International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 12 | Dec - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

Gesture Recognition Using Machine Learning

Anona Borges[1], Priya Hiremath[2], Valentina Kerwadkar[3], Prof. Vijeth Swadi[4]

[123]*Student, Department of Computer Science and Engineering, KLS Vishwanathrao Deshpande Institute of Technology, Haliyal, Karnataka, India*

[4]*Asst.Prof, Department of Computer Science and Engineering, KLS Vishwanathrao Deshpande Institute of Technology, Haliyal, Karnataka, India*

Abstract - Hand gesture recognition is also an important application of human–computer interaction because it enables devices to interpret and answer the motions which users perform with their palms users perform with their palms. Development in the field of deep learning and computer vision has helped the detection of gestures more precisely There are multiple layers in CNN which allow the image to be very clear and accurate to detect the gesture. Our system identifies the gestures and translates the gesture into selected language and also displays the text Here the text consists of the gesture that is detected for example if we are showing the gesture hi through the webcam The Real time frames are considered and high is displayed on the screen along with the audio output for it.

Our system has been trained on various data sets which has helped the system to Detect the gesture with high accuracy and flexibility for various users also we have integrated the gesture recognition system with translation tools which can detect the gesture in five different languages This feature of a project removes the barrier of verbal and written communication The languages selected are Mostly Indian languages which can help people who might not know English to understand the gestures in their own mother tongue. Our Project mainly focuses on mainly removing linguistic barriers and providing a flexible interface for gesture recognition among the different language-based people.

Ultimately, it demonstrates how artificial intelligence can enhance inclusivity, facilitate real- time interactions, and promote user-friendly technology in everyday scenarios. Key concepts include gesture recognition, CNN, human- computer interaction, multilingual translation, image classification, and deep learning.

Keywords: Gesture recognition, CNN, Human Computer Interaction, Multilingual Translation, Image Classification, Deep Learning.

I INTRODUCTION

Good communication isn't just about what you say. The way you say it—your tone, your gestures, even how you move—shapes how people understand you.

Good communication plays important role in sign language because many people might not know it.

These things matter when you're expressing feelings, giving directions, or just sharing information. Now, with AI and deep learning getting smarter, gesture recognition has become a big deal. Computers can pick up on what we mean, even if we never say a word out loud. Thanks to convolutional neural networks (CNNs), machines are way better at reading complex visual cues, so they can spot gestures and body movements more accurately.

You see this tech everywhere now—video games, robotics, healthcare, and all sorts of adaptive devices. Let's face it, old tools like keyboards and screens just get in the way when we want to interact naturally with technology. Imagine someone say "hi" in their languages Which is difficult to interpret to the people who speak different languages or are Unknown from different languages This is the idea behind the approach where we have trained the system on gesture data from different sign languages across the world. CNN his perfect match for understanding the patterns more effectively and they're also very well trained in understanding the movements of the gesture which we show through our hand.

First the system grabs the visual data then figures out what is happening and then matches the motion to its gesture or we can say that it can match each motion to its meaning once the system understands the gesture it translates it to the selected language using NLP. This project is not only about making technology smarter but also helping people Through different kinds of languages connecting This can help the digital space more effective and it can help everyone feel like they are included and also may not feel the Linguistic barrier while approaching new technology.

II LITERATURE REVIEW

1. "Real-Time American Sign Language Recognition using Deep Learning" by S. Sharma et al., 2020

This paper explains how CNN models are used to classify American Sign Language along with good accuracy. It demonstrated that deep learning implements it way better than traditional. image processing methods but highlighted the need for large datasets and better generalization across lighting conditions.

2."Multilingual Sign Language Translation Using Neural Networks" by L. Zhang and K. Chen, 2021

The authors presented a framework for interpreting sign-based gestures and translating them into multiple natural languages. The study emphasized the importance of combining vision

models with NLP-based translation systems for multilingual support.

3.“Gesture Recognition Using CNN and OpenCV” by A. Patel et al., 2019

This work focused on static gesture recognition using CNNs and image preprocessing techniques such as thresholding and contour detection. It achieved real-time recognition but lacked multilingual capabilities.

4.“Vision-Based Hand Gesture Recognition for Human-Computer Interaction” by D. Kumar and P. Singh, 2022

The study proposed a hybrid CNN model that improved gesture classification accuracy. It explored gesture control applications but did not incorporate translation or cross- lingual communication.

5.“Cross-Lingual Communication via Gesture-to- Speech Systems” by R. Gupta et al., 2023

This recent work demonstrated gesture-to-speech translation using AI models. It motivated integrating multilingual support for accessibility tools, like the goals of the present project.

III PROPOSED SYSTEM

The proposed system introduces an intelligent gesture recognition model integrated with multilingual translation and output generation. with multilingual translation and output generation Functional Requirements

The system captures real-time hand gesture images or videos using a webcam and preprocesses them through cropping, resizing, and background removal. It then uses a trained CNN model to classify the gestures accurately. The recognized gestures are translated into multiple languages such as English, Hindi, and Tamil, and the output is either displayed or spoken.

The processed gestures' frames employ various transformations for elimination of noise and enhancement before analysis. Adjust background sounds, balance lighting levels, and scale all photos uniformly to an aspect ratio of sixty-four-by-sixty-four pixels.

Applying convolutional neural networks enables automatic extraction of both spatial and temporal features in data analysis tasks. Extract structural elements directly from gestures using automated methods. The CNN framework categorizes hand movements by assigning them to specific classes through the use of a SoftMax function.

A. Non-Functional Requirements

The system is designed to recognize gestures with minimal delay, ensuring a response time of less than one second. It aims to achieve an accurate rate of above 90% for all trained gestures, providing reliable and precise recognition. The user interface is built to be simple, intuitive, and accessible to users of all backgrounds. The system is also scalable, allowing new gestures or languages to be added easily without major changes. It ensures consistent performance in different lighting and background conditions, enhancing reliability. Security steps are implemented so that only recognized people can access training or data modules.

Hardware Requirements

•High-definition webcam.
•Processor: Intel i5 /i7
•RAM: Minimum 8 GB for desktop-based
•GPU: NVIDIA
•Optional: Microphone and speakers for speech output

b. Software Requirements

•Operating System: Windows 10 / Ubuntu 20.04 /Raspbian OS.
•Programming Language: Python 3.9 or higher.

•Libraries/Frameworks:
•TensorFlow / Keras for CNN implementation.
•OpenCV for image capture and preprocessing.
•NumPy, Pandas, Matplotlib for data handling and visualization.
•Google Translate API for multilingual translation.
•PyAudio / gTTS for speech output.
•Flask / Streamlit for GUI development (optional)
•Display: Standard monitor for output visualization.

IV SYSTEM DESIGN

A. system architecture description

The users can interact with the system by showing the hand gestures in front of the web camera if we want to fix a selected language, we can use the web page buttons And select the buttons of a preferred language

The webcam captures a video and the converts it into video frames which are ready for the preprocessing stage. Here in the preprocessing stage the captured frames Are processed by cutting out the background switching the colors to black and

white and by trimming its edges Now the data going into the CNN model is clean and accurate

Here in the CNN model the gestures are matched and text and audio is generated through translation service All libraries like gTTS the database holds everything that the data set which we have used for training and testing the CNN also the library of languages. It means that the database also contains the language data, also the gesture data set. The preprocessed image before deciding its output passes through the CNN layers after that finally the output that is text and audio is displayed on the screen.

Data Flow Summary:

Gesture → Camera Capture → Preprocessing → CNN Classification → Language Mapping → Translation → Text/Speech Output.

System Workflow Overview

The implementation of the Gesture Recognition Using Multiple Languages system follows a modular and layered architecture. Each layer is designed to handle a specific task in the pipeline, from data collection to real-time prediction and translation. The workflow can be broken down into the following stages:

1. Data Acquisition: Hand gestures are captured using a high-definition webcam. The dataset includes both static and dynamic gesture images that represent letters, numbers, and simple commands like "Hello," "Yes," "No," and "Thank You."

2. Data Preprocessing: The captured gesture images go through several preprocessing steps to remove background noise, normalize lighting, and resize the images to a standard dimension (for example, 64×64 pixels).
3. Feature Extraction using CNN: A deep Convolutional Neural Network automatically extracts spatial
and structural features from the gesture images, eliminating the need for manual feature engineering.
4. Classification: The CNN model classifies the gestures into predefined categories using SoftMax activation.
5. Language Mapping and Translation:
Once the gesture is recognized, it is mapped to a text label
(for example, “Hello”), which is then translated into the user-selected language using translation APIs (Google Translate or DeepL).
6. Output Presentation:
The translated output is displayed on the screen as text or converted into audio using a text-to-speech (TTS) module.

RESULTS AND DISCUSSION

The experimental results demonstrated the system’s ability to effectively recognize and translate gestures in real time across multiple languages. The CNN-based Model worked and provided much good performance as compared to the traditional one.
image- processing models that relied on manual feature extraction.

A. Performance comparison

A baseline comparison was made between CNN, SVM, and KNN classifiers:

Model	Accuracy (%)	Inference Time (ms)
SVM	78.2	250
KNN	81.4	300
CNN	92.4	120

The CNN model worked better than other methods both in functionality and execution time due to its good feature extraction power and parallel GPU computation.

B. Translation and speech output performance

Integration with Google Translate API provided accurate translations in more than four languages, with an average
Translation time of 0.8 seconds per query. The gTTS (Google Text-to-Speech) module produced clear, natural- sounding speech output in the selected language, successfully bridging the communication gap between hearing-impaired users and multilingual listeners.

C. Discussion

The system performed exceptionally well in real-world testing scenarios.
Key observations include:

- The model maintained high correctness even with average lighting changes.
- Real-time translation was almost instant with minimal delay.
- Users found the interface simple and engaging, which improved accessibility. However, accuracy dropped slightly in poor lighting or when gestures were partially blocked, indicating room for communication upgrade through better preprocessing or depth-sensing cameras.

CONCLUSION

This research endeavor titled "Gestural Identification Through Multilingual Techniques Employing Convolutional Neural Networks" showcases its significance in technology advancement. The innovative fusion of computer vision, advanced machine learning techniques like deep learning, and sophisticated text analysis methods is aimed at improving human capabilities. computer communication.
System's capability for interpreting hand movements and converting them instantly across various languages. demonstrates both technical sophistication and social utility. This study successfully utilizes Convolutional Neural Networks to automatically extract hand gestures' features classification, significantly outperforming traditional classifiers.
Integrating translation and speech components significantly enhances functionality by facilitating smooth communication across various languages interaction, especially for the hearing and speech-impaired community. The testing phase validated that the CNN-based model So, the real-time validation stuff showed like, almost no lag which is great, and users said it was pretty easy to use...that makes it useful for everyday stuff. And like, big picture-wise, this system we're proposing kind of helps with tech inclusivity, which is super important for breaking down communication barriers in a world where everyone speaks different languages.
It kind of shows how deep learning and AI translation can

be mixed, creating communication setups that are accessible and respond intelligently, like it is designed for user inputs.
Future enhancements
1. Dynamic Gesture Stuff: Future versions could maybe have ongoing gesture recognition using RNN or LSTM models, idk.
2. Edge Deployment Optimization: For deploying on mobile or like Raspberry Pi for on-the-go usage.
3. Offline Translation, maybe? Add offline translation libraries to kind of reduce the leverage on internet.
4. Better Lighting Adaptation: Like, using infrared or depth sensors to make detection better, whatever the lighting.
5. More Gestures: Collect a bigger dataset, including regional sign languages for better inclusivity. Achieves a good level of functionality (above 92%) and keeps accurate performance under varying conditions.
In summary, this system successfully demonstrates how AI-driven gesture recognition combined with multilingual translation can revolutionize communication for individuals across linguistic and physical boundaries. It lays the foundation for future human-computer interaction systems that are smarter, faster, and more inclusive.

REFERENCES

[1] Simonyan, K., & Zisserman, A. (2015). Very Deep Convolutional Networks for Large-Scale Image Recognition.

[2] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. IEEE CVPR.

[3] Chollet, F. (2017). Exception: Deep Learning with Depth wise separable convolutions

[4] Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press.

[5] American Sign Language Dataset, Kaggle (2023), Retrieved from: https://www.kaggle.com/

[6] Google Translate API Documentation. (2024). Retrieved from: https://cloud.google.com/translate

[7] OpenCV Library Documentation. (2023). Open-Source Computer Vision Library. https://opencv.org

[8] TensorFlow Documentation. (2024). Deep Learning Framework for AI Applications. https://www.tensorflow.org