.; Khan, R. Speaking system for deaf and mute people with flex sensors. In Proceedings of the 2022 6th International Conference on Trends in Electronics and Informatics, ICOEI 2022, Tirunelveli, India, 28–30 April 2022; pp. 168–173.