International Journal of Computer Science Engineering Techniques – Volume 7 Issue 3, May - June- 2023

Sign Language Detector Using Convolutional Neural Network

Vennam Ratna Kumari
Computer science and engineering
ratnab@nigama.org
Nigama Engineering College
Karimnagar Telangana

Swamy Gachikanti
swamygachikanti@gmail.com
Vignana Bharathi Institute of Technology &
Hyderabad

Dr.Mula Veera Hanumantha reddy,
Dept of Computer science and engineering ,
Vivekananda institute of technology and
sciences,Hanuman.vits@gmail.com

ABSTRACT:

The paper utilizes computer vision and machine learning to interpret real-time sign language gestures. It processes video input from a webcam using OpenCV to detect and isolate hand movements. A trained Convolutional Neural Network (CNN) classifies these gestures based on a dataset of American Sign Language (ASL) signs and British Sign Language (BSL) signs. The system then translates recognized gestures into text or spoken language outputs, displayed through a user-friendly interface. This technology aims to bridge communication gaps for deaf or hard-of-hearing individuals and promote inclusivity in educational, professional, and social settings.

KEYWORDS:

Computer vision, Natural Language Processing, Image Processing, Hand Tracking, Recurrent Neural Networks, Long Short Term Memory.

1. INTRODUCTION

This paper aims to develop a deep learning tool that bridges communication gaps for hearing-impaired individuals by recognizing and translating sign language gestures into text or speech in real-time, thereby enhancing inclusivity and accessibility. Addressing the problem of limited communication tools for the hearing-impaired, it focuses on creating an accessible system capable of accurately translating sign language into spoken or written language. Using convolutional neural networks (CNNs), the system processes hand gestures from live video feeds or images with high precision. The workflow begins with input capture through a camera, followed by preprocessing to enhance quality and isolate hand regions using computer vision techniques. Key landmarks, such as fingertips and knuckles, are extracted for gesture analysis, which is crucial for understanding hand shapes and positions. The system leverages labeled datasets to train CNN models for robust gesture classification and ensuring scalability . Applications span in various domains providing practical solutions to bridge the communication gap. While existing systems have made significant advancements in static gesture recognition, they face limitations such as insufficient dataset diversity, challenges in handling complex gestures, and hardware constraints. This project addresses these limitations and aims to incorporate future advancements like 3D pose estimation, natural language processing (NLP) for context-aware translations, and augmented reality (AR) for interactive learning, further expanding its real-world impact and ensuring seamless communication for the Deaf community

2. LITERATURE REVIEW:

1. The focuses on developing an Indian Sign Language Recognition System (ISLRS) for dynamic signs using a Convolutional Neural Network (CNN)[5]. It aims to facilitate communication for the deaf community by interpreting sign language gestures. The model was trained and tested on video clips of dynamic signs, achieving a training accuracy of 70%. Challenges include data collection and handling variations in lighting and backgrounds. The study emphasizes the potential of deep learning techniques, particularly CNNs, to improve accessibility and bridge educational gaps for the hearing-impaired community.

2. the addresses bidirectional sign language translation, converting gestures to text/speech and vice versa, with a focus on Indian Sign Language (ISL)[6]. It introduces a reversible CNN model for gesture recognition and NLP techniques for text-to-sign conversion. The system achieves 70% accuracy, emphasizing dataset expansion and multilingual support for improvement. Challenges include grammatical complexities and dataset diversity. Future goals involve sentence-level translation and regional language adaptability.

3. This study [2] develops a real-time system for American Sign Language (ASL) recognition using Media Pipe for feature extraction and LSTM networks for gesture interpretation. It achieves near-perfect accuracy on a diverse dataset, handling variations in lighting and backgrounds. Data preprocessing includes normalization and augmentation to improve robustness. The system excels in recognizing complex gestures and temporal patterns. Future work focuses on integrating natural language processing for two-way communication.

4. The paper [3] presents a Convolutional Neural Network (CNN)-based system for recognizing static American Sign Language (ASL) gestures, focusing on improving human-computer interaction and accessibility for the hearing-impaired. It involves training on a dataset of 2,000 static images of ASL alphabets captured under varying conditions, using CNNs for feature extraction and classification. The system demonstrates high accuracy in recognizing hand signs and discusses challenges like variability in gestures and lighting conditions. Future directions include expanding the system to recognize dynamic gestures and numerical signs, enhancing real-world applications for accessibility and communication.

5. The paper [4] provides a comprehensive review of image-based Arabic Sign Language Recognition (ArSLR), highlighting methods for recognizing Arabic alphabet signs, isolated words, and continuous sign sequences. It discusses various techniques like Hidden Markov Models, neuro-fuzzy systems, and neural networks, emphasizing their strengths and limitations. The study identifies challenges in vocabulary size, recognition accuracy, and real-life applications while suggesting future directions, such as fusion methods and bidirectional translation systems for improved communication between deaf and hearing individuals. This survey serves as a foundational resource for advancing ArSLR technologies.

6. The paper [6] proposes a sign language recognition system using Long Short-Term Memory (LSTM), a variant of Recurrent Neural Networks (RNN), to capture long-term dependencies in sequential data. The system focuses on recognizing American Sign Language alphabets and employs preprocessing techniques and feature extraction to ensure accuracy. Using 70% of the data for training

and 30% for testing, the model achieves a high accuracy of 98.13%. Challenges include handling environmental variations and expanding the system to recognize gestures beyond alphabets. Future work aims to include dynamic gestures and facial expressions for broader communication.

7. This paper [7] proposes an automated sign language captioning system using TensorFlow, Keras, and LSTM layers. Media Pipe Holistic extracts key points from sign language videos, which are processed for real-time gesture recognition. The system enhances communication accessibility for hearing-impaired individuals. It emphasizes accuracy and adaptability to diverse environments. Future work includes expanding datasets and improving robustness.

8. This research [8] presents a video chat system with real-time sign language recognition using TensorFlow Object Detection and WebRTC. The system translates hand gestures into text to aid deaf and mute users. It uses MobileNet and SSD for efficient object detection and OpenCV for data acquisition. The prototype enhances accessibility and scalability. Future improvements aim at wider language support and dataset expansion.

9. This paper [9] develops a real-time sign language translation system for Indian Sign Language, achieving 92.25% accuracy using Region-based Convolutional Neural Networks (R-CNN). The system bridges communication gaps for the deaf and hard-of-hearing, utilizing multimodal data sources to handle gesture variability. It highlights applications in education, healthcare, and public services, with scalability to diverse languages. The study emphasizes inclusivity and offers a foundation for future assistive technologies.

10. This study [10] introduces a real-time system for recognizing American Sign Language (ASL) alphabets using Convolutional Neural Networks (CNN), with a 99% accuracy. Implemented with Python and TensorFlow, it translates gestures into text for accessible communication. The robust methodology minimizes overfitting, ensuring reliable performance. While dataset limitations exist, the work lays a foundation for recognizing words and phrases in future applications.

3. METHODOLOGY:

Proposed systems: The system aims to detect and classify hand gestures into predefined sign language labels (e.g., "Hello," "Thank you," "Yes") in real-time. It uses a webcam to capture hand gestures, processes the images using OpenCV, detects hand features using the HandDetector module, and classifies the gestures using a Convolutional Neural Network (CNN) model implemented with TensorFlow/Keras.

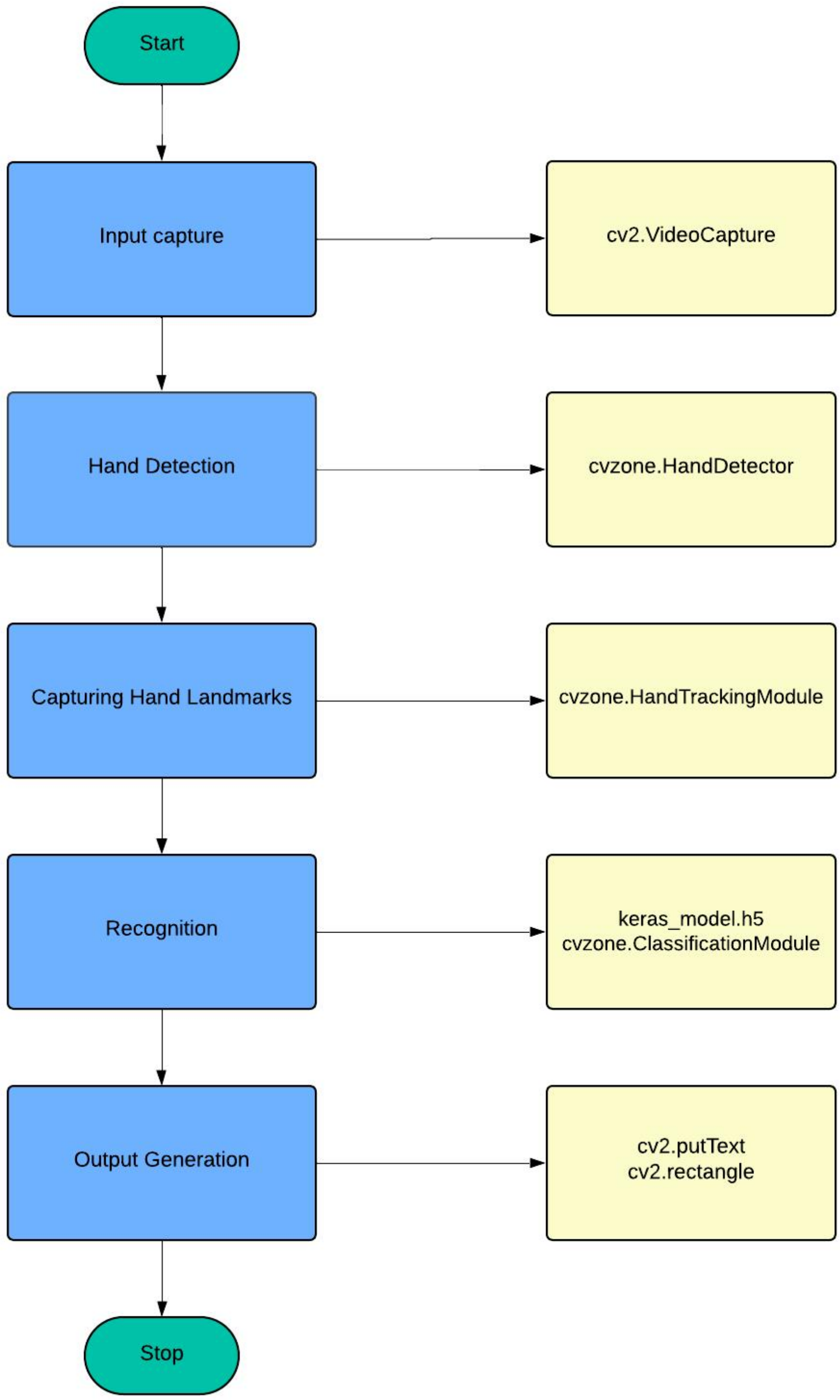

Fig 1: Architecture

A CNN-based sign language detection system starts by capturing real-time video or images of hand gestures through a camera. Tools like OpenCV process the input to isolate the hand region, removing background noise. Key hand landmarks, such as fingertips and knuckles, are identified using machine learning models for accurate gesture representation. These landmarks are fed into a CNN trained on gesture datasets to classify them into sign language symbols or words. Deep learning is utilized to interpret visual information effectively. The recognized gestures are then translated into text or speech, enabling communication for those unfamiliar with sign language. Real-time feedback ensures a seamless user experience, with system accuracy relying on high-quality input, preprocessing, and robust training.

K-NN is a simple algorithm used for classifying gestures based on extracted features such as hand landmarks or shape descriptors. It compares the input gesture with labeled examples and assigns the label of the closest neighbors in the feature space. While effective for small datasets, k-NN is computationally expensive for larger datasets and does not leverage deep feature learning like CNNs. Backpropagation is essential for training neural networks, including CNNs, by calculating the error between predicted and actual labels. It propagates this error backward through the network to adjust weights and improve accuracy. Backpropagation allows deep learning models to automatically learn complex features from raw data, making it ideal for tasks like sign language recognition, though it requires large datasets and significant computational resources.

The dataset for sign language recognition consists of two parts: a publicly available dataset from Kaggle and a custom dataset created by capturing images of hand gestures. The Kaggle dataset provides a diverse collection of labelled sign language gestures, which helps in training the model on standard hand gestures. Additionally, a custom dataset was created by capturing images of specific hand gestures, ensuring that it aligns with the target sign language system. This combination of existing and custom data enhances the model's ability to generalize and improve recognition accuracy across varied gesture types and conditions.

4. RESULTS AND DISCUSSION

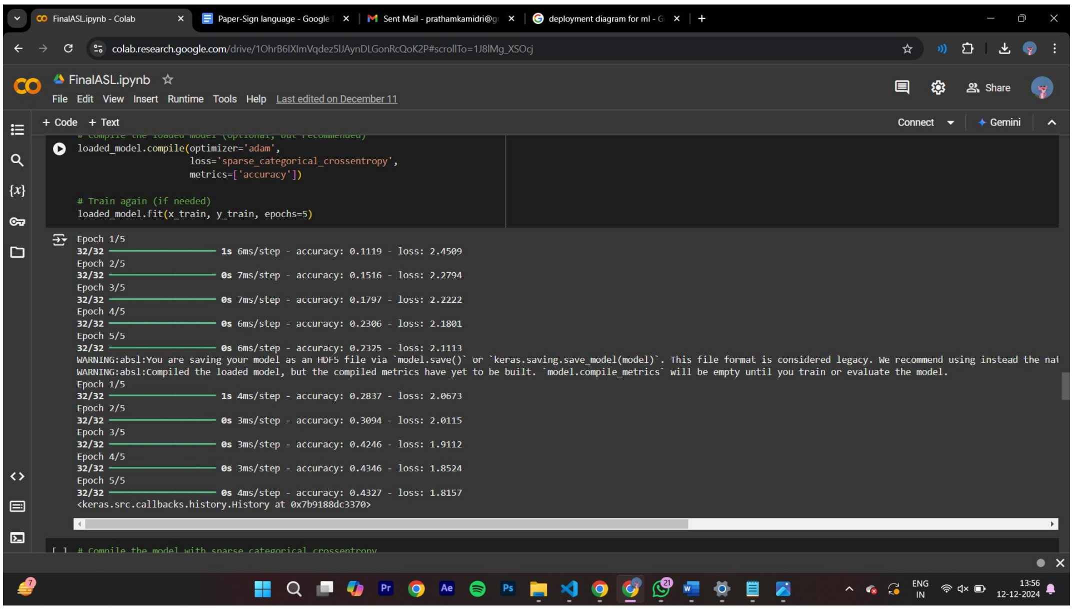

Fig 4 Model Training

Model: "sequential_1"

Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 62, 62, 32)	896
max_pooling2d (MaxPooling2D)	(None, 31, 31, 32)	0
conv2d_1 (Conv2D)	(None, 29, 29, 64)	18,496
max_pooling2d_1 (MaxPooling2D)	(None, 14, 14, 64)	0
conv2d_2 (Conv2D)	(None, 12, 12, 128)	73,856
max_pooling2d_2 (MaxPooling2D)	(None, 6, 6, 128)	0
flatten (Flatten)	(None, 4608)	0
dense (Dense)	(None, 128)	589,952
dropout (Dropout)	(None, 128)	0
dense_1 (Dense)	(None, 37)	4,773

Total params: 687,973 (2.62 MB)
Trainable params: 687,973 (2.62 MB)
Non-trainable params: 0 (0.00 B)

Fig.6 CNN Layers

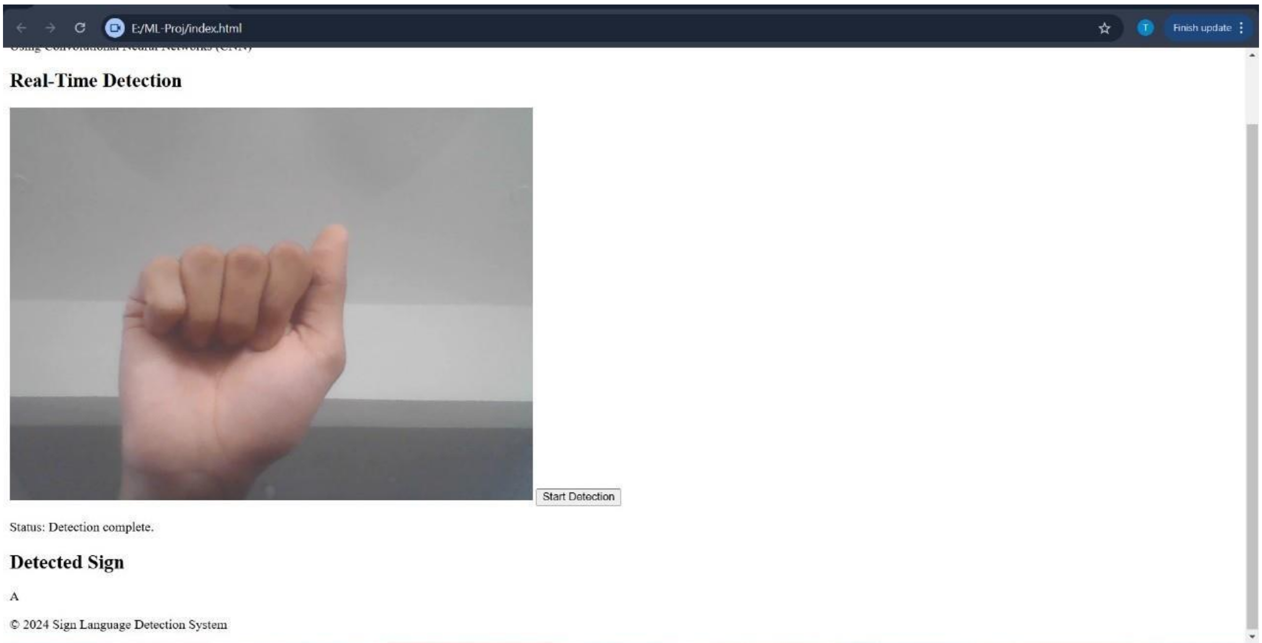

Fig 7 HTML page

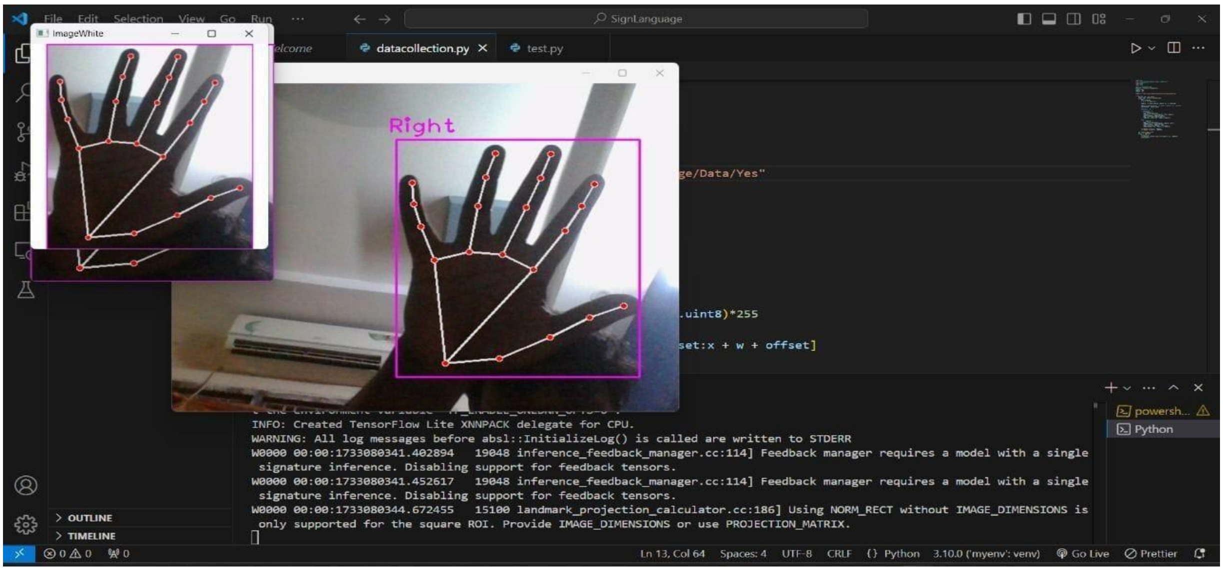

Fig 9 Right hand representation

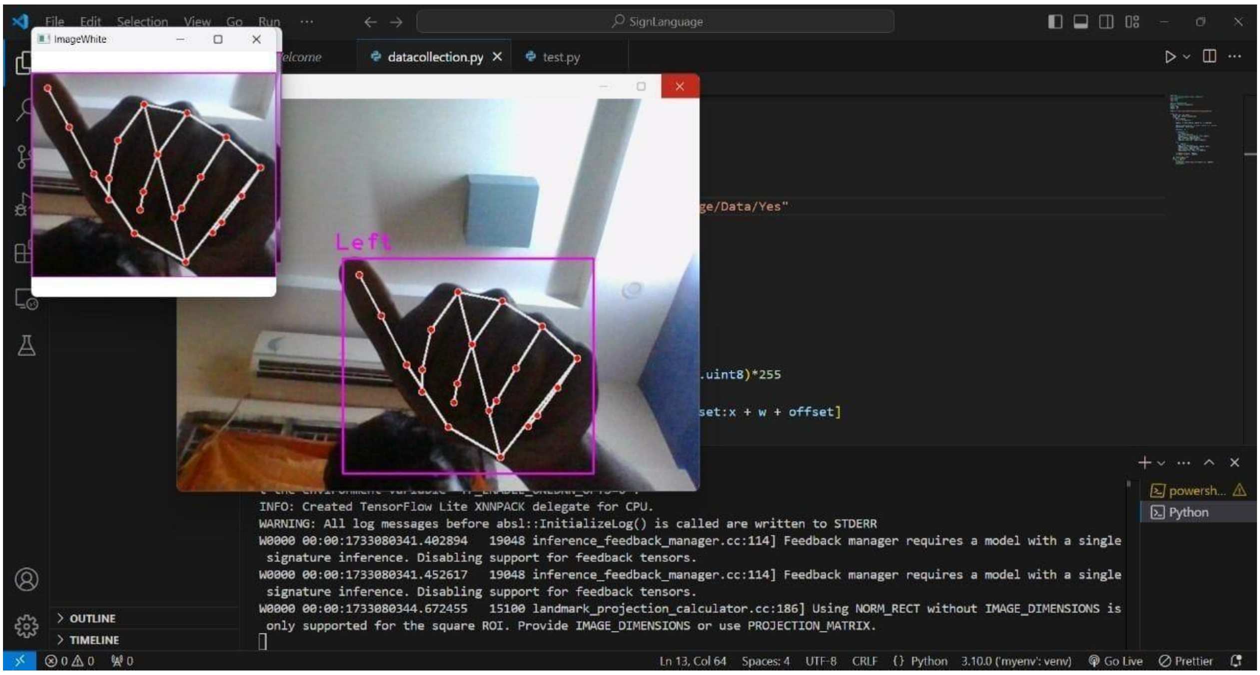

Fig 4.5 Left hand representation

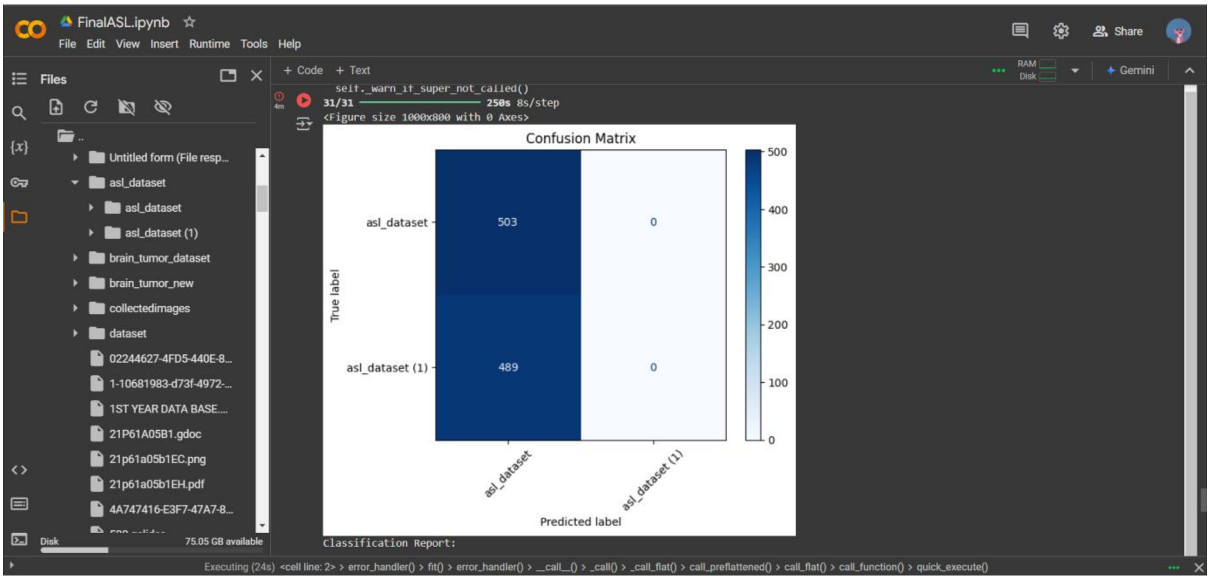

Fig 4.6 Confusion Matrix

Classification Report:

	precision	recall	f1-score	support
asl_dataset	0.51	1.00	0.67	503
asl_dataset (1)	0.00	0.00	0.00	489
accuracy			0.51	992
macro avg	0.25	0.50	0.34	992
weighted avg	0.26	0.51	0.34	992

Fig 4.7 Classification Report

5. CONCLUSION AND FUTURE SCOPE

Deploying a convolutional neural network (CNN) for sign language detection has yielded promising results in accurately identifying and interpreting static hand gestures, effectively bridging a crucial communication gap and offering a tool for facilitating interactions between sign language users and those unfamiliar with it. By utilizing a well-curated dataset and implementing effective preprocessing and optimization techniques, the model achieved impressive accuracy and robustness, highlighting the potential of deep learning and computer vision technologies in creating inclusive communication solutions. Looking ahead, there is ample potential for enhancing and scaling the project, such as incorporating dynamic gesture and full sentence recognition to expand its practical applications. Real-time deployment on devices and integration with video processing systems can make the system more accessible and user-friendly. Additionally, including multimodal inputs like facial expressions and lip movements can offer a comprehensive approach to sign language interpretation. Expanding the dataset to include diverse hand shapes, skin tones, and environmental conditions will further improve the system's adaptability and accuracy. Moreover, developing a user-friendly interface and offering extensive training materials can facilitate widespread adoption and usage of the technology. Partnerships with educational institutions and organizations serving the deaf and hard-of-hearing community can also play a crucial role in the deployment and enhancement of the system. Overall, this project establishes a solid foundation for advancing assistive technologies aimed at promoting inclusivity and accessibility.

The future scope for sign language detection using CNNs is extensive. Enhancing the system to recognize dynamic gestures and full sentences can make it more practical. Real-time deployment on various devices and integrating advanced video processing systems can improve accessibility and user experience. Incorporating multimodal inputs like facial expressions and lip movements can offer a more comprehensive approach to sign language interpretation. Expanding the dataset to include diverse hand shapes, skin tones, and environmental conditions will boost adaptability and accuracy.

Collaborations with educational institutions and organizations for the deaf and hard of hearing can refine and promote the technology, paving the way for more inclusive communication solutions. Integrating AI with technologies like augmented reality (AR) and virtual reality (VR) can offer immersive learning and communication experiences. Improving model efficiency and reducing computational costs can facilitate widespread adoption while developing robust privacy and security measures will ensure user trust and data protection.

REFERENCES

[1]. K. S. Sindhu, Mehnaaz, B. Nikitha, P. L. Varma and C. Uddagiri, "Sign Language Recognition and Translation Systems for Enhanced Communication for the Hearing Impaired," 2024 1st International Conference on Cognitive, Green and Ubiquitous Computing (IC-CGU), Bhubaneswar, India, 2024, pp. 1-6, doi: 10.1109/IC-CGU58078.2024.10530832.

[2]. J. Debnath and P. J. I R, "Real-Time Gesture Based Sign Language Recognition System," 2024 International Conference on Advances in Data Engineering and Intelligent Computing Systems (ADICS), Chennai, India, 2024, pp. 01-06, doi: 10.1109/ADICS58448.2024.10533518.

[3]. P. Kumar, P. Chanakya and S. Velusamy, "Sign Language Recognition System Using Convolutional Neural Network," 2024 IEEE International Conference on Information Technology, Electronics and Intelligent Communication Systems (ICITEICS), Bangalore, India, 2024, pp. 1-4, doi: 10.1109/ICITEICS61368.2024.10625127.

[4]. M. Mohandes, Junzhao Liu and M. Deriche, "A survey of image-based Arabic sign language recognition," 2014 IEEE 11th International Multi-Conference on Systems, Signals & Devices (SSD14), Barcelona, Spain, 2014, pp. 1-4, doi: 10.1109/SSD.2014.6808906.

[5]. A. Singh, A. Wadhawan, M. Rakhra, U. Mittal, A. A. Ahdal and S. K. Jha, "Indian Sign Language Recognition System for Dynamic Signs," 2022 10th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), Noida, India, 2022, pp. 1-6, doi: 10.1109/ICRITO56286.2022.9964940.

[6]. M. Soundarya, M. Yazhini, N. Thirumala Sree, N. Sornamalaya and C. Vinitha, "Sign Language Recognition Using Machine Learning," 2024 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI), Chennai, India, 2024, pp. 1-7, doi: 10.1109/ACCAI61061.2024.10602025.

[7]. R. G. Sakthivelan, N. Kumar A and R. G, "Sign Language Caption Generation Using LSTM," 2024 IEEE 9th International Conference for Convergence in Technology (I2CT), Pune, India, 2024, pp. 1-5, doi: 10.1109/I2CT61223.2024.10543890.

[8]. S. K. K L, S. M, S. A, S. J. U, and V. S. P V, "Video Chat using WebRTC with In-built Sign Language Detection," 2023 Third International Conference on Ubiquitous Computing and Intelligent Information Systems (ICUIS), Gobichettipalayam, India, 2023, pp. 167-170, doi: 10.1109/ICUIS60567.2023.00036.keywords: {Sign language;Visualization;Tensors;Prototypes;Object detection;Web conferencing;Ubiquitous computing;Web Real-Time Communications;Tensor flow object detection;Video chat;Sign detection}.

[9]. E. Sharon, G. J. Leelipushpam Paulraj, I. Johnraja Jebadurai and C. Merlin, "Sign Language Translation to Natural Voice Output: A Machine Learning Perspective," 2024 International Conference on Cognitive Robotics and Intelligent Systems (ICC - ROBINS), Coimbatore, India, 2024, pp. 332-339, doi: 10.1109/ICC-ROBINS60238.2024.10533916.

[10]. S. Kurundkar, A. Joshi, A. Thaploo, S. Auti and A. Awalgaonkar, "Real-Time Sign Language Detection," 2023 2nd International Conference on Vision Towards Emerging Trends in Communication and Networking Technologies (ViTECoN), Vellore, India, 2023, pp. 1-4, doi: 10.1109/ViTECoN58111.2023.10157784.