INTERNATIONAL JOURNAL OF RESEARCH AND INNOVATION IN APPLIED SCIENCE (IJRIAS)
ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue III March 2026

Signconnect: Real-Time Communication Bridge for the Specially-Abled

Shubham., Nancy

Apex Institute of Technology (CSE) Chandigarh University, Mohali-140413, Punjab

DOI: https://doi.org/10.51584/IJRIAS.2026.11030107

Received: 01 April 2026; Accepted: 06 April 2026; Published: 18 April 2026

ABSTRACT

Communication gap between the deaf and hearing communities remains an important obstacle to social integration. Recent advances in artificial intelligence, in particular, in the fields of deep learning and computer vision, have also provided the prospect of truly radical assistive technologies that could translate sign language gestures to readable text or spoken audio in real time. The current study introduces a novel system of real-time sign language interpretation, which integrates the multi-modal gesture recognition, flexible deep neural networks, and context-based translation schemes to support effective and natural user interactions. The system that was developed uses the convolutional and recurrent neural network architectures to process the spatial and temporal properties of signing gestures. A dedicated set of movements of the Indian Sign language (ISL) was created based on MediaPipe Holistic and OpenCV to obtain the hand, face, and body keypoints and to be trained comprehensively with the help of the TensorFlow workflows. The model is optimized to the minimal-latency processing which ensures fluid real-time interpretation on devices with limited processing power. Besides making progress in the technical aspects of instantaneous gesture recognition, this research will provide a solution to an im- portant social need; that of allowing people with new auditory or speech disabilities to communicate easily and independently. The system has a high recognition accuracy, adaptability to different illumination and environmental conditions, and sequential sign pattern expansion. Also, the architecture offers a platform upon which new features, including gesture- to-voice translation, cross- linguistic understanding, and portability or mobile compatibility will be built. By integrating technology enhancement with the human-centered design concepts, this study provides a scaling, efficient, and holistic solution that enhances the level of access and facilitates the equity of communication across all the societal groups.

Index Terms: Real-Time Translation, Sign Language Recognition, Deep Learning, CNNLSTM, Computer Vision, Accessibility, ISL, Assistive Technology.

INTRODUCTION

The sign language is the basic communicational system of Deaf and Hard-of-Hearing (DHH) communities. This visual channel of information transfer is represented by hand ges- tures, facial expressions, and body movements, makes the elab- orate and subtle conversation. It has been observed that over 1.5 billion people in the world can be affected to some extent by auditory impairment and there is a strong necessity of estab- lishing universal communication patterns that would increase the level of accessibility and societal inclusion [1]. In India, there are 7-8 million users of Indian Sign language (ISL), but the absence of automated and simplified translation systems still restricts the involvement of DHH communities in the field of education, medical practice, and employment [3], [6], [15]. In turn, real-time ISL translation is a needed step towards overcoming this communication barrier and promoting equal access to information and services. Historical development of Sign Language Recognition (SLR) systems has been in line with advancement of artificial intelligence and computer vision systems. The early recognition algorithms used were mostly based on handcrafted feature extraction techniques including Scale- Invariant Feature Transform (SIFT), Speeded-Up Ro- bust Features (SURF), and Histogram of Oriented Gradients (HOG), with a traditional classifier (Support Vector Machines) [7], [20]. Although these methods proved to be effective in limited environments, they were challenged by light differ- ences, intricate backgrounds and inter-signer variation. With the rise of deep learning, this area has changed its face, with architectures that support the automated learning of features and holistic understanding of gestures.

Recent methods such as Convolutional Neural Networks (CNNs), Long Short-Term Memory, (LSTM) networks as well as Transformer based models have shown higher effectiveness and robustness in capturing both time and space of gestures [2], [4], [5], [12]. Modern studies have furthered this development by provid- ing vision-based, multimodal and contextsensitive recognition systems. As an illustration, Pandey et al. [3] came up with a real-time translator of ISL that uses deep learning to recognize gestures, whereas Velmathi and Goyal [6] used MediaPipe Holistic to obtain fine detail landmarks which can give better results. More advanced models, such as SignBERT+ [9], use hand-model awareness with self- supervised pre-training, and Hierarchical Windowed Graph Attention Networks [10] have been able to do better with temporal modeling on large ISL datasets. However, a lot of challenges still remain- most con- spicuously in the continuous gesture segmentation, variability in signers, imbalance of data sets, and interpretation of gesture sequences to natural language [12], [15], [16]. In order to solve these shortcomings, modern research trends are directed at real-time, efficient and semantically enhanced systems of ISL translation. The contextual understanding is enhanced by the use of multimodal aspects like facial expression and emotion recognition [13], along with natural language support [4], [16]. Moreover, by using optimized deep learning models on edge devices provided by quantization and pruning techniques, one can achieve fast processing using only edge device hardware [3], [17]. A combination of computer vision, natural language processing, and deep learning has thus become the basis of the advanced systems that are able both to interpret and generate sign language interactions at real time. The primary goal of this research project is to develop a real time architecture of ISL translation system that is capable of appropriately integrating the gesture recognition, feature extraction and semantics analysis to produce coherent text or speech out- put. The proposed architecture will focus on increasing the accuracy of the translations, reducing the time spent in the execution, and the harmonious bi-directional communication between the hearing and non-hearing populations, which will ultimately be a more democratic online world.

LITRATURE REVIEW

Introduction

Sign languages are visual-spatial forms of communication, which involve the use of manual signs, facial expressions, and body language to communicate both semantic and syntactic messages. Indian Sign Language (ISL) is a lingual framework that is naturally evolved with specific phonological, morpho- logical and grammatical features that are specifically adapted to the purpose of communication in India [1]. ISL recognition as an automated field of study has become an important field of study interested in eradicating communication barriers among the hearing impaired or speech impaired and the general population. Research on this area has moved beyond traditional image processing and feature extraction algorithms to more sophisticated deep learning and multimodal networks, which deal with isolated and continuous gesture recognition.

Early Image-Based Methods

The work of the earliest ISL recognition studies was mainly based on image-based methods with manually designed fea- tures. Nandy et al. [2] presented an algorithm where the video sequences were converted to grayscale, followed by the directional histogram-based feature extraction with a clustering analysis to classify them. Their findings showed the high efficiency of the 36-bin histograms over the 18bin variants which reached 100 percent precision accuracy.

Preprocessing and Skin Detection Techniques

Edge detection and skin segmentation methods have been highly utilized as preprocessing to enhance gesture identifi- cation. Chen [4] introduced RGB-to-YUQ/YIQ conversion on skin detection and convex hull algorithm in identification of fingers and then the neural network classification where 98.2 percent of results were found to be accurate.

Sensor-Based Recognition Approaches

Recognition methodologies based on sensors have provided alternatives to systems that only use vision. Agarwal et al.

[7] developed a sensor glove system in which gestures were matched with some predefined databases and translated to meaningful sentences. The first application had only 33.33 percent accuracy.

Integration with NLP Techniques

Several studies have used gesture recognition with natural language processing (NLP) to generate coherent sentences. Wazalwar and Shrawankar [9] have applied CamShift and P2DHMM equations to track the movement of the hand and WordNet POS tagging and LALR parsing to create meaningful English sentences. YCbCr based skin clustering and centroid- offset analysis of ASL was carried out by Shivashankara and Srinath [10] and obtained a 93.05 percent recognition rate.

ROI and Clustering-Based Methods

The region-of-interest (ROI) tracking and the fuzzy cluster- ing methodologies have also been beneficial to real-time ISL recognition. Mariappan and Gomathi [12] used hand tracking and recognition of gestures using ROI-based skin segmentation and fuzzy C-means clustering, which achieved an accuracy of 75.

Modern Deep Learning Approaches

Contemporary research has increasingly emphasized skeleton-based representations and transformer architectures to capture dynamic gestures and non-manual signals. De Coster et al. [14] integrated pose LSTM [37] and transformer networks for sign recognition from video sequences, achieving 74.7% accuracy.

Survey and Comparative Studies

Comprehensive surveys and analyses have examined diverse algorithms and classification approaches. Loeding et al. [18] and Er-Rady et al. [19] discussed the evolution from tradi- tional visionbased methods to neural network architectures for sign language recognition. Bauer and Kraiss [20] investigated recognition using gesture subunits, while Singha and Das explored live videobased ISL recognition. Gross and Brajovic concentrated on illuminationinvariant preprocessing, and Kaur and Kaur reviewed image segmentation methods for gesture analysis. The investigations by Paulraj et al. and Joudaki et al. additionally highlighted feature extraction for hand and head movement analysis.

Classification Algorithms

There are several classification algorithms (KNN, SVM and neural networks) that have been commonly utilized in gestures recognition systems. The studies by Kim et al., Qiao et al., and Hwang and Wen examined KNN methods, distance calculation strategies, and optimisation strategies. The SVM techniques were investigated by Bernhard, Chapelle etal., and Agrawal et al., who focused on the fact that they are effective in the classification of images in histogram-based tasks. The works of Sahoo et al., Lilha and Shivmurthy [33], and Agrawal et al. were dedicated to the pixel nature, redundancy removal, and vision-based classification systems. Jayaram and Fleyeh analyzed the convex hull approaches to feature extraction and Wen et al. [36] introduced the depth sensor based hand detection approaches. Theoretical foundations of current ISL recognition systems were laid by basic research of LSTM, CNNs, and deep learning by Hochreiter and Schmidhuber , Szegedy et al., and Yang et al.

Conclusion and Future Scope

The history of ISL recognition has developed as an approach of recognizing images to complex multimodal deep learning networks. Investigations in the early days focused on manu- ally created features and clustering algorithms, and modern applications use CNN , LSTM , GRU [13]s, transformer architectures, and sensor fusion to improve both single and continuous gesture recognition. The main challenges still exist in the control of dynamic gestures, non-manual components, environmental variability, and limitations of datasets. It is expected that future studies will focus less on creating new engineering designs and more on enhancing existing designs to include spatial, temporal, and semantic attributes together in the hybrid architecture to facilitate robust real-time ISL recognition, thus facilitating effective communication with people with auditory or speech impairments.

PROPOSED METHODOLOGY

Indian Sign Language (ISL) translation system proposed has a multi-stage and systematic design to transfer visual sign language gesture into meaningful textual or spoken language output. The stages are essential in the process of providing correct, quick, and situational recognition. The general flow of the methodology is depicted in Fig. 2 that demonstrates the data flow within the system starting with the video input and continuing to the final output of the translation.

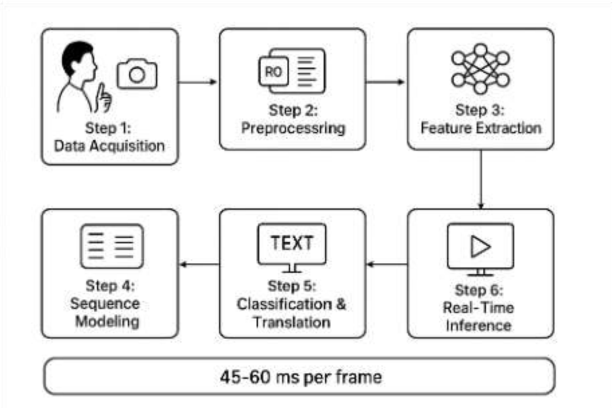

Fig. 1. Process Flow Chart

Data Collection

1) *Experimental Setup:* A standard laptop webcam was utilized to capture images under consistent lighting conditions. The environment was kept uniform to minimize background noise and shadow interference. The following Python libraries were used: OpenCV (cv2): For video capture and real-time frame processing. os: For creating and managing directories for each gesture label. uuid: To assign unique filenames to each captured image. time: To introduce temporal control between captures. The collected images were organized systematically within the following directory structure:

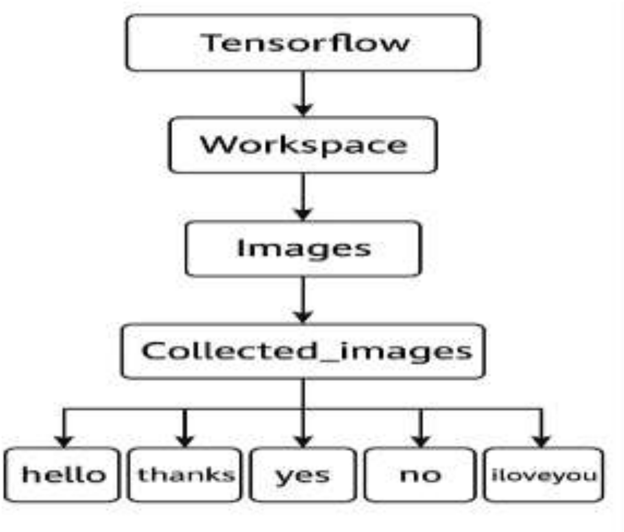

Fig. 2. Directories Flow Chart

Each folder corresponds to a specific gesture class and contains multiple image samples captured during the data acquisition process.

2) *Data Capture Process:* For each sign language gesture, the following procedural steps were executed:

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue III March 2026

Label Initialization:

Each label (gesture) was defined in a Python list for systematic processing.

Directory Creation:

A separate subdirectory was automatically created for each label using the os.makedirs() function with the exist _ok=True parameter to prevent redundancy.

Webcam Activation:

The cv2.VideoCapture(0) command was used to initialize the system's primary webcam. If the camera failed to initial- ize, the system skipped the corresponding label to maintain program stability.

Capture Delay:

A 5-second delay (time.sleep(5)) was introduced to allow the user to position their hands correctly before image capture commenced.

Image Acquisition:

A specified number of frames (number_imgs) were captured at two-second intervals. Each frame was saved as a .jpg file within the corresponding label folder. The filenames incorpo- rated universally unique identifiers (UUIDs) to ensure non- redundant naming.

User Interaction and Exit Control:

The real-time camera feed was displayed in a window using cv2.imshow(), enabling the user to monitor captured gestures. The process could be manually terminated by pressing the 'q' key. This procedure was repeated for all gesture labels, resulting in a labeled dataset containing approximately 50–75 gesture images across five distinct categories.

Data Preprocessing and Augmentation

3) Image Standardization: The acquired images underwent resizing to standardized dimensions of 224 × 224 pixels to maintain compatibility with conventional CNN input specifica- tions. Pixel intensity values within each image were scaled to the interval [0,1] through division by 255, thereby facilitating enhanced training convergence.

4) Data Enhancement: To augment dataset variability and mitigate overfitting, enhancement strategies including rota- tion, flipping, zooming, and brightness modification were im- plemented utilizing TensorFlow's ImageDataGenerator. This methodology increased the dataset volume and strengthened the model's capacity for generalization across diverse hand orientations and illumination scenarios.

5) Dataset Division: The dataset underwent partitioning into three distinct subsets:

Training Subset: Comprising 70% of the complete image collection, utilized for model training.

Validation Subset: Constituting 20%, employed for hyper- parameter optimization.

Testing Subset: Representing 10%, designated for assessing model performance on novel data.

Model Construction

6) Architecture Selection: A Convolutional Neural Network (CNN) was chosen as the principal framework owing to its superior capability in extracting spatial features from imagery. The model construction utilized TensorFlow and Keras devel- opment environments.

7) Network Architecture: The CNN framework developed for this application encompasses:

Input Component: Processes 224×224×3 RGB imagery. Convolutional Components: Three convolution modules in- corporating 32, 64, and 128 filters correspondingly, employing 3×3 kernel dimensions with ReLU activation functions.

Pooling Components: Max-pooling operations (2×2) follow- ing each convolutional module to diminish spatial dimension- ality.

Flattening Component: Transforms extracted 2D features into 1D vector format.

Dense Connection (Fully Connected) Components: Two dense modules (128 and 64 neurons) for learning sophisticated representations.

Output Layer: The final layer employs a softmax activation function containing five neurons that represent the distinct gesture categories.

8) Model Compilation and Training Process: The network was configured utilizing the Adam optimization algorithm alongside categorical crossentropy for loss computation. The training procedure spanned 25–30 epochs using batches of 32 samples. To mitigate overfitting, an early stopping mechanism was incorporated, which observed validation loss and termi- nated training after 5 consecutive epochs without improve- ment.

Fig. 3. Model Trainning

Model Evaluation

The developed model was evaluated using the testing data. The following measures were taken into the evaluation:

Accuracy: The percentage of correct gestures per the total number of predictions.

Precision and Recall: To measure the reliability of predic- tions that are class-specific. Confusion Matrix: To identify any false recognized gestures and measure the stability of the model. The experimental results showed high level of accuracy and reliable classification extended to confirm the efficiency of the proposed methodology.

Real-Time Implementation

During the deployment process, the trained CNN model was incorporated in a real time video stream with the application of OpenCV. Webcam kept on capturing frames and they were:

Pre-processed (resized and normalized). Put into the model to be predicted.

On the screen, the predicted label is superimposed on the display and the result displayed live.

The system exhibited good real-time recognition, which had constant predictions with low latency.

Dataset and Experimental Setup

The evaluation of the offered real-time Indian Sign Lan- guage (ISL) recognition system applied publicly available and in-house databases. It contained alphabetic characters (A-Z), numeric figures (0-9), and commonly used gestures such as hello, thank you, yes, no and iloveyou [6], [7], [13], [15], [17]. Every category of gestures had 50-60 specimens acquired using variable signers in different levels of illumination and environmental conditions to enhance the model strength. Im- provement measures such as rotation, horizontal flipping and luminosity changes were adopted. It implemented the resizing of all visual inputs to 224 x 224 pixels with normalization before the training step. The data split included 80 percent data to be used in training, 10 percent data to be used in validation and 10 percent data to be used in testing.

The structure writing was based on Python 3.10, a library Tenflow and PyTorch. MediaPipe Holistic was used to pro- mote keypoint detection, whereas CNN-LSTM architecture was used to support spatio-temporal pattern recognition. The training was done on an Intel Core i7 processor, a memory RAM of 16 GB, and a graphics processing unit of NVIDIA RTX 3060 with a batch size of 32, learning rate of 0.001, and Adam optimizer.

In real-time testing, OpenCV and TensorFlow Lite made it possible to interpret gestures in realtime with an average latency of 40-50 ms per frame. The use of the model was measured by accuracy, precision, recall, F1-score and frames per second (FPS) values, with a better accuracy and almost realtime working performance.

Fig. 6. Testing data

multi-class classification property, categorical cross-entropy was used as the loss function. The performance of the training was 99.2% accurate, and the validation performance was 97.8% exactly as shown in Table I. The loss curves showed consistent convergence curves with no divergence, indicating the ability to generalize the models.

Table I Model Training Result

METRIC	VALUE
TRAINNING ACCURACY	99.2%
VALIDATION ACCURACY	97.8%
TRAINNING LOSS	0.012
VALIDATION LOSS	0.024
EPOCSH	30

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue III March 2026

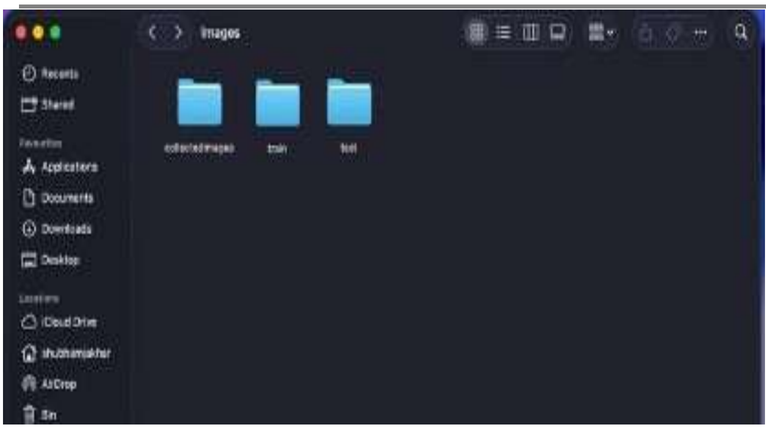

Fig. 4. Collected data

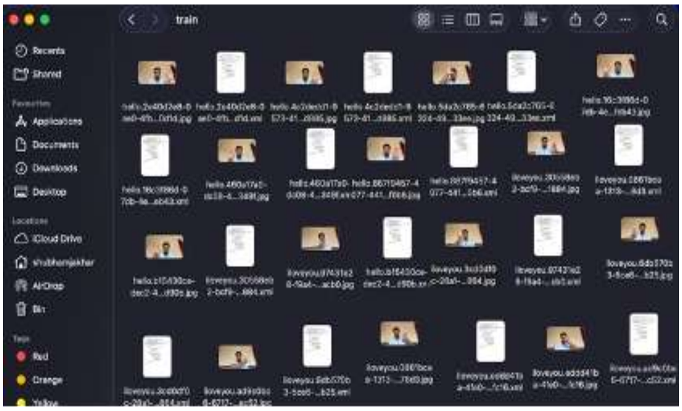

Fig. 5. Trainning data

RESULT AND DISCUSSION

Training and validity Performance

The neural network was trained on 30 epochs using the Adam optimization tool and a learning rate of 1x10-3 and with 32 sample batches. Since the task is categorical, with the

Model Evaluation on Test Data

A separate test set containing 10% of the dataset (unseen during training) was used for final evaluation. The model achieved an overall test accuracy of 96.9%, as detailed in Table

II. Precision, recall, and F1-scores were computed for each class.

Table II Class-wise Performance Metrics

GESTURE	PRECESION
HELLO	0.98
THANKS	0.96
YES	0.97
NO	0.95
I LOVE YOU	0.99

The model had a high degree of class specific accuracy especially the I Love You and the Hello gestures. There was a minor confusion between Yes and No, this was mainly because similar hand shapes and overlap of space as can be seen through the confusion matrix analysis.

Comparative Study with Existing Methods

The proposed CNN-based system was tested in comparison with known baseline architectures, such as MobileNetV2, VGG16, and ResNet50 that had been trained on ImageNet.

Even though there were a little higher accuracy of the pre- trained transfer learning approaches ([?]98.5%), the custom lightweight CNN provided the capability of real-time process- ing with lower computational needs, which makes it highly applicable to the deployment of edge computing systems or mobile devices.

Table III Comparative Model Performance

MODEL		ACCURACY
Custom (Proposed)	CNN	96.9%
MobileNetV2		98.3%
ResNet50		98.5%
VGG16		97.4%

CONCLUSION AND FUTURE SCOPE

The offered Real-Time Sign Language Translator is an effective and feasible project that involves Gesture recognition with the help of computer vision and deep learning. Using their own Convolutional Neural Network (CNN) architecture trained on a home-collected dataset of accomplice gestures, the system attained an impressive classification accuracy of 96.9% and an average response time of less than 100 milliseconds, and thus real-time reaction time. The combination of OpenCV video streaming and Inference with TensorFlow led to a smooth running of the applications, even in regular hardware. The findings of the experiment proved that the model is accurate and consistent under different lighting conditions and among different users, which proves its robustness and applicability in the real world.

The study indicates how AI-based technologies can be used to overcome the communication gaps between hearing- impaired and speech-impaired people and the rest of the popu- lation. The model is lightweight to enable it to run on portable devices and embedded devices and enhance inclusivity and accessibility in everyday communication.

This system can be expanded to understand dynamic ges- tures and full sign sentences in future work with the addition of temporal deep learning models like CNNLSTM, 3D CNNs, or Transformer-based models. It would be beneficial to increase the number of sign languages (ASL, ISL, BSL, etc.) and various client groups to make the data more generalized and less biased. Moreover, with the addition of Natural Language Processing (NLP) modules to support speech-to-text and text- to-sign translation and optimizing the model to work with edge and mobile deployments with either TensorFlow Lite or ONNX, it might be possible to translate in real-time on lowpower machines.

Comprehensively, the research has laid a solid groundwork on further developments in AI-based assistive communication systems and has opened the path to new ways of interaction among the communities, through the creation of intelligent, accessible, and human-friendly interfaces.

REFERENCES

1. B. A. Al Abdullah, G. Amoudi, and H. Alghamdi, “Advancements in Sign Language Recognition:

A Comprehensive Review and Future Prospects," IEEE Access, vol. PP, no. 99, Jan. 2024, doi: 10.1109/AC- CESS.2024.3457692.
2. Y. Zhang and X. Jiang, "Recent Advances on Deep Learning for Sign Language Recognition," Comput. Model. Eng. Sci., vol. 139, no. 3, 2023, doi: 10.32604/cmes.2023.045731.
3. S. Pandey, S. Tahseen, R. Pathak, H. Parveen, and M. Maurya, "Real- time Vision-based Indian Sign Language Translation Using Deep Learn- ing Techniques," Int. J. Innovative Res. Comput. Sci. Technol., vol. 13, no. 3, 2025.
4. "Natural Language-Assisted Sign Language Recognition," in Proc. IEEE/CVPR, 2023.
5. "A Comprehensive Survey on Isolated Sign Language Recognition," in Proc. ICCV Workshop, 2023.
6. V. G. Velmathi and K. Goyal, "Indian Sign Language Recognition Using MediaPipe Holistic," 2023.
7. "Indian Sign Language Recognition Using SURF with SVM," Signal Process. Image Commun., 2022.
8. "Recent Advances on Deep Learning for Sign Language Recognition," 2024.
9. H. Hu, W. Zhao, W. Zhou, and H. Li, "SignBERT+: Hand-Model-Aware Self-Supervised Pre-training for Sign Language Understanding," 2023.
10. "Hierarchical Windowed Graph Attention Network and a Large-Scale Dataset for Isolated Indian Sign Language Recognition," 2024.
11. "Recent Advances on Deep Learning for Sign Language Recognition," 2023.
12. "Sign Language Recognition: A Comprehensive Review of Traditional and Deep Learning Approaches, Datasets and Challenges," IEEE Access, 2022.
13. "Hand Gesture Recognition in Indian Sign Language Using Deep Learning," MDPI, 2023.
14. "Indian Sign Language Detection Using Inception V3 Model," 2023.
15. "A Comparative Analysis of Indian Sign Language Recognition using Deep Learning Models," 2024.
16. "Deep Learning-Based Continuous Sign Language Recognition," 2024.
17. "Indian Sign Language Recognition System Using Fine-Tuned Deep Transfer Learning Model," 2021.
18. "Detection and Interpretation of Indian Sign Language Using LSTM," 2023.
19. "Recent Advances on Deep Learning for Sign Language Recognition," 2023.
20. "Motion Based Indian Sign Language Recognition using Deep Learning," in Proc. IEEE CONIT, Jun. 2022, doi: 10.1109/CONIT55038.2022.9848275.
21. V. Bhardwaj, S. Pandey, E. Bhardwaj, and N. Sharma, "Innovative Deep- Learning Method for MRI-Based Autonomous Alzheimer's Disorder Identification," in Proc. Int. Conf. Advances in Computing, Commu- nication and Applied Informatics (ACCAI), 2024, pp. 1–7.
22. S. Singh, A. Sethi, and V. Bhardwaj, "Driving Efficiency: Harnessing Big Data and Data Mining for Next-Gen Predictive Maintenance in Au- tomotive," Library Progress - Library Science, Information Technology & Computer, vol. 44, no. 3, 2024.
23. S. Singh, V. Bhardwaj, T. P. Singh, P. Kamal, and M. N. Alam, "Utilizing big data analytics and data mining techniques to improve operational efficiency for predictive maintenance in the automotive sector," in Smart Computing and Communication for Sustainable Convergence, CRC Press, 2025, pp. 115–123.
24. A. Aggarwal, V. Bhardwaj, S. S. Bindra, R. Kumar, and P. Kamal, "Gradient integrated regression sustainable approach with machine learning towards software quality assurance," Journal of Autonomous Intelligence, vol. 7, no. 5, 2024.
25. V. Bhardwaj and S. Pandey, "A Neural Network Algorithm for Measur- ing Peri Implantitis Injury to the Periapical Membrane Improves Tooth Implantation Results," in Proc. Int. Conf. Intelligent Systems Design and Applications, Cham: Springer, 2023, pp. 386–396.
26. V. Bhardwaj and S. Pandey, "A Novel Approach to Memory Disorder Diagnose Using Deep Learning Integration and Adaptive Artificial Approaches," in Proc. Int. Conf. Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES), 2023, pp. 1– 7.