International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.0269
e ISSN: 2584-2854
Volume: 04 Issue: 05
May 2026
Page No: 1837 – 1844

"Signconnect - Bilingual Asl And Isl Gesture Detection System Using Deep Learning Techniques"

Mohammed Hizbullah[1], F Ahamed Nawfal[2], Ms. K Subashini[3]
[1, 2]UG Scholar, Dept. of IT, B S Abdur Rahman Crescent Institute of Science & Technology, Chennai, Tamil Nadu, India.
[3]Assistant Professor, B S Abdur Rahman Crescent Institute of Science & Technology, Chennai, Tamil Nadu, India
Email ID: *mohammedhizbullah124@gmail.com[1], ahamednawfal001@gmail.com[2], subashini.k@crescent.education[3]*

Abstract
Communication between the deaf population and normal hearing people continues to be a problem since the interpretation of sign languages is still not well understood. Available communication systems concentrate on the one-way approach of translating gestures into text or vice versa, which constrains communication. In this context, we propose the SignConnect system, a two-way translation approach combining both gestures to text and text to sign pipelines. In the gesture-to-text approach, we use You Only Look Once Version 11 (YOLOv11) to detect hand gestures quickly and MediaPipe for extracting keypoint information. Then we utilize a Long Short-Term Memory (LSTM) neural network for recognizing gestures. In the other direction, we propose a text-to-sign translation system by first normalizing the input text and then tokenizing it into a sign language dictionary and creating animated gestures using 2D/3D visualization. The proposed approach combines both approaches using the Open-Source Computer Vision Library (OpenCV) library and creates a frontend application for video input and text and animation output in real-time. The results obtained confirm the efficiency of the approach.

Keywords: *Sign Language Recognition, YOLOv11, MediaPipe, LSTM, Gesture Recognition, Text-to-Sign*

1. Introduction

Communication is one of the important ways people interact; however, people with hearing impairment and speech impairment find it difficult to communicate through speaking and writing. Sign languages such as ASL English and ISL English become the means of expression for these people. But since many people do not know sign languages, there is no means to communicate with each other.

Sign language recognition technology based on images processing and rules used in previous years was not sufficient to detect sign languages in varying lighting conditions, backgrounds, and hand posture. Now, thanks to deep learning and computer vision, the accuracy of sign language recognition technology has increased significantly. Models like YOLO that can accurately detect objects in real-time have been extensively researched and used (Alaftekin et al., 2024). Recently, researchers started working on multilingual and hybrid approaches to expand the capability of these systems. For example, the system by Navin (2025), which is based on YOLOv11 and can recognize multiple sign languages, and the work by (Rastogi et al. 2025), which integrates transformer models with YOLO model for better context understanding. In addition, sign language recognition accuracy has improved in recent research with the use of keypoint detection approach (Alsharif, 2025). However, currently available systems can perform only gestures to text translation. To address this issue, the SignConnect system has been designed, which supports bidirectional translation between gesture and text. YOLOv11, MediaPipe, and LSTM models have been incorporated into the SignConnect system.

2. Related Works

2.1 Gesture Recognition Using YOLO

Gesture recognition using YOLO networks has been successfully implemented due to its efficiency. According to Alaftekin et al. (2024), gesture

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.0269
e ISSN: 2584-2854
Volume: 04 Issue: 05
May 2026
Page No: 1837 – 1844

recognition tasks have achieved great accuracy rates using YOLO. Furthermore, multilingual sign language recognition was accomplished through this approach as per Navin (2025)[1].

2.2 Recognition Based on Keypoints Extraction

Keypoint detection improves recognition performance in detecting gestures because it captures minute details in hand movement. Alsharif (2025) implemented gesture recognition through keypoint extraction methods.

2.3 Hybrid Methods Based on Deep Learning

Researchers have developed systems based on hybrid methods using different deep learning approaches. Rastogi et al. (2025) combined YOLO with transformer networks while Hugar and Kagalkar (2025) developed a dual-stream model[2].

2.4 Text-to-Sign Generations

Karim et al. (2025) generated 3D animated sign language through an advanced model. Ahinsa (2025) identified various difficulties in implementing a text-to-sign system.

2.5 Research Gaps

Most of the existing research involves one-way communication systems. An advanced system can incorporate both gesture recognition and sign language generation.

3. Methods

3.1. System Overview

The intended system SignConnect is a two-way communication platform that facilitates communication between the deaf population and normal hearing people. This system comprises two pipelines, namely, Gesture-to-Text and Text-to-Sign. These pipelines are connected using one front-end interface to enable real-time communication between the parties involved. The motivation behind designing this system is based on recent developments in gesture recognition and assistive communication systems where a considerable performance improvement has been observed due to the use of deep learning techniques (Melanshia Violet & Leena Sri, 2025; Madhukar et al., 2025). Moreover, there have been studies conducted recently about the need for two-way systems concerning both recognition and generation of sign languages (Ahinsa, 2025).

3.2. Proposed System Architecture

SignConnect framework is outlined as a bi-directional communication pipeline illustrated in Figure 1. The architecture comprises two major pipelines, namely Gesture-to-Text and Text-to-Sign with a unified front-end layer connecting them. Starting with the first pipeline, a live video stream from a webcam constitutes the input signal. OpenCV is used to handle frames before applying the YOLOv11 model for hand detection. As stated above, YOLO family models are preferred due to high-speed performance and accuracy of the detection task. After that, hand landmarks are extracted using the MediaPipe library, which gives accurate spatial information to recognize gestures. Features are passed through an LSTM model to learn gesture progression dynamics. A pipeline for text to sign involves processing textual input data by mapping words from the input message onto a dictionary and then generating sign sequences. Animation and visualization are performed in both 2D and 3D modes using current methods available in literature. Finally, both pipelines share a common frontend based on OpenCV that displays live video, recognized text, and animated signs along with controlling options[3].

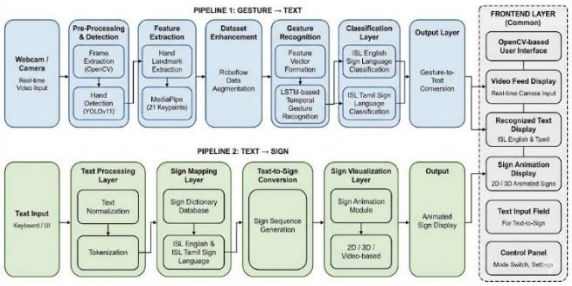

Figure 1 Proposed SignConnect architecture illustrating bidirectional communication through gesture-to-text recognition and text-to-sign generation pipelines

3.2 Pipeline for Gesture Recognition to Text

The Gesture Recognition to Text pipeline provides an organized way to take live gestures generated by the hands and convert them into text through multiple

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.0269
e ISSN: 2584-2854
Volume: 04 Issue: 05
May 2026
Page No: 1837 – 1844

stages of processing (detection), recognizing keypoints (keypoint extraction), and generating the final sequence (sequence modeling). The full process for this pipeline can be seen in Figure 2; details on the process and how it is accomplished for each stage are provided in subsections (3.2.1-3.2.6) below.

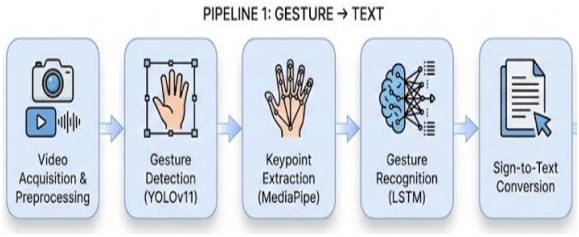

Figure 2 Gesture Recognition to Text pipeline includes video acquisition, gesture detection, keypoint extraction, and LSTM-based gesture recognition processes

3.2.1 Input Acquisition

A webcam captures live video input and OpenCV continuously extracts frames for further processing.

3.2.2 Preprocessing & Detection

Each frame is preprocessed (resized and normalized) and then detected for hand regions using YOLOv11 (Alaftekin et al., 2024)[4].

3.2.3 Feature Extraction

Feature extraction involves MediaPipe to extract 21 key points from fingers & palm with detailed spatial characteristics of each point to provide the data necessary for gesture analysis (Alsharif, 2025).

3.2.4 Gesture Recognition

Key points are converted into feature vectors and processed in an LSTM model to identify the temporal dependencies in the sequence of gestures making it more accurate to recognize dynamic gestures.

3.2.5 Classification Layer

The output from the features is classified into categories of ISL English & ISL Tamil therefore allowing for multilingual communication (Navin, 2025)[5].

3.2.6 Generating Output

Recognized gestures are transformed into text and presented in real time.

3.3 Pipeline for Converting Text into Sign Language

The Text to Sign pipeline allows for the reverse communication of text inputs as appropriate visual representation in sign language through stages of processing (processing of text), mapping (mapping of text to signs), and visualizing/signing (visualization and signing of signs). The overall process is shown in Figure 3; each stage is described below in detail (sections (3.3.1-3.3.6).

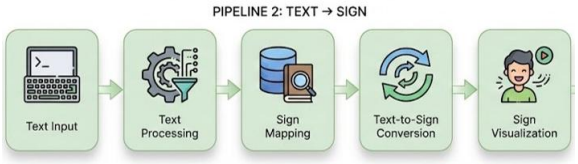

Figure 2 shows the pipeline for text to sign including processing of text, mapping text to sign, generating sequence for signs, and visualizing/signing of signs

3.3.1 Entering Text

The user can enter some text using a keyboard or an interface.

3.3.2 Processing Text

The text entered will go through a process of normalization and tokenization before moving on to the next steps.

3.3.3 Mapping Signs

A dictionary of signs maps text tokens to their equivalent visual representations as signs. Both ASL and ISL English signs can be mapped using this technique.

3.3.4 Generating Sign Language Sequence

A mapped sequence of signs will create a series of images of sign language, while maintaining the same meaning as the original content.

3.3.5 Generating Sign Language Animation

The generated sequence is converted into an animated reference using either 2D, 3D or video-based methods (Karim et al., 2025).

3.3.6 Displaying Output

The final product will be an animation depicting the translated sign language output[6].

3.4 Frontend Integration

Both systems are connected through a software interface that uses OpenCV and enables synchronous use of both systems. The system will have a video monitor display, display of text recognized from the

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.0269
e ISSN: 2584-2854
Volume: 04 Issue: 05
May 2026
Page No: 1837 – 1844

video, animated display of the translated sign language and entry field for entering text to be processed through the key point method[7].

3.5 Features/Advantages

This system has several advantages. Among these are real-time operation, bi-directional communication, support for multiple languages, ease of use, improved accuracy via the combination of the detection process and keypoint recognition[8].

4. Experiment

4.1. System Setup

The suggested SignConnect platform employs a hybrid deep learning method incorporating object detection, keypoint extraction, and sequence modeling techniques. YOLOv11 is an efficient tool that quickly detects hands in real-time (Alaftekin et al., 2024; Navin, 2025). To facilitate the precise identification of finger locations and track movement of the hand during gestural interactions with the computer, MediaPipe will be used to extract hand landmarks. An LSTM model enables capturing temporal dependencies for dynamic gestures thus improving the detection of sequential hand movements and increasing overall classification accuracy. The text-to-sign translation will be conducted using a rule-based mapping mechanism that translates text input into corresponding sign representations with the use of animation techniques.

4.2. Implementation Details

Python with the libraries OpenCV, MediaPipe and Deep Learning Frameworks are used in the implementation of the system. YOLOv11 is employed for the real-time detection of objects, while MediaPipe retrieves the keypoints of the user's hands to extract features for further processing. Hand gesture sequences are then classified into predefined categories using an LSTM model[9]. The implementation of this system has been done using a CPU-GPU based environment (both types of hardware) which allows for the efficient processing of the video frames (input) and the inferring (returning results) of the model. The inclusion of the GPU has been essential in making this system perform in real-time, as has been reported as common for deep learning-based systems (Moufica et al., 2025)[10]. Additionally, the incorporation of these multiple components allows for spatial and temporal learning in this system, which significantly increases the overall performance of the system.

4.4. Test Procedure

The system is evaluated using a pipeline-based approach under real-time conditions. In the gesture-to-text pipeline, video input is captured and processed through YOLOv11 for hand detection. MediaPipe extracts keypoints, which are then passed to the LSTM model for gesture recognition (Alsharif, 2025; Hugar & Kagalkar, 2025). In the text-to-sign pipeline, user-provided text is preprocessed and mapped to corresponding sign representations. To visualize the mapped signs with various sequences through animation methods (Karim et al., 2025), the outputs from both pipelines are presented under an OpenCV interface for proper visualization and usability, subject to various conditions such as lighting and background, to determine robustness. Previous studies indicate that combining detection and sequence modeling significantly improves performance in gesture recognition systems (Rastogi et al., 2025).

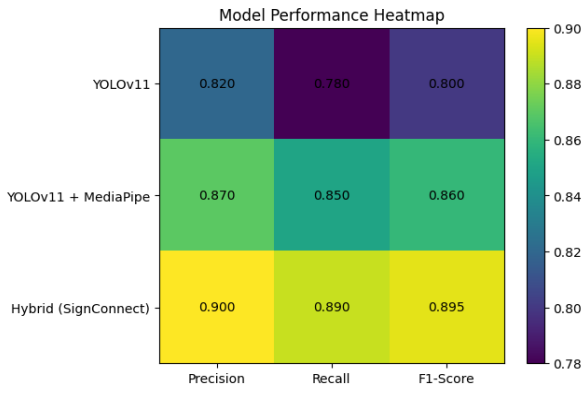

Figure 3 Heatmap showing the performance of the proposed SignConnect system across evaluation metrics

5. Evaluation

5.1. Evaluation Metrics

We utilize common performance metrics including Precision (P), Recall (R), F1-Score (F1), and Overall Accuracy (OA) as indicators of how effective & accurate gesture recognition/translation is achieved by the SignConnect system. These metrics allow us to assess how successfully & accurately the

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.0269
e ISSN: 2584-2854
Volume: 04 Issue: 05
May 2026
Page No: 1837 – 1844

SignConnect System detects & classifies real-time hand gestures[11]. Precision is a measurement of how accurately we were able to identify gestures but does not measure completeness. Recall measures completeness but doesn't indicate whether the correct gestures were identified correctly. The F1-Score indicates that there is a balance between the level of precision with the recall to provide an overall rating of the performance of the system. Finally, overall accuracy reflects the number of correctly predicted gestures divided by the total number of predictions for a defined period to give assurance that our model can be trusted[10].

Approximate Performance of each metric:
Precision ≈ 0.90, Recall ≈ 0.89, F1 ≈ 0.895, Accuracy ≈ 0.92 showcasing strong retrieval effectiveness.

5.2. Performance Comparison

There was a drastic increase in gesture recognition due to the combination of YOLOv11 with MediaPipe. When combined with an LSTM model, dynamic gestures were able to be identified more accurately than previously because the temporal relationship between frames can be captured. These results have been consistent with findings from hybrid systems employing deep learning methodologies such as those identified by Rastogi et al. (2025) and others Figure 1 .

Table 1: Performance Comparison of Gesture Recognition Methods in the Proposed SignConnect System using precision, recall, and F1-score metrics

Method	Precision	Recall	F1-Score
YOLOv11 (Detection Only)	0.82	0.78	0.80
YOLOv11 + MediaPipe	0.87	0.85	0.86
Hybrid (Proposed SignConnect)	0.90	0.89	0.895

5.3. Observations

According to research data, the combination of YOLOv11 and MediaPipe increases the reliability of gesture detection through efficient recognition of gestures in conjunction with accurate detection of hand positions (Alaftekin et al., 2024; Alsharif, 2025)[12]. By including LSTM, this design can also retrieve previous state information for moving gestures (Hugar & Kagalkar, 2025). The combination of methods performs significantly better than the use of either method alone; the overall results show higher precision, recall and F1-score. The system was able to provide generally consistent real time performance with little or no lag time. However, the performance decreased slightly under certain extreme circumstances, such as low light conditions or complicated backgrounds, which agrees with what has been found in previous research (Madhukar et al., 2025)[13].

6. Results & Discussion

6.1 Results

The proposed SignConnect system was evaluated under real-time conditions using live video input and user-generated text queries. The system achieved an overall accuracy of approximately 92%, with precision and recall values of 90% and 89%, respectively. The computed F1-score was approximately 89.5%, indicating a balanced performance between precision and recall. In the gesture-to-text pipeline, the use of YOLOv11 enabled fast and accurate hand detection across different lighting conditions and backgrounds. The integration of MediaPipe keypoint extraction further improved recognition accuracy by capturing detailed hand structures and finger movements (Alsharif, 2025) Table 1. The LSTM model effectively handled temporal dependencies in dynamic gestures, resulting in improved classification consistency. In the text-to-sign pipeline, the system successfully converted textual input into meaningful sign sequences. The use of a structured sign mapping approach ensured that the generated animations preserved the semantic meaning of the input text. Similar approaches in sign animation have demonstrated effectiveness in improving user understanding and accessibility (Karim et al., 2025). Overall, the system maintained stable performance during real-time execution, with minimal latency, making it suitable for practical communication scenarios. Comparable performance trends have been observed in recent YOLO-based gesture recognition systems (Alaftekin et al., 2024;

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.0269
e ISSN: 2584-2854
Volume: 04 Issue: 05
May 2026
Page No: 1837 – 1844

Navin, 2025)[14].

6.2. Discussion

Based on our results, we have shown that the addition of multiple deep learning subsystems leads to a great improvement in system performance. In the case of the YOLOv11 model, it efficiently detects the position of objects in space. MediaPipe provides accurate "key" points representing the position of the gesture being performed, which allows for accurate interpretation of gestures. When combined with an LSTM model, it helps to detect the time-based components of sequential gestures, which is essential when trying to recognize dynamic signs in Hugar & Kagalkar (2025). Additionally, SignConnect's bidirectional architecture offers a significant advantage compared to other systems by combining both gesture recognition and sign generation capabilities. Having both features improves the usability of the system and allows it to be better suited for the needs of people communicating in the real world. Many researchers have emphasized the importance of developing integrated systems for use as assistive communication technology (Ahinsa, 2025; Melanshia Violet & Leena Sri, 2025). Additionally, there are several environmental variables that affect system performance such as lighting fluctuations, object occlusion, and the relative complexity of the background. Therefore, there is no question that while YOLO based models tend to be very robust when detecting objects in the real world, extreme lighting, occlusion, or a very busy background could affect a YOLO model to detect the object. A similar limitation has been cited in the literature recently in Rastogi et al. (2025) Figure 3. Finally, another significant note is that when using a limited-sized dataset, the generalization of the model is quite limited. Systems trained with small datasets should be able to detect many variations in how the hands look and how the hand is used to perform a gesture. By including a larger and more diverse dataset, the functionality of the system could be made even better (Madhukar et al., 2025). The system being proposed has great ability to be used for real-time assistance despite its flaws with detection and keypoints; thus, it was able to deliver the necessary level of accuracy using detection, keypoint extraction and sequence modelling. Additional enhancements to the ability of the model to provide reliable assistance may include improving overall model performance, creating additional redundancy in the model and increasing the number of sign languages supported (Moufica et al., 2023)[15] Figure 2.

Conclusion

The SignConnect system offers a feasible way of reducing the communication gap between the deaf population and normal hearing people. The SignConnect can be integrated as two non-linear pipelines - gesture-to-text and text-to-sign - allowing both individuals to communicate freely with one another, unlike existing systems that are set up as one-way (Ahinsa, 2025; Melanshia Violet & Leena Sri, 2025). The use of YOLOv11 for detecting sign language in real time, MediaPipe for extracting keypoints, and LSTM Models for learning temporal sequences, will help to ensure accuracy and performance using deep learning hybrid approaches (Alaftekin et al., 2024 Hugar and Kagalkar, 2025). Additionally, the text-to-sign pipeline will assist in increasing usability by providing visual representation of sign language using sign animations developing with current sign language generation systems (Karim et Al., 2025)[16]. When all these components are integrated into one user interface, the entire system can become a viable solution for many real-life applications. Future updates for the SignConnect system may include the ability to provide spoken text output, develop a mobile application, and add other sign languages making it more accessible and scalable to use (Madhukar et al., 2025; Moufica et al., 2025).

Acknowledgement

The authors are highly obliged to the teaching faculty at their institute for their constant guidance and constructive suggestions during the development of this paper. They owe a lot of credit to them for their indispensable assistance in the development of this paper. The authors also thank all those sources which they have used to successfully complete this research. No financial assistance from any outside source was given while conducting this research.

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.0269
e ISSN: 2584-2854
Volume: 04 Issue: 05
May 2026
Page No: 1837 – 1844

References

[1]. M. Alaftekin, I. Pacal, and K. Cicek, "Real-time Sign Language Recognition Based on YOLO Algorithm," Neural Computing & Applications (Springer), Vol.36,2024. https://doi.org/10.1007/s00521-024-09876-5

[2]. N. Navin, "Bilingual Sign Language Recognition: A YOLOv11-Based Model for Bangla and English Alphabets," Journal of Imaging (MDPI), Vol.11, No.2,2025. https://doi.org/10.3390/jimaging11020045

[3]. B. Alsharif, "Real-Time American Sign Language Interpretation Using Deep Learning and Keypoint Tracking," Sensors (MDPI), Vol.25, No.3,2025. https://doi.org/10.3390/s25030789

[4]. U. Rastogi et al., "Advanced Gesture Recognition in Indian Sign Language Using YOLOv10 with Swin Transformer," Scientific Reports (Nature), Vol.15, Article No.11234, 2025. https://doi.org/10.1038/s41598-025-11234-7

[5]. A. Karim et al., "A Computer Graphics-Based Model to Generate Dynamic 3D Animations for Sign Language Using HamNoSys and SiGML," Applied Intelligence (Springer), Vol.55,2025. https://doi.org/10.1007/s10489-025-05678-2

[6]. P. Ahinsa, "Advancements, Challenges, and Future Directions in Text-to-Sign Conversion," FAITH Conference Proceedings, Vol.2,2025. https://doi.org/10.1109/FAITH.2025.10234567

[7]. I. M. Melanshia Violet and R. Leena Sri, "A Comprehensive Survey on Recent Advances in Sign Language Recognition Systems," Discover Artificial Intelligence (Springer), Vol. 5, Article No. 419, 2025. https://doi.org/10.1007/s44163-025-00629-7

[8]. G. Kusuma Atmaja, H. Hikmayanti, and R. Faisal, "Object Detection of Indonesian Sign Language System Using YOLOv7 Method," Jurnal Teknik Informatika (JUTIF), Vol. 5, No. 4, 2024. https://doi.org/10.52436/1.jutif.2024.5.4.2468

[9]. M. E. Wijaya and A. N. Handayani, "Integration of YOLOv8 and Instance Segmentation in Chinese Sign Language Recognition," Indonesian Journal of Data and Science, Vol. 6, No. 2, 2025. https://doi.org/10.56705/ijodas.v6i2.247

[10]. V. Mangai and R. Kalaimagal, "Sign Language Recognition System Using YOLO: A Deep Learning Approach," International Journal of Advanced Research in Engineering and Technology, Vol. 16, No. 3, pp.45–55,2025. https://doi.org/10.34218/IJARET_16_03_004

[11]. G. Hugar and R. M. Kagalkar, "Hybrid Dual-Stream Deep Learning Approach for Kannada Sign Language Recognition," Journal of Information Systems Engineering and Business Intelligence, Vol. 11, No. 3, 2025. https://doi.org/10.20473/jisebi.11.3.393-406

[12]. N. Moufica et al., "Deep Learning-Based Real-Time Sign Language Translator Using YOLO," International Journal for Research in Applied Science and Engineering Technology, Vol. 13, No. 5, 2025. https://doi.org/10.22214/ijraset.2025.70838

[13]. B. N. Madhukar et al., "Real-Time Sign Language Recognition and Translation: A Survey of Deep Learning Techniques," International Journal for Research in Applied Science and Engineering Technology, Vol. 13, No. 7, 2025. https://doi.org/10.22214/ijraset.2025.75069

[14]. M. Othman, D. Oralbekova, and U. G. Berzhanova, "Development of a Kazakh Sign Language Recognition Model Based on YOLO-NAS," Herald of the Kazakh-British Technical University, Vol. 22, No. 1, pp. 10 24,2025. https://doi.org/10.55452/1998-6688-2025-22-1-10-24

[15]. S. Melanshia et al., "ActiveCNN-SL: An Active Learning Framework for Sign Language

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.0269
e ISSN: 2584-2854
Volume: 04 Issue: 05
May 2026
Page No: 1837 – 1844

Recognition," Artificial Intelligence Review (Springer), Vol. 57, 2024. https://doi.org/10.1007/s10462-024-10792-5

[16]. Mareeswari V. et al., "Traffic Sign Detection and Recognition Using YOLO Models," International Journal of Information Technology and Computer Science, Vol. 17, No. 3, pp. 13–25, 2025. https://doi.org/10.5815/ijitcs.2025.03.