Deaf And Mute Language Identification Using Machine Learning: A Review

Vaishnavi Yelnare*, Santosh Gaikwad**, Arshiya Khan***, R.S. Deshpande****
Department Of Computer Science and Application
JSPM UNIVERSITY
Email: vaishnaviyelnare5@gmail.com
**Associate Professor
Faculty of Science and Technology, JSPM University Pune
Email: santosh.gaikwadcsit@gmail.com
***Assistant Professor
Faculty of Science and Technology, JSPM University Pune
Email: ak.scos@jspmuni.ac.in
****Professor and Dean, Faculty of Science and Technology
JSPM University Pune
Email: dean-fst@jspmuni.ac.in

***Abstract*—Deaf and mute individuals face unique communication barriers that limit access to education, healthcare, and employment opportunities. Machine Learning (ML) has emerged as a transformative technology that enables intelligent systems to learn from data, offering promising solutions for automatic sign language recognition and gesture interpretation. This review paper explores the application of ML techniques in identifying and interpreting sign languages used by deaf and mute individuals. It covers current models, datasets, real-time systems, challenges, and future research directions. The paper also presents real- world use cases, ethical considerations, and the role of NGOs and governments in promoting inclusive communication technologies.**

I. Introduction

The inability to speak or hear can significantly hinder communication, especially in a society where verbal and auditory cues are dominant. Deaf and mute individuals primarily rely on sign languages, gestures, and facial expressions for daily interactions. Unfortunately, most people do not understand sign language, leading to social and communication exclusion.

Sign language, a rich visual language with unique grammar and syntax, varies from one region to another, making its recognition a complex task [2]. The recent advancement of computer vision and ML algorithms has enabled systems that can translate these signs into spoken or written language in real time [1], [3].

Traditional systems used gloves or sensors to detect gestures but were inconvenient for regular use [6]. With improvements in image processing and video analysis, systems now use normal webcams or smartphone cameras to capture hand movements and facial expressions [3], [8]. Deep learning models like CNNs, RNNs, and Transformer architectures have improved accuracy and flexibility [16], making it possible to interpret continuous sign sequences.

This paper provides a thorough review of ML approaches for sign language recognition, datasets, model performance, system designs, real-world deployments, and challenges. It also discusses the ethical implications and the need for in- clusive and fair systems.

II. Background and Motivation

According to the World Health Organization (WHO), over 70 million people globally are hearing or speech impaired. They often face difficulties in communicating with the gen- eral population due to a lack of interpreters or accessible technologies. Automatic sign language recognition using ML can bridge this gap, allowing deaf and mute individuals to communicate more freely and access services that were previously inaccessible [17].

III. Sign Language Modalities

Sign language includes multiple components that must be processed together for accurate interpretation [13], [14]:

- **Manual Features:** Hand shape, palm orientation, hand movement.
- **Non-Manual Features:** Facial expressions, head nods, and body posture.
- **Temporal Features:** Sequence of signs forming complete phrases or sentences.

IV. Machine Learning Models for Recognition

A. *Convolutional Neural Networks (CNNs)*

CNNs are widely used for image-based gesture recognition [11]. They extract spatial features from input frames, identifying key patterns related to hand shapes and positions.

B. *Recurrent Neural Networks (RNNs) and LSTM*

RNNs are designed to handle sequential data, making them suitable for dynamic gesture recognition [12]. LSTMs address vanishing gradient problems and capture long-term dependencies in sign sequences.

C. *3D-CNN and Transformer Models*

3D-CNNs extend standard CNNs to include temporal dimensions, analyzing frame sequences simultaneously [14]. Transformer models offer attention mechanisms, improving continuous sign language translation accuracy [16].

V. Feature Extraction Techniques

- **OpenPose:** Detects body, hand, and facial keypoints for pose estimation [18].
- **MediaPipe:** Google's cross-platform library for real-time hand and face tracking [18].
- **Optical Flow:** Captures movement patterns for dynamic sign recognition [15].
- **HOG Descriptors:** Extract orientation and gradient-based features [15].

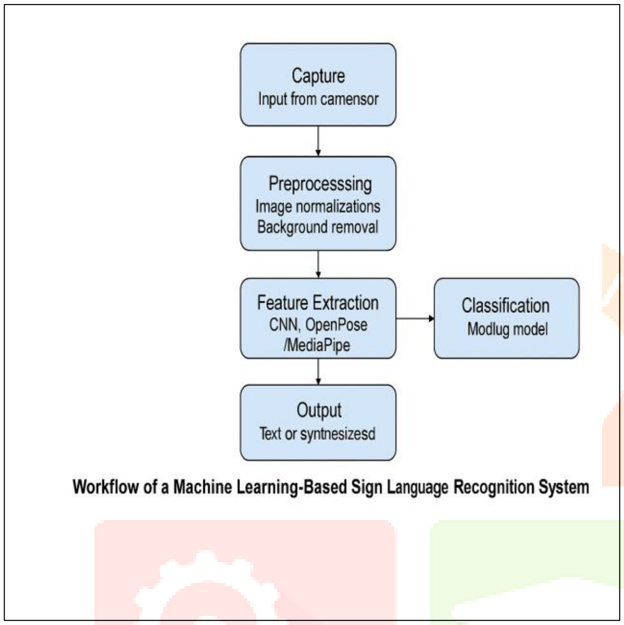

Fig. 1: Workflow of a Machine Learning-Based Sign Language Recognition System

VI. Datasets for Training and Evaluation

These datasets form the foundation of model training and evaluation across global sign language research [4], [5], [9], [10], [20].

TABLE I: Popular Datasets for Sign Language Recognition

Dataset	Language	Type
ASLLVD	American Sign Language	Video
RWTH-PHOENIX	German Sign Language	Video
LSA64	Argentine Sign Language	Image
SIGNUM	German Sign Language	Audio-Visual
ISL Dataset	Indian Sign Language	Video/Image

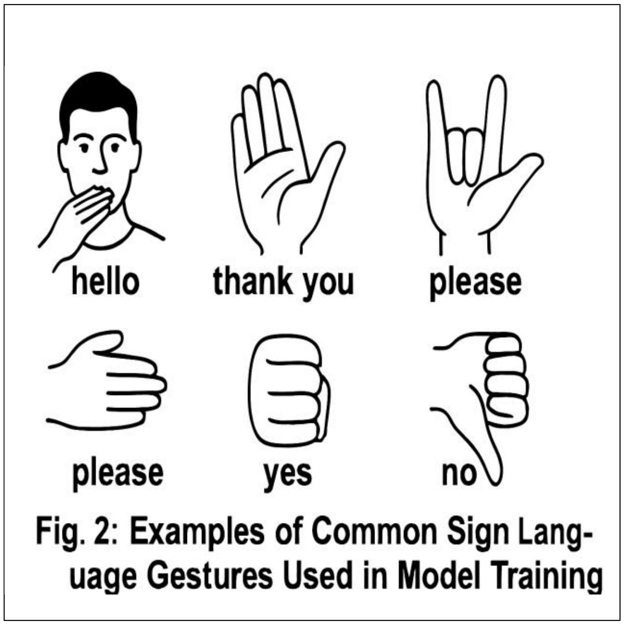

Fig. 2: Examples of Common Sign Language Gestures Used in Model Training

VII. System Architecture

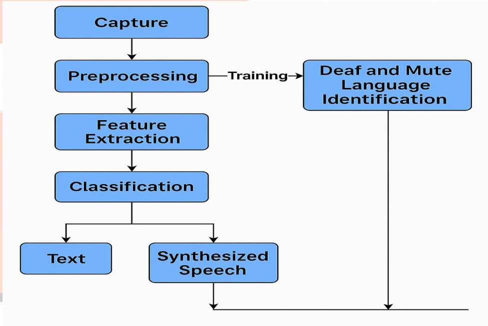

Fig. 3: Flowchart illustrating the architecture of a machine learning-based sign language recognition system.

A real-time sign recognition system generally follows these steps:

1) **Capture:** Input from camera or sensor.
2) **Preprocessing:** Image normalization, background removal.
3) **Feature Extraction:** Using CNN/OpenPose/MediaPipe.
4) **Classification:** Sign classification using ML model.
5) **Output:** Converted to text or synthesized speech.

VIII. Applications

- **Healthcare:** Enabling communication between doctors and patients [19].
- **Education:** Making learning content accessible [18].
- **Customer Service:** Kiosks and chatbots for sign language users [7].
- **Translation Tools:** Mobile apps for real-time sign to speech/text conversion [19].

IX. Evaluation Metrics

- **Accuracy:** Correct classification rate.
- **Precision/Recall:** Useful for class-imbalanced datasets.
- **F1 Score:** Harmonic mean of precision and recall.
- **Latency:** Important for real-time deployment [19].

X. Challenges

- **Signer Independence:** Performance drops with new users [17].
- **Background Noise:** Dynamic environments affect accuracy [3].
- **Data Scarcity:** Limited annotated datasets for regional sign languages [17], [20].
- **Computation Cost:** Requires efficient models for edge devices [19].

XI. Comparative Analysis

TABLE II: Comparison of ML Techniques

Model	Accuracy	Real-Time	Resource
CNN	High	Yes	Moderate
RNN	High	No	High
LSTM	Very High	Yes	High
Transformer	Very High	Yes	Very High
SVM	Medium	Yes	Low

The above table provides a comparative evaluation of various machine learning models used for sign language recognition. Convolutional Neural Networks (CNNs) are effective in extracting spatial features and are suitable for real-time applications with moderate resource requirements. Recurrent Neural Networks (RNNs), although accurate, are not ideal for real-time systems due to their sequential processing and high computational load. Long Short-Term Memory networks (LSTMs) offer better accuracy and real-time performance by handling long-term dependencies effectively. Transformer models surpass others in accuracy and speed, but they require significantly more computational resources. Support Vector Machines (SVMs) are lightweight and efficient for small-scale or static gesture datasets, but they lack the accuracy of deep learning approaches. This comparison highlights the trade- off between performance, speed, and resource consumption in selecting a model for real-time sign language recognition systems.

XII. Case Studies

A. SignAll

Uses computer vision and NLP for ASL to English translation in real time [7].

B. Google's Teachable Machine

Allows training ML models for gestures without coding, useful for sign training [19].

C. Indian Startups

Platforms like "SignAble" and "SayItSign" are developing ISL tools for accessibility [20].

XIII. NGO and Government Initiatives

- **ISLRTC:** Government initiative for Indian Sign Language research [20].
- **WFD:** Global advocacy for deaf rights and technological inclusion.
- **Accessible India Campaign:** Promoting inclusive technology in India.

XIV. Ethical Considerations

ML systems must be inclusive and fair. Key considerations:

- **Bias Mitigation:** Ensure diversity in training data [17].
- **Privacy:** Protect video inputs and user information.
- **Transparency:** Explain model decisions to build trust.

XV. Future Scope

- **Integration of Emotion Recognition for Context-Aware Communication:** Future systems can enhance sign interpretation by detecting emotional cues through facial expressions. This helps in understanding the intent and tone behind gestures, improving the overall accuracy and naturalness of communication [18].
- **Multi-Modal Systems Combining Audio, Video, and Skeletal Data:** Combining hand gestures, facial expressions, audio (if available), and skeletal tracking can create more robust recognition systems. These multi- modal approaches improve generalization across users and environments [16], [17].
- **Use of Federated Learning to Train Models On-Device Without Data Sharing:** Federated Learning enables training models directly on users' devices, preserving privacy by keeping data local. This technique is promising for secure, real-time recognition in sensitive applications such as healthcare [19].

XVI. Limitations and Research Gaps

- **Regional sign languages remain underrepresented.** [20]
  While considerable research has focused on widely recognized sign languages such as American Sign Language (ASL) and British Sign Language (BSL), many regional sign languages, such as Indian Sign Language (ISL), remain significantly underexplored. This lack of representation leads to biased models that are not inclusive of diverse linguistic communities. Consequently, systems developed using these limited datasets may not generalize well across cultures and regions.
- **Lack of longitudinal studies for user adaptation.** [17]
  Most studies in sign language recognition assess model

performance using short-term datasets collected under controlled environments. However, they do not evaluate how users adapt to these systems over time or how model accuracy holds up in real-world, long-term usage scenarios. Without longitudinal studies, it's difficult to measure consistency, learning curves, or user satisfaction in dynamic environments.

- **Limited support for dialects and informal gestures.** [13]
  Sign languages are highly expressive and often contain dialectal variations and informal gestures that are used in daily communication but are not present in formal datasets. Current models typically fail to recognize these informal signs, which reduces the effectiveness of recognition systems in practical use. To achieve better accuracy and user acceptance, it is essential to incorporate diverse and realistic gesture data.

XVII. Conclusion

Machine learning has immense potential to improve communication for the deaf and mute community. With the development of robust models, real-time systems, and inclusive datasets, sign language recognition can become widely accessible. Continued research, combined with ethical deployment and policy support, will drive this technology towards a more inclusive future.

Furthermore, the integration of emerging technologies such as edge computing and federated learning can enable more secure and real-time gesture recognition on portable devices, reducing dependency on centralized servers. Collaboration between researchers, linguists, NGOs, and governments is essential to create standardized sign language corpora and ensure models are inclusive of regional and cultural variations.

Public awareness campaigns and open-source contributions can also accelerate the development and adoption of assistive technologies. As machine learning continues to evolve, the focus must remain on building equitable systems that empower and uplift the deaf and mute community globally.

References

[1] Camgoz, N.C. et al., "Neural Sign Language Translation," CVPR, 2018.
[2] Albanie, S. et al., "BSL-1K: Scaling Sign Language Recognition," ECCV, 2020.
[3] Zhao, H. et al., "Skeleton-based Sign Recognition," IEEE Access, 2021.
[4] RWTH-PHOENIX Dataset, 2014T.
[5] Boston University ASLLVD Dataset.
[6] Koller, O. et al., "Deep Sign: Hybrid CNN-HMM," BMVC, 2017.
[7] Bragg, D. et al., "Sign Language Generation," ACM SIGACCESS, 2019.
[8] Neverova, N. et al., "ModDrop for Gesture Recognition," TPAMI, 2014.
[9] LSA64 Dataset, University of Buenos Aires.
[10] Signum Dataset, University of Hamburg.
[11] Ko, J. et al., "American Sign Language Recognition with Convolutional Neural Networks," Stanford CS231n Report, 2019.
[12] Cui, R. et al., "Recurrent Convolutional Neural Networks for Continuous Sign Language Recognition," ACM MM, 2017.
[13] Huang, J. et al., "Video-based Sign Language Recognition without Temporal Segmentation," AAAI, 2018.
[14] Sincan, M. and Keles, H.Y., "Autonomous Sign Language Recognition Using 3D CNNs," Neural Computing and Applications, 2020.
[15] Pfister, T. et al., "Recognising Gesticulation in Sign Language using HOG and Optical Flow," BMVC, 2013.
[16] Cihan Camgoz, N. et al., "Sign Language Transformers: Joint End-to-End Sign Language Recognition and Translation," CVPR, 2020.
[17] Adaloglou, N. et al., "A Comprehensive Study on Sign Language Recognition Methods," Journal of Imaging, 2021.
[18] Arti, S. et al., "Hand Gesture Recognition using Deep Learning for Sign Language," Procedia Computer Science, 2022.
[19] Wang, J. et al., "Deep Learning for Sign Language Recognition on Mobile Devices," IEEE Transactions on Industrial Informatics, 2020.
[20] Indian Sign Language Research and Training Center (ISLRTC) Datasets, Government of India.