Communications on Applied Nonlinear Analysis
ISSN: 1074-133X
Vol 32 No. 9s (2025)

Real-Time Sign Language Recognition Using Deep Learning

Mrs.V.Vimala Dheekshanya[*1], Dharshani .M[*2], Jeeviga .M [*3], Janani .T [*4], Dharshini K.T [*5]

[*1] Assistant Professor, Department Of Information Technology,

Manakula Vinayagar Institute Of Technology, Puducherry, India.

[*2, *3,* 4, *5] UG Student, Department Of Information Technology,

Manakula Vinayagar Institute Of Technology, Puducherry, India.

Email: Vimaladheekshanya1304@Gmail.Com

Article History:

Received: 03-07-2025

Revised: 16-08-2025

Accepted: 20-09-2025

Abstract:

This project aims to improve communication challenges faced by the deaf and dumb peoplepeople.. Itdoes this by creating a system that turns hand signs into spoken words and written text. The paper describes a real-time sign language recognition system built with deep learning. It uses Convolutional Neural Networks (CNN) to identify handgestures and Google Text-to-Speech (GTTS) to produce voice output. The system captures images of hand signs with a camera, classifies the gestures with CNN, and then uses GTTStoconverttherecognized signs into speech. The system works in real time, making it easier for people with communication difficulties to connect with others. This approach promotes inclusion and helps reduce language and cultural barriers. It makes communication simpler for everyone, no matter their physical abilities.

Keywords—Convolutional Neural Networks (CNN),Recurrent Neural Networks (RNN), and Long ShortTerm Memory (LSTM)

Introduction

Communication is a key part of how humans connect, allowing us to share thoughts, ideas, and feelings. For people with speech and hearing difficulties, this basic activity can be very hard. Sign language is a visual form of expression that helps millions of people worldwide understand each other. However, many people do not understand or know how to analyze the respective sign languages. This creates a barrier in conversations, making it hard for sign language users to be understood by those unfamiliar with it.

So we need a tool or an interface which fixes this communication problem between physically challenged people. Such a tool would make everyday communication easier. It would also promote inclusion in schools, workplaces, and public services. Older methods of sign language recognition depended on carefully selectedfeaturesfromgesturedata.Thesemethods involved manually finding patterns in gestures. While helpful, they often fell short in accuracy, flexibility, and working in different settings orwith various sign languages.

The rise of deep learning, especially Convolutional Neural Networks (CNNs), has changed how we recognize gestures. CNNs can learn complex patterns from visual data without needingmanualinput.Thismakesthemperfectfor understanding the detailed movements in sign language. Modern deep learning models can now give highly accurate results in real time, even in lively and changing conditions.

Tomakesignlanguagemoreaccessible,a real-time recognition system can include Google Text-to-Speech (GTTS). This tool converts recognized signs into spoken words. Itallowssign language users to "talk" directly to people who do not understand sign language. This makes the process of speaking and listening quicker andmore inclusive. Forexample,someonesigningcan have their message turned into speech by the system, so others can hear and understand.

The goal is to build a simple, flexible, andeasy-to-use platform. It should work well in many different settings. Important features include:

1. Instant Recognition: Detects signs and converts them into the respective word or sentence.
2. Support for Many Languages: Can recognize various signs and gestures which are in practice.
3. Customizable: Easy to add to schools, workplaces, or public offices.
4. Room for Growth: Designed so it can includemoresignlanguages,dialects,or even use gestures for other tasks lateron.

I. Literature Survey

Deep learning methods are useful in manyfields like medical science and defense. In this study, XAI is applied to recognize signlanguages. The approach combines attention mechanismswith ensemble learning toimprovetheaccuracyof predictions. ThemethodusesResNet50alongwith a Self Attention model to build the ensemble system. This combined method reached an accuracy of 98.20 percent.

Sign LanguageRecognition(SLR)helpspeople to communicate their feelings and expressions easily withotherswhoarenotawareoftheirsigns. It aims to go beyond using human interpreters. To improve this, a new system called multistage GCAR is proposed. It uses two key features: attention mechanisms that look at the importance of different parts of the body and a graph convolutional network that captures movementand position. This system looks at jointpointsand motion patterns to fully understand sign language gestures. These features are used in two separate streams to improve accuracy.

Thispaperdiscussestheproblemsofconverting the signs into text or voice. Sign languages have different grammar rules, which makes creating computers that understand them difficult. The translation system converts recognized signs into spoken words, but it faces issues due to complex grammar and meanings. Building datasets that include allregionalsignsanddialects,likethosein India, adds to the challenge. The paper reviews recent technological progress and solutions for both recognition and translation, helping make communication more inclusive.

This paper presents a system designed for Indian Sign Language (ISL) that focuses on easeof use and inclusivity. The system uses an LSTM neural network to recognizesixdynamicsigns,six emergency signs, and 26 static alphabet signs in ISL.Thishelpsuserscommunicatemoreeasily.

The system can also send emergency alerts when verbal communication is impossible. It quickly informs rescue teams, helping improve safety for people with hearing disabilities in India.

This project offers a new way to translatesigns using tools like TensorFlow, Keras, and LSTM. First, it uses MediaPipe Holistic tofindkeypoints on the body during sign language videos. Then, it builds a translation model using LSTM layers.This model can interpret signs in real time by reading the key points and predicting gestures..

II. Problem Statement

The challenge of accurately converting signs or gestures into voice or text is a significant barrier. Despite advancements in artificial intelligence, complexity of sign language characterizedbygesturesandsubtlevariations—makes it difficult for existing models to achieve high levels of accuracy.

We need a system which delivers precise translations but also provides clear

explanationsfor its decision-making process to build reliability initsoutputs.Thisstudyaddressesthisproblemby developing an innovative approach that not only provide the equivalent text of the sign language , but also provide the audio to enhance both the accuracy and interpretability of sign language recognition systems.

III. EXISTING SYSTEM

Deep learning has significantly advanced artificial intelligence, surpassing traditional machine learning in areas like Natural Language Processing, Computer Vision Human-Computer Interaction and Robotics. However, these models are often seen as "black boxes," making it hard to understand their inner workings. To foster trustand accountability, especially in important applications, it is crucial for deep learning models to provide not only accurate predictions but also clear explanations for their choices. ExplainableAI (XAI) offers methods to clarify the decisions made by neural networks, particularly in fields such as healthcare and defense.

TheapproachmergesResNet50withaSelf-Attentionmechanism,forminganensemble architecture that achieves a high accuracy of 98.20%. To enhance interpretability, a framework named SignExplainer is introduced, which shows the relevance of predictions in percentages. This framework outperforms other XAI models in the same area.

Despite its strengths, the current system faces several challenges. The integration of ResNet50 and Self-Attention increases computational complexity, making it resource-heavy and requiringadvancedhardwarefortrainingand real-time usage.

While SignExplainer improves understanding,it does not fully explain every decision made by the model, especially in critical fieldslikemedical diagnostics. The system's accuracy heavilydepends on the training data's quality and variety, making it tough to manage differences in sign language gestures caused by individual users, backgrounds, or lighting.

The model's complexity also introduceslatency issues, reducing its effectiveness in real-time settings and making it less suitable for quick communication. These limitations highlight the need for strategies that balance accuracy, interpretability, andefficiencytoensurethesystem can work well in practical applications.

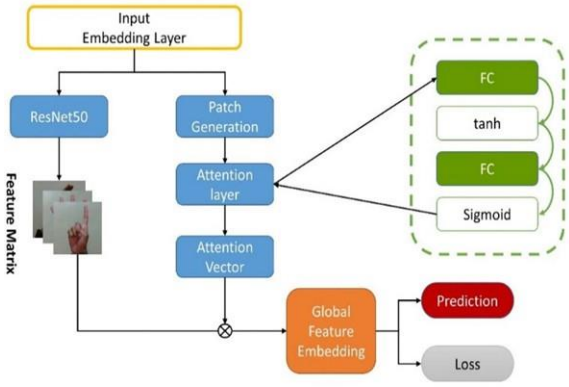

Fig.1ExistingSystem

IV. PROPOSED SYSTEM

The existing system architecture is used to analyze the user's sign and gestures in a live camera feed and turn them into text. It has five mainpartsthatworktogethertotranslategestures

smoothly and accurately into sound and words. The goal is to make the system efficient, easy to use, and able to grow, helping improve communication for everyone.

Itstartswithcapturingimages.Ahigh-resolution webcam records videos of hand gestures as they happen. These videos are thensent to the preprocessing step, where the images are cleaned and prepared for feature extraction. During this step, images are resized to a standard size, noise is reduced with filters, and techniques like rotation or scaling are used to make thesystem more reliable even if gestures look different.

Once cleaned, the images move to the feature extraction and classification part. Here, a Convolutional Neural Network (CNN) picks out important detailslikehandshapesandmovements. This process helps the system understand exactly what gesture is being made

The CNN's fully connected layer maps these features to corresponding gesture labels, while the softmax layer assigns probability scores to ensure accurate classification. Once a gesture isidentified, itismappedtoitscorrespondingtextual representation in the text conversion module. For example, the gesture for "Thank you"isconverted to the text "Thank you." Finally, the text ispassed tothevoiceoutputmodule,whichemploysGoogle Text-to-Speech (GTTS) to generate an audio file. This audio file is played in real-time, enablingnon-sign language users to hear the spoken equivalent of the gesture, thus completing the communication process.

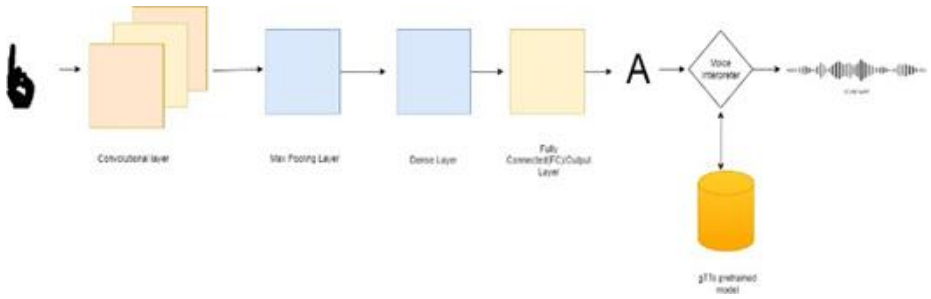

Fig.2HandGestureRecognitionandVoiceInterpreterSystem

V. METHODOLOGY

1. *Capturing Hand Gestures:* The system starts by capturing live video or images from a camera.Itprocessestheuser'shandgestures,

which represent symbols from sign language, includinglettersorwords.OpenCV,anopen-source tool for computer vision, is used to handle real-time image processing. During preprocessing, methods like removing the background, converting to grayscale, andapplying thresholdinghelpisolatethehand.Thisensuresthe gestures are accurately recognized.

2. *Convolutional Layer (Feature Extraction)*: After capturing the hand gesture image, it is sent through a Convolutional Neural Network (CNN)to pull out key features. This step's goal is to identify edges, curves, and patterns in the gesture image. The early layers of the CNN function as feature detectors, recognizing hand shapes and contours. Filters slide over the image to extract characteristics, creating featuremapsthatillustrate different elements of the gesture.

3. *Max Pooling Layer (Dimensionality Reduction)*:The features are then processed in a Max Pooling Layer, which simplifies the data and lowers its complexity. This step is crucial for easing the computational demands while keeping essential features for recognition. The pooling operation highlights the strongest features from small portions of the image,

reducing overall size but retaining important details. It also helps avoid overfitting by focusing on only the most meaningful features for further analysis.

4. *Dense Layer (Further Processing):* Once the features are extracted and reduced, the data enters the Dense Layer, or FullyConnectedLayer. This stage isvitalforanalyzingandclassifyingthe hand gesture based on the features identified. The Dense Layer assigns weights to the features, allowing the system to recognize patterns in the signlanguagegestureandconnectittotherelevant text.

5. *Fully Connected (FC) / Output Layer (Classification):* The last step in the classification process happens in the Fully Connected Layer. This layer determines the output category, identifying the corresponding letter, number, or word of the gesture. It selects the neuron with the highest activation value as the predicted class, whichisthenlinkedtoaspecificletterorword.

For example, if the system sees the gesture for"A," it will output the letter "A."

6. *Voice Interpreter*: After the classification, the identified text is sent to a Voice Interpreter to produce an audio output. The Voice Interpreter processes the classified text and turns them into understandable audio in the user native language.

7. *Text-to-Speech Conversion (GTTSModel):* The system employs Google Text-to-Speech (GTTS), a pre-trained model, to create clear voice output from the recognized text. This model uses advanced techniques to transform the text into audible speech. The resulting sound wave allows listeners to easily comprehend the recognized gestures.

VI. Project Implementation

1. *Dataset Preparation:* The dataset comprises labeled data(images) of various signs, these images are stored in the respective labelled folders for training.

2. *Data Preprocessing:* Each image is preprocessed to remove the noise and inorder to increase the clarity.

3. *Data Augmentation*: The dataaugmentation involves the following processwhichinclude:

- Rotation: Slightly rotating the images to account for variations in hand orientation.
- Flipping: Mirroring theimageshorizontally to simulate left and right-hand gestures.
- Scaling: Adjusting the size of the gestures within the image to make the model robust to size variations.

4. Model Training:TheCNNmodelistrained based on the following params:

- The loss function helps the model getcloser tothetruegesturelabelsbyreducing the gap between what it predicts and the actual data.
- The optimizer selected is Adam. Ithandles large amounts of data efficiently and adjuststhelearningrateastraininggoes on.Thisleadstoquickerlearningandbetter accuracy.

- The model trains for 30 epochs. Eachepoch involves one complete run through the training data. This helps the modellearn the patterns without overfitting.

5. *Real-Time Testing*: After training, the CNN model is deployed in a real-time application where it processes live video input and generates auditory feedback:

- Live Input Processing:The webcam continuously captures video frames, which are preprocessed in real-time (grayscale conversion, resizing, and normalization) before

being fed into the trained CNN model for classification.

- Gesture Classification:The model predicts the category of the gesture in each frame using the Softmaxoutputlayer.Thesystem ensures low latency to provide instantaneous feedback.
- Text Mapping and Voice Output:The classified labelled is mapped to the corresponding index which is mappedwith a word or sentence that we have used for training the dataset. The GTTS module converts the text into audio, which isplayed through the speaker. This feature makes the system particularly useful for individuals with communication disabilities, as it provides a seamless translation of gestures into auditory communication.

CONCLUSION

The model accuracy and efficiency are, measured through accuracy, precision, recall, and F1-score, shows it is strong and reliable. This shows the model can efficiently and consistently recognize different gestures, making it suitablefor gesture recognition uses.

Adding Google Text-to-Speech (GTTS) greatly improves the system's usefulness by turning recognized gestures into spoken words. This feature provides clear voice output, making it easierforuserstounderstand.Ithelpsconnect gesture recognition withsound,whichishelpfulin many settings.

Moreover,theintegrationofGoogleText-to-Speech (GTTS) plays a critical role in enhancing the system's practicality by seamlessly converting recognized gestures into speech. This feature ensures smooth and efficient communication.

The strong performance metrics of this model, combined with the real-world utility provided by GTTS integration, underscore its potential for deployment in practical scenarios. It not only achieves technical excellence but also addresses real-world challenges, making it a significant contribution to gesture recognition and assistive technologies. Ultimately, this project not only enhances communication but also contributes to greater inclusivity and accessibility in various aspects of daily life

REFERENCES

[1] Jia, Wanjun, and Changyong Li. "SLR-YOLO: An improved YOLOv8 network for real-timesignlanguage recognition." Journal ofIntelligent&FuzzySystems46, no. 1 (2024): 1663-1680.

[2] Rautaray, Siddharth S., and Anupam Agrawal. "Vision based hand gesture recognition for human computer interaction: a survey." Artificial IntelligenceReview43, no.1(2015):1-54.,DOI10.1007/s10462-012-9356-9 , SpringerScience+BusinessMediaDordrecht2012

[3] Varshney, Pankaj Kumar, Shrawan Kumar Kumar, and BhartiThakur."Real-TimeSignLanguageRecognition." In Medical Robotics and AI-Assisted Diagnostics for a High-Tech Healthcare Industry, pp. 81-92. IGI Global, 2024.

[4] MithunGeorgeJacob,JuanPabloWachs, "Context-based hand gesture recognition for the operating room,"inPatternRecognitionLetters,Volume 36, 15 January 2014, Pages 196- 203, ISSN 0167-8655

[5] Cabana, Elisa. "A real-time Artificial Intelligencesystem for learning Sign Language." arXiv preprint arXiv:2404.07211 (2024).

[6] Sharma, Vaidehi, Abhishek Sharma, and SandeepSaini. "Real-time attention-based embedded LSTM for dynamic sign language recognition on edge devices." JournalofReal-TimeImageProcessing21,no.2(2024): 53.

[7] Bakariya, Brijesh. "Sign Language Recognition-Based Machine Learning Model for Hearing Disabilities Person."InAppliedAssistiveTechnologiesand

InformaticsforStudentswith Disabilities, pp. 113-133. Singapore: Springer Nature Singapore, 2024.

[8] Alaftekin,Melek,IshakPacal,andKenanCicek. "Real-time sign language recognition based on YOLO algorithm." Neural Computing and Applications 36, no. 14 (2024): 7609-7624.

[9] Priya, K., and B.J.Sandesh."Developinganofflineand real-time Indian sign language recognition system with machine learning and deep learning." SN Computer Science 5, no. 3 (2024): 273.

[10] Parveen, M. Shabana, R. G. Keerthana, S. Shanmathi, and M. Shajitha. "Sign LanguageDetectionUsingOpen Cv." In 2024 International Conference on Communication, Computing and Internet of Things (IC3IoT), pp. 1-5. IEEE, 2024.

[11] Rastgoo, Razieh, Kourosh Kiani, and Sergio Escalera. "Sign language recognition: A deep survey." Expert Systems with Applications 164 (2021): 113794.

[12] Cooper, Helen, Brian Holt, and Richard Bowden. "Sign language recognition." In Visual Analysis of Humans: Looking at People, pp. 539-562. London: Springer London, 2021.

[13] Wadhawan, Ankita, and Parteek Kumar. "Signlanguage recognition systems: A decade systematic literature review." Archives of computational methods in engineering 28 (2021): 785-813.

[14] Khang, Adam Wong Yoon, Jamil Abedalrahim Jamil Alsayaydeh, Johar AkbarMohamatGani,TaraniBascar, Azman AwangTeh,JaysumanPusppanathan,andAndrii Oliinyk. "Portable American Sign Language System using RF Signals and IoT Technology for Deaf-Blind People." Journal of Advanced Research in Applied Sciences and Engineering Technology 45, no. 1 (2025):
79-89.

[15] Giovanelli, Elena, Gabriele Gianfreda, Elena Gessa, Chiara Valzolgher, Luca Lamano, Tommaso Lucioli, Elena Tomasuolo, Pasquale Rinaldi, and Francesco Pavani. "The effect of face masks on sign language comprehension: performance and metacognitive dimensions."Consciousnessand Cognition 109 (2023):
103490.

[16] JuhyeonHong;EungSupKim;Hyuk-JaeLee,"Rotation invariant hand posture classification with a convexity defect histogram," in Circuits and Systems (ISCAS), 2012 IEEE International Symposium on , vol., no., pp.774-777, 20-23 May 2012, doi: 10.1109/ISCAS.2012.6272153

[17] Liu, Jianbo, Ying Wang, ShimingXiang,andChunhong Pan. "Han: An efficient hierarchical self-attention network for skeleton-based gesturerecognition."Pattern Recognition (2025): 111343.

[18] Bankar, Sanket, Tushar Kadam, Vedant Korhale, and Mrs AA Kulkarni. "Real timesignlanguagerecognition using deep learning." International Research Journal of Engineering and Technology 9, no. 4 (2022): 955-959.

[19] Tolentino, Lean Karlo S., RO Serfa Juan, August C. Thio-ac, Maria Abigail B. Pamahoy, Joni Rose R. Forteza, and Xavier Jet O. Garcia."Staticsignlanguage recognition using deep learning." International JournalofMachineLearningandComputing9,no.6(2019):
821-827.

[20] Bantupalli, Kshitij, and Ying Xie. "American sign language recognition using deep learning and computer vision." In 2018 IEEE international conference on big data (big data), pp. 4896-4899. IEEE, 2018.

[21] Shirbhate, Radha S., Vedant D. Shinde, Sanam A. Metkari, Pooja U. Borkar, and Mayuri A. Khandge. "Sign language recognition using machine learning algorithm." International Research Journal of EngineeringandTechnology(IRJET) 7, no. 03 (2020):
2122-2125.

[22] Kothadiya, Deep, ChintanBhatt,KrenilSapariya,Kevin Patel, Ana-Belén Gil-González, and Juan M. Corchado. "Deepsign: Sign language detection and recognition usingdeeplearning."Electronics11,no.11(2022):
1780.

[23] Sahoo, Ashok Kumar."Indiansignlanguagerecognition using machine learning techniques." InMacromolecular symposia, vol. 397, no. 1, p. 2000241. 2021.

[24] Wadhawan, Ankita, and Parteek Kumar. "Deeplearning-based sign language recognition system for static signs." Neural computing and applications 32,no. 12 (2020): 7957-7968.