Real-Time Sign Language Recognition and Text Conversion Using LSTM for Enhanced Communication Accessibility

Anish S*, Dharani K**, Jayasri B***, Mr. R. Rajkumar†

*UG Scholar, **UG Scholar, ***UG Scholar
Department of Computer Science and Engineering
Chettinad College of Engineering and Technology, Puliyur, Karur, Tamil Nadu, India

†HOD, Department of Computer Science and Engineering
Chettinad College of Engineering and Technology, Puliyur, Karur, Tamil Nadu, India

*anishscse2003@gmail.com**, *dharaniker@gmail.com***, *srijai2352@gmail.com****, *rajkumar@chettinadtech.ac.in*†

ABSTRACT

Sign language recognition has gained significant attention in recent years as an essential tool for bridging the communication gap between individuals with hearing impairments and the broader society. Traditional approaches relying on Convolutional Neural Networks (CNNs) have shown promise in recognizing static hand gestures, but they struggle with sequential and dynamic sign language interpretation. This survey paper explores recent advancements in Long Short-Term Memory (LSTM)-based deep learning models, which excel at temporal sequence processing, making them more suitable for real-time sign language recognition and text conversion. The paper reviews state-of-the-art methodologies used in sign language translation, focusing on the effectiveness of LSTM networks in capturing gesture sequences. A comparative analysis is conducted on existing datasets, including Indian Sign Language (ISL), American Sign Language (ASL), and Bengali Sign Language (BdSL), highlighting their limitations and applicability to deep learning models. Furthermore, key challenges such as gesture complexity, real-time processing constraints, dataset availability, and multilingual sign recognition are discussed. This survey aims to provide a comprehensive evaluation of existing research, offering insights into hybrid deep learning models, transformer-based approaches, and multimodal recognition techniques as potential solutions for improving accuracy and efficiency in AI-driven sign language translation systems. By addressing these challenges, the study seeks to guide future research in developing scalable, real-time, and accessible sign language recognition technologies.

KEYWORDS

Real-Time Sign Language Recognition, LSTM-Based Gesture Recognition, Sign Language to Text Conversion, Deep Learning for Sign Language, Sequential Gesture Processing.

I. INTRODUCTION

Sign language serves as a fundamental mode of communication for individuals with hearing and speech impairments. These systems aim to bridge the communication gap by converting sign gestures into text and speech, thereby improving accessibility for the hearing-impaired community.

Existing research has explored various approaches to sign language recognition, including Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), Vision Transformers (ViTs), and sensor-based gesture recognition. Studies have also investigated the use of 3D modeling, text-to-sign generation, and real-time sign detection. Notably, the Hybrid InceptionNet-Based Enhanced Architecture for Isolated Sign Language Recognition, which serves as the foundation of this survey, has demonstrated high accuracy (98.46%) in recognizing isolated signs. However, it primarily focuses on single-word gestures rather than continuous sign sequences, highlighting a key research gap in real-time sentence-level recognition.

This survey reviews recent advancements in sign language recognition and translation, analyzing different methodologies such as computer vision-based detection, machine learning classification, 3D avatar-based sign generation. By examining these approaches, this paper aims to identify the strengths, limitations, and future directions for building more accurate, scalable, and real-time sign language translation systems.

A. *Core Themes of This Survey*

The paper is structured around four core themes:

- **Traditional vs. AI-Based Methods:** Shift from sensor-based techniques to deep learning models.
- **Deep Learning Architectures:** CNNs, RNNs, transformers, and hybrid models improve accuracy.
- **Datasets and Evaluation Metrics:** Benchmark datasets and performance measurement techniques.
- **Challenges and Future Directions:** Addressing real-time processing, dataset limitations, and model efficiency.

By consolidating existing research and advancements, this survey explores the shift from traditional sensor-based methods to AI-driven deep learning models. It also examines key challenges like real-time processing, dataset limitations, and model efficiency to guide future research.

II. Review of Existing Research Papers

Sign Language Recognition (SLR) has evolved significantly, shifting from sensor-based techniques to deep learning-driven approaches. Early systems relied on glove-based methods with motion sensors, but these were costly and restrictive. The emergence of computer vision techniques introduced feature extraction methods like edge detection and skin segmentation, though they struggled with varying lighting and backgrounds.

Deep learning models, particularly Convolutional Neural Networks (CNNs), improved static sign recognition by learning spatial patterns, but lacked efficiency in dynamic gesture sequences. Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks addressed sequential gesture recognition but suffered from high computational costs.

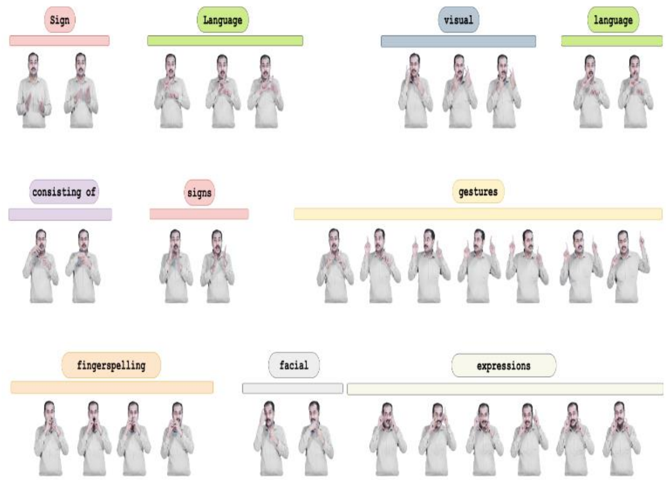
More recently, transformer-based models and hybrid architectures like InceptionNet and ResNet have demonstrated superior accuracy and robustness. These models leverage attention mechanisms and ensemble learning, making SLR systems more scalable and efficient for real-world applications. However, challenges like dataset limitations, real-time processing constraints, and high computational requirements still need to be addressed for widespread adoption.

A. *Comparison of Past Methodologies*

Early sign language recognition systems relied on glove-based methods, where sensors captured hand movements and gestures. While these systems provided high accuracy in controlled environments, they were costly, uncomfortable, and limited by hardware dependencies. Later, computer vision techniques replaced gloves, using edge detection, skin color segmentation, and shape analysis to recognize signs. However, these approaches struggled with lighting variations and required extensive preprocessing.

With the rise of deep learning, Convolutional Neural Networks (CNNs) became popular for static sign recognition, excelling at feature extraction but failing to handle sequential gestures effectively. To address this, Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTMs) were introduced to capture temporal dependencies in dynamic sign sequences. Despite their effectiveness, these models suffered from high computational costs and vanishing gradient issues.

Recently, transformer-based models have emerged, leveraging self-attention mechanisms to improve accuracy in recognizing continuous gestures. However, they require large datasets and significant processing power. To further enhance recognition, hybrid architectures combining CNNs with RNNs, and ensemble learning methods like InceptionNet and ResNet, have been developed. These models achieve higher accuracy and robustness, but their increased complexity and computational demands pose challenges for real-time applications.



A sample from ISLTranslate: "Sign Language is a visual language consisting of signs, gestures, fingerspelling and facial expressions."

III. Strengths and Weaknesses of Existing Approaches

A. *Strengths*

- **Improved Accuracy:** Deep learning models (CNNs, RNNs, Transformers) achieve high recognition accuracy, especially hybrid models like InceptionNet (98.46% in your base paper).
- **Automation & Scalability:** AI-driven systems eliminate the need for specialized gloves, making them more scalable and adaptable across different sign languages.
- **Robust Feature Learning:** Advanced architectures capture both spatial and temporal features, improving recognition of isolated and continuous gestures.
- **Better Human-Computer Interaction (HCI):** AI-based models enhance accessibility, making real-time applications like sign-to-text and sign-to-speech translation more feasible.

B. *Weaknesses*

- **High Computational Cost:** Transformer-based and ensemble models require large datasets and powerful GPUs, making real-time deployment difficult.
- **Real-World Limitations:** Variations in lighting, hand orientation, background noise, and occlusions affect recognition accuracy.
- **Data Dependency:** Models require large, diverse datasets for training, which are often limited or unavailable for many sign languages.
- **Latency in Real-Time Applications:** Deep learning models are often slow in processing real-time gestures, making them less practical for instant communication.

IV. Deep Learning-Based Models

Deep learning has significantly improved Sign Language Recognition (SLR) by enabling automated and accurate gesture classification. Convolutional Neural Networks (CNNs) are widely used for static sign recognition, as they effectively extract spatial features from images. However, CNNs struggle with dynamic gestures, which require sequential processing. To address this, Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks have been introduced, allowing models to capture temporal dependencies in continuous sign sequences.

More recently, transformer-based models have emerged, leveraging self-attention mechanisms to improve contextual understanding of sign gestures. These models, such as SignTransformer and Vision Transformer (ViT), outperform RNNs in long-sequence recognition but demand high computational power. Hybrid models combine CNNs with RNNs or transformers, enhancing recognition accuracy by learning both spatial and temporal features. Additionally, ensemble learning approaches, like the Hybrid InceptionNet model from your base paper, integrate multiple deep learning architectures to improve robustness, achieving 98.46% accuracy.

V. Hybrid Models Combining Multiple Techniques

Hybrid models in Sign Language Recognition (SLR) combine multiple deep learning techniques to improve accuracy and robustness. Traditional CNNs are effective for extracting spatial features from images, while RNNs and LSTMs handle sequential dependencies in dynamic gestures. However, these models struggle with long-term dependencies and complex hand movements. To overcome these limitations, transformer-based models are integrated with CNNs to enhance contextual understanding.

InceptionNet-based hybrid architectures, like the one in your base paper, optimize feature extraction and learning through ensemble techniques, achieving higher recognition accuracy (98.46%). Despite their effectiveness, hybrid models come with higher computational costs, making real-time deployment challenging. Future advancements should focus on reducing model complexity, improving efficiency, and integrating multimodal learning for real-world applications.

VI. Challenges in Real-Time Sign Language Recognition

Despite significant advancements in deep learning-based Sign Language Recognition (SLR), real-time implementation remains a major challenge. Most models, including CNNs, RNNs, and transformers, require high computational power for processing complex hand gestures, making them difficult to deploy on edge devices. Variability in lighting, backgrounds, hand orientations, and signer differences further affects recognition accuracy.

Additionally, large dataset requirements limit model generalization, as many sign languages lack sufficient annotated training data. To address these challenges, future research should focus on lightweight neural networks, dataset augmentation techniques, and hardware-efficient deep learning models to ensure faster and more reliable real-time SLR systems.

VII. Advantages and Limitations of Various Approaches

A. Glove-Based Systems

Advantages: High accuracy in controlled environments, directly captures hand movements.
Limitations: Expensive, uncomfortable, and limited flexibility due to hardware dependency.

B. Computer Vision-Based Methods

Advantages: No need for specialized hardware, works with standard cameras.
Limitations: Affected by lighting variations, background noise, and requires complex preprocessing.

C. Deep Learning Models (CNNs, RNNs, Transformers)

Advantages: Automatically learns features, provides high accuracy for static and dynamic signs.
Limitations: Requires large datasets, high computational power, and struggles with real-time processing.

D. Hybrid and Ensemble Models

Advantages: Combines strengths of multiple architectures, improves recognition accuracy and robustness.
Limitations: Computationally expensive, complex training process, and challenging real-time implementation.

VIII. Discussion on Real-World Applicability

A. Assistive Technologies for the Hearing Impaired

SLR systems can be integrated into mobile applications, smart gloves, and wearable devices to enable real-time communication between hearing-impaired individuals and the general public.

B. Human-Computer Interaction (HCI) and Smart Assistants

AI-powered sign language recognition can enhance voice assistants, chatbots, and virtual interpreters, making digital communication more inclusive.

C. Education and Learning Platforms

SLR can be used in online learning tools and smart classrooms to assist students with hearing disabilities in understanding educational content more effectively.

D. Public Services and Accessibility

Governments and organizations can implement SLR systems in hospitals, banks, airports, and customer service centers to improve accessibility for the hearing-impaired community.

E. Challenges in Deployment

Despite its potential, real-world implementation faces challenges such as hardware limitations, environmental variations, dataset scarcity, and real-time processing constraints. Future advancements should focus on developing lightweight models, optimizing real-time performance, and expanding dataset diversity.

IX. Current Limitations in the Field

A. High Computational Requirements

Deep learning models like CNNs, RNNs, and transformers require powerful hardware, making real-time deployment difficult on mobile and edge devices.

B. Limited and Imbalanced Datasets

Many sign languages lack large, diverse datasets, leading to bias in recognition models and reduced accuracy for underrepresented gestures.

C. Real-Time Processing Challenges

High latency and computational complexity prevent instantaneous sign recognition, affecting usability in live communication scenarios.

D. Variability in Sign Language and Gestures

Different regions use distinct sign languages and variations, making it hard to develop a single model that works universally.

E. Environmental and Practical Constraints

Changes in lighting, background clutter, hand occlusions, and camera angles reduce model performance in real-world conditions.

X. Potential Research Opportunities

A. Development of Lightweight Models

Optimizing deep learning architectures to run efficiently on low-power devices and mobile platforms for real-time sign language recognition.

B. Expansion of Multimodal Learning

Integrating gesture, facial expression, speech, and text recognition to create more comprehensive and context-aware SLR systems.

C. Creation of Large-Scale, Diverse Datasets

Developing standardized, multi-language datasets covering different sign languages, signer variations, and real-world conditions to improve model generalization.

D. Enhancing Real-Time Processing and Deployment

Researching techniques like quantization, pruning, and edge computing to reduce computational complexity and enable faster, real-time recognition.

E. Ethical AI and Privacy-Preserving SLR Systems

Exploring methods for secure data collection, user consent mechanisms, and bias mitigation to ensure fair and ethical AI applications in sign language recognition.

XI. Emerging Technologies That Could Improve Existing Methods

A. Transformer-Based Architectures

Advanced models like Vision Transformers (ViT) and SignTransformers improve context understanding and sequential gesture recognition, outperforming traditional CNNs and RNNs.

B. Edge AI and Low-Power Computing

Deploying sign language recognition on edge devices and mobile platforms using model compression techniques (e.g., quantization, pruning) for real-time performance.

C. Multimodal AI Integration

Combining gesture recognition with speech, text, and facial expression analysis to create more natural and inclusive communication systems.

D. 3D Pose Estimation and Depth Sensing

Using technologies like LiDAR, depth cameras (e.g., Microsoft Kinect), and 3D hand tracking to enhance gesture recognition accuracy, even in challenging environments.

XII. Real-Time Applications and Practical Implementation Issues

A. Sign Language Translation Systems

Real-time sign-to-text and sign-to-speech systems can bridge communication gaps, but they require fast and efficient processing for seamless interaction.

B. Integration with Assistive Technologies

SLR models can be embedded into mobile apps, smart glasses, and virtual interpreters, but hardware limitations and processing delays affect usability.

C. Challenges in Real-World Environments

Factors like lighting variations, background noise, hand occlusions, and signer differences reduce recognition accuracy in uncontrolled settings.

D. Computational Constraints on Edge Devices

Deploying SLR on smartphones, IoT devices, and embedded systems is challenging due to high computational costs and energy consumption.

E. *Need for Standardized Benchmarking and Evaluation*

Lack of universal performance metrics and real-world testing frameworks makes it difficult to measure and compare different SLR models effectively.

XIII. SUMMARY OF KEY FINDINGS

This survey paper explored the integration of Machine Learning (ML) and Geospatial Artificial Intelligence (GeoAI) for flood hotspot prediction, emphasizing the development of a real-time Flask-based web application for data visualization and mapping. Key findings include:

- **Evolution of Sign Language Recognition (SLR):** SLR has advanced from sensor-based and computer vision methods to deep learning-based CNNs, RNNs, transformers, and hybrid models, improving accuracy and automation.

- **Strengths and Weaknesses of Existing Models:** While deep learning models achieve high recognition accuracy, they face challenges in real-time processing, dataset diversity, and computational efficiency.

- **Emerging Technologies Enhancing SLR:** Innovations like Vision Transformers, 3D pose estimation, Edge AI, and multimodal learning offer promising solutions to improve real-world usability.

- **Real-Time Implementation Challenges:** SLR deployment in mobile devices, IoT systems, and public services is limited by high computational costs, environmental variability, and lack of standardized benchmarks.

- **Future Research Directions:** Advancements in lightweight models, dataset expansion, and privacy-preserving AI are essential to make SLR more accessible, scalable, and practical for everyday use.

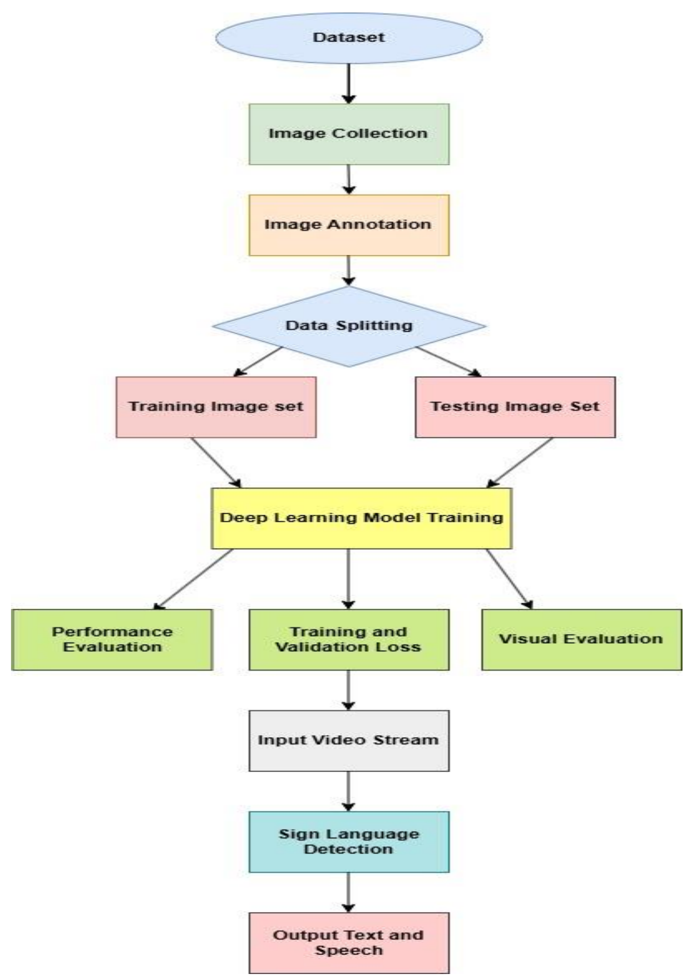
XIV. FINAL THOUGHTS ON ADVANCEMENTS IN THE FIELD

Sign Language Recognition (SLR) has made significant strides with deep learning-based models, achieving high accuracy and improved automation. The shift from sensor-based systems to AI-driven approaches has enabled better scalability and accessibility. However, challenges in real-time deployment, dataset availability, and computational efficiency still hinder widespread adoption.

Emerging technologies like Vision Transformers, Edge AI, and multimodal learning offer promising solutions for enhancing recognition accuracy and usability. Moving forward, research should focus on developing lightweight, real-time SLR models, improving dataset diversity, and ensuring ethical AI practices. With continued advancements, SLR can become an essential tool for inclusive and barrier-free communication in everyday life.

XV. FUTURE PERSPECTIVES AND POSSIBLE RESEARCH DIRECTIONS

- The future of Sign Language Recognition (SLR) lies in developing more efficient, accurate, and accessible systems.
- Research should focus on lightweight deep learning models optimized for real-time performance on mobile and edge devices.
- Multimodal learning, integrating gesture, facial expression, and speech recognition, can enhance contextual understanding.
- Expanding large-scale, diverse datasets is crucial to improve generalization across different sign languages and signer variations.
- Federated learning and privacy-preserving AI can address ethical concerns in data collection.
- Future research should aim to bridge the gap between high-performance AI models and real-world usability, ensuring seamless communication for the hearing-impaired community.



XVI. CLOSING REMARKS

Sign Language Recognition (SLR) has evolved significantly, leveraging deep learning and AI to break communication barriers for the hearing-impaired community. While CNNs, RNNs, and transformer-based models have improved recognition accuracy, challenges in real-time processing, dataset diversity, and practical deployment remain.

Emerging technologies like Edge AI, multimodal learning, and 3D gesture tracking offer promising solutions for overcoming these limitations. Future research should focus on making SLR more efficient, accessible, and widely adopted in real-world applications. With continued advancements, SLR has the potential to become an essential tool for inclusive communication, bridging the gap between sign language users and the broader world.

REFERENCES

[1] Kothadiya, D. R., Bhatt, C. M., Kharwa, H., & Albu, F. (2024). Hybrid InceptionNet Based Enhanced Architecture for Isolated Sign Language Recognition. IEEE Access.

[2] Yadav, P., Sharma, P., Khanna, P., Chawla, M., Jain, R., & Noor, L. (2024). Harnessing AI to Generate Indian Sign Language from Natural Speech and Text for Digital Inclusion and Accessibility. International Journal of Advanced Computer Science and Applications, 15(4).

[3] Yadav, B. K., Jadhav, D., Bohra, H., & Jain, R. (2021). Sign Language to Text and Speech Conversion. International Journal of Advance Research, Ideas and Innovations in Technology, 7(3).

[4] Adewale, V. A., & Olamiti, A. O. (2018). Conversion of Sign Language to Text and Speech Using Machine Learning Techniques. Journal of Research and Review in Science, 5, 58-65.

[5] Kumar, M. N. B. (2018). Conversion of Sign Language into Text. International Journal of Applied Engineering Research, 13(9), 7154-7161.

[6] Khan, A. B., Chandru, M., & Kalaiselvan, R. (2023). Sign Language Detection and Conversion to Text and Speech Conversion. International Journal of Scientific Research in Engineering and Management.

[7] Jim, A. A. J., Rafi, I., Tiang, J. J., Biswas, U., & Nahid, A. A. (2024). KUNet-An Optimized AI-Based Bengali Sign Language Translator for Hearing Impaired and Non-Verbal People. *IEEE Access*, 12, 155052-155063.

[8] Amrutha, B. L., Harshitha, H. M., Nisarga, G. P., Sindhu, S. R., & Sowmya, G. V. (2024). Speech to Sign Conversion Using Natural Language Processing. *International Journal of Scientific Research in Engineering and Management (IJSREM)*, 8(5).

[9] Patil, D., Patil, A., Patil, A., More, A., & Shinde Patil, S. (2024). Sign Language to Text Conversion. *International Journal for Research in Applied Science & Engineering Technology (IJRASET)*, 12(5).

[10] Sebastian, M., M, M., Sabiq, M., Krishnan, Y. U., & Rakesh. (2024). Sign Language Translator. *International Journal of Creative Research Thoughts (IJCRT)*, 12(5).

[11] Annapoorna, E., Nikhil, B. J., Kashyap, B., Abhishek, J., & Vadlapatla, T. S. S. (2023). Hand Gesture Recognition and Conversion to Speech for Speech Impaired. *E3S Web of Conferences*, 391, 01148.

[12] Joshi, A., Agrawal, S., & Modi, A. (2023). ISLTranslate: Dataset for Translating Indian Sign Language. *Findings of the Association for Computational Linguistics: ACL 2023*, 10466–10475.

[13] Phatale, A. P., & Malode, V. B. (2023). Indian Sign Language: Study of Existing Approaches. *International Journal of Scientific Research and Engineering Trends*, 9(5), 956-963.