ISSN: 2582-7219

International Journal of Multidisciplinary Research in Science, Engineering and Technology

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Impact Factor: 8.206

Volume 8, Issue 11, November 2025

© 2025 IJMRSET | Volume 8, Issue 11, November 2025| DOI:10.15680/IJMRSET.2025.0811004

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Real-Time Sign Language Conversion into Text and Audio

Prof. Dr. Aniruddha S Rumale[1], Prof. Dr. Dipak D Bage[2], Vishakha Sarnaik[3], Anushka Gosavi[4], Aryan Chapne[5], Pratap Tupe[6]

Assistant Professor, Dept. of IT, Sandip Institute of Technology and Research Centre, Nashik, India[1]

Assistant Professor & Head, Dept. of IT, Sandip Institute of Technology and Research Centre, Nashik, India[2]

PG Students, Dept. of IT, Sandip Institute of Technology and Research Centre, Nashik, India[3, 4, 5, 6]

ABSTRACT: Sign Language Recognition is a majorly growing sphere of technology and research. Sign Languages are mostly used for communication amongst the deaf and dumb. This paper presents a novel approach for translation of sign language action analysis, recognition and conversion of it into a text and audio speech. The proposed system consists of six modules such as: data acquisition, pre-processing of data, feature extraction, model training, sign recognition and text or audio output. The system uses the Convolutional Neural Network (CNN) modules as it plays a **vital role** in sign language conversion into text because the task relies on recognizing **visual patterns** in hand signs, finger positions, and movements.

KEYWORDS: Image Processing, Convolutional Neural Networks (CNN), Sign Recognition, Sign language conversion, Sign language to audio, Sign language to text, Long Short-Term Memory (LSTM), MediaPipe, OpenCV, Google Text-to-Speech (GTTS)

I. INTRODUCTION

Communication is a fundamental human need, essential for building interpersonal relationships in a society in our daily lives. Effective communication is essential in all aspects of life, and for every individual. A communication barrier arises for the individuals with hearing and speech ailments. Sign language is an expressive and natural way of communication for the impaired people. It is the only form of communication for deaf and mute people. However, it is still difficult to interact with such people without any aid as public are not eager to learn the sign language. People without hearing impairments, tend to overlook learning sign language, thus limiting the ability to communicate with the deaf individuals. As a result the deaf and mute become isolated from the society. Nevertheless, the development in technologies and more research in this field makes it possible to overcome the obstacles and close the communication gap. Hence, a system or a software needs to be programmed in such a way that it can convert the sign language into text and audio.

There are about 300 different sign languages that are being used around the globe. The reason behind it being individuals from various ethnic groups of the world naturally creating their own sign language. ISL abbreviated as the Indian Sign Language is derived from both the British Sign Language (BSL) and the French Sign Language (FSL). The Indian Sign Language makes use of both of the hands to represent any alphabet or sign. While many of the researchers studying in this field focus on the recognition of American Sign Language (ASL) because most of it's signs are made with the help of a single hand and therefore complexity is less. Another reason why ASL is most widely used is because it already consists of a standard database that is easily available for use. The Indian Sign Language when compared to ASL uses both hands making it more complex for recognition. However, there have been new initiatives taken to standardize the Indian Sign Language. The Indian Sign Language vocabulary contains thousands of sign just like words. Thus, it is very easy to get multiple different signs mixed up, which ends up leading to miscommunication.

The ISL is an independent language. It does not depend on any verbal language be it English or any regional language. English alphabets are extensively used by the deaf people in India rather than finger spelling of other spoken languages like Hindi, Tamil and Telugu. Some of the educational institutes which are implementing the signing systems approach under total communication also making use of manual alphabet of the regional languages. Because of

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

the extensive access towards the international sign language and also due to international conferences and social media, ISL has started to integrate fewer signs from other sign languages too.

The proposed system is developed to recognize various alphabets and hand signs of Indian Sign Language that will give accurate text conversions of input signs. This system is designed to be simple, systematic and accurate. The process of the system is very simple and efficient. The system will first capture hand signs through a capturing device such as a webcam. The captured signs are further processed and the images is segmented. Then, the key features are extracted from the processed image. Further, a comparison of the extracted features with the entries in the testing database is made. This process is done in order to calculate and evaluate the accuracy of sign recognition system. At the end a text representation of the sign is expected as an output. The converted text can also be converted to speech. Convolutional Neural Network (CNN) modules are utilized with the purpose of segmenting, processing and recognizing the sign. This system will facilitate communication between deaf-mute and vocal people with efficient use of algorithms and accurate categorization of hand signs. The system aims to emit the communication barriers between the hearing and the non-hearing individuals.

II. RELATED WORKS

Sign languages (e.g., ASL – American Sign Language, ISL – Indian Sign Language, BSL – British Sign Language) are complete natural languages consisting their own grammar, syntax, and semantics. It serves as the basic and fundamental mode of communication for the individuals who are unfortunately not able to hear. Over the years, one of the major challenges faced by the this community has been interacting with the vocal persons. Traditionally, human interpreters bridged this gap as an aid, but human interpreters are not always available. Before the introduction of computer vision and AI, researchers heavily relied on data gloves and sensors to capture hand signs, positions, and orientations. These were connected to systems that would then translate the signs into words or text.

From the early 2000s advancements in computer vision and technologies allowed the researchers to access the cameras instead of gloves. The earlier systems developed then, used image processing (edge detection, skin color segmentation, background subtraction) to identify hand shapes. However, they use to often struggle with complex movements, background noise and lighting variations.

During the 2010s, the rapid advancements in the field of machine learning and deep learning, especially the Convolutional Neural Network (CNN), led to significant improvement in the accuracy of sign recognition. The researches started to build large datasets of sign signs to train models for recognition. The systems also started to handle dynamic signs. It was noticed that conversion is not just sign recognition, but also the sign must be translated into grammatically correct text.

Modern technologies include 3D Sensors, Depth Cameras, Smartphones and AI apps that improve hand movement capturing and real time translation of sign language. There are also efforts being made to support multiple sign languages since each country has it's own. The education, healthcare, and also government services are beginning to adopt such technologies to improve accessibility.

Yet there are challenges that are still faced such as variability of sign languages, complexity of signs involving hand shape, orientation, motion, facial expressions and body postures. And also the scarcity of datasets i.e. the lack of large and labeled sign language datasets.

III. DESIGN ISSUES

Several critical design challenges are faced for the development of AI-Based Real-Time Sign Language into Text and Speech Conversion which may affect the accuracy and efficiency of the system.

A. Sign Representation

A critical issue faced by the development of sign language conversion systems is accurate and precise representation of the hand signs of the user. Same signs might convey a different meaning depending on the context in which it is used. For example, the signs for "Where are you from ?" and "What is your name ?"

© 2025 IJMRSET | Volume 8, Issue 11, November 2025| DOI:10.15680/IJMRSET.2025.0811004

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

B. Real-Time Processing

It is difficult to maintain low latency while achieving high accuracy. It is necessary for the system to be optimized to run efficiently on various devices while maintaining high recognition speed.

C. Handling Transitional Movements Between Signs

The system should be able to distinguish between meaningful signs and transitional movements. It must avoid capturing signs during transitions between two different signs, as these movements do not contain meaning and may otherwise lead to incorrect recognition.

D. Variability in Sign Languages

Different countries use different sign languages, even within one language dialects and individual styles exist. Thus, making it difficult to build an universal system for all users to use.

IV. METHODOLOGY

The methodology of AI-based real-time sign language conversion initiates with data acquisition, where signs using cameras or sensors. The acquired data is pre-processed in order to get rid of noise, enhance visual clarity and normalize the size of the sign. In the next step, feature extraction techniques are applied for detection. These extracted features are used in the model training phase, where machine learning or deep learning algorithms learn to classify different signs accurately. Once trained, the system can recognize signs in real time**.** Henceforth, the signs recognized by the system are converted into text or speech. It enables effortless communication between signers and non-signers.

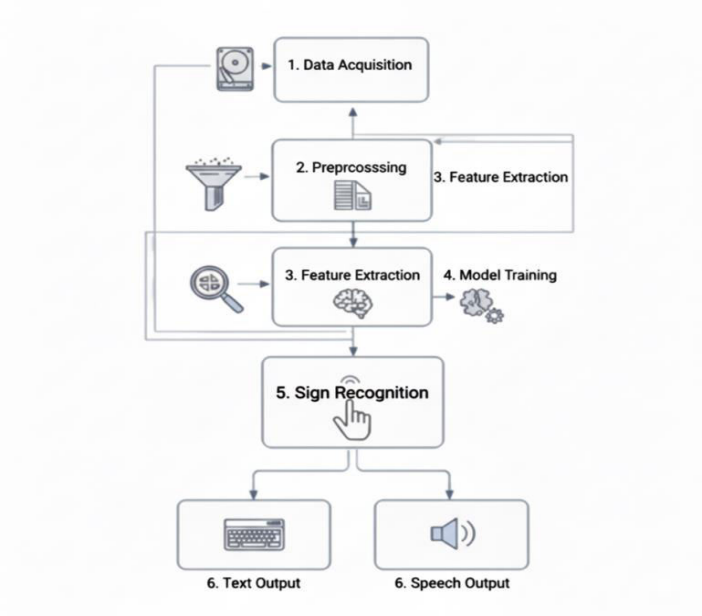

Fig.1. Sign Recognition System Work Flow

The major steps that are to be followed in the process of sign language conversion are data acquisition, image segmentation, pre-processing, data modelling, sign recognition and text - audio conversion.

A. Data Acquisition

The main objective behind data acquisition is to gather the data in the form of static images or video recordings. The webcams and sensors make sure to capture the hand signs or movements. The public datasets provide with the ASL Alphabets or Indian Sign Language Datasets. CNNs usually require large and labeled datasets for training. Also variability in hand shapes, signer styles, lighting and background sometimes makes data acquisition critical. Each frame/image captured acts as an input to CNN for feature learning. OpenCV serves as an open source library which

© 2025 IJMRSET | Volume 8, Issue 11, November 2025| DOI:10.15680/IJMRSET.2025.0811004

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

offers a comprehensive set of computer vision and machine learning functions which are used to process image frames in real-time and for the detection of signs using CNN algorithm. It includes modules which can be used for dynamic and static image processing, 3D reconstruction. The proposed system will make use of the OpenCV to acquire the hand signs by accessing the web camera and detect the patterns.

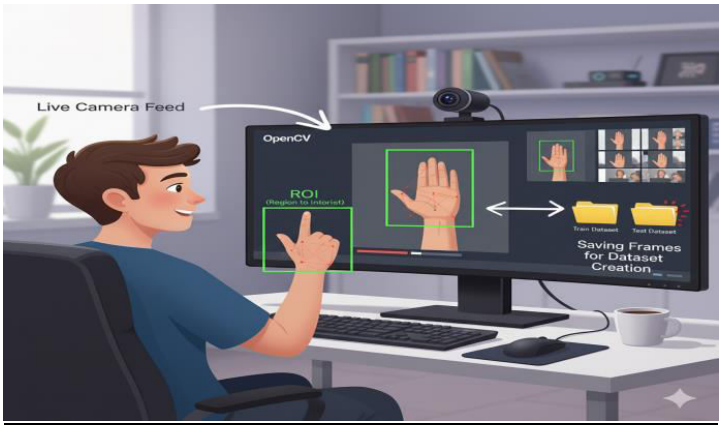

Fig.2. Data Aquisition

B. Image Segmentation

In image segmentation the hand and relevant body parts are separated from the background to reduce noise. It faces various challenges in doing so due to different lighting conditions, background clutter and overlapping objects. However, the system makes use of methods like skin-color detection and background subtraction. The system also makes use of CNN based segmentation modules such as Mask R-CNN, that learns spatial features of hand regions and provide a clean and focused input to later CNN layers for sign recognition. The proposed system makes use of MediaPipe to segment the detected sign. MediaPipe is cross-platform framework which serves the purpose of building multimodal applied machine learning. It helps to separate the hand section from the remaining or unwanted background. As a result, there is a reduction in background noise and improvement in the image quality.

Fig.3. Image Segmentation

C. Pre-processing

This stage standardizes and enhances the input images for CNN. It resizes all images to a fixed size, scales the pixel values and removes noise by filtering or background suppression. Pre-processing ensures input data fit's the CNN's fixed input dimension and augmentation ensures CNN doesn't overfit to limited signing samples. The essential preprocessing steps are resizing of the image, conversion color space RGB to grayscale or HSV (Hue, Saturation,

© 2025 IJMRSET | Volume 8, Issue 11, November 2025| DOI:10.15680/IJMRSET.2025.0811004

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Value) for better skin tone detection which enhances image quality and prepares it for future analysis. It uses KNN algorithm for background subtraction which results into isolation of signer`s hand from the background.

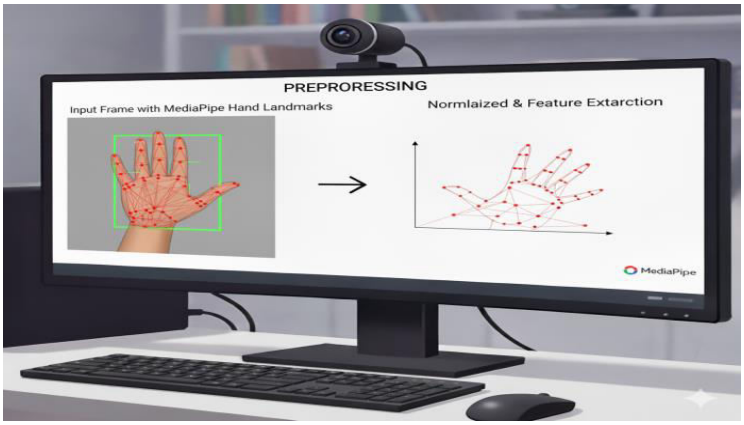

Fig.4. Preprocessing

D. Data Modelling

The system designs and trains a CNN model that can classify signs. It makes use of the complete architecture of the CNN. The Convolutional Layers detect low-level (edges, curves), mid-level (finger orientation) and high-level features (hand shapes and signs). The Fully Connected Layers map extracted features into sign classes. And finally Activation Functions (ReLU, Softmax) introduce non-linearity and produce class probabilities. LSTM i.e. Long Short-Term Memory is utilized in the proposed approach to capture temporal relationships and translate the sequences of signs into corresponding text or speech. The main purpose is to train the CNN model to classify the extracted features.

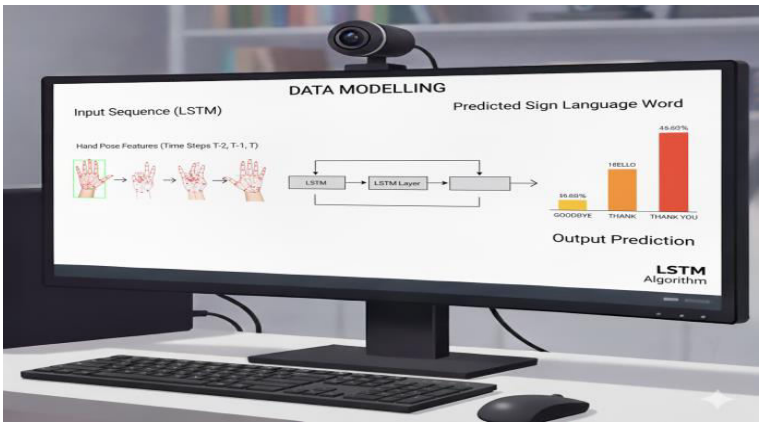

Fig.5. Data Modelling

E. Sign Recognition (Sign Classification)

The system in this stage identifies which sign the user is making. For this to happen the CNN outputs a probability distribution over sign classes. For example, if an image is provided as input, the CNN outputs ("HELLO": 0.92, "THANK YOU": 0.05, "YES": 0.03). The class which is able to achieve the highest probability is selected i.e. HELLO. The CNN module provides a robust classification accuracy by learning spatial features.

© 2025 IJMRSET | Volume 8, Issue 11, November 2025| DOI:10.15680/IJMRSET.2025.0811004

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

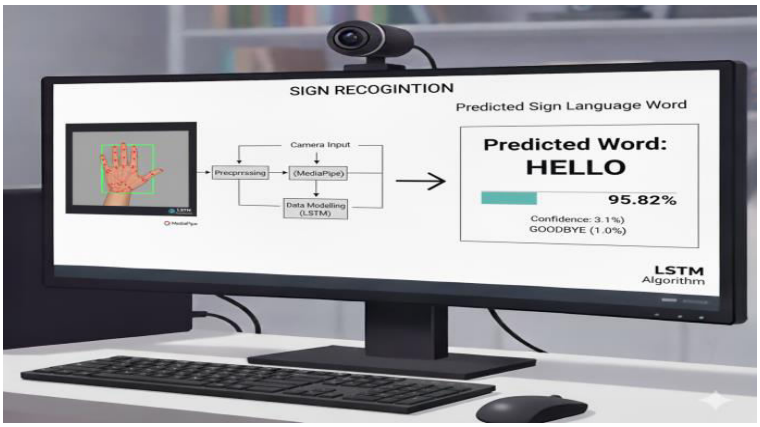

Fig.6. Sign Recognition

F. Text and Audio Output

Finally, after classification the system displays the recognized sign label as text. In case of continuous signing, the recognized signs are sequenced to words. The CNN module ensures that the sentences formed are grammatically correct. Also for audio output the GTTS i.e. Google Text To Speech is used. GTTS system synthesize natural speech.

V. EXPECTED RESULTS

The major objective of the sign language conversion system is to ensure that the communication is made easier and smoother by interpreting the signs, and then converting the signs into text and speech. The sign once identified is displayed on the screen, while the text-to-speech engine produces audio output. The process bridges the communication gap between hearing and hard of hearing individuals and takes down the communication barriers by allowing for real-time, two-way communication. The expected result of the sign language conversion system is the conversion of a person's hand signs into a form that can be understood by those who don't understand sign language. The model is converting text output as the recognized signs are displayed as a word or phrase on a screen and speech output as the recognized signs are converted into spoken words or sentences using text-to-speech technology. Because of this process, the communication barriers between the hearing and hard of hearing individuals can be taken down as it allows for real-time, two-way communication for the non-hearing community. In the Real-Time Translation the system is able to process signs as they are performed and provide a near-instantaneous translation. This is crucial for natural conversation. The model is expected to have a high recognition rate, correctly identifying signs with minimal errors. Many challenges arise, specifically for continuous, fluid signing and non-manual features like facial expressions for interpretation. Using Natural Language Generation, allows these systems to produce outputs which are properly structured and grammatically correct. This is especially important for translating between sign languages and spoken languages.

VI. ANALYSIS

The CNN model is able to automatically learn critical spatial features from images of hand signs, reducing the need for handcrafted descriptors. The outputs of CNN layers is fed to fully connected layers, yielding a probability distribution over sign classes (e.g., letters, words). For dynamic signing, hybrid architectures, like CNN-SA-LSTM (Convolutional Neural Network - Self-Attention - Long Short-Term Memory), are used to handle temporal dependencies and word/sentence formation. Recognized sign labels are converted to text strings, which are then passed to text-to-speech modules for audio conversion.

VII. CONCLUSION

Most existing research works focus particularly on the classification and recognition of static signs of ISL by making the use images or video recordings that have been recorded under controlled conditions. The application of CNN algorithm for sign recognition reduces the dimensionality, thereby reducing noise and improving accuracy. The system

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

utilizes various principles and concepts of image processing and fundamental properties. CNN algorithms utilization has led to successful and accurate recognition of sign. Every individual holds significance in the society, remembering this fact, the aim if the system is to try to ensure the inclusion of hearing impaired people in our day to day life and live together.

REFERENCES

[1] Jayanthi P., P. R. K. S. Bhama, and B. Madhubalasri, "Sign Language Recognition using Deep CNN with Normalised Keyframe Extraction and Prediction using LSTM," Journal of Scientific & Industrial Research (JSIR), 2023.
[2] K. L. Cheng, Z. Yang, Q. Chen, and Y.-W. Tai, "Fully Convolutional Networks for Continuous Sign Language Recognition," arXiv preprint arXiv:2007.12402, 2020.
[3] Authors, "An Adam-based CNN and LSTM approach for sign language recognition in real time for deaf people," Bulletin of Electrical Engineering and Informatics, 2024.
[4] V. G. Velmathi and K. Goyal, "Indian Sign Language Recognition Using Mediapipe Holistic," arXiv preprint arXiv:2304.10256, 2023
[5] S. Dasgupta, J. Lloret Mauri, J. H. Abawajy, et al., "A Study of CNN Architectures over Two Hand Indian Sign Language Dataset," in Applied Soft Computing and Communication Networks, Springer, 2020.
[6] F. Zhang, V. Bazarevsky, A. Vakunov, A. Tkachenka, G. Sung, C.-L. Chang, and M. Grundmann, "MediaPipe Hands: On-device Real-time Hand Tracking," arXiv preprint arXiv:2006.10214, 2020.
[7] Authors, "Isolated Video-Based Sign Language Recognition Using a Hybrid CNN-LSTM Framework Based on Attention Mechanism," Electronics, vol. 13, no. 7, p. 1229, 2024.
[8] Y. Yaseen, O.-J. Kwon, J. Kim, S. Jamil, J. Lee, and F. Ullah, "Next-Gen Dynamic Hand Gesture Recognition: MediaPipe, Inception-v3 and LSTM-Based Enhanced Deep Learning Model," Electronics, vol. 13, no. 16, p. 3233, 2024.
[9] Authors, "Sign Language Recognition Based on Geometric Features Using Deep Learning," Jurnal Nasional Pendidikan Teknik Informatika (JANAPATI), 2024.
[10] H.-H. Li and C.-C. Hsieh, "Dynamic Hand Gesture Recognition Using MediaPipe and Transformer," Engineering Proceedings, vol. 108, no. 1, p. 22, 2025.
[11] Authors, "Emergency Sign Language Recognition from CNN and LSTM models," International Journal of Advances in Intelligent Informatics, vol. 10, no. 1, pp. 64–78, 2024.
[12] M. Author et al., "An optimized automated recognition of infant sign language using enhanced CNN and deep LSTM," Multimedia Tools and Applications, 2023.
[13] Authors, "Chinese Sign Language Recognition Based on Two-Stream CNN and LSTM Network," International Journal of Advanced Networking and Applications, 2023.
[14] R. Li and L. Meng, "Multi-View Spatial-Temporal Network for Continuous Sign Language Recognition," arXiv preprint arXiv:2204.08747, 2022.
[15] Y. C. Bilge, R. G. Cinbis, and N. Ikizler-Cinbis, "Towards Zero-shot Sign Language Recognition," arXiv preprint arXiv:2201.05914, 2022.
[16] G. Sung, V. Bazarevsky, F. Zhang, and M. Grundmann, "On-device Real-time Hand Gesture Recognition," arXiv preprint arXiv:2111.00038, 2021.
[17] Indriani, M. Harris, and A. S. Agoes, "Applying Hand Gesture Recognition for User Guide Application Using MediaPipe," in Proc. ISSAT, 2021.
[18] A. Authors, "An Integrated CNN-LSTM Model for Bangla Lexical Sign Language Recognition," in Proc. Int. Conf. on Trends in Computational and Cognitive Engineering, Springer, 2020.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH

IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | ijmrset@gmail.com |

www.ijmrset.com