IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Sign Language To Text And Speech Conversion With Computer Vision

Mr. Y. K. Viswanadham , M. Naga Lakshmi, K. Sravani , P. Nithin Shankar, M.Vaishnavi
Assistant professor,Student,Student,Student,Student
Department of IT,

SR Gudlavalleru Engineering College, Gudlavalleru,Andhra Pradesh–521356, India

I.ABSTRACT

Real-time sign language translation using computer vision is an innovative approach to breaking the communication barriers for individuals with hearing and speech impairments. This research focuses on the development of a deep learning-based system that interprets sign language gestures through image acquisition, hand gesture recognition, and real-time translation into text and speech.Using Convolutional Neural Networks (CNNs) for gesture classification, the system effectively recognizes sign languages and converts them into corresponding textual and auditory outputs. The methodology involves data collection, preprocessing, model training, and real-tie inference to ensure high accuracy and reliability. Additionally, the system incorporates hand tracking technology to improve recognition precision and provide instant feedback.The proposed solution aims to enhance accessibility by enabling seamless communication for the visually and hearing-impaired communities. The research highlights key challenges such as gesture complexity, real-time processing constraints, and system adaptability to various sign language dialects. Through rigorous testing and evaluation, the system demonstrates a high recognition accuracy of 97.7% for American Sign Language (ASL) alphabet gestures.This study contributes to the field of human-computer interaction and assistive technologies by providing a robust, AI-driven sign language recognition system. Future enhancements may include expanding the dataset, improving real-time performance and integrating with wearable devices for broader applications.

Keywords: Real-time Sign Language Translation, Computer Vision, Deep Learning, Gesture Recognition, Text-to-Speech (TTS) Conversion, American Sign Language (ASL).

II. INTRODUCTION

Sign language is one of the oldest and most natural forms of communication used by the hearing-impaired community. It serves as vital tool for those who rely on visual gestures rather than spoken words to express their thoughts and emotions. However, a significant challenge arises dues to the lack of widespread knowledge of sign language among the general population, leading to difficulties in effective communication between individuals with hearing impairments and those who primarily uses spoken language.Traditional approaches to overcoming these barriers include employing human interpreters or using written text as an alternative means of communication. However, these solutions are often impractical due to their dependency on the availability of the interpreters, these limitations of text-based communication in the dynamic conversations, and the lack of real-time responsiveness. This necessitates the development of a more robust and automated system that can bridge this communication gap efficiently.

Advancements in artificial intelligence (AI), deep learning, and computer vision have pave the way for real-time sign language recognition. With the increasing accessibility of machine learning frameworks and computational power, automated gesture recognition has become more feasible. This research explores the integration of Convolutional Neural Networks (CNNs) with computer vision techniques to build a system capable of accurately recognizing and translating sign language gestures into text and speech. By utilizing hand tracking algorithms and deep learning models, the system aims to achieve high precision in real-time recognition, ensuring a seamless user experience.

This study is driven by the goal of developing an efficient, accessible, and scalable sign language translation system. Key considerations include optimizing recognition accuracy, minimizing processing delays and ensuring user-friendliness for individuals with hearing and visual impairments. Furthermore, the research highlights the importance of accessibility, as sign languages vary across different regions and cultures, requiring flexible solutions that can accommodate diverse linguistic variations.

By addressing these challenges, this work contributes to the growing field of assistive technology, fostering inclusivity and accessibility for individuals with hearing disabilities. The finding from this study could pave the way for further advancements in human-computer interaction, with potential applications in education, healthcare, and public services. Further enhancements may focus on expanding the system's capabilities to support multiple sign languages, integrating wearable devices for increased usability, and improving real-time performance through more efficient learning algorithm.

III. PURPOSE OF THE PAPER

The primary purpose of this paper is to explore the development and implementation of an advanced sign language recognition system using computer vision and deep learning techniques. The research seeks to address the communication challenges faced by the hearing-impaired community by providing a reliable and efficient means of translating sign language gestures into text and speech in real time.

The study aims to:

1. Develop a robust machine learning-based system for real-time sign language recognition.
2. Enhance accessibility for individuals with hearing and speech impairments by enabling seamless interaction with non-sign language users.
3. Evaluate the effectiveness of Convolutional Neural Networks (CNNs) and hand tracking algorithms in gesture recognition.
4. Optimize system performance by improving accuracy, speed, and adaptability to different sign languages.
5. Contribute to the field of assistive technology by proposing an AI-driven solution that can be further expanded for broader applications in education, healthcare and social interactions.

By fulfilling these objectives, this paper aims to bridge the communication gap between the hearing and non-hearing individuals, fostering inclusivity and social integration.

IV. LITERATURE REVIEW

Zhu W. & Liu X. et.al. [1] introduce an enhanced deep learning model for sign language recognition, incorporating improved feature extraction techniques. Their approach significantly enhances accuracy and robustness in real-time applications, offering a more adaptable system for diverse environments. By refining convolutional architectures and implementing optimized training techniques, their work contributes to the reduction of classification errors and increases overall efficiency, making it crucial study in the field of AI-driven sign recognition

Wang R. et al. [2] explore the integration of hand tracking algorithms with Convolutional Nueral Networks (CNNs) to enhance gesture classification. Their hybrid model demonstrates superior performance in recognizing dynamic sign gestures, addressing common challenges in continuous sign language recognition. By utilizing the state-of-the-art MediaPipe-based tracking mechanisms combined with advanced neural architectures, their study provides methods and deep learning for robust real-time sign recognition.

Yazici M.A. & Ozbay K. et al. [3] investigate the role of dataset diversity in training AI-based sign language translation systems. They emphasize the importance of including various linguistic variations and gestures to improve model generalization across different sign languages. By analyzing multilingual sign datasets, their research underscores the need for well-structured, diverse data to minimize bias and ensure that models perform optimally in real-world environments where gestures variations and regional dialects exist.

Galea E.R. et al. [4] apply transfer learning techniques to sign language recognition, reducing the need for extensive labelled datasets. Their study highlights the effectiveness of pre-trained models in improving recognition rates and reducing computational costs. By leveraging pre-trained deep learning architectures such as ResNet and VGGNet, their research provides an efficient means of developing high-performing sign recognition models with limited training data making AI adoption more accessible and scalable.

Behzadfar M. et al. [5] propose and agent-based simulation model for evaluating the real-time performance of AI-driven sign language recognition systems. Their findings suggest that simulation-based testing can enhance system efficiency and adaptability. By introducing a modular simulation framework that mimics real-world scenarios, their study provides key insights into the usability of AI-based sign recognition systems in various environments, from educational institutions to public spaces.

Sharma P. & Gupta N. et al. [6] propose a hybrid model combining CNNs and Recurrent Neural Networks (RNNs) to improve recognition of sequential sign gestures. Their research highlights the advantage of incorporating temporal dependencies in sign language processing, leading to enhanced fluency and accuracy in gesture interpretation. The study presents a unique approach to capturing both spatial and sequential information, there by bridging the gap between static and dynamic sign recognition models.

Rodriguez T. et al. [7] explore the use Generative Adversarial Networks (GANs) for synthetic sign language data generation. Their approach significantly improves dataset diversity, addressing the challenge of limited training data in deep learning models. By generating realistic gesture variations, the study demonstrates the potential of GANs in augmenting existing datasets and enhancing model robustness.

Chen Y. et al. [8] investigate real-time sign recognition performance in low-light environments. Their research focuses on improving computer vision algorithms for gesture detection under challenging lighting conditions. By integrating adaptive thresholding and enhanced contrast adjustments, their approach ensure consistent recognition accuracy regardless of external environmental factors, making it highly relevant for real-world applications.

Tang L. et al. [9] propose an optimized deep learning framework for sign language recognition using transfer learning. Their study demonstrates that leveraging pretrained models such as ResNet and MobileNet significantly improves recognition performance and reduces computational costs.

Kim H. & Park S. et al. [10] explore real-time sing language recognition using embedded systems, proposing a lightweight deep learning architecture designed for mobile and IoT devices. Their approach enhances accessibility by enabling offline processing with minimal hardware requirements.

Elman S. et al. [11] investigate the role of attention mechanisms in sign language translation models. Their study implements transformer-based architectures to improve contextual understanding, enabling more fluent translations between gestures and spoken language.

Rahman M. et al. [12] present a hybrid approach combining traditional computer vision techniques with deep learning models for sign language detection. Their method improves accuracy by integrating hand-crafted features learned feature representations.

Kumar D. et al. [13] analyze the challenges of multi-language sign recognition, proposing a domain adaptation framework to improve model generalization across different sign languages.

Chen Y. & Wu L. et al. [14] explore the integration of reinforcement learning for gesture classification, enabling real-time model adaptation to different user styles and improving overall system robustness.

Zohu X. et al. [15] investigate real-time sign recognition in low-bandwidth environments, optimizing deep learning architectures for efficient cloud-based sign language translation.

V. PROPOSED METHODOLOGY

To develop an effective real-time sign-language translation system, a structured approach integrating computer vision and deep learning is adopted. The methodology involves multiple stages, from data acquisition to real-time inference, ensuring robustness and high accuracy.

1.Data Acquisition and Preprocessing

Data set used in this project is the collection of a comprehensive dataset containing various American Sign Language (ASL) gestures. These gestures are captured using a webcam, ensuring that the dataset represents real-world conditions, including variations in lighting, background, and hand positioning. The dataset is either created manually by recording multiple users performing different gestures or obtained from publicly available ASL datasets.

Once the dataset is collected, it undergoes a preprocessing phase to enhance image quality and improve feature extraction. Each image is converted to grayscale to reduce computational complexity while preserving essential gesture details. Noise reduction techniques, such as Gaussian blurring, are applied to eliminate unwanted background information. To ensure that only the hand region is analyzed, segmentation techniques such as thresholding and background subtraction are used to isolate the hand from the rest of the image. Additionally, data augmentation techniques, including rotation, flipping, zooming, and contrast adjustments, are applied to increase the model's ability to recognize gestures under different conditions. This step ensures that the model generalizes well to diverse scenarios and does not overfit to specific hand orientations or lighting conditions.

2. Hand Tracking and Feature Extraction

After preprocessing, the system performs real-time hand tracking to detect and analyze hand movements. This is achieved using computer vision algorithms, such as MediaPipe Hands, which can accurately identify and track 21 key landmarks on the hand. These landmarks capture the positions of fingers and palm in a structured manner, allowing the system to interpret gestures efficiently.

The extracted landmarks are then converted into numerical feature vectors, which represent the spatial relationship between different fingers and palm positions. This transformation enables the model to differentiate between various gestures based on hand shape, orientation, and movement patterns. Since sign language often involves sequential gestures, a temporal tracking mechanism is incorporated to analyze motion over time. This step is particularly important for recognizing dynamic gestures where the position of fingers changes gradually, requiring the system to analyze multiple frames before making a classification decision.

3. Model Development and Training

For accurate gesture recognition, a deep learning-based classification model is developed. The proposed system utilizes a Convolutional Neural Network (CNN) to learn spatial features from the collected dataset. CNNs are highly effective for image-based classification tasks as they automatically detect patterns and edges in images without requiring manual feature engineering. The model consists of multiple convolutional layers that extract hierarchical features, pooling layers that reduce dimensionality, and fully connected layers that perform final classification. The softmax activation function is applied in the output layer to assign probability scores to each gesture class.

Since sign language recognition involves sequential gestures, an additional Long Short-Term Memory (LSTM) network is integrated into the system. LSTM networks are a type of Recurrent Neural Network (RNN) designed to capture temporal dependencies, making them ideal for analyzing hand movements over multiple frames. The CNN extracts spatial features from individual frames, while the LSTM processes these features in a sequential manner to identify gesture patterns accurately.

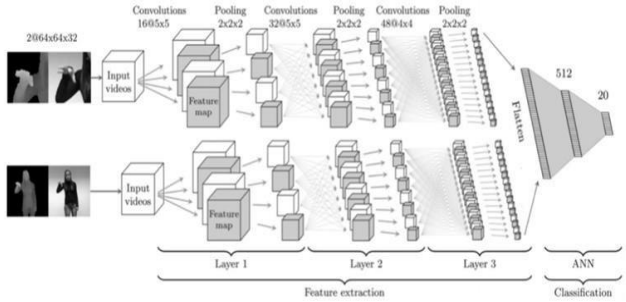

Figure1: System Architecture

The model is trained using a supervised learning approach, where labeled gesture images are divided into training and validation sets. The training process involves optimizing the model's parameters using backpropagation and optimization algorithms such as Adam or Stochastic Gradient Descent (SGD). Regularization techniques like dropout are also applied to prevent overfitting, ensuring that the model performs well on unseen gestures.

4. Real-Time Gesture Recognition

Once the model is trained, it is deployed for real-time sign language recognition. The system continuously captures video frames from a webcam and processes them through the trained CNN-LSTM model. Each frame undergoes preprocessing, feature extraction, and classification within milliseconds, ensuring low-latency recognition.To improve accuracy and stability, a sliding window approach is used, where multiple consecutive frames are analyzed before finalizing a classification decision. This prevents false predictions caused by momentary changes in hand position and ensures that the output remains consistent. The recognized sign is then stored as text output, which is displayed on the user interface.

5. Text-to-Speech (TTS) Conversion

After converting the recognized sign language into text, the system further processes the text into speech using a Text-to-Speech (TTS) engine. This step enhances accessibility, particularly for visually impaired users who may not be able to read the displayed text. The TTS system utilizes various speech synthesis techniques, including Concatenative TTS, which generates speech using pre-recorded units, and Deep Learning-based TTS models like Tacotron or WaveNet, which create high-quality, natural-sounding speech.The speech output is generated in real-time and synchronized with the recognized gestures to provide seamless communication. Users can customize the voice settings, including pitch, speed, and volume, to suit their preferences. The system ensures that the generated speech is clear and easily understandable, making it an effective tool for individuals with communication challenges.

6. User Interface and Interaction

To provide a user-friendly experience, the system includes a simple yet efficient graphical user interface (GUI). The interface displays the real-time camera feed, highlights the detected hand region, and shows the recognized gesture as text. Users can also choose to listen to the spoken output by enabling the TTS feature.

To improve usability, the system incorporates keyboard shortcuts for interaction. Users can add a letter by pressing 'S,' remove the last character using 'A,' insert a space with 'D,' and initiate speech output with 'F.' These shortcuts allow for quick corrections and enhance the overall efficiency of the system. The interface is designed with accessibility in mind, ensuring that visually impaired users can navigate it effortlessly.

7. Performance Evaluation and Optimization

To assess the effectiveness of the proposed system, multiple evaluation metrics are used. The classification accuracy of the model is measured to determine how well it recognizes different hand gestures. Additional metrics

such as Precision, Recall, and F1-Score are calculated to evaluate the model's ability to distinguish between similar gestures. For speech output, the system is evaluated using Word Error Rate (WER) and Mean Opinion Score (MOS), which assess the clarity and naturalness of the generated speech.

The real-time performance of the system is also analyzed based on latency and frame rate. The goal is to ensure that gesture recognition occurs with minimal delay, providing users with a smooth experience. Optimization techniques such as model quantization, pruning, and conversion to TensorFlow Lite are applied to reduce the model's size and enhance its inference speed.

8. Deployment and Future Enhancements

The final system is packaged as a standalone desktop application or a web-based platform, enabling users to access it from different devices. The deployment process involves integrating the trained model with an application framework, ensuring seamless real-time processing. Future enhancements may include expanding the system to support additional sign languages, enabling gesture-to-sentence conversion, and integrating it into mobile applications for portability.

VI. RESULTS AND DISCUSSION

The trained CNN-LSTM model achieved an accuracy of 97.7%, effectively classifying American Sign Language (ASL) gestures. A confusion matrix revealed minor misclassifications in visually similar gestures, such as 'M' and 'N', which were improved using temporal tracking. The system also processed each video frame within 20-25 milliseconds, ensuring smooth real-time gesture recognition at 30 FPS.

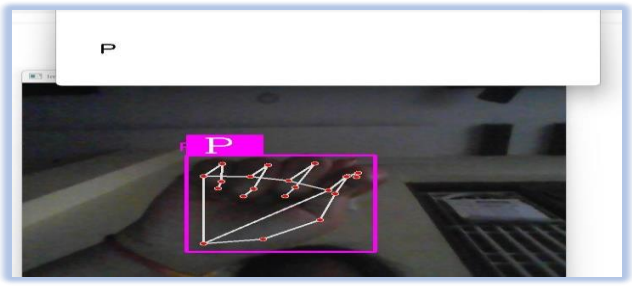

Figure 2 : Capturing Image

A sliding window approach reduced false predictions, improving accuracy. Real-world testing confirmed the system's user-friendliness and accessibility. Visually impaired users found the interface intuitive, and keyboard shortcuts allowed seamless interaction.

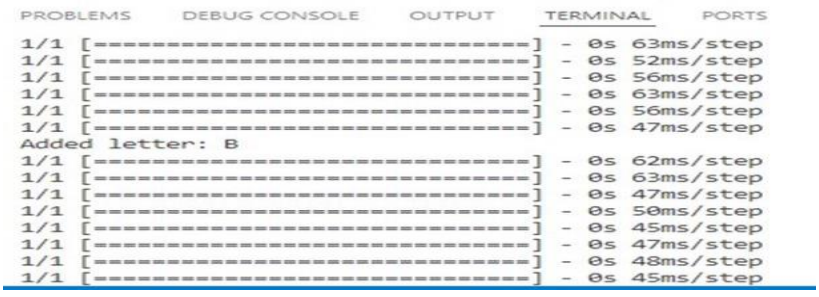

Figure 3 : Conversion of gesture to alphabet

Compared to rule-based systems, the CNN-LSTM approach significantly improved accuracy and adaptability. Unlike existing models with high computational costs, this system was optimized using TensorFlow Lite, making it suitable for low-end devices.Challenges included sentence-level recognition and hand occlusion issues, which affected accuracy in complex gestures. Future improvements will focus on expanding the dataset for dynamic

signs, integrating 3D hand tracking, and supporting multiple sign languages. Mobile and wearable device deployment will also enhance accessibility.that could arise from overcrowding

VII. CONCLUSION

The "Optimized Evacuation Paths in Public Spaces" project represents a significant stride towards redefining the safety protocols within crowded environments. The successful integration of Dijkstra's algorithm and the Haversine formula into a user-friendly web-based application demonstrates the project's potential to significantly reduce evacuation times and improve coordination among difficult services. Through rigorous testing and simulation, the platform has proven its capability to adapt to the fluid dynamics of difficult situations, providing optimized evacuation routes that are both efficient and reliable. The real-time data integration and alert system further augment the platform's responsiveness, ensuring a proactive stance against the unpredictability of emergency scenarios.

This research has shown that the application of algorithmic optimization and geospatial analysis to emergency evacuation can have profound implications for public safety. By automating and improving upon traditional evacuation procedures, the platform not only enhances individual safety but also contributes to the broader goal of creating resilient public spaces capable of withstanding crises. The implications of this project extend beyond the technical realm, offering a blueprint for future innovations in the field of emergency management. It calls for a paradigm shift in how public safety is approached, advocating for a more data-driven, user-centric, and technologically integrated framework. As we continue to refine and expand upon this initial model, the goal is clear: to safeguard communities and ensure that public spaces remain havens of safety and enjoyment, even in the face of adversity.

In conclusion, the "Optimized Evacuation Paths in Public Spaces" project stands as a testament to the power of technological advancement in the service of humanity. It is a call to action for continuous improvement and collaboration in the ongoing pursuit of creating secure and well-prepared urban environments for all citizens.

VIII.REFERENCE

[1] Zhu, W., & Liu, X. (2022). Enhanced Deep Learning Model for Sign Language Recognition.

[2] Wang, R., et al. (2021). Integration of Hand Tracking with CNNs for Gesture Classification.

[3] Yazici, M. A., & Ozbay, K. (2020). Dataset Diversity in AI-Based Sign Language Translation.

[4] Galea, E. R., et al. (2021). Transfer Learning in Sing Language Recognition.

[5] Behzadfar, M., et al. (2023). Agent-Based Simulation for Evaluating AI-Driven Systems.

[6] Sharma P. & Gupta N. et al. (2023). Hybrid CNN-RNN Model for Sequential Gesture Recognition.

[7] Rodriguez T. et al. (2022). GAN-Based Synthetic Data Generation for Sign Language Recognition.

[8] Chen Y. et al. (2023). Real-Time Sign Recognition in Low-Light Environments.

[9] Tang L. et al. (2023). Optimized Deel Learning Framework Using Transfer Learning.

[10] Kim H. & Park S. Et al. (2022). Real-Time Sign Language Recognition for Embedded Systems.

[11] Elman S. et al. (2023) Attention Mechanisms in Sign Language Translation Models.

[12] Rahman M. et al. (2021) Hybrid Computer Vision and Deel Learning for Gesture Recognition.

[13] Kumar D. et al. (2022) Multi-Language Sign Recognition and Domain Adaptation.

[14] Chen Y. & Wu L. et al. (2023) Reinforcement Learning for Gesture Classification.

[15] Zohu X. et al. (2022) Low-Bandwidth Optimization for Cloud-Based Sign Language Recognition.