ISSN:2230-9993

INDIAN SIGN LANGUAGE RECOGNITION USING YOLOV5

ANU ROSE JOY
Dept. of Computer Science & Engineering
Amal Jyothi College of Engineering
kottayam, India
anurosejoy@amaljyothi.ac.in

ANSU MARIUM SHIBU
Dept. of Computer Science & Engineering
Amal Jyothi College of Engineering
kottayam, India
ansumariumshibu2024@cs.ajce.in

ANSPIN TITUS
Dept. of Computer Science & Engineering
Amal Jyothi College of Engineering
kottayam, India
anspintitus2024@cs.ajce.in

ASNA AZEEZ
Dept. of Computer Science &Engineering
Amal Jyothi College of Engineering
Kottayam, India
asnaazeez2024@cs.ajce.in

FATHIMA P.S
Dept. of Computer Science & Engineering
Amal Jyothi College of Engineering
Kottayam, India
fathimaps2024@cs.ajce.in

***Abstract*—In our rapidly advancing technological era, characterized by the ubiquity of home automation and a demand for streamlined solutions, a project unfolds with the mission to address communication challenges for individuals with hearing and speech impairments. Sign language, a vital mode of expression for the deaf and mute, forms the focal point of this initiative. Utilizing sophisticated Deep Learning algorithms including YOLOv5 the project aims to analyze and interpret sign language gestures from input images. The ultimate goal is to seamlessly translate these gestures into text and, subsequently, into audio, thereby providing an encompassing communication solution. A diverse dataset, encompassing English letters, numbers, and words, enhances the system's proficiency. This endeavor not only embraces technological progress but, more importantly, champions inclusivity by breaking down communication barriers for those who have long faced challenges in expressing themselves effectively.**

Keywords: YOLOv5, ISL, Words Recognition, Static Gesture Recognition

I. INTRODUCTION

In the dynamic landscape of assistive technology, the pursuit of comprehensive sign language recognition and detection systems has emerged as a critical endeavor, promising to revolutionize communication for individuals with hearing impairments. Sign language, a rich and nuanced form of expression, serves as a primary mode of communication for the deaf community. However, unlocking its full potential in an interconnected world requires advanced technologies capable of accurately interpreting and translating sign language gestures. This paper delves into the transformative realm of sign language recognition and detection, focusing on the utilization of sophisticated algorithms, notably YOLOv5, to propel these technologies into new frontiers.

Sign language is a visual language that uses movements to convey meaning and convey emotions. When a person with a disability want to share their ideas and opinions with the broader society, there is a communication gap. These two groups currently depend mostly on cumbersome and expensive human interpreters. There are numerous sign languages used throughout the world, including American, Indian, British, and Japanese sign languages. Every language has its own distinctive characteristics, such as the use of one or both hands and the expressions made on the face to indicate emotion.

The majority of signs in Indian Sign Language are made using both hands. This makes it difficult to recognize Indian Sign Language since it requires keeping track of a person's hand orientation and location. A complete expert system is necessary to make daily communication easier for those with special needs. The development of an automatic sign-to-text translator could benefit greatly from this kind of technology.

Sign language, with its diverse vocabulary of gestures and expressions, presents a unique set of challenges for recognition and detection systems. Traditional methods often fall short in capturing the subtleties of sign language, limiting the effectiveness of communication tools for the deaf. Enter YOLOv5, a cutting-edge algorithm renowned for its prowess in real-time object detection. The integration of YOLOv5 into the realm of sign language recognition holds great promise,

ISSN:2230-9993

offering the potential to identify and interpret sign language gestures with remarkable precision and efficiency.

In this study, hand signs from the Indian Sign Language are categorized using YOLOv5.This study suggests a system that can correctly categorize and identify hand motions from a wide terms from Indian Sign Language vocabulary. The primary driving force behind this effort is the model's ability to be quickly and simply implemented on edge devices to develop further solutions that will significantly benefit the hearing impaired population. Additionally, the model makes it easier for the community to communicate with the wider public.

One of the key advancements facilitated by YOLOv5 is its ability to seamlessly convert detected sign language gestures into audible speech, thus bridging the communication gap between the deaf and the hearing world. This integration not only enhances accessibility for the deaf community but also contributes to a more inclusive and interconnected society.

II. LITERATURE SURVEY

Creating a machine learning-based system for continuous word-level sign language recognition is the aim of the research described in article [1]. The system uses SVM with mediapipe and YOLOv4 to classify hand signs from Indian Sign Language. The project's self-created dataset consists of eighty static indicators. The technique correctly recognizes a larger repertoire of words, which facilitates communication with the general public for people with hearing problems. The two recommended methods must be combined to create an expert system in order to boost overall accuracy.

The development of Internet of Things-based technology for hearing, vision, and voice aid in [2]. Using a Raspberry Pi and the Google Cloud Vision API, the device can speak text to speech. It also has a camera for taking images of printed documents. Additionally, it describes the operation of the blind module, which entails taking a picture of the text, uploading it to the Google Cloud Vision API for text extraction, and subsequently using the gTTS API to transcribe the text into speech. It also discusses potential developments and applications of IoT technology in the future to support those with disabilities

The study of [3] compares the performance of four pre-trained deep models with a customized three-layered CNN model using five gradient-based optimizers and their corresponding hyperparameters. The proposed CNN model achieved outstanding recognition accuracies for ISL gestures, outperforming the pre-trained models. The findings have important implications for deep learning models ISL identification performance as well as potential ramifications for the development of assistive technology for Indian sign language users.

It [4] investigates how hand gesture detection and wearable technology can enhance communication between people, especially in difficult auditory settings with deaf individuals. While noting the revolutionary potential of wearable technology to inspire creative human-computer interactions, it also draws attention to shortcomings in earlier research concerning wearability and practicality. The analysis underscores how important it is to capture both static postures and dynamic activities in order to have a thorough knowledge of gestures. By highlighting the necessity of embedded learning systems using acceleration and strain detection for real-time pose and gesture classification, it sets the stage for the contributions made by the paper. Overall, the analysis emphasizes wearable technological improvements for successful communication and interaction, laying a framework for present efforts.

The application of a multi-stream 3D Convolutional Neural Network (CNN) for automatic sign language translation is examined in this [5] study. It presents the results, discusses the implementation of this strategy, and provides explanations and examples of the models used. The study emphasizes the value of gesture recognition in sign language translation by focusing on the benefits of artificial depth maps and the potential for generalization across different signers. The authors further state that they have no conflicting interests and express gratitude for the encouragement to continue working on their project.

For the benefit of people who are deaf-dumb, this [6] study offers a method for identifying the Tamil Sign Language Alphabet. The proposed approach converts binary integers—which represent finger locations—into decimal values, which are subsequently translated into Tamil letters. The process includes training and testing stages, feature point extraction, palm extraction, and feature extraction for both static and dynamic sign language photos. In order to improve communication, future studies will look at various hand positions, backgrounds, and the potential to convert Tamil text to speech. The precision of the system appears good.

The research [7] presents a novel technique that uses a lightweight deep learning model to identify hand motions from low-resolution thermal pictures. For data processing, a MEMS thermal camera is linked to a Raspberry Pi. Discussion is had regarding the machine learning model, dataset specifics, experimental design, and outcomes. The technology intends to improve the efficiency and accuracy of hand gesture recognition using thermal images, and it will enable contactless communication in a variety of industries, including entertainment, medical diagnostics, and crisis management.

This paper presents a thorough [8] analysis of deep learning based methods for sign language recognition. A novel RGB+D dataset for Greek sign language is presented, implemented architectures are described, sequence training requirements are suggested, experimental results are reported, lessons learned from the experiments are discussed, and future research objectives are outlined. A comparative experimental evaluation of computer vision-based methods is also included. The paper evaluates various deep neural network topologies and training conditions for sign language identification using cutting-edge findings and fresh perspectives.

A [9] deep learning-based method that can identify Indian Sign Language (ISL) motions and help hearing-impaired people in an emergency. It uses cutting-edge techniques for object recognition and classification, such as a VGG + LSTM model for better gesture classification and promising outcomes for dynamic gesture object detection. The study consists of a review of the literature, an examination of films showing emergency signs, and a comparison of model performance.

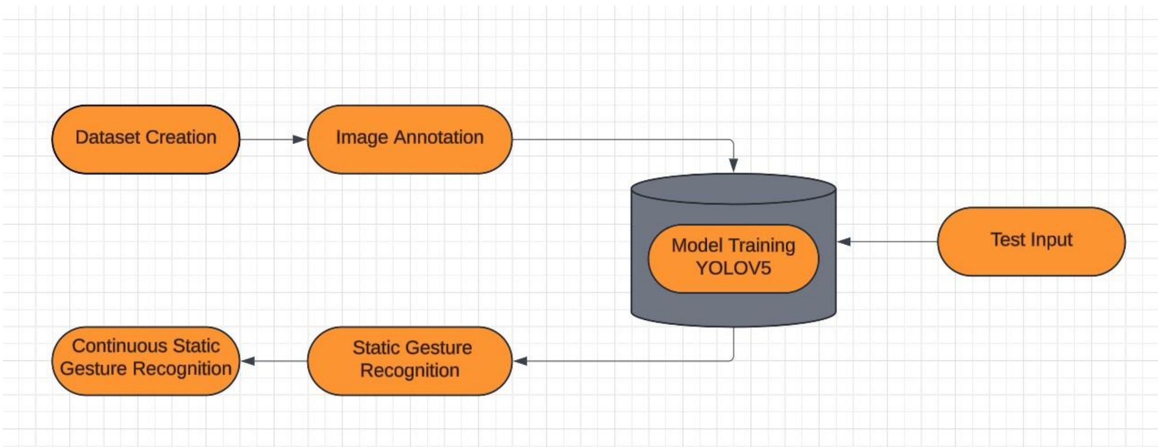


Fig. 1. Block Diagram of the Proposed System

The study covers model construction, experimental results, conclusions, deep learning network techniques, and gesture detection methods.

A hybrid method [10] combines a random forest classifier and a deep transfer learning model to recognize Bangla sign language. Its objective is to improve the current strategy. strategies with reference to recall, accuracy, precision, and f1score. A comprehensive methodology comprising a backdrop elimination module, assessment metrics, experimental analysis, findings, and future research goals is used in this review of the literature on sign language recognition systems. It highlights how important effective communication is for both normal and hearing-impaired people, and it implies that the suggested technique can help with this, leading to significant socioeconomic changes.

III. METHODOLOGY

A. Dataset Creation

Dataset is a comprehensive and meticulously curated collection designed to advance the development of sign language recognition technologies. It contains diverse vocabulary of words commonly used in everyday communication. The chosen dataset consists of 18 different words of ISL for sign language detection. It offers a focused collection of sign language expressions representing key phrases used in various communicative contexts. The dataset includes distinct sign gestures for common expressions such as ”Hello”, ”Thanks”, ”No”, ”Yes”, ”Attendance”, ”Bad”, ”Beautiful”, ”Call”, ”Drink”, ”Love”, ”Smile”, ”Study”, ”Truth”, ”Please”, ”Break”, ”Iloveyou”, ”Brush”, ”You” are all integral to effective communication in ISL.

B. Image Annotation

Image Annotation is the process of labelling the image and identifying the features of the object. It helps in identifying the hand gestures given in real-time. Labellmg is used to annotate each sign’s image in this project.

C. Model Training Using YOLOv5

YOLOv5, or ”You Only Look Once” version 5 is a real-time object detection system in the YOLO family. It takes a distinct tack by approaching object detection as a regression problem. The approach allows for real-time processing by directly predicting bounding boxes and class probabilities from the input image in a single forward pass. In order to improve its adaptability to a variety of datasets, YOLOv5 uses anchor boxes to make it easier to forecast bounding boxes for objects with varying sizes and forms. The design uses a YOLO head for accurate bounding box predictions, PANet for feature aggregation, and CSPDarknet53 as the backbone. With its several model sizes (YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x), YOLOv5 gives customers the flexibility to select the best balance between speed and accuracy for their particular requirements. YOLOv5, which is implemented in PyTorch, is a flexible and popular tool for real-time object identification in robotics, autonomous systems, and surveillance since it allows for training on bespoke datasets.

D. Static Gesture Recognition

The YOLOv5 model predicts the static gestures shown by the users. The predicted object contains the confidence values and labels. These confidence values are used to predict the hand gestures in real-time.

ISSN 2230-9993

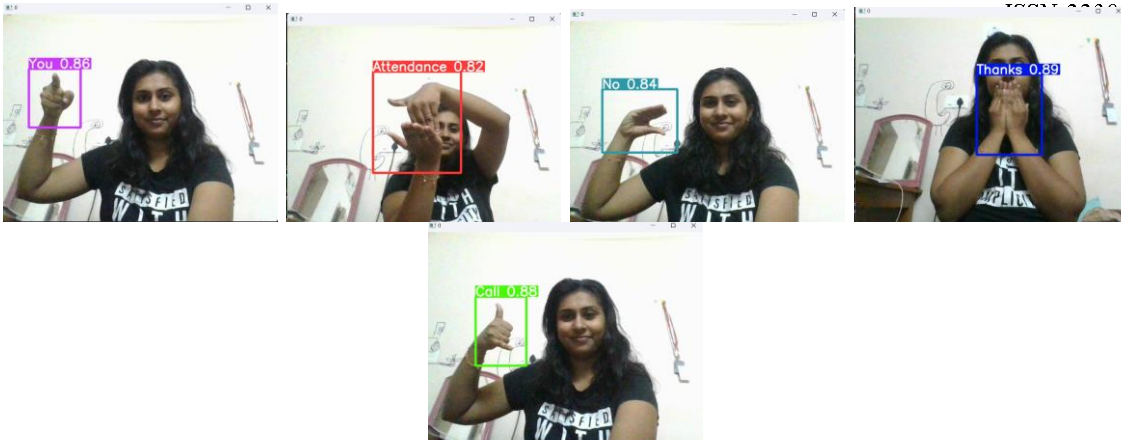


Fig. 2. Output from YOLOv5

E. Continuous Static Gesture Recognition

The predicted output from the static gesture is stored in a variable in the program and it display the sentence with the words. Figure 1 displays the suggested algorithm's block diagram.

The algorithm's steps are as follows:

1. A dataset was created for each of the hand gestures.
2. Labellmg was used to annotate each sign's image.
3. The YOLOv5 algorithm was then used to train the model.
4. Static gestures were then used to test the algorithm.
5. After that, these motions were fed into the program and forms the sentence.

IV. RESULTS AND DISCUSSION

A. Evaluation Measures

The performance of the suggested model is assessed using evaluation measures. These measures will assist us in evaluating the precision and efficacy of our model for detecting sign language.

1) Accuracy: It is determined by dividing the total number of predictions the model makes by the number of accurate predictions.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

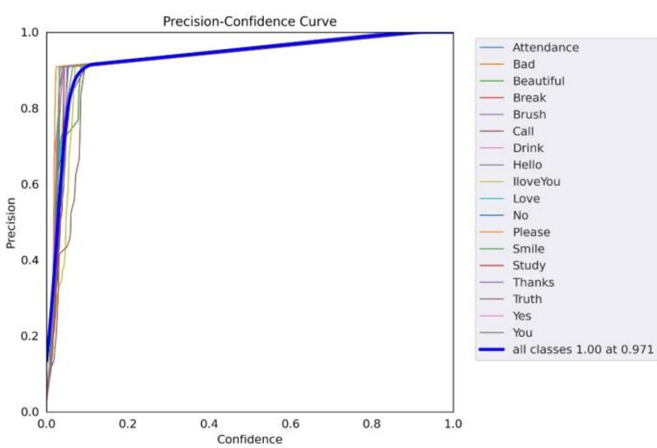


Fig. 3. Precision-Confidence Curve

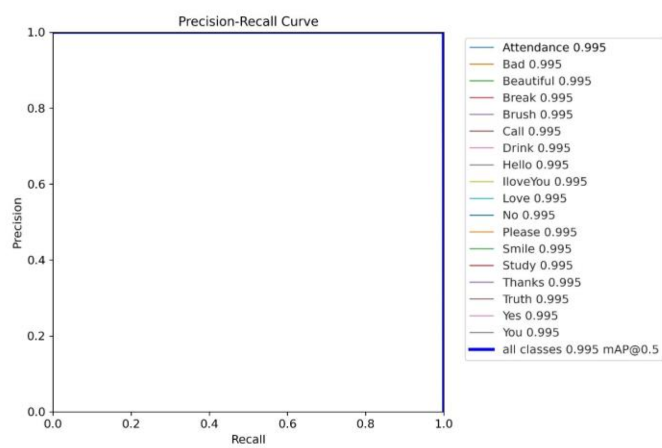
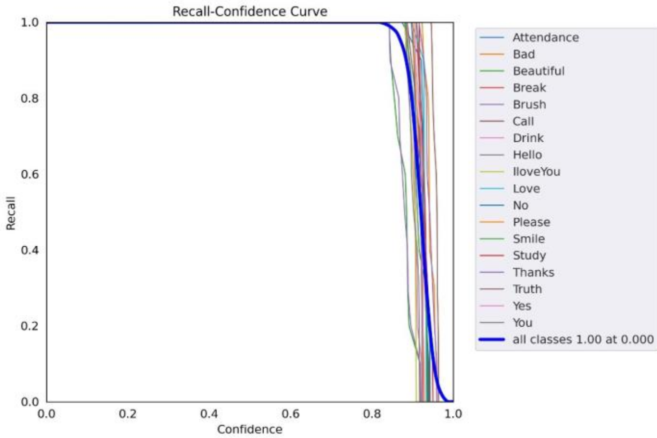


Fig. 5. Precision-Recall Curve

ISSN:2230-9993



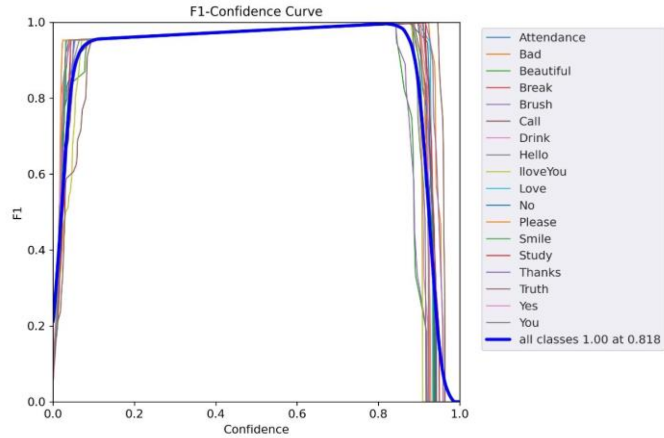


Fig. 4. Recall-Confidence Curve

Fig. 6. F1-Confidence Curve

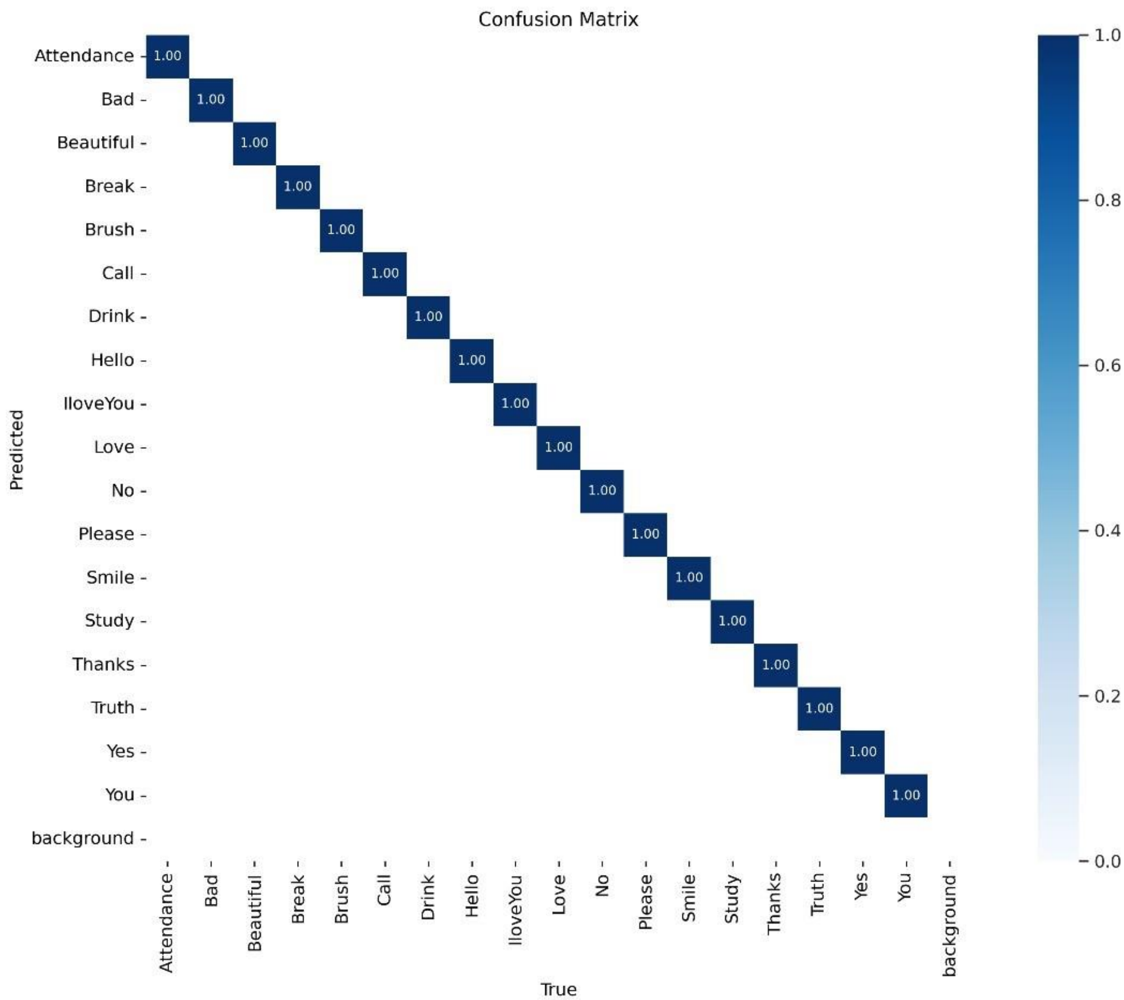


Fig. 7. Confusion Matrix of the Proposed System

ISSN:2230-9993

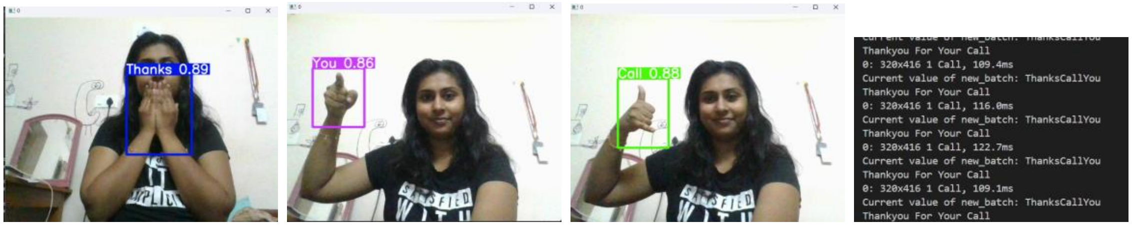


Fig. 8. Output predicted from the proposed model

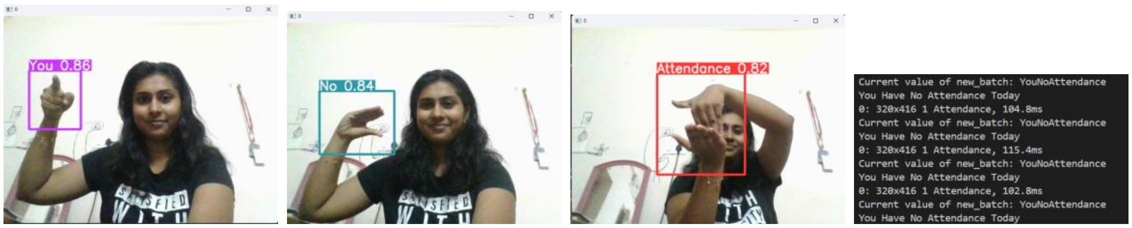


Fig. 9. Output predicted from the proposed model

2) Precision: It assesses the proportion of true positive predictions among all positive predictions, indicating the model's ability to avoid false positives. Fig 3 represents the precision confidence curve.

$$\text{Precision} = \frac{TP}{TP + FP}$$

3) Recall: It measures the proportion of true positive predictions among all actual positive instances, reflecting the model's ability to capture all positive instances. Fig 4 represents the recall-confidence curve.

$$\text{Recall} = \frac{TP}{TP + FN}$$

4) F1-Score: It the harmonic mean of precision and recall, provides a balanced assessment of the model's performance. Fig 6 represents the F1-confidence curve.

$$\text{F1-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

Here, TP represents true positives, TN represents true negatives, FP represents false positives, and FN represents false negatives. Fig 5 represents the precision-recall curve.

B. Confusion Matrix

The confusion matrix provides a vital assessment tool for evaluating the performance of the proposed model. This matrix is constructed using the predicted and actual classes of our dataset.

Fig 6 represents the true class label and the predicted class label. This comprehensive assessment aids in refining our model, enhancing its accuracy and efficacy in accurately detecting sign language gestures.

C. Model Output

The output in Fig 9 is displayed as ”You Have No Attendance” when the hand motions ”you,” ”no,” and ”attendance” are used. Fig 8 is displayed the output as ”Thank You For Your
Call”.

V. CONCLUSION

This paper represents a significant advancement in hand gesture recognition technology by transitioning from traditional static gesture recognition to a comprehensive approach encompassing both static and continuous gestures. This system can interpret continuous hand gestures and generate fully formed sentences, thereby enhancing communication accessibility for users. Additionally, the development of a database containing 18 Indian Sign Language words with 500 images contributes to the expansion and improvement of gesture recognition models. Leveraging the YOLOv5 algorithm for hand sign detection and classification further enhances the efficiency and accuracy of the system. Overall, this project paves the way for more inclusive and effective gesture-based communication systems, with broad applications in diverse fields such as assistive technology, human-computer interaction, and education.

REFERENCES

[1] R Sreemathy, MP Turuk, S Chaudhary, K Lavate, A Ushire, S Khurana,Continuous word level sign language recognition using an expert system based on machine learning,International Journal of Cognitive Computing in Engineering, Volume 4,2023,Pages 170-178,ISSN 2666-3074,https://doi.org/10.1016/j.ijcce.2023.04.002.

[2] A Karmel, Anushka Sharma, Muktak pandya, Diksha Garg,IoT based Assistive Device for Deaf, Dumb and Blind People,Procedia Computer Science,Volume 165,2019,Pages 259-269,ISSN 1877-0509,https://doi.org/10.1016/j.procs.2020.01.080.

[3] Prachi Sharma, Radhey Shyam Anand,A comprehensive evaluation of deep models and optimizers for Indian sign language recognition,Graphics and Visual Computing,Volume 5,2021,200032,ISSN 26666294, https://doi.org/10.1016/j.gvc.2021.200032.

[4] J. DelPreto, J. Hughes, M. D'Aria, M. de Fazio and D. Rus, "A Wearable Smart Glove and Its Application of Pose and Gesture Detection to Sign Language Classification," in IEEE Robotics and Automation Letters, vol.7, no. 4, pp. 10589-10596, Oct. 2022, doi: 10.1109/LRA.2022.3191232.

[5] Giulia Zanon de Castro, Rubia Reis Guerra, Frederico Gadelha´ Guimaraes,˜ Automatic translation of sign language with multistream 3D CNN and generation of artificial depth maps, Expert Systems with Applications, Volume 215,2023,119394,ISSN 0957-4174, https://doi.org/10.1016/j.eswa.2022.119394. (https://www.sciencedirect.com/science/article/pii/S0957417422024125)

[6] P. Subha Rajam, G. Balakrishnan,Recognition of Tamil Sign Language Alphabet using Image Processing to aid Deaf-Dumb People,Procedia Engineering,Volume 30,2012,Pages 861-868,ISSN 18777058, https://doi.org/10.1016/j.proeng.2012.01.938.

[7] D. S. Breland, S. B. Skriubakken, A. Dayal, A. Jha, P. K. Yalavarthy and L. R. Cenkeramaddi, "Deep Learning-Based Sign Language Digits Recognition From Thermal Images With Edge Computing System," in IEEE Sensors Journal, vol. 21, no. 9, pp. 10445-10453, 1 May1, 2021, doi: 10.1109/JSEN.2021.3061608.

[8] N. Adaloglou et al., "A Comprehensive Study on Deep LearningBased Methods for Sign Language Recognition," in IEEE Transactions on Multimedia, vol. 24, pp. 1750-1762, 2022, doi: 10.1109/TMM.2021.3070438.

[9] Areeb, Mohammad & Maryam, Ms & Nadeem, Mohammad & Alroobaea, Roobaea & Anwer, Faisal. (2022). Helping Hearing-Impaired in Emergency Situations: A Deep Learning-Based Approach. IEEE Access. PP. 1-1. 10.1109/ACCESS.2022.3142918.

[10] Das, Sunanda & Imtiaz, Samir & Neom, Nieb & Siddique, Nazmul & Wang, Hui. (2022). A hybrid approach for Bangla sign language recognition using deep transfer learning model with random forest classifier. Expert Systems with Applications. 213. 118914. 10.1016/j.eswa.2022.118914.