Sign language Detection using Deep Learning

1: Prof. Sonali Sonvane
Dept. Of Artificial Intelligence
GHRCEM, Pune
sonali.sonavane@raisoni.net

2: Aditi Kannawar
Dept. Of Artificial Intelligence
GHRCEM, Pune
aditi.kannawar.ai@ghrcem.raisoni.net

3: Pranali Patil
Dept. Of Artificial Intelligence
GHRCEM, Pune
line 4: City, Country
pranali.patil.ai@ghrcem.raiosni.net

4: Rani Manawar *Dept. Of Artificial Intelligence*
GHRCEM, Pune
rani.manwar.ai@ghrcem.raisoni.net

5: Tanay Mapare
Dept. Of Artificial Intelligence
GHRCEM, Pune
tanay.mapare.ai@ghrcem.raisoni.net

***Abstract*— The predominant means of communication is speech; however, there are persons whose speaking or hearing abilities are impaired. Communication presents a significant barrier for persons with such disabilities. The use of deep learning methods can help to reduce communication barriers. This paper proposes a deep learning-based model that detects and recognizes the words from a person's gestures. Deep learning models, namely, LSTM and GRU (feedback-based learning models), are used to recognize signs from isolated Indian Sign Language (ISL) video frames. The four different sequential combinations of LSTM and GRU (as there are two layers of LSTM and two layers of GRU) were used with our own dataset, IISL2020. The proposed model, consisting of a single layer of LSTM followed by GRU, achieves around 97% accuracy over 11 different signs. This method may help persons who are unaware of sign language to communicate with persons whose speech or hearing is impaired.**

I. INTRODUCTION

There are various ways of communication or expression, but the predominant mode of human communication is speech; when it is hindered, people must use a tactile-kinaesthetics mode of communication instead. In India, the overall percentage of persons with this disability in the population was 2.2 percent from July 2018 to December 2018, according to the National Statistical Office survey report 76th round of the National Sample Survey (NSS) . Sign language is one of the greatest adaptations for persons with speech and hearing impairments. It is also known as a visual language. Generally, it has five basic parameters: hand shape, orientation, movement, location, and components such as mouth shape and eyebrow movements. Research has been conducted on voice generation using smart gloves, which could give a voice to sign language movements. However, those who do not know sign language usually undervalue or reject persons with such an impairment because of the lack of proper communication between them. Hence, this paper proposes a system aimed at removing the communication gap and giving every individual a fair and equal chance. It involves taking a video of the person making hand gestures, processing it, and passing it to the proposed model, which predicts words one by one. The system then generates a meaningful sentence out of those words that can then be converted into the language selected by the communicator.

II. LITERATURE SURVEY

1. Hindi sign language recognition based on computer vision [1]. This paper presents a system that uses CNN and RNN to recognize Hindi language to aid communication with people with speech disorders. Challenges include large data sets and instant analysis. Purpose: to expand knowledge, increase accuracy, define a word and translate it into Hindi.

2. Using CNN for Hindi language search [2] Data collection includes OpenCV character detection, prioritization, CNN training and evaluation. Mission: Help deaf people communicate, solve problems accurately and increase punctuality.

3. Using convolutional neural networks for Indian sign language recognition [3]. Article using CNN for Indian sign recognition: covers image acquisition, segmentation and CNN training. Designed to improve communication, accessibility and freedom. Competition: special character replacement, image name predictions. The future: advanced image processing, two-way communication and instant translation.

4. Using customizable neural networks for language recognition [4]. Language recognition: including data collection, prioritization, CNN generation, accuracy assessment of GUI and time about real-time operations and product removal. Challenges include learning more to increase accuracy. The future: Reduced reality, alternative learning and computer vision.

5. Gesture recognition based on convolutional neural networks - Bidirectional longterm memory network for attaching wrist sensors with multi-walled carbon nanotubes/cotton fabric [5] to the wrist, recording movements and testing accuracy. Flexible and sensitive sensors made from MWCNT/CF composites. Challenges include limited movements and real-world applications. It aims to promote advances in human-computer interaction, healthcare, service technology, intelligent information and business applications.

6. Signature Recognition System Using Deep Neural Networks [6]. Factsheet Note: Classification using Python using 2-layer CNN. It can recognize up to six sign languages with high accuracy and knows how to control equipment. There is a lack of discussion of CNN's usage issues. The future: improved

performance, video recognition and integrated device management.

7. Gesture recognition based on static annotations using convolutional neural networks [7]. Gesture recognition: including hand segmentation, dataset creation, CNN design, training/testing, testing and measuring time. Get high-precision, real-time output and provide customizable datasets. Challenges include lighting, static movements, timing, background and Indian Sign Language. Focus on the future: Improving CNN in terms of quality, multilingualism, lighting and translation.

8. Gesture-like language recognition using CNN [8]. Includes hand segmentation, dataset creation, CNN design, training/testing, testing and evaluation time. High precision, instant output. Limitations: lighting, static movements, timer, background, Indian Sign Language. Future: Dynamic movements, multilingual, advanced CNNs for illumination and translation.

9. Mudra: Convolutional Neural Network Based Banking Indian Sign Language Translator [9]. Mudra: CNN based Banking Indian Sign Language Translator. Involved in handsharing, data generation, CNN design, training, immediate testing and evaluation. It aims to bridge the communication gap with deaf people and solve problems arising from the complexity of Indian Sign Language. Future steps include dataset expansion, sentence insertion, automation, text-to-speech integration, and pattern discovery.

10. Hindi VARNAMALA Language recognition using CNN [10]. Hindi Varnamala Language Recognition: Includes data generation, CNN training, prediction and distribution as a web application. It aims to promote unity and communication in India. Focus on providing interesting information, accuracy and continuous improvement to improve knowledge of Indian Sign Language (ISL). Future steps include improving accuracy, expanding data, and marketing for broader adoption.

11. Multi-task sign language recognition system using Wi-Fi [11]. Multi-task sign language recognition: Wi-SignFi system uses CSI data, eight-layer CNN and KNN. Leverage Wi-Fi detection for broad coverage, fewer privacy concerns, and cost-effectiveness. Among the challenges is the fact that it is affected by the access problem. The future: Internet of Things integration, knowledge discovery and real-time embedded applications.

12. Dynamic sign language and auto-coding time series neural network known as CBAM [12]. Dynamic sign language recognition: CBAM's convolutional auto-coding network is used for recognition and prioritization. Thanks to continuous exposure, 89.90% accuracy and efficiency were achieved.

III. ALGORITHM

A. *Initialization*

Load required libraries (e.g., TensorFlow, OpenCV, MediaPipe). Load pre-trained deep learning models (CNN for feature extraction, LSTM/GRU for sequence modeling).

B. *Dataset Preparation*

Load and preprocess the Sign Language MNIST dataset. Augment data to increase robustness (e.g., rotations, scaling).

Fig. 1. Dataset Collection

C. *Hand Detection and Recongnition*

Capture real-time video feed from the camera.Use MediaPipe library to detect and track the hand. Extract hand region from each frame

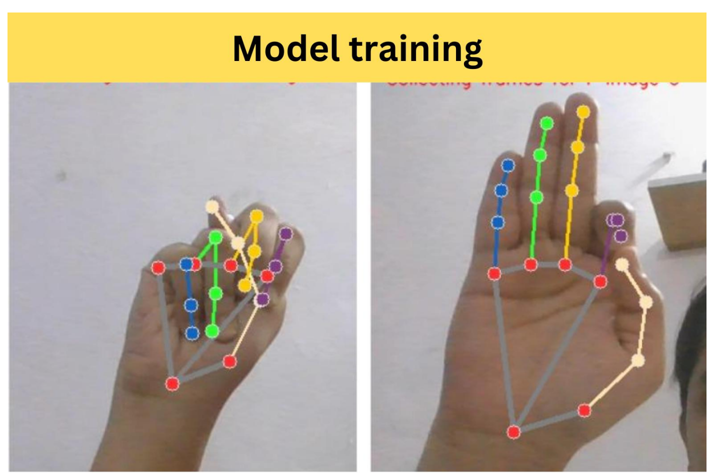

Fig. 2. Model Training using MediaPipe

D. Preprocessing

Resize the image to the input size required by the model (e.g., 28x28 for MNIST). Normalize pixel values to the range [0, 1]

E. Feature Extraction

Pass the preprocessed hand image through the CNN model. Extract relevant features (e.g., shape, size, orientation).

F. Sign Recognition

Use LSTM/GRU to analyze the sequence of features extracted from frames. Classify each frame into a specific sign language.

G. Output

Convert the recognized gestures into text.

IV. METHODOLOGY

Sign language detection involves the recognition and interpretation of hand gestures, movements, and facial expressions to translate sign language into text or speech.

It plays a crucial role in facilitating communication for individuals with hearing impairments, allowing them to interact more effectively with others who may not know sign language.

A. Deep Learning Model

- Long-Term Memory (LSTM): The LSTM model is used to recognize and interpret hand movements in video frames, resulting in better temporal memory.
- Gated Recurrent Unit (GRU): The GRU model is used to recognize and interpret hand gestures and enables the system to understand complex sequences.
- Convolutional Neural Network (CNN): CNN plays an important role in video extraction and classification of hand gestures, which can be converted into English alphabet.
- Hybrid approach of LSTM and GRU: This project increases overall recognition by combining LSTM and GRU models to better capture the progression of hand movements.
- Using computer vision technology to improve accuracy: Hand tracking using pipelines and a dual camera setup provides different perspectives to increase accuracy andrecognition of symbols in real time.

B. System Architecture

Data flow diagram (DFD) is important for visualizing the flow of data in a system. It usually consists of processes, data storage, data flow, and other resources. A system flowchart shows all the higher order interactions between system components and other entities. It captures how data moves through the system by identifying inputs and outputs. Module plans break down the system into smaller, manageable modules, showing how the modules interact with each other. Itshows the relationship and dependency of different parts of the system. Execution order diagrams focus on the order of operations, showing the order in which operations in the system are executed. This picture gives a clear picture of the good behavior of the body, helping to understand the progress of the system through which information and many activities are carried out. Together, these diagrams help define, design and understand systems, enabling efficient data processing and process execution. They also help identify potential vulnerabilities and improve performance by clarifying interactions and data dependencies.

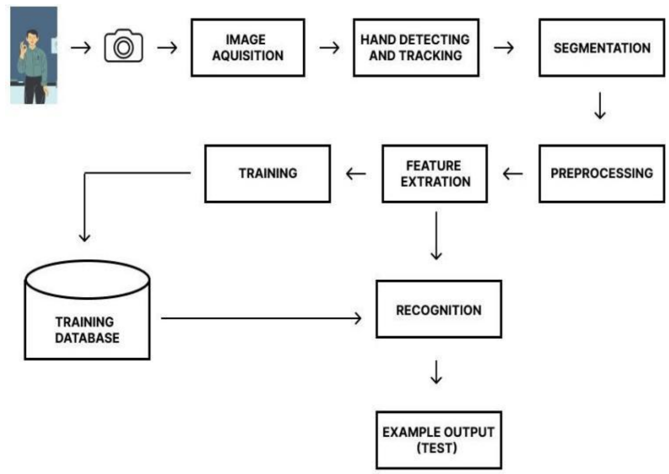

Fig. 3. System Architecture

C. Implementation

- In the follow-up phase, various models were developed to meet the needs of different module levels.
- This involves translating the design into actual language-specific code and functionality. This stage is very important because it ensures that the design and the necessary work for inspection and delivery are completed and the project is implemented.

These algorithms are integrated into an integrated system that can process videos or images in real time. Additionally, machine learning models such as convolutional neural networks (CNN) are trained to recognize and classify different aspects. to run.

D. Training

- Training a language search model for the Deaf involvesa set of cognitive skills sufficient to interpret and recognize signs in images. In this case, the training model focuses only on the detection of images that are important for the correct understanding of sign language and its translation into written or spoken text.
- Raise your hand. These images are recorded to show the location of interest, providing the ground truth necessary for training the model.
- This file contains many mesh types, orientations and

movements to ensure the model can be made well. Data set. CNNs are especially good at image recognition tasks because they can learn hierarchical features directly from pixel values.

E. Testing

- Testing plays an important role in ensuring the validity and reliability of descriptive diagnostic methods developed for the deaf, especially when focusing on visual diagnosis.
- This important phase in the development of the software life cycle must complete various tests to carefully evaluate the performance and functionality of the model without exceeding limitations and restrictions. This includes confirming that the model definition is correct and recognizing different movements in the image.
- These test cases are designed to cover a wide range of hand positions, orientations, and movements to ensure that the model can be useful for different descriptions. Performance testing also evaluates the model's ability tohandle real-world conditions such as changes in lighting conditions or background noise that may affect its performance.

F. Result

- The proposed deep learning model performed well, achieving approximately 97% accuracy on 24 different symbols representing English letters (ISL) in video frames.
- This remarkable fact demonstrates a good standard of knowledge in detecting and identifying characters basedon human movements.
- This decision is important to reduce communication problems for people with speech or hearing disabilities.Self-sufficiency in every situation.

  Whether in the classroom, in the workplace, or in daily interaction, a high standard of work ensures effective and efficient communication. Demonstrates the ability to improve inclusivity and accessibility across multiple contexts.

The model could enable better and more effective interactions for people with disabilities by reducing reliance on traditional means of communication, such ashandwriting or lip reading

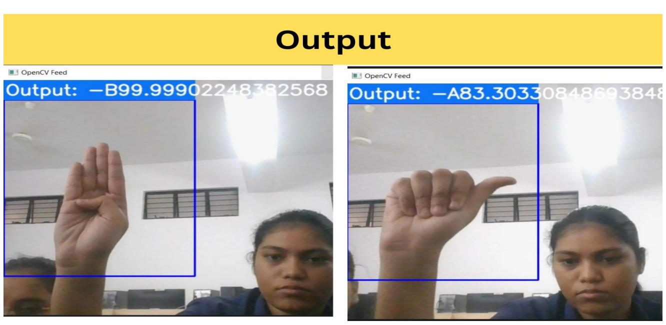

Fig. 4. Result of Model

V. MODEL ACCURACY COMPARISON

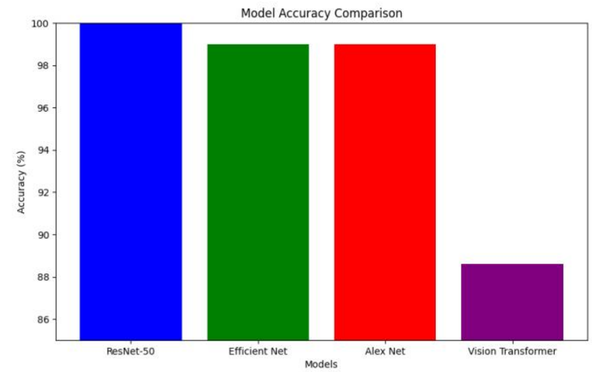

Fig. 5. Model Accuracy Comparison

- **ResNet-50**: Achieved an exceptional accuracy rate of 99.988%.
- **Efficient Net and Alex Net**: Demonstrated commendable accuracy levels of over 99%.
- **Vision Transformer model**: Showed a comparatively lower accuracy rate of 88.59%

VI. FEATURE IMPORTANCE

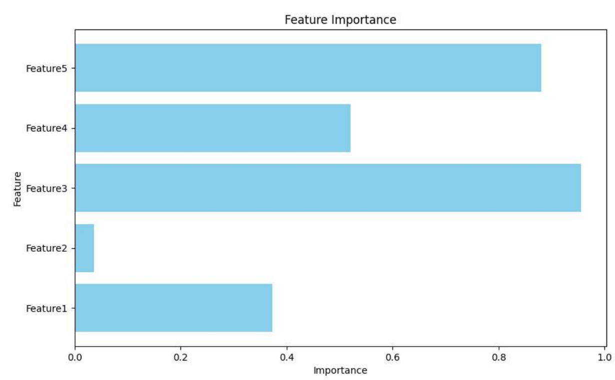

Fig. 6. Feature Importance

- Feature Importance Graphs showing the importance of different features extracted by the model, helping to understand which features are most significant for accurate gesture recognition.
- A feature importance graph for the CNN model, highlighting the key features that contribute to gesture recognition.

VII. ROC GRAPH

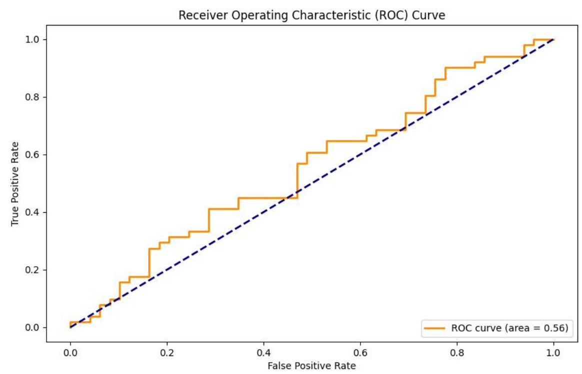

Fig. 7. ROC Graph

- Receiver Operating Characteristic (ROC) curve for assessing the diagnostic ability of the classification model, plotting the true positive rate against the false positive rate.
- ROC curve for the CNN model to evaluate its performance in distinguishing between different hand gestures.

VIII. TRAINING AND VALIDATION ACCURACY

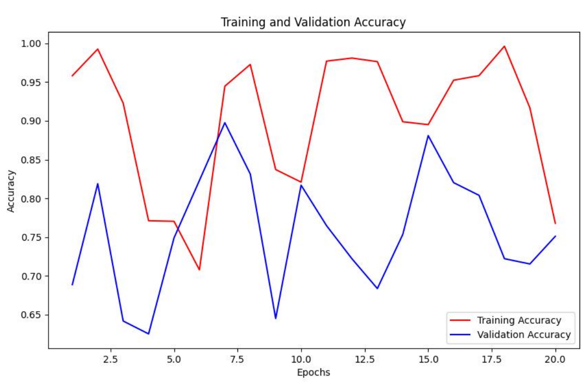

Fig. 8. Training and Validation Accuracy

- Graphs showing the performance metrics of different models, such as accuracy, precision, recall, and F1-score, across various epochs or iterations of training.
- Line graphs depicting the accuracy and loss of the CNN model during training and validation phases.

IX. CONFUSION MATRIX

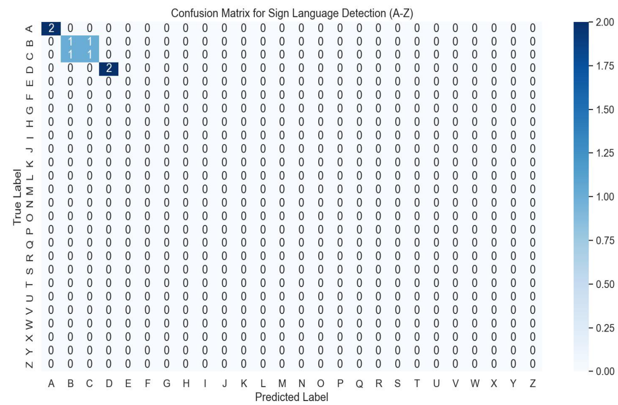

Fig. 9. Confusion Matrix

The confusion matrix visualizes the performance of a sign language detection model, representing predictions against actual classes from 'A' to 'Z'. Each cell displays the count of instances where a true label (rows) was predicted as a specific class (columns). Higher counts on the diagonal indicate accurate predictions, while off-diagonal elements represent misclassifications. The heatmap's color intensity reflects the frequency of predictions, aiding in identifying areas of confusion. This visual summary helps assess the model's effectiveness, revealing which classes are frequently confused and guiding improvements in classification accuracy for each sign language gesture.

X. TECHNOLOY USED

A. Python

Python is a high-level programming language known for its simplicity and readability. Web development, data science, machine learning, etc. It is widely used in many areas including. Python's extensive standard library and large ecosystem of third-party packages make it a popular choice for developing a variety of applications. Python provides powerful libraries and functions for machine learning and computer vision, such as TensorFlow and OpenCV. Perception-based AI including body recognition, hand measurement, face detection and more. It provides advanced training and integration models for efficient processing of multimedia files such as images, video and audio.

B. Mediapipe

MediaPipe's modular design and functionality make it suitable for real-time applications, allowing developers to quickly build and implement AI-driven solutions without the need for technical expertise in machine learning or computer vision.

C. Tensorflow

TensorFlow is a popular solution. It provides a great ecosystem of tools, libraries, and resources for developing and implementing learning models, especially deep learning models. TensorFlow provides flexibility and scalability by supporting high-level APIs like Keras for simple modeling and low-level APIs for better control of model design and training.

D. OpenCV

OpenCV is written in C++ but provides Python bindings, making it easier for Python developers to access and use. It is widely used in many applications such as robotics, augmented reality, medical imaging, surveillance and autonomous vehicles. OpenCV's rich set of tools and techniques make it an importanttool for developing computer vision applications, especially when combined with deep learning such as TensorFlow. It provides a comprehensive set of tools and capabilities for building, debugging, and deploying software applications across multiple platforms and programming languages, including Python.

E. Visual Studio Code

Visual Studio Code provides a rich development environment with functions such as code editing, project management, integration management, debugging tools and more. By supporting many programming languages and frameworks, it offers many options to developers working on different projects, including machine learning and computer vision implemented in Python using patterns such as TensorFlow and OpenCV.

XI. CONCLUSION

This project represents a commitment to transforming communication and accessibility for the deaf through the output of technology. The project leverages advanced deep learning models such as convolutional neural networks (CNN), long-term memory (LSTM), and gated recurrent units (GRU), as well as advanced computer vision techniques such as manual pointing and dual camera setups. design. instantly understand and interpret hand movements. This collaboration demonstrates a commitment to using the power of artificial intelligence and computer vision to break down communication barriers and enable deaf people to communicate to their benefit. Collection, training standards, and performance evaluations demonstrate a commitment to continuous improvement.

ACKNOWLEDGEMENT

In the preliminary project report on Sign Language Recognition Using Deep Learning, the acknowledgment section expresses gratitude to Prof. Sonali Sonavane and Prof. Rachna Sable for their guidance and support throughout the project. The internal guide, Prof. Sonali Sonavane, and the Head of the Artificial Intelligence Department, Prof. Rachna Sable, played crucial roles in providing assistance and valuable suggestions that significantly contributed to the successful completion of the project. Their support and insights were instrumental in the project's development and completion.Authors and Affiliations.

FUTURE WORK

The aim of the project is to improve linguistic knowledge through a variety of strategies and is expected to lead to significant advances in the accessibility and communication of words for the deaf. An important approach involves expandingthe database to include various gestures, dynamic characters, and faces. By diversifying the dataset, deep learning models can be more general and more accurate in explanation, thus improving overall knowledge.

REFERENCE

[1] Shital Pawar, Yashashree Shastri, Syeda Zarah Aiman, "Bidirectional Sign Language Assistant with MediaPipe Integration", 2024 International Conference on Emerging Smart Computing and Informatics (ESCI), pp.1-8, 2020

[2] Snitik Swaroop, Kurivella Yatindar Prasad, G Sai Srikar Reddy, Mahathi Rachavelpula, Usha A. Jogalekar, "Real Time Sign Language Detection", *2024 International Conference on Artificial Intelligence, Computer, Data Sciences and Applications (ACDSA)*, pp.1-7, 2024.
NagaJyothi Devabathini, P Mathivanan, "Sign Language Recognition Through Video Frame Feature Extraction using Transfer Learning and Neural Networks", *2023 International Conference on Next Generation Electronics (NEleX)*, pp.1-6, 2022

[3] Md.Zahirul Islama and Md.Shahadat Hossain, "Static Hand Gesture Recognition " 2019 IEEE International Conference paper, doi: 10.1109/ICIEV.2019.8858563.

[4] Mukesh Kumar Makwana, "Sign language recognition using convolutional neural networks." 2014.

[5] K R Prabha, B Nataraj, B Thiruthanikachalam, S Surya NarayananS, M Vishnu Prasath, "Audio to Sign Language Translation using NLP", *2023 Third International Conference on Smart Technologies, Communication and Robotics (STCR)*, vol.1, pp.1-4, 2023.

[6] Sujata Joshi, Bhaskar Sharma, Dharmender, Divyansh Tripathi, Akshat Shah, "Hand Sign Language Detection Using Deep Learning and Computer Vision", *2023 International Conference on Ambient Intelligence, Knowledge Informatics and Industrial Electronics (AIKIIE)*, pp.1-4, 2023

[7] Syeda Fariha Naqvi, Mahek Zahid Suleman, Kavya Dev Verma, Prachi Dhanawat, "ISL Animated Translator and Desktop Virtual Assistant", *2023 3rd Asian Conference on Innovation in Technology (ASIANCON)*, pp.1-6, 2023.

[8] K Anitha, R Naveen Karthick, K C Varsheni, A Kalaiselvi, "Gesture based Sign Language Recognition System", *2023 2nd International Conference on Advancements in Electrical, Electronics,*

Communication, Computing and Automation (ICAECA), pp.1-6, 2023.

[9] S Vijaya Shetty, Ajith Ram, Arjun Bhargav, A Harshinee, Kanchan Pathak, "The Voiceless Assistant for The Specially Abled", *2023 International Conference on Applied Intelligence and Sustainable Computing (ICAISC)*, pp.1-7, 2023.

[10] K. Bavani, P. Viswanath Reddy, P.V. Siddhartha, B. Tarasingh Naik, Y.V.S. Srihari, "Sign Language Recognizer: A Deep Learning Approach", *2023 2nd International Conference on Vision Towards Emerging Trends in Communication and Networking Technologies (ViTECoN)*, pp.1-6, 2023

[11] Aditi Deshpande, Ansh Shriwas, Vaishnavi Deshmukh, Shubhangi Kale, "Sign Language Recognition System using CNN", *2023 International Conference on Intelligent and Innovative Technologies in Computing, Electrical and Electronics (IITCEE)*, pp.906-911, 2023.

[12] Hitesh Sri Sai Kaushik Vezzu, Sunny Nalluri, Kakumanu Swetha, Vadapalli SaiKumar, "Hand Sign Recognition using Image Processing", *2023 International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT)*, pp.288-293, 2023.