MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 1, 2026, Pages: 2876-2883

Real-Time Sign Language Recognition and Translation

CH, Bhupati[1*]
assistant professor, Department of IoT,
Koneru lakshmaiah education foundation,
bhupati@kluniversity.in

Garre Gayatri[2]
Department of IoT,
Koneru lakshmaiah education foundation,
2200100025@kluniversity.in

Vudata Harika[3]
Department of IoT,
Koneru lakshmaiah education foundation,
2200100076@kluniversity.in

Vemula Thanuja[4]
Department of IoT,
Koneru lakshmaiah education foundation,
2200100009@kluniversity.in

Abstract:

Communication is the primary way to interact with each other.It is difficult for people suffering with hearing impaired individuals for communication.To help close this gap, this paper introduces a model which works in real time that makes communication and translation between signers and non- signers smooth and natural. Instead of relying on expensive gloves or sensors, this system simply uses a camera and a deep learning model built with TensorFlow and OpenCV. The model, based on SSD MobileNet V2, can quickly recognize different hand gestures both still and moving — from a live video feed.Every frame of the video is processed in real time, and the identified gesture is immediately converted to text and voiced through a Text-to- Speech (TTS) module.

Keywords: Sign Language Recognition, Deep Learning, TensorFlow, SSD MobileNet V2, OpenCV, Text-to-Speech, Real-Time Translation, Indian Sign Language (ISL), Human with computer Interaction.

1. INTRODUCTION

Communication is the core of human interaction, but millions of people worldwide who are hearing-impaired or speech- impaired still experience barriers in communicating their ideas to people who are not conversant with sign language. Sign languages, which are made up of organized hand gestures, movements, and facial expressions, are the main tools of communication for the hearing-impaired or speech-impaired. But since most people are not conversant with sign language, this has resulted in a major communication barrier that has led to social isolation and limited access to vital services. To solve this problem, the project named "Sign Language Recognition and Translation" presents a real-time vision- based system that can recognize sign language gestures and translate them into both text and speech. Using computer vision and deep learning, the system facilitates smooth two- way communication between the hearing-impaired and the hearing, without the need for wearable sensors or the use of special gloves. Developed using TensorFlow and OpenCV, the system uses an SSD MobileNet V2 model that has been trained on a carefully selected dataset of hand gestures for efficient real-time recognition. After recognition, the gesture is translated into corresponding text, which is then spoken out through Text-to-Speech (TTS) synthesis. In contrast to the conventional glove system, which is expensive, restrictive, and often faced with mechanical constraints, the proposed camera system is more affordable, flexible, and easily implementable on a laptop, smartphone, or IoT-enabled platform. Through the combination of machine learning, natural language processing (NLP), and computer vision, the proposed system not only demonstrates a technological innovation but also promotes so- cial inclusion by enabling the differently abled. The system has various applications in education, healthcare, public services, and smart environments, hence promoting an inclusive digital society.

2. LITERATURE REVIEW

In the early stages of research, traditional feature-based ap- proaches dominated the field, while current developments leverage the power of deep learning and pose estimation to enable real-time and accurate gesture recognition and trans- lation. Sign language recognition has undergone substantial development in the past decade, shifting from animation- based approaches to highly advanced deep learning-based translation systems. The literature reviewed above presents a thorough background understanding of the developments in sign language recognition and translation. Abhishek and Sumanathilaka [1] presented an end-to-end sign language recognition system with a focus on developing energy-efficient deep learning for real-time systems. Their system combines optimized CNN-based feature extraction, sequence modeling, and model compression strategies like pruning and quan- tization to decrease computational complexity. While the main focus remains on enhancing efficiency and deployability rather than increasing linguistic capabilities, their contribution presents a vital milestone towards developing sustainable and edge-deployable SLR systems.Ahmed et al. [2] proposed one of the first interactive systems with the objective of filling the gaps in communication using 3D animated sign language, which enabled the conversion of spoken or typed messages into visual representations using signs. Although the proposed system does not specifically address gesture recognition, it remains a crucial precursor to bidirectional communication systems. Cassim et al. [3] proposed a low-cost, portable sign language to speech translation system that seeks to improve accessi- bility for people with hearing and speech impairments. The system uses flex sensors and motion sensors to detect hand gestures, which are then processed using a microcontroller- based recognition algorithm to interpret the signs and produce corresponding speech output. The system's emphasis on porta- bility and real-time translation makes it a viable assistive tech- nology solution. Although the system relies on hardware-based gesture recognition and not machine learning algorithms, the system is a significant contribution to the development of sign- to-speech systems that are inclusive. Chatterjee [4] addressed the problem of computational complexity in the processing of sign language videos by providing approaches that can minimize delays in processing. The paper highlighted the need for real-time efficiency, which has become the trend in the development of models for light detection.Based on recognition and language modeling, Chatterjee and Abinaya [5] developed a hybrid CNN-LSTM approach that could translate word-level sign language sequences to mean- ingful English sentences. By exploiting both spatial and tem- poral information, their approach had the potential to translate continuous gesture inputs into linguistically valid outputs. Chen and Di [6] developed a lightweight intelligent sign language recognition model by optimizing the Region Convo- lutional 3D (R-C3D) framework to improve both accuracy and efficiency. Their optimized R-C3D framework comprises fea- ture optimization modules and temporal attention mechanisms to efficiently extract dynamic gesture information with lower complexity. The system has high recognition accuracy with low training time complexity and memory complexity, making it ideal for real-time processing.Even though the research contribution is mainly focused on model optimization and not multimodal fusion, it is an important step towards developing efficient, scalable, and intelligent SLR systems.

Cooper et al. [7] proposed a sub-unit-level approach for sign language recognition, breaking down complex signs into smaller units that can be reused to improve recognition accu- racy. By applying machine learning algorithms to recognize and classify these sub-units, such as hand shape, movement, and location, the approach promotes generalization across signs and can be achieved without requiring large amounts of annotated data. This approach was one of the early milestones in the development of compositional SLR systems and paved the way for future advances in hierarchical modeling of sign language recognition. Das et al.[8] proposed a sign language translation system in Bangla using a deep Long Short-Term Memory (LSTM) network to aid communication for the deaf and speech-impaired. The proposed system uses sequential gesture input from the camera and translates signs into text and speech outputs. The deep LSTM network structure is very effective in modeling the temporal dependencies in gesture sequences, thereby ensuring high recognition accuracy on var- ious Bangla sign language datasets. Although the application is limited to the Bangla linguistic community, the system illustrates the efficacy of deep recurrent networks in modeling the temporal complexity of sign language translation tasks.

Fregoso et al. [9] developed an optimization methodology for CNN architectures based on Particle Swarm Optimization (PSO) to enhance the efficiency of sign language recognition systems. The proposed method can automatically optimize the hyperparameters of CNNs, including the depth of layers, kernel size, and learning rate, to obtain optimal accuracy and efficiency. The experimental results demonstrate that the PSO- optimized CNNs perform better than the manually designed CNNs in terms of accuracy and efficiency. Although this research paper does not address the multimodal and temporal issues of SLR, it presents a reliable evolutionary optimization method that helps in the development of high-performance deep learning models for gesture recognition.

Garg et al. [10] have contributed an important resource to the field by creating a specialized dataset for sign language detection. The dataset includes variations in the background, lighting, and orientation of the hands, which helps in robust training of deep learning models.

A significant paradigm change in real-time sign recognition was brought about by the development of Google's MediaPipe library [11], which enables efficient hand and pose landmark detection. The efficiency and robustness of MediaPipe have made it a central component of contemporary vision-based gesture recognition systems. Theoretical advances in sequence modeling were achieved by Graves et al. [12], who proposed deep recurrent neural networks and bidirectional LSTMs for speech recognition. Although these ideas are not exclusive to sign language, they have since become central to continuous gesture recognition and video learning. A significant advance in the field was achieved by Koller et al. [13], who proposed a statistical model for continuous sign language recognition by multiple individuals. Their model facilitated vocabulary recognition and the feasibility of scaling systems for com- munication.

Li, Liu, and Cheung [14] analyzed the impact of intermodal transfer, which refers to the translation of meaning from one modality to another, such as text, gesture, or pictorial representation, on linguistic simplification. The research study analyzes the impact of intermodal transfer on the form and complexity of communication, which helps in understanding how humans communicate language while transferring from spoken, written, or sign language. The authors use experimen- tal linguistic analysis and modeling to analyze the patterns of simplification in intermodal translation tasks. Although this research study is not relevant to the recognition of sign language, it provides theoretical knowledge about cross-modal processing.

Paneru et al. [15] proposed an AI-based translation system that translates American Sign Language (ASL) gestures into Nepali speech by using convolutional neural networks (CNNs) for gesture recognition and text-to-speech (TTS) technology for audio output. The proposed system takes hand gestures as input in the form of images or videos, recognizes them using a deep CNN model, and then translates them into Nepali text and speech. The paper showcases the progress made in cross- lingual and multimodal translation, showing the possibility of combining vision and speech technologies for human- computer interaction. Although the size of the dataset and the variety of signs are still constraints, this work has significantly contributed to multilingual and assistive SLR translation.

Early machine learning-based solutions for ISL were inves- tigated by Reshna and Jayaraju [16], who employed skin segmentation and SVM classification for gesture recognition. Although rudimentary compared to deep learning, their solu- tion indicated that conventional image processing techniques could still provide functional results, especially in resource- limited settings. S. S. R. et al. [17] designed a sign-to-sign translation system that facilitates the translation of signs from various national sign languages using deep learning and gesture synthesis. The system is able to translate signs from one language and synthesize signs from another language, thus improving accessibility between languages. Despite the system having a limited vocabulary, it is a major milestone towards achieving the objective of global sign language compatibility. Sharma and Anand [18] have provided a comparative study of deep learning architectures for Indian Sign Language, indicating that light models like MobileNet can also work well with a small dataset. Their study confirms the effectiveness of efficient CNN models in real-time applications.

Sharma and Singh [19] provided a comprehensive review of vision-based recognition systems, from sensor-dependent systems to camera-based deep learning models. The review emphasized the accuracy improvements, usability, and scal- ability, positioning deep learning at the top of modern SLR systems. Further improving real-time translation, Shetty et al.[20] described a gesture-to-text system integrating MediaPipe landmark detection and LSTM-based temporal modeling. The proposed system showed high accuracy in different envi- ronmental conditions, proving the effectiveness of landmark- based systems. Sreemathy et al. [21] analyzed hybrid recog- nition systems through an expert system integrating rule- based systems and machine learning approaches. The pro- posed system improved error correction and provided more reliable recognition for continuous word-level signing patterns. Tripathi et al. [22] described an intelligent sign language recognition system translating gestures into real-time text output to assist speech and hearing-impaired people. Using computer vision and deep learning approaches, the system provides efficient gesture recognition and accurate translation performance sufficient for real-time translation. The study emphasizes real-time assistive communication with a focus on user accessibility. Wijaya et al. [23] assessed the quantitative performance of sign language translation through an artificial neural network (ANN) model.

In an extended form of their research, Xue et al. [24] suggested the use of improved feature fusion and attention to further enhance the accuracy of recognition, especially for complex sentences and signers. Zhang and Jiang [25] conducted a survey on the latest trends in the use of deep learning methods for sign language recognition, including the innovative use of Transformers, multimodal fusion, transfer learning, and edge optimization. The survey emphasized the trend towards context-rich and resource-optimized models that are suitable for real-time execution on mobile and IoT platforms.

MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 1, 2026, Pages: 2876-2883

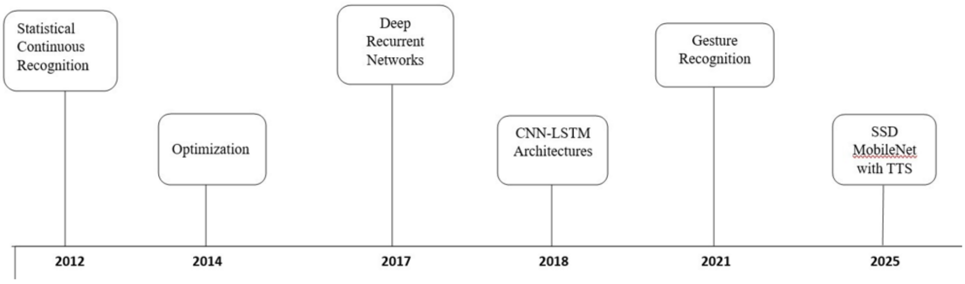

Fig. 1. Evolution of Research in Sign Language Recognition and Translation Systems.

3. METHODOLOGY

The proposed system of Real-Time Sign Language Recogni- tion and Translation would be used for the recognition and interpretation of hand gestures and would be translated into text and speech for the purpose of effective communication. The proposed system has been divided into eight major modules, and each module would have a major role to play in the proposed system.

A. System Architecture Overview

The system architecture combines computer vision, deep learning, and tensorflow together in one framework for real- time sign recognition and translation. The system architecture consists of several key modules that work together to ensure smooth operation:

- **Data Acquisition and Preprocessing:** Raw sign data is acquired from the webcam or laptop camera, and frames are preprocessed by resizing, noise elimination, and normalization. The dataset is then transformed into TFRecord format for efficient data processing using Ten- sorFlow.
- **Sign Detection and Classification:** SSD MobileNet V2, a lightweight CNN model, is employed to detect hands and classify static and dynamic signs in real-time.
- **Translation and Text-to-Speech (TTS):** Recognized signs are translated to text using a predefined sign-text mapping database. The text is then converted to speech using a TTS engine (Google TTS or pyttsx3).

B. Dataset Preparation and Labeling

The data has been gathered using OpenCV, and different signs were made to train and test the detection model.

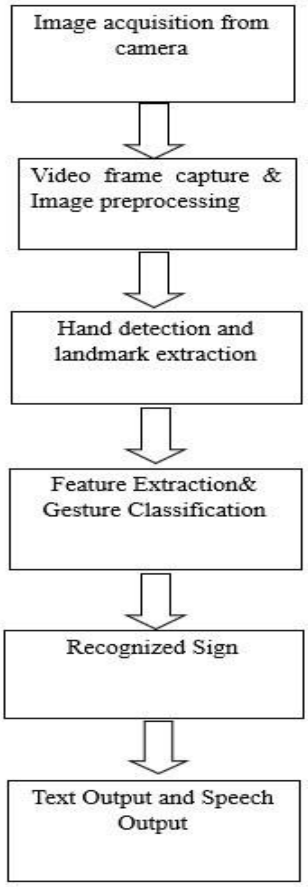

Fig. 2. Workflow of the proposed real-time sign language recognition and translation system.

- The dataset is made more robust by adding variations of lighting, background, skin color, and hand orientation.
- The images are annotated using the LabelImg software and saved in TFRecord format for efficient processing with TensorFlow. Data augmentation techniques such as rotation, flipping, scaling, and brightness are also applied to increase the size of the dataset and prevent overfitting.
- To check the efficiency of the model, the data is divided into 70 % for training, 15% for validation, and 15% for testing purposes.

This preprocessing step helps the model generalize well for different users and environments.

MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 1, 2026, Pages: 2876-2883

C. Model Training

The SSD MobileNet V2 model was chosen because of its efficiency and effectiveness in performing real-time object detection tasks. The training process followed these steps:

- Transfer learning was performed using a pre-trained SSD MobileNet V2 model and fine-tuned it on the custom gesture dataset.
- The loss function was a combination of Localization Loss and Confidence Loss to optimize the accuracy of object detection and classification.
- The training process was performed using the Adam optimizer and early stopping to prevent overfitting during training.
- The training metrics used during training were Mean Average Precision, F1-score, and latency.

D. Real-Time Gesture Recognition and Translation

The trained model was then applied in a real-time setting for the recognition and translation of gestures from real-time video streams. The working process involves:

1) Frame Capture: The frames captured from the webcam are resized to the required dimensions of the model.
2) Sign Detection: SSD MobileNet V2 detects the hand region and predicts the class of the gesture.
3) Recognition Display: The predicted sign class is dis- played on the screen in real-time.
4) Text-to-Speech Conversion: The recognized gesture text is converted to speech output.

This real-time pipeline allows users to communicate instantly, with little delay between gesture execution and speech output.

E. Key Generation and Feature Extraction Process

The real-time pipeline uses a key generation mechanism to enhance classification accuracy and distinguish between visually similar gestures.

- Features such as finger positions, joint angles, and palm centers are extracted from detected hand regions.
- These features are represented as feature vectors that serve as unique keys for each gesture. inference, the new Sign vectors are matched with the keys of the trained model for correct classification.
- This helps in improved recognition irrespective of the background conditions and lighting.

F. Key Generation Process

For the purpose of secure communication and data integrity between the system components, a cryptographic key gener- ation mechanism has been implemented. This will facilitate encrypted data communication, device authentication, and pro- tection of sensitive user data such as gesture-to-text mappings and speech outputs.

G. General Performance Dashboard

The efficiency of the proposed Sign Language Recognition and Translation System have been tested by evaluating its performance on various parameters. The system has been tested in different environmental settings using a standard webcam and computer.

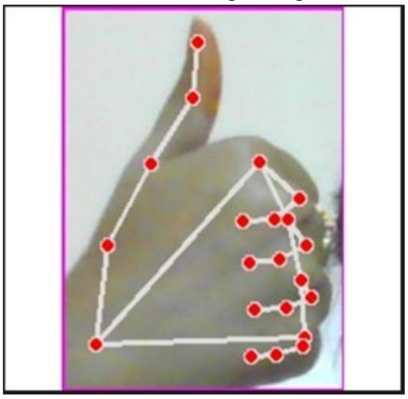

Fig. 3. Sign captured using MediaPipe for the dataset

4. WORKFLOW PROCESS

A. Data Collection and Preprocessing

- Gather images or short video clips of various sign lan- guage gestures (static and dynamic).
- Annotate each gesture with its corresponding meaning (letters, words, or phrases).
- Annotate files from XML to CSV and then to TFRecord for efficient TensorFlow training.
- Preprocess the images (resize, remove noise, and adjust lighting) to maintain image quality.

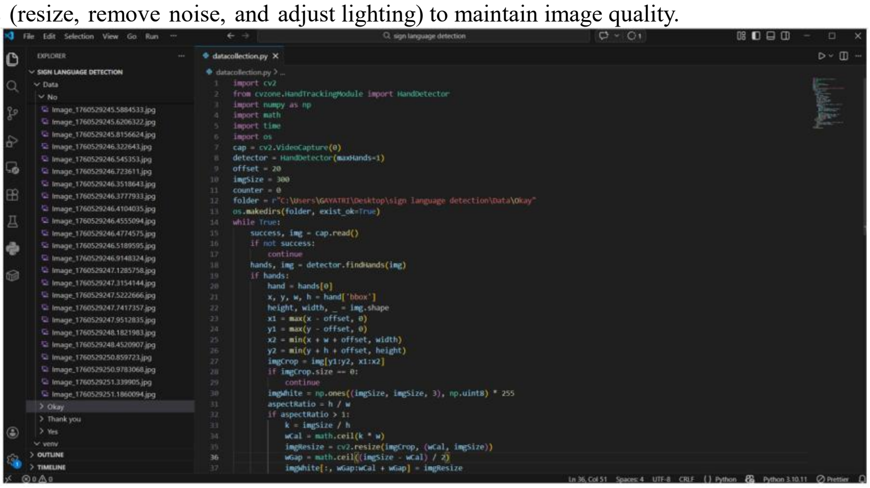

Fig. 4. data collected and labelled using labelMg; bounding boxes to focus on particular region.

MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 1, 2026, Pages: 2876-2883

B. Model Training

- Train the system using SSD MobileNet V2, a light-weight and efficient deep learning model.
- Use the preprocessed TFRecord dataset as input for the TensorFlow Object Detection API.
- The model will learn to detect hand shapes, motion patterns, and boundaries of gestures.
- Evaluate the model using accuracy, precision, and loss functions, and fine-tune it for optimal performance

Table I: Performance Comparison of Different Models

Model	Accuracy (%)	Precision (%)	Recall (%)
SVM	82.4	80.1	78.6
CNN	89.7	88.9	87.5
Faster R-CNN	93.8	94.1	92.9
YOLOv5	92.6	91.8	90.9
SSD MobileNet V2	94.6	94.2	95.1

C. Real-Time Gesture Detection

- Record live video using a webcam.
- Every frame in the video is then processed using OpenCV to identify and extract the region of the hand.
- The model is able to detect and classify the gesture in real time.
- Bounding boxes and labels of gestures are displayed on the live video feed.

D. Gesture Recognition and Translation

- The detected gesture is translated into text output that corresponds to the detected sign.
- The TTS engine is used for converting the text into speech that can be heard by non-signers.
- Both the text and voice outputs are displayed instantly on the interface.

E. Integration and Interaction

- The system's modular design allows easy connection to IoT devices, mobile platforms, and smart assistants.
- It can be used in classrooms, hospitals, or public areas for inclusive communication.
- The process runs continuously, enabling real-time, natural interaction between signers and non-signers.

F. Overview and Architecture

The Real-Time Sign Language Recognition and Translation System is intended to fill the gap between the hearing or speech impaired and the rest of the population. The system employs computer vision and deep learning algorithms to rec- ognize hand gestures in real-time video, translate the gestures into text, and finally convert the text into speech using a TTS engine.

G. Comparative Result of the Analysis

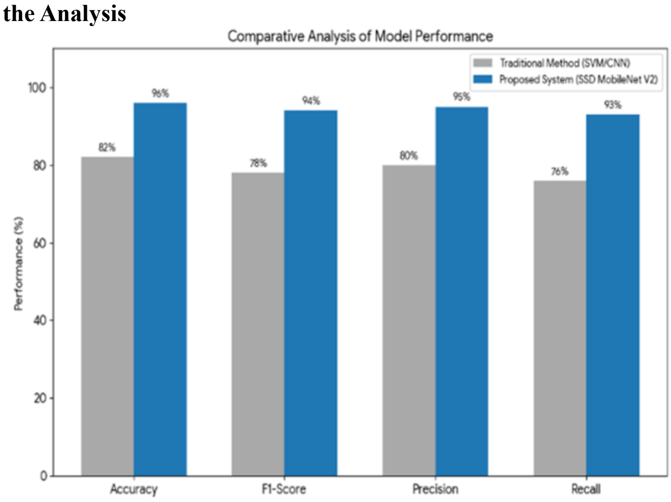

Fig. 5. showing comparison between SVM/CNN and SSD MobileNet V2 in accuracy, F1- score, precision and recall

H. Tools and Technologies used before implementation

OpenCV (4.8.0) used for real-time video capture, frame pro- cessing, and image preprocessing; TensorFlow (2.11.0) and Keras used to train and deploy the SSD MobileNet V2 deep learning model; NumPy (1.23.5) and Pandas (1.5.3) used to handle datasets, image arrays, and performance analysis; time and datetime libraries used to handle real-time frame timestamps during gesture recognition. IPython.display used to present visual and audio outputs such as recognized gesture text and Text-to-Speech (TTS) playback within the interface. Matplotlib (3.6.2 used to visualize training results, accuracy, and loss curves; Plotly (5.11.0) used to create interactive dashboards showing model performance metrics and gesture recognition results.

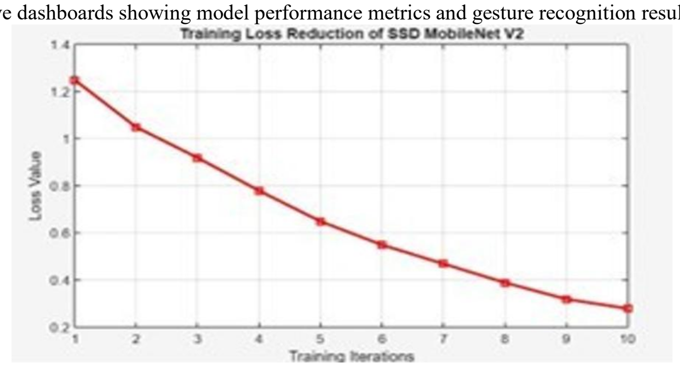

Fig. 6. Illustrates the reduction in training loss over successive iterations. The decreasing trend indicates effective optimization and stable convergence of the proposed deep learning model.

MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 1, 2026, Pages: 2876-2883

I. Evaluation Metrics and Validation Strategy

We have calculated the model performance using various metrics like accuracy, precision, recall. These metrics are employed to examine the extent to which the system is capable of identifying the signs correctly. The Mean Average Precision (MAP) provides us with an insight into the correctness of the model in identifying different signs.

I. Accuracy: Accuracy is the number of correctly classified samples out of the total samples.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

II. Precision: Precision is the measure of how many of the positive samples that the model has picked up are actually correct.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (2)$$

III. Recall: Recall is the number of actual positive samples that are correctly picked up by the model.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (3)$$

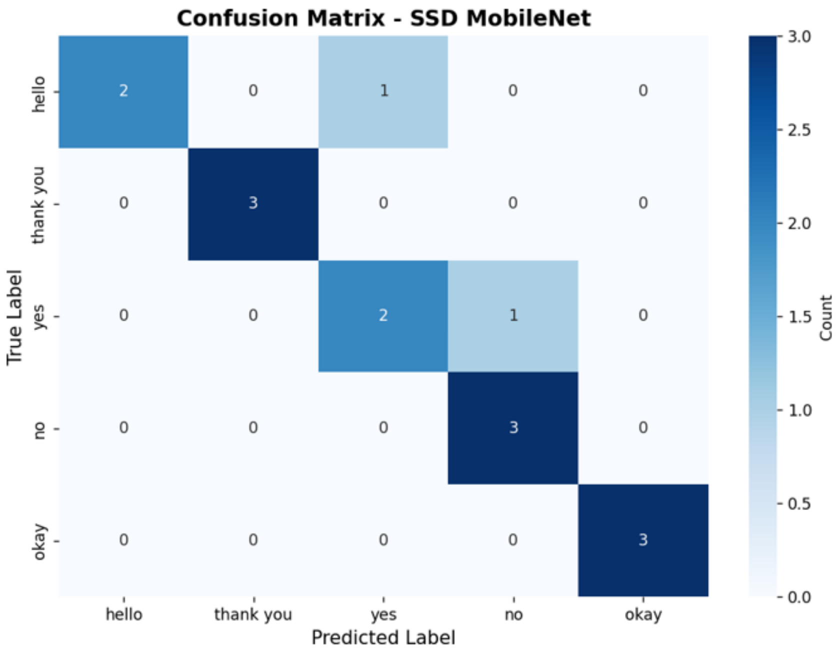

Fig. 7. Confusion Matrix of SSD MobileNet

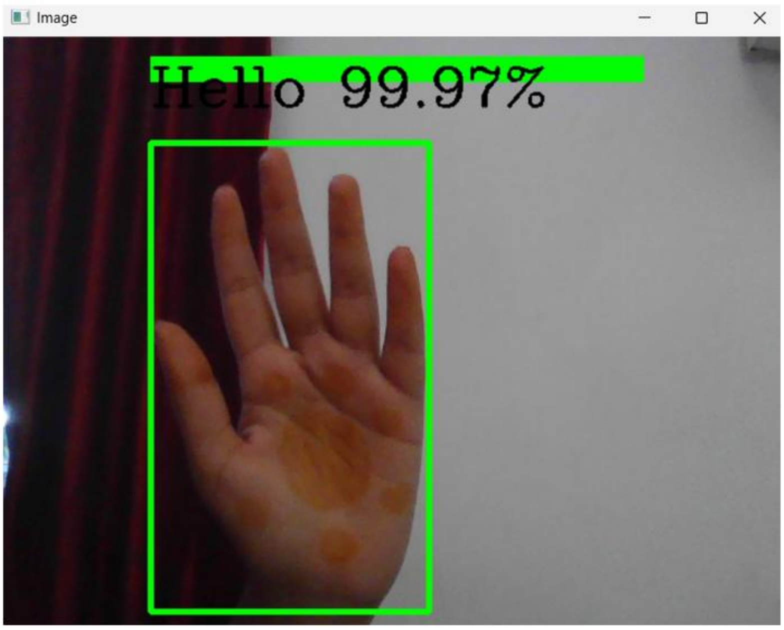

Fig. 8. Real Time Gesture Recognition

MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 1, 2026, Pages: 2876-2883

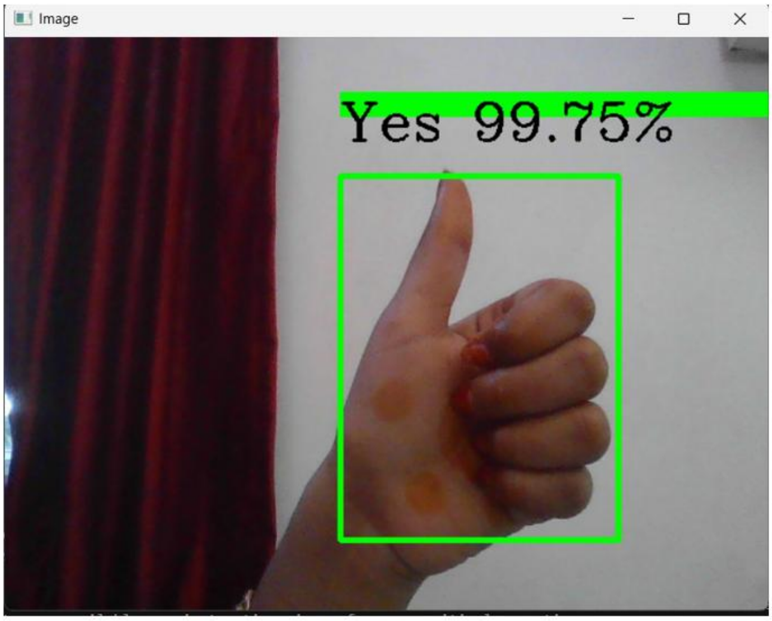

Fig. 9. Sign Recognized

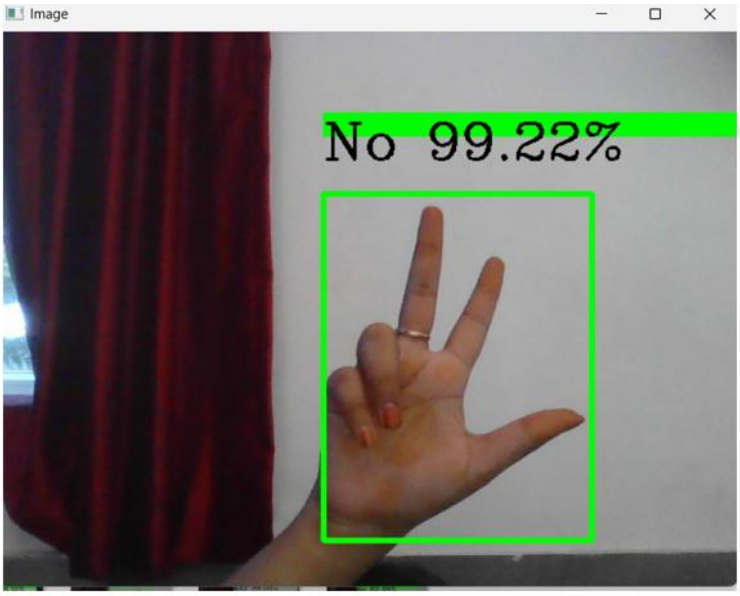

Fig. 11. Sign interpreted and accuracy rate

5. CONCLUSION

Communication is the most important parts of human inter- action, but for hearing or speech-impaired people, the process of communicating their ideas and feelings can sometimes be tough in a world that doesn't understand sign language. This project has been undertaken to fill this gap by devel- oping a real-time sign language recognition and translation system that can help make communication easier, faster, and more accessible.By using computer vision, deep learning, and blockchain-based authentication, this system not only provides accurate recognition of signs but also provides a secure and verifiable experience for all interactions. Using TensorFlow and OpenCV, this system provides efficient detection and translation of signs from live video streams to text and speech, enabling hearing-impaired individuals to "speak" freely with others in real-time. The system is designed to be cost-effective, flexible, and user-friendly, requiring only a simple webcam and a computer. Its connectivity with IoT platforms and cloud storage enables it to be scalable for use in a classroom setting, hospitals, public places, and even smart devices. However, the true worth of this project lies not in its accomplishment but in its use.In a way, this project seeks to accomplish a more connected and compassionate world by enabling a better understanding between the hearing and non-hearing commu- nities. Looking forward, this system can be further developed to incorporate continuous sign language recognition, multi-language translation, and cloud-based learning for real-time adaptability.

6. FUTURE SCOPE

The potential of this research to move forward is in the achievement of higher accuracy, scalability, and real-time intelligence of the Sign Language Recognition and Translation framework integrated with Blockchain and IoT systems. Future models can be developed to incorporate continuous gesture recognition and context-aware translation, allowing the system to recognize entire sentences and emotions rather than indi- vidual gestures. New optimization methods may improve the efficiency of the SSD MobileNet V2 model, further improving processing latency and energy use, particularly for mobile and edge computing devices. Additional approaches to improve system robustness and adaptability will be the incorporation of Artificial Intelligence (AI) and Machine Learning (ML) methods for gesture prediction, automatic retraining, and adap- tive error correction in different environmental settings. The compatibility of heterogeneous IoT systems and cloud-edge hybrid architectures can provide for seamless communication between devices and faster processing of real-time gesture information across platforms. Furthermore, the incorporation of blockchain-based authentication with quantum-secure cryp- tographic methods can future-proof the system against emerg- ing cybersecurity risks, ensuring secure data management and privacy. Finally, practical implementation of the system in ed- ucational, healthcare, smart IoT, and communication systems will help in the validation of the performance of the system and the inclusion of the hearing and speech-impaired population.

REFERENCES

[1.]Y. Abhishek and D. Sumanathilaka, "End-to-end sign language recogni- tion pipeline: Towards energy efficient modeling," *Procedia Computer Science*, vol. 265, pp. 483–490, 2025, doi: 10.1016/j.procs.2025.07.208.
[2.]M. Ahmed, M. Idrees, and Z. U. Abideen, "Deaf Talk Using 3D Animated Sign Language," in *Proc. IEEE Int. Conf. Smart Innovations (SAI)*, 2016, pp. 1–6, doi: 10.1109/SAI.2016.7556034.
[3.]M. R. Cassim, J. Parry, A. Pantanowitz, and D. M. Rubin, "Design and construction of a cost-effective, portable sign language to speech translator," *Informatics in Medicine Unlocked*, vol. 30, p. 100927, 2022, doi: 10.1016/j.imu.2022.100927.
[4.]P. S. Chatterjee, "An approach for minimizing the time taken by video processing for translating sign language to simple sentence in English," in *Proc. IEEE Int. Conf. Innovations in Communication and Embedded Systems (CINE)*, 2015, pp. 1–5, doi: 10.1109/CINE.2015.40.
[5.]S. Chatterjee and S. Abinaya, "Word level LSTM CNN based sign language to sentence translator," *Procedia Computer Science*, vol. 231, pp. 1–10, 2025, doi: 10.1016/j.procs.2025.04.276.
[6.]H. Chen and C. Di, "Lightweight sign language intelligent recognition model based on improved R-C3D," *Egyptian Informatics Journal*, vol. 32, p. 100801, 2025, doi: 10.1016/j.eij.2025.100801.
[7.]H . Cooper, E.-J. Ong, N. Pugeault, and R. Bowden, "Sign language recognition using sub-units," *Journal of Machine Learning Research*, vol. 13, pp. 2205–2231, 2012.
[8.]A. Das, A. D. Aishi, M. U. Toha, and M. F. Kader, "Bangla sign language translator for deaf and speech impaired people using deep LSTM," *International Journal of Speech Technology*, vol. 28, no. 2, pp. 351–368, 2025, doi: 10.1007/s10772-025-10184-2.
[9.]J. Fregoso, C. I. Gonzalez, and G. E. Martinez, "Optimization of convolutional neural networks architectures using PSO for sign language recognition," *Axioms*, vol. 10, no. 3, p. 139, 2021, doi: 10.3390/ax- ioms10030139.
[10.]B. Garg, M. Kasar, and P. Paygude, "Sign language detection dataset: A resource for AI-based recognition systems," *Data in Brief*, vol. 50, Art. no. 111703, 2025, doi: 10.1016/j.dib.2025.111703.
[11.]Google Research, "MediaPipe framework for real-time machine learning solutions," 2021. [Online]. Available: https://mediapipe.dev
[12.]A. Graves, A. Mohamed, and G. Hinton, "Speech recognition with deep recurrent neural networks," in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)*, 2013, pp. 6645–6649, doi: 10.1109/ICASSP.2013.6639343.
[13.]O. Koller, J. Forster, and H. Ney, "Continuous sign language recogni- tion: Towards large vocabulary statistical recognition systems handling multiple signers," *International Journal of Computer Vision*, vol. 113, no. 2, pp. 85–103, 2015, doi: 10.1007/s11263-015-0837-7.
[14.]R. Li, K. Liu, and A. K. F. Cheung, "Exploring the impact of intermodal transfer on simplification," *Language Sciences*, vol. 110, p. 101726, 2025, doi: 10.1016/j.langsci.2025.101726.
[15.]B. Paneru, and K. N. Poudyal, "Advancing human–computer interaction: AI-driven translation of American sign language to Nepali using convolutional neural networks and text-to-speech conversion ap- plication," *Systems and Soft Computing*, vol. 6, p. 200165, 2024, doi: 10.1016/j.sasc.2024.200165.
[16.]S. Reshna and M. Jayaraju, "Spotting and recognition of hand gesture for Indian sign language recognition system with skin segmentation and SVM," in *Proc. IEEE WiSPNET*, 2017, pp. 1–6, doi: 10.1109/WiSP- NET.2017.8299830.
[17.]S. S. R., S. S., A. K. G. M., P. Siddhartha, and N. K., "Sign language to sign language translator," *Procedia Computer Science*, vol. 260, pp. 373–381, 2025, doi: 10.1016/j.procs.2025.03.213.
[18.]P. Sharma and R. S. Anand, "A comprehensive evaluation of deep models and optimizers for Indian sign language recognition," *Journal of Visual Communication and Image Representation*, vol. 75, Art. no. 200032, 2021, doi: 10.1016/j.gvc.2021.200032.
[19.]S. Sharma and S. Singh, "Vision-based sign language recognition system: A comprehensive review," in *Proc. IEEE Int. Conf. Innovative Computing and Communication Technologies (ICICT)*, 2020, pp. 1–6, doi: 10.1109/ICICT48043.2020.9112432.
[20.]S. Shetty, E. Hirani, A. Singh, and R. Koshy, "Gesture-to-text: A real-time Indian sign language translator with pose estimation and LSTMs," *Procedia Computer Science*, vol. 219, pp. 1–10, 2024, doi: 10.1016/j.procs.2024.04.253.
[21.]R. Sreemathy, M. P. Turuk, S. Chaudhary, K. Lavate, A. Ushire, and S. Khurana, "Continuous word level sign language recognition using an expert system based on machine learning," *International Journal of Cognitive Computing in Engineering*, vol. 4, pp. 1–9, 2023, doi: 10.1016/j.ijcce.2023.04.002.
[22.]A. Tripathi, D. S. V. N. R. Kiran, D. Jaya Chandra, M. B. Kumar, A. S. S. Rama, S. Anjankar, K. Suthendran, and N. Malviya, "Intelligent sign language recognition for real-time text conversion to aid speech and hearing impaired," *Procedia Computer Science*, vol. 259, pp. 1472–1478, 2025, doi: 10.1016/j.procs.2025.04.102.
[23.]F. Wijaya, L. Dahendra, E. S. Purwanto, and M. K. Ario, "Quantitative analysis of sign language translation using artificial neural network model," *Procedia Computer Science*, vol. 245, pp. 998–1009, 2024, doi: 10.1016/j.procs.2024.10.328.
[24.]S. Xue, L. Gao, L. Wan, and W. Feng, "Multi-scale context-aware network for continuous sign language recognition," *Virtual Real- ity & Intelligent Hardware*, vol. 5, no. 4, pp. 1–12, 2023, doi: 10.1016/j.vrih.2023.06.011.
[25.]Y. Zhang and X. Jiang, "Recent advances on deep learning for sign language recognition," *Computer Modeling in Engineering & Sciences*, vol. 139, no. 2, pp. 1–25, 2024, doi: 10.32604/cmes.2023.045731.