Deep Learning for Sign Language Interpretation with Conventional Neural Network

A. Senthilselvi
Department of CSE
SRMIST, Ramapuram
Chennai, India
senthila3@srmist.edu.in

Ballela Avinash Reddy
Department of CSE
SRMIST, Ramapuram
Chennai, India
br6152@srmist.edu.in

Charan Deepesh Kristam
Department of CSE
SRMIST, Ramapuram
Chennai, India
ck3378@srmist.edu.in

Rachaputi Yeswanth Vardhan
Department of CSE
SRMIST, Ramapuram
Chennai, India
rv3396@srmist.edu.in

Senthil Pandi S
Department of CSE
Rajalakshmi Engineering College,
Chennai, India
mailtosenthil.ks@gmail.com

2024 International Conference on Smart Technologies for Sustainable Development Goals (ICSTSDG) | 979-8-3315-1628-4/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICSTSDG61998.2024.11026679

Abstract— Deep Learning for Sign Language Interpretation With Conventional Neural Network project aims to create a sign language recognition system in real-time using CNNs, which includes voice output integration. The system can effectively identify hand movements that represent letters of the alphabet and translate them into spoken words, improving communication for those who cannot speak or have difficulty hearing. The recognition model is educated with a strong dataset of American sign language (ASL) and incorporated with TensorFlow Lite for efficient operation on mobile devices, guaranteeing portability and convenience. The user interface enables easy interaction, including the ability to add gestures, listen to them, add spaces, and delete the input. Instant translation of sign language gestures into spoken words is made possible with real-time processing, ensuring a smooth user experience. Although the system shows excellent precision in controlled environments, additional efforts are necessary to enhance its effectiveness in different settings, hand positions, and lighting situations. This project is a major advancement in making communication easier with sign language recognition and voice output.

Keywords—Sign language, American sign language, conventional neural network

I. INTRODUCTION

Sign language is an important mode of communication for numerous people worldwide, particularly for those who are deaf or have difficulty hearing. It is crucial in enabling these individuals to communicate, access information, get education, and engage in various aspects of daily life. Sign language conveys ideas through hand gestures [1,2], body movements, and facial expressions, in contrast to spoken languages that rely on sound for communication. It is an advanced and purposeful method of articulating thoughts and emotions, able to communicate nuanced sentiments and intricate ideas. Knowing sign language allows individuals to establish personal relationships and participate actively in society, promoting inclusiveness and mutual understanding among different communities. Despite the significance of sign language, individuals who depend on it frequently encounter substantial communication obstacles when engaging with individuals unfamiliar with this unique language. Conventional techniques, like lip-reading or trying to understand facial expressions, frequently do not suffice, resulting in miscommunications or ineffective interactions. These barriers can restrict sign language users' participation in social, educational, and professional environments, leading to a separation between the hearing and non-hearing populations. With these difficulties in mind, there is a growing need for technological solutions that can help bridge the communication barrier between sign language users and non-signers. Meeting this need is essential for promoting inclusivity and enhancing the quality of life for people who depend on sign language.

Our research is centered on creating a strong, flexible, and effective system that can identify sign language in real-time by using Convolutional Neural Networks (CNNs) [3,4]. This system is created to meet the urgent demand for a tool that can correctly understand sign language movements and transform them into understandable or audible results, facilitating better communication between sign language users and those who do not use sign language. Our goal is to develop a solution that excels at recognizing and categorizing sign language gestures, particularly those of American Sign Language (ASL), by utilizing the capabilities of CNNs, a type of deep learning architecture renowned for its effectiveness in processing visual data [5]. We aim to create a tool that translates hand gestures into text or speech in real-time to improve communication for sign language users and remove barriers to interaction.

Recognizing sign language involves multiple disciplines such as computer vision, natural language processing (NLP), pattern recognition, and machine learning. The difficulty is in correctly understanding the visual information of hand movements, which can differ in form, motion, and position, while also taking into account other elements such as facial expressions and body gestures. Our method relies on CNNs to examine this intricate visual information, enabling the system to gain insights from the patterns in the hand gestures shown in images and videos. This allows the system to distinguish minor variations in gestures, which is essential for correctly identifying the wide variety of signs used in ASL. When creating the system, we adhered to multiple essential stages to guarantee the model's precision and dependability. Data preprocessing is the initial crucial stage of the process. This step guarantees that the input data is devoid of errors and is uniform, ensuring efficient model training. We utilized a large dataset from Kaggle, consisting of numerous ASL hand gesture images, in the training of our CNN model. Moreover, we gathered additional information using surveys and experiments, enhancing the dataset. Preparation tasks involve resizing images, normalizing, and augmenting data. Resizing images is important to maintain uniform size for input into the CNN. Normalization adjusts the pixel values to a defined range, which accelerates the training process. Data augmentation, including random rotations, flips, and brightness adjustments, is employed to increase the dataset artificially and enhance the model's capacity to generalize to unfamiliar data.

979-8-3315-1628-4/24/$31.00 ©2024 IEEE

Segmentation is the next vital step following data preparation. During this stage, the system recognizes the primary components in every image, with a special emphasis on the hands as they play a critical role in accurate sign language identification. Segmentation helps the model focus on important parts of the image by removing unnecessary background noise and irrelevant details. For example, a picture could have background items or different lighting that might confuse the model during training. By concentrating exclusively on the hands the system can more effectively recognize and classify the gestures in sign language. Once the key highlights are recognized, the CNN design is utilized to find outstanding designs within the visual information. CNNs are extraordinary at this work since they are talented at getting various leveled information representations through numerous layers of convolutional channels. The demonstration starts by recognizing basic highlights such as edges and lines within the starting layers, at that point advances to identifying complex designs like shapes and surfaces within the afterward layers. This approach makes a difference the CNN gets it the complexities of hand signals, counting little changes in area, development, and frame. By analyzing these designs, the CNN learns to recognize a wide extend of ASL signs, indeed in challenging conditions ASL is the most sign language utilized by deaf and hard-of-hearing individuals within the United States and a few ranges of Canada. It has its own linguistic use, sentence structure, and word arrangement that are unmistakable from English, making it a totally normal language. Sign language depends on hand sign and gestures. The language is more than fair signals; it could be an advanced framework that can pass on basic commands as well as complex contemplations and sentiments. The complexity and abundance of ASL's visual and expressive highlights show a troublesome however satisfying objective for computerized acknowledgment frameworks.

This article presents an interesting strategy for improving the accuracy of frameworks for recognizing sign language. The proposed arrange strategy employs language models to help with the acknowledgment strategy, coordination and early sign language detection. Module matched with a content rectification module, an unused improvement found in earlier ponders. The content adjustment module is used for adjustment purposes to adjust the estimates made some time recently the sign language module for sign acknowledgment is the ultimate organizer of the process. model for recognizing dialects. By combining these components, the accuracy of the demonstration for recognizing sign language modules is incredibly made strides.

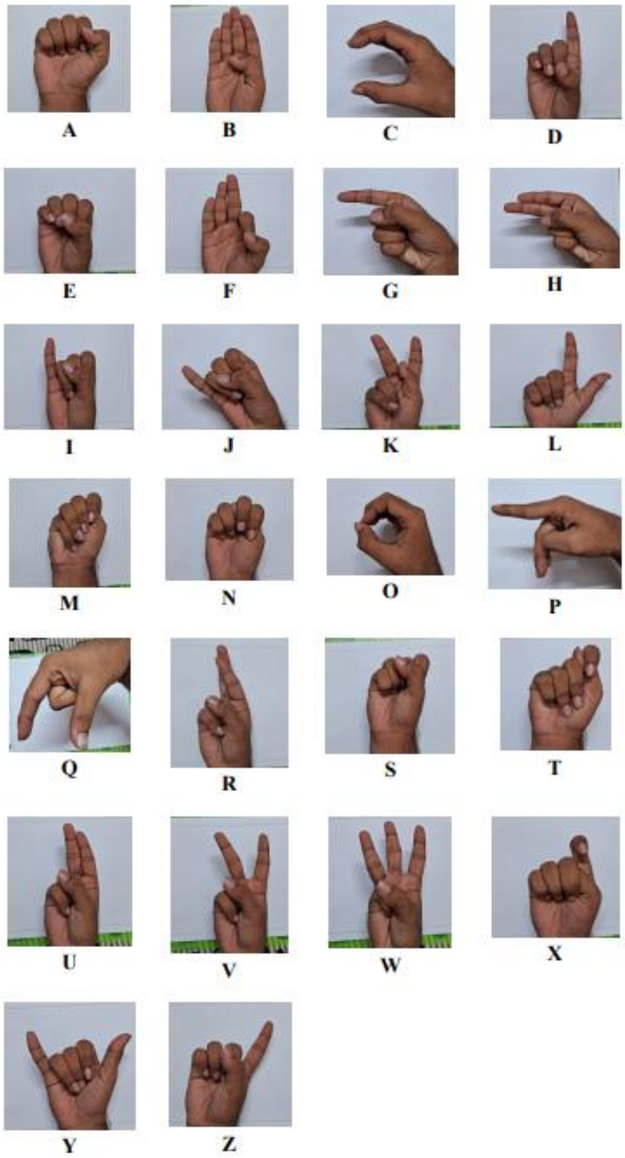

Fig.1: Guide for American Sign Language

II. LITERATURE SURVEY

Kambhampati Sai Sindhu et al.[6] Sign Language Translation (SLT) systems, with an emphasis on transforming sign language into text/speech and vice versa. It presents a model called a "reversible CNN" for precise recognition of sign language, while also reconstructing original inputs for better analysis. The research highlights the importance of having varied datasets, especially for Indian Sign Language, and explores the challenges of converting the distinct grammatical patterns of sign languages. The outcomes of the study emphasize the potential of the model, while also recognizing areas that need Improvement, such as mitigating biases and increasing the variety of data sets used. Mohamed Mohandes et al. [7] Arabic Sign Language using images, with a focus on alphabet, single word, and continuous recognition. It examines different techniques like support vector machines, neural networks, and Hidden Markov Models, emphasizing their precision and obstacles in practical use. The paper also highlights the distinct features of Arabic sign language and outlines potential research areas for enhancing ArSLR systems.

Xuebin Xu et al.[8] sign language recognition technology by incorporating sophisticated linguistic modeling. Rather than converting GLOSS to text directly, the method incorporates a text correction module. Initially, the module utilizes a system for recognizing sign language to create initial predictions, and subsequently enhances these predictions by adjusting the GLOSS sequence. This revised GLOSS helps in achieving a more precise final recognition outcome. Significant improvements in recognition accuracy were demonstrated by evaluating the framework with the RWTH-PHOENIX-Weather-2014-T and CSL datasets

Rufus reno et al. [9] The main aim of this project is to develop a complex system that can understand ASL gestures by combining ASL gesture recognition with audio feedback. The primary objective is to correctly identify and understand both still and in-motion ASL gestures to improve communication for the deaf and hard of hearing. The system uses machine learning techniques to enhance engagement and usability, concentrating on the unique linguistic and expressive features of ASL. Sivamohan s et al. [10] A innovation that employments CNNs to analyze hand signals in ISL live. The innovation captures hand developments in live video and deciphers them into discourse, bridging the communication hole between sign dialect clients and those who don't know sign dialect. It has the capacity to precisely recognize different ISL motions with a tall accuracy rate of 97.85%, illustrating its strength in different foundations and lighting conditions. Mechanical headways advance inclusivity by empowering smooth communication between individuals with diverse dialect and hearing capacities, making a difference to break down communication boundaries.

Arun singh et al. [11] CNNs within the ISLRS are utilized for recognizing energetic sign dialect motions displayed in video clips. The framework, made to upgrade hard of hearing community communication, come to a preparing exactness level of 70%. Distinctive forms like planning information, extricating highlights, and coordinating designs are combined to permit for real-time acknowledgment and move forward interaction between humans and computers. The extend points to shut the communication boundary and help with instructive advance for individuals who have hearing challenges. Challenges comprise of collecting a assortment of information choosing the leading show for exact recognizable proof. Hira Hameed et al. [12] Contact-free BSL acknowledgment framework utilizing radar innovation and profound learning calculations. By analyzing radar-based micro-Doppler marks of hand motions with the VGG16 show, the framework is able to precisely identify six feelings - perplexity, discouragement, joy, scorn, depression, and pity - with a 93.33% victory rate. This approach addresses concerns almost security and lighting issues related with conventional camera-based frameworks, improving the location of moving sign dialect signals. Jeet Bibnath et al. [13] Utilizing computer vision and profound learning strategies to recognize American Sign Dialect (ASL) signals in real-time. The framework successfully tracks and comprehends ASL developments, changing over them into composed or talked words utilizing Python, OpenCV, MediaPipe All encompassing, and LSTM systems. The point is to decrease communication impediments between ASL clients and non-users with tall exactness and negligible slack time. Its capability to work with NLP for two-way communication illustrates the potential to move forward openness and inclusivity since of its quality.

Sneha Sharma et al. [14] a sign dialect acknowledgment framework for Indian Sign Dialect (ISL) that utilizes YOLOv4 to interpret motions into content in real-time. After altering a pre-existing organize with information enlargement on a bigger ISL dataset, the framework comes to 98.4% cruel normal accuracy (mAP). This real-time arrangement with tall precision improves communication for those with hearing and discourse challenges in down to earth scenarios.

Wuyang Qin et al. [15] utilized Video Transformer Net (VTN) to create a sign language interpretation organize that's lightweight and outlined to address challenges in vision-based sign dialect acknowledgment such as restricted information and moderate preparing speed. The CSL_BS dataset was made for this reason, with VTN illustrating superior comes about in precision and speed compared to the Expanded 3D (I3D) demonstrate for recognizing confined and nonstop sign dialect. VTN is able to realize 87.9curacy in recognizing person signs and upgrades acknowledgment speed by 46.8%. On the other hand, VTN_seq2seq accomplishes a 73.5curacy in deciphering nonstop sign dialect and forms speedier than I3D.

III. METHODOLOGY

A precise approach is pivotal for the effective execution of a solid and compelling framework in your sign language recognition project utilizing Convolutional Neural Systems (CNN). This approach comprises of three primary stages:gathering dataset from Kaggle, preparing models with TensorFlow Lite and OpenCV, and creating an Android app in Android Studio using Java.

A. Collection of dataset

The favorable outcome of a machine learning project is based on the data's quality and amount. A dataset for a sign language recognition framework as a rule incorporates pictures outlines that appear distinctive motions related to different letter sets or words in sign dialect. Kaggle, a well-known stage for competitions and datasets in information science, gives a run of sign language datasets that are open source to the public. Firstly, you wish to discover a dataset that meets the necessities of our project. One well known choice may well be a dataset counting labeled pictures of hand signals that symbolize different letters. The information is ordinarily isolated into three parts: preparing data, training data, and testing data. The CNN demonstration is prepared with the training set, hyperparameters are tuned with the validation set, and final evaluation is done utilizing the testing set. It is imperative to preprocess the dataset, which seems to incorporate standardizing picture sizes, normalizing pixel values, and keeping up reliable names.

B. Model training

The most important of the project is making and instructing a CNN demonstration particularly outlined to recognize hand signals in sign language. CNNs are exceptionally great at recognizing pictures since they can capture spatial progressions in pictures utilizing convolutions and pooling layers. TensorFlow, an open-source machine learning system, is the favored choice for this extent of its solid backing for profound learning models and its capacity to work with TensorFlow Lite for versatile arrangement.

The introductory stage includes choosing or making a CNN structure appropriate for the work. You've got the alternative to form a personalized CNN or utilize a pre-trained show such as MobileNet, which is both lightweight and appropriate for versatile apps. Fine-tune pre-trained models on your dataset to utilize exchange learning, which brings down preparing time and improves precision.

Preparing the demonstration consists of contributing the preprocessed pictures into the CNN and altering the model's parameters to decrease the classification blunder. Utilizing GPU speeding up with TensorFlow can incredibly increase the speed of preparation due to the tall computational request of the method. Amid the preparing period, it is imperative to track estimations like precision and misfortune on both the preparing and approval information sets in order to affirm that the show is learning proficiently without encountering overfitting. After effective preparing and assembly execution necessities, the demonstration must be changed into a TensorFlow Lite organize. TensorFlow Lite is made for running machine learning models on portable gadgets, giving improvements to diminish demonstration measure and upgrade induction speed, imperative for applications like sign dialect acknowledgment.

OpenCV, a computer vision library that's open-source, serves as a complementary component amid this arrange. OpenCV is appropriate for different picture handling purposes like snatching live video outlines, pre-processing images for input into CNN, and overseeing post-processing obligations like laying out signals with bounding boxes. Coordination OpenCV with TensorFlow Lite empowers the creation of a pipeline that can productively handle video input, execute the acknowledgment demonstrate, and show the recognized signals in real-time.

C. Android Application Development

The final step of this project is to making an Android app that utilizes a pre-existing Convolutional Neural Network (CNN) demonstrate to distinguish and track sign language motions in real-time. Android Studio serves as the most IDE for Android app advancement, advertising essential tools and help. The method of advancement starts with the creation of a modern Android project in Android Studio. The project imports the prepared CNN show that has been changed over to TensorFlow Lite format as an asset in the android studio. TensorFlow Lite is chosen for its capability to execute machine learning models on versatile gadgets with small computational burden, empowering speedier deduction. The venture joins OpenCV libraries for taking care of picture handling assignments. The app include capturing signs from the camera, analyzing each outline with OpenCV to recognize and partitioned hand developments, and sending the prepared outlines to the TensorFlow Lite demonstrate for recognizing sign dialect. The application translates the developments and gives moment reaction by appearing the distinguished motion on the screen. The app's client interface (UI) is made to be direct and user-friendly, empowering simple client interaction with the framework. The interface appears a real-time bolster from the camera, as well as visual signals appearing the identified motion. Additional alternatives can comprise of buttons to examined the written substance utilizing signs, to delete the current discovery, embed spaces with space button or adjust camera settings for progressed usefulness. These characteristics ensure that the application can effortlessly alter to different necessities for sign dialect acknowledgment.

It is fundamental to perform testing and optimization amid this organize to ensure the app capacities easily on a run of Android gadgets, such as smartphones and tablets with changing equipment setups. Assessment of execution measurements like deduction time, exactness of signal acknowledgment, and utilization of assets (battery, CPU, and memory) is done to ensure the productive working of the app on different gadgets with shifting execution capabilities. Successful optimization ensures the application's speed and responsiveness, indeed on low-powered gadgets.

IV. RESULT AND DISCUSSION

Fig. 2: Home page

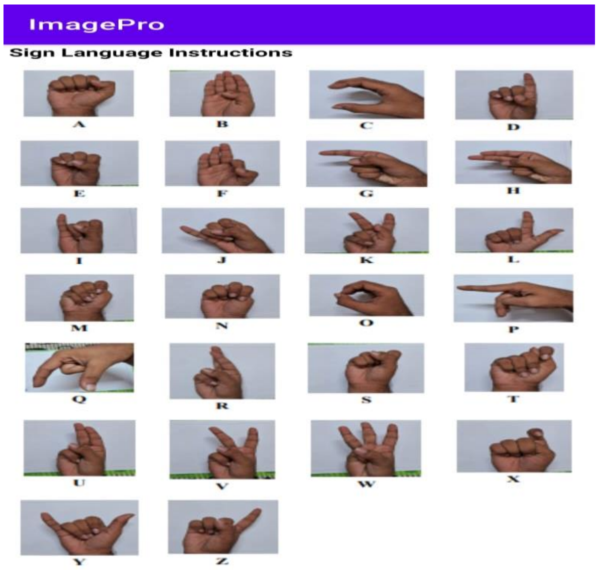

Fig. 3: Sign language instructions in the developed application

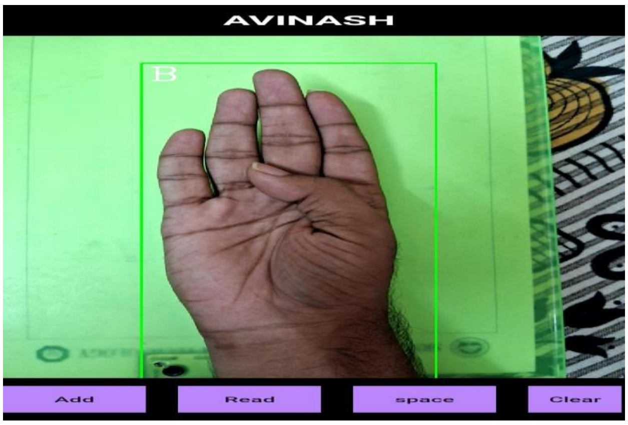

Fig. 4: Sign language recognition

Using Convolutional Neural Networks (CNN), the project effectively developed a system that recognizes sign language in real-time. The system can effectively recognize gestures, demonstrated by accurately identifying the letter "B." Voice output integration enhances this feature by offering auditory feedback, improving accessibility for communication with people who are non-verbal or have hearing impairments. The interface of the application is user-friendly, providing choices such as "Add," "Read," "Space," and "Clear," enabling smooth interaction for users. The system is designed to work well on mobile devices by using TensorFlow Lite, which is important for practical use due to its efficiency and ease of portability. Nevertheless, though the system shows good performance with the letter "B," it is crucial to test its precision with all gestures under different conditions like various lighting or hand positions. Future research may emphasize enhancing the model to achieve faster processing and improved accuracy in various settings, possibly through refining the CNN structure or employing more advanced optimization methods. In general, the project shows a major advancement in developing a practical sign language recognition tool that works well on mobile devices.

V. CONCLUSION

The project effectively created a real-time sign language recognition system using CNNs, emphasizing accessibility and practical use. The system effectively recognizes sign language gestures, as shown by its accurate identification of the letter "B," and incorporates voice output to aid communication for individuals who are non-verbal or have hearing impairments. The user-friendly mobile app interface utilizes TensorFlow Lite to enable efficient operations on mobile devices. Although the system is performing effectively now, additional evaluation is required with a wider variety of gestures and conditions to maintain consistent accuracy. Future enhancements could concentrate on improving the model to ensure quicker and more precise processing, enhancing its functionality in various real-world settings. In general, this project is a major accomplishment in making sign language recognition technology more available and feasible for everyday usage.

REFERENCES

[1] T. Kumaragurubaran, S. S P, M. Tharun and M. V. Devi, "Unlocking Sign Language Communication: A Deep Learning Paradigm for Overcoming Accessibility Challenges," 2024 IC3IoT, Chennai, India, 2024, pp. 1-6, doi: 10.1109/IC3IoT60841.2024.10550356.

[2] S. S Pandi, J. Sutha, P. Shanmugam and P. Kumar, "A Novel Approach for Sign Language Detection Using Learning Algorithms," 2024 ICCIGST, Vijayawada, India, 2024, pp. 1-5, doi: 10.1109/ICCIGST60741.2024.10717614.

[3] S. S P, D. Paulraj, D. Mithun and N. K. A, "Object Detection using Learning Algorithm and IoT," 2023 RMKMATE, Chennai, India, 2023, pp. 1-6, doi: 10.1109/RMKMATE59243.2023.10369005

[4] S. S. P, D. Paulraj, N. K. A and N. K. B, "A Novel Approach for Image Background Elimination," 2023 RMKMATE, Chennai, India, 2023, pp. 1-6, doi: 10.1109/RMKMATE59243.2023.10369121

[5] S. S P, K. D. Kumar, K. Raja and A. Senthilselvi, "Rice Plant Leaf Disease Classification using Machine Learning Algorithm," 2024 InCACCT, Gharuan, India, 2024, pp. 59-63, doi: 10.1109/InCACCT61598.2024.10551249.

[6] Hridoy Adhikari, Md. Sakib Bin Jahangir, Israt Jahan, Md. Solaiman Mia, Md. Riad Hassan "A Sign Language Recognition System for Helping Disabled People" 2023 5th International Conference on Sustainable Technologies.

[7] Mohamed Mohandes, Junzhao Liu, Mohamed Deriche "A Survey of Image-Based Arabic Sign Language Recognition" 978-1-4799-3866-7/14/$31.00 ©2024 IEEE.

[8] Xuebin Xu, Jun Fu "A two-stage sign language recognition method focusing on the semantic features of label text" 2024 20th CSI International Symposium on Artificial Intelligence and Signal Processing (AISP).

[9] Rufus Reno, R.Chitra "Gesture Talk: Multi Model Sign Language Recognition system" 2024 international conference recent advances in electrical, electronics, ubiquitous communication and computational intelligence.

[10] Sivamohan S, Anslam Sibi S, Divakar "Hand Gesture Recognition and Translation for International Sign Language Communication" 2024 2nd International Conference on Advancement in Computation & Computer Technologies (InCACCT).

[11] Arun Singh, Ankita Wadhawan, Manik Rakhra, Usha Mittal, Ahmed Al Ahdal, Shambhu Kumar Jha "Indian Sign Language Recognition System for Dynamic Signs" 2022 ICRITO.

[12] Hira Hameed, Muhammad Usman, Muhammad Zakir Khan, Amir Hussain, "Privacy-Preserving British Sign Language Recognition Using Deep Learning" 2022 44th EMBC.

[13] Jeet Debnath, Praveen Joe I R "Real-time Gesture Based Sign Language Recognition System" 2024 ADICS.

[14] Sneha Sharma, R Sreemathy, Mousami Turuk, , Soumya Khurana "Real-Time Word Level Sign Language Recognition Using YOLOv4" 2022 INCOFT Karnataka, India. Nov 25-27, 2022.

[15] Wuyang Qin, Xue Mei, Yuming Chen, , Shi Hu "Sign Language Recognition and Translation Method based on VTN" 2021 International Conference on Digital Society and Intelligent Systems (DSInS).