https://doi.org/10.5281/zenodo.15013597

ISSN: 2456-1134 www.isjcresm.com
Vol-09 Issue-01 Mar 2024

Sign Language Detector using Convolutional Neural Network

S. Pothalaiah[1], M Amru[2], Kranti Kumar Appari[3],
[123]Department of ECE, CMR Engineering College
[1]Vignana Bharathi Institute of Technology, Hyderabad, India
E-mail: pothalaiahs@gmail.com malothuamru@gmail.com , krantikumara@gmail.com

ABSTRACT:

In the paper the study employs computer vision and machine learning to analyze real-time sign language gestures. It captures video input from a webcam and utilizes OpenCV to detect and track hand movements. A pre-trained Convolutional Neural Network (CNN) then classifies these gestures based on a dataset containing American Sign Language (ASL) and British Sign Language (BSL) signs. The system converts the recognized gestures into text or speech, which is displayed through an intuitive user interface. This technology seeks to enhance communication for deaf or hard-of-hearing individuals, fostering inclusivity in educational, professional, and social environments..

KEYWORDS:

Natural Language Processing, Image Processing, Hand Tracking, Recurrent Neural Networks, Long ShortTerm Memory.

1. INTRODUCTION

This paper proposes the development of a deep learning tool designed to bridge communication barriers for hearing-impaired individuals by recognizing and translating sign language gestures into text or speech in real-time. The objective is to enhance accessibility and inclusivity by providing an efficient and accurate communication system. Addressing the limitations of existing tools, this research focuses on creating an accessible solution capable of accurately converting sign language into written or spoken language.

The system utilizes convolutional neural networks (CNNs) to process hand gestures captured from live video feeds or images with high precision. The workflow begins with capturing input via a camera, followed by preprocessing techniques to enhance image quality and isolate hand regions using computer vision. Key landmarks, such as fingertips and knuckles, are extracted to analyze hand shapes and positions for gesture interpretation. Labeled datasets are used to train CNN models, ensuring robust and scalable gesture classification.

This technology finds applications across multiple domains, offering practical solutions to bridge the communication gap. While existing systems have significantly advanced static gesture recognition, they still face challenges such as limited dataset diversity, difficulties in recognizing complex gestures, and hardware constraints. This project addresses these limitations and explores future enhancements, including 3D pose estimation, natural language processing (NLP) for context-aware translations, and augmented reality (AR) for interactive learning. By integrating these advancements, the system aims to further improve real-world communication accessibility for the Deaf community.

ISSN: 2456-1134 www.isjcresm.com
Vol-09 Issue-01 Mar 2024

2. LITERATURE REVIEW:

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.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.

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.

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.

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.

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.

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.

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., 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 in real-time. It capture hand gestures, processes the images, detects hand features using the HandDetector module, and classifies the gestures using a Convolutional Neural Network (CNN) model implemented.

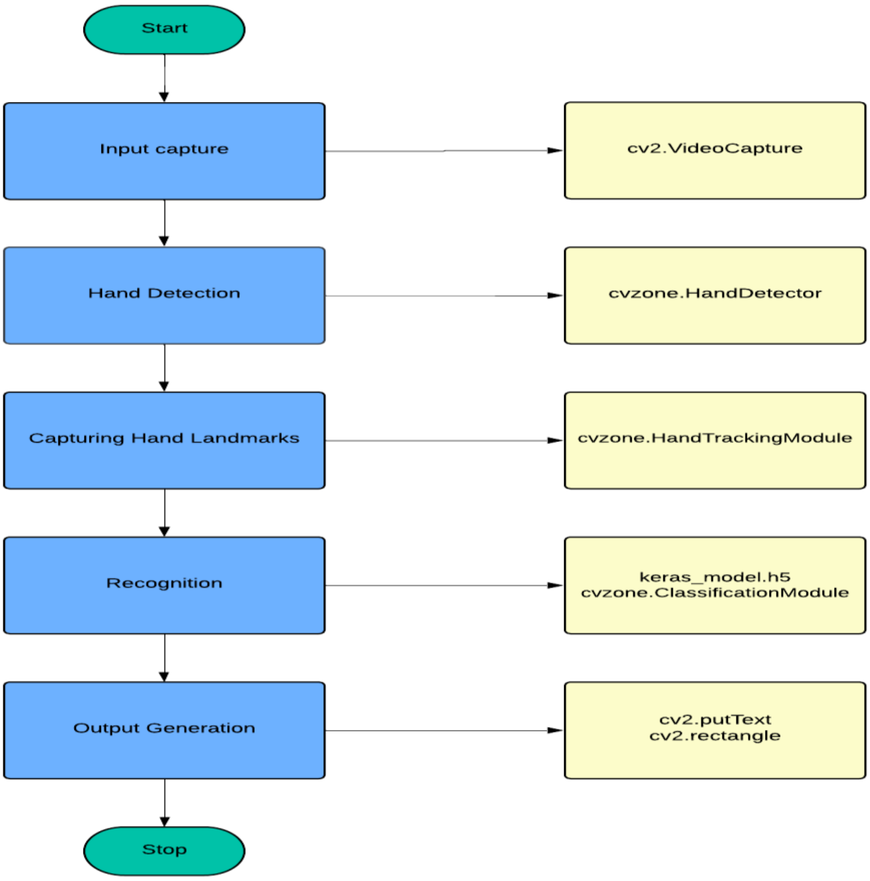

Fig 1: Architecture

A CNN-based sign language detection system begins by capturing real-time video or images of hand gestures using a camera. OpenCV processes the input to isolate the hand region and eliminate background noise. Machine learning models identify key hand landmarks, such as fingertips and knuckles, ensuring accurate gesture representation. These extracted landmarks are then fed into a CNN trained on gesture datasets, enabling classification into sign language symbols or words. Deep learning facilitates effective interpretation of visual data. Once recognized, the gestures are converted into text or speech, aiding communication for individuals unfamiliar with sign language. Real-time feedback enhances user experience, with system accuracy depending on high-quality input, preprocessing, and robust training.

K-Nearest Neighbors (k-NN) is a simple classification algorithm that assigns labels to gestures by comparing extracted features—such as hand landmarks or shape descriptors—with labelled examples. It determines the closest matches within the feature space and classifies gestures accordingly. While k-NN is effective for small datasets, it becomes computationally expensive for larger datasets and lacks the deep feature learning capabilities of CNNs. In contrast,

ISSN: 2456-1134 www.isjcresm.com
Vol-09 Issue-01 Mar 2024

backpropagation is a crucial technique for training neural networks, including CNNs, by computing the error between predicted and actual labels. This error is propagated backward through the network to adjust weights, improving accuracy. Backpropagation enables deep learning models to automatically learn complex patterns from raw data, making them well-suited for sign language recognition, though they require large datasets and significant computational resources.

The dataset for sign language recognition comprises two components: a publicly available dataset from Kaggle and a custom dataset created by capturing images of hand gestures. The Kaggle dataset offers a diverse collection of labelled sign language gestures, providing a strong foundation for training the model on standard hand signs. Additionally, a custom dataset was developed to include specific hand gestures aligned with the target sign language system. The integration of both datasets improves the model's ability to generalize, enhancing recognition accuracy across a variety of gesture types and conditions.

4. RESULTS AND DISCUSSION

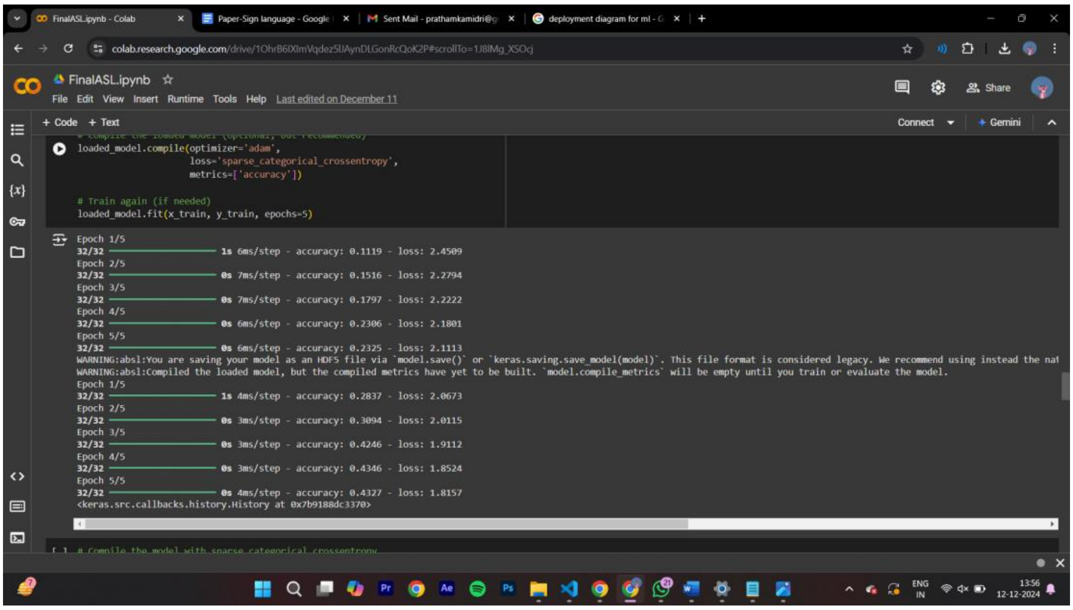

Fig 4 Model Training

ISSN: 2456-1134 www.isjcresm.com
Vol-09 Issue-01 Mar 2024

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

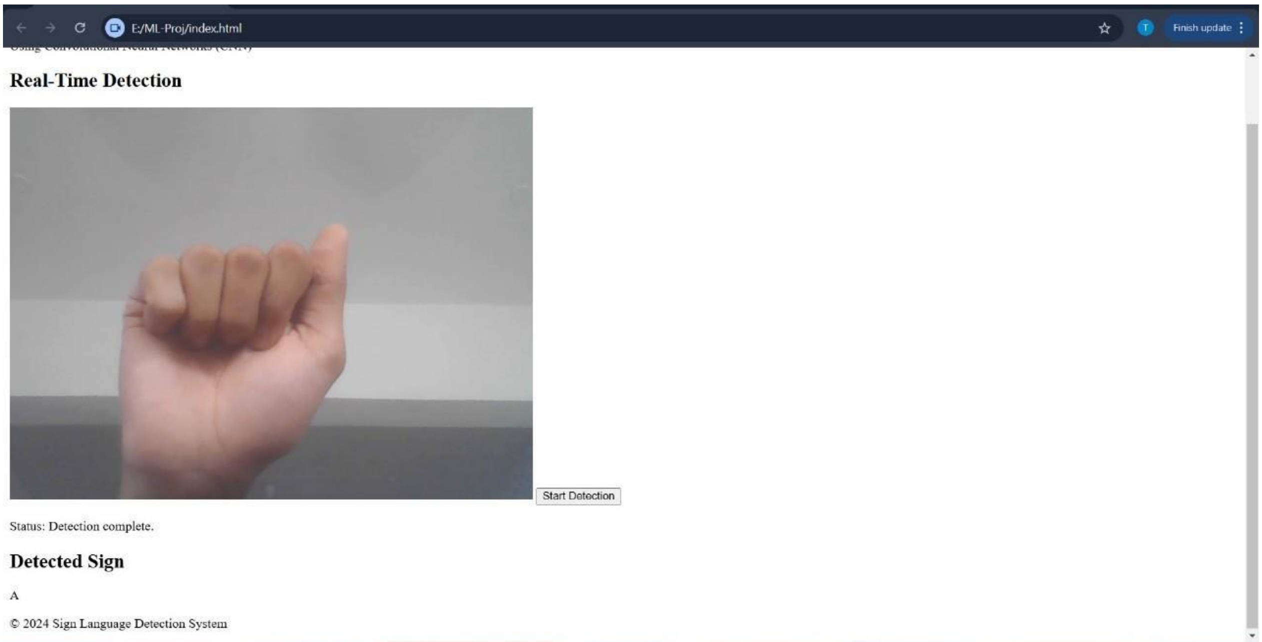

Fig 7 HTML page

ISSN: 2456-1134 www.isjcresm.com
Vol-09 Issue-01 Mar 2024

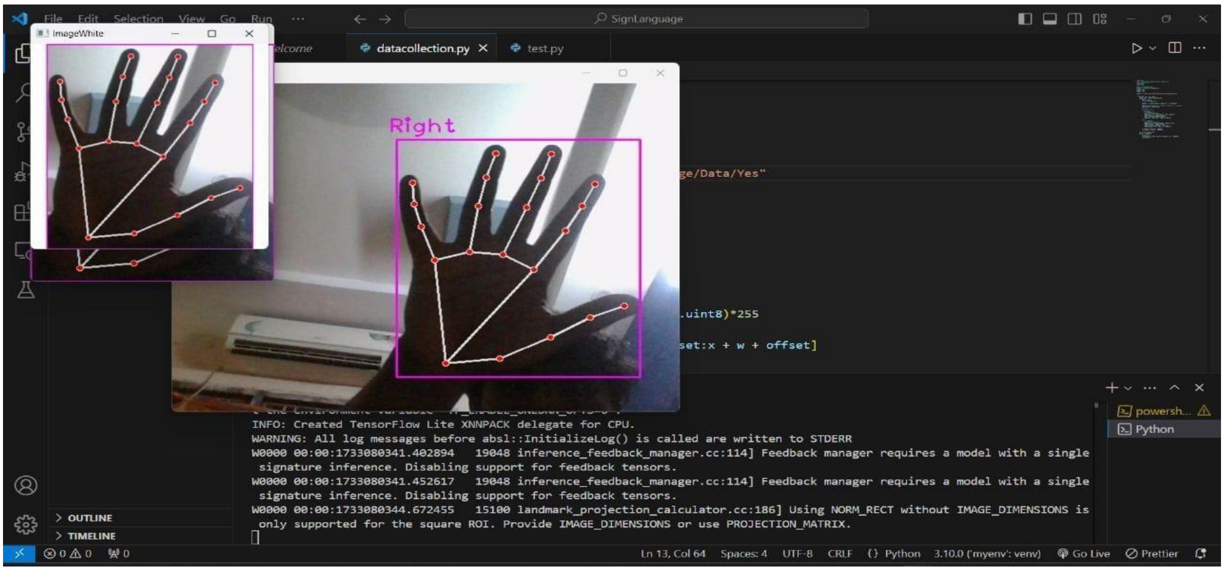

Fig 9 Right hand representation

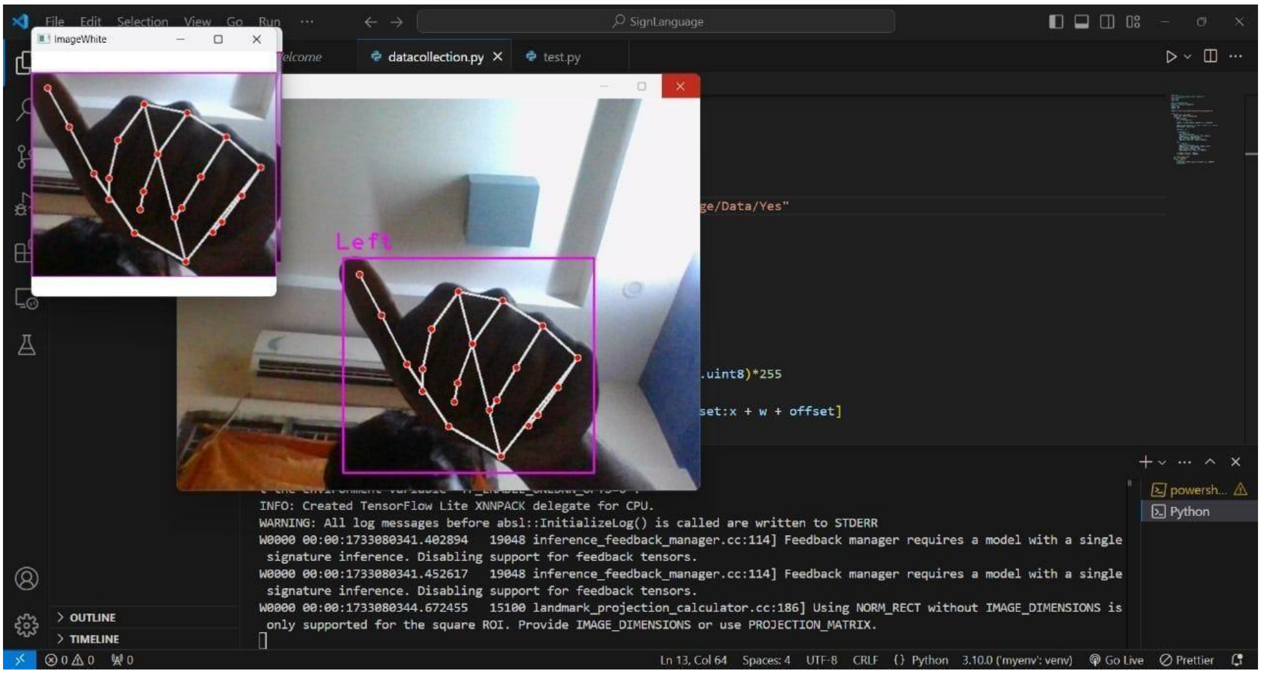

Fig 4.5 Left hand representation

ISSN: 2456-1134 www.isjcresm.com
Vol-09 Issue-01 Mar 2024

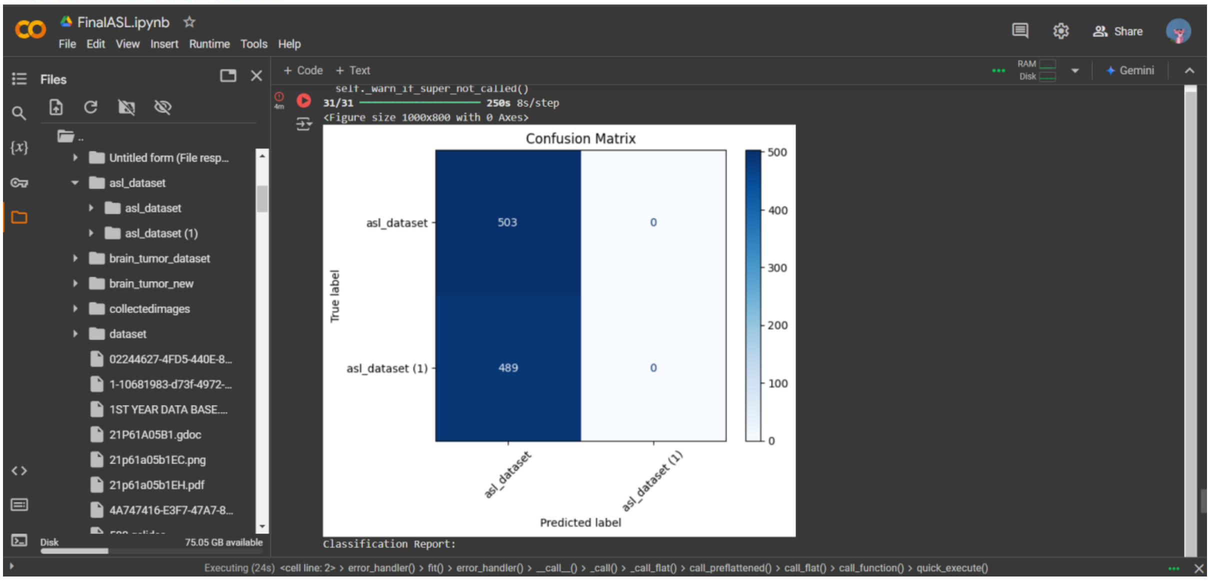

Fig 4.6 Confusion Matrix

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 demonstrated significant success in accurately recognizing and interpreting static hand gestures. This advancement helps bridge a crucial communication gap, providing a valuable tool for interactions between sign language users and those unfamiliar with it. By leveraging a well-curated dataset and applying effective preprocessing and optimization techniques, the model has achieved remarkable accuracy and robustness, underscoring the potential of deep learning and computer vision in fostering inclusive communication solutions.

Looking ahead, there are numerous opportunities to enhance and scale the system. Incorporating dynamic gesture and full sentence recognition can broaden its practical applications. Real-time deployment on various devices and integration with video processing systems will enhance accessibility and

ISSN: 2456-1134 www.isjcresm.com
Vol-09 Issue-01 Mar 2024

usability. Additionally, incorporating multimodal inputs, such as facial expressions and lip movements, can provide a more comprehensive approach to sign language interpretation. Expanding the dataset to include a wider range of hand shapes, skin tones, and environmental conditions will further improve adaptability and recognition accuracy.

Developing a user-friendly interface and providing extensive training materials can encourage widespread adoption. Collaborations with educational institutions and organizations supporting the deaf and hard-of-hearing community can drive further refinement and promotion of the technology. Furthermore, integrating AI with emerging technologies like augmented reality (AR) and virtual reality (VR) can create immersive learning and communication experiences. Enhancing model efficiency and minimizing computational costs will enable broader adoption, while implementing robust privacy and security measures will ensure user trust and data protection.

Overall, this paper lays a strong foundation for advancing assistive technologies that promote inclusivity and accessibility, paving the way for more effective and widespread communication solutions.

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, Idia, 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 (AICS), Chennai, India, 2024, pp. 01-06, doi: 10.1109/ADICS58448.2024.10533518.

[[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.

[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.