A Survey on Machine and Deep Learning Approaches in Sign Language Recognition: Techniques and Future Trends

Harshitha C
Dept. of Computer Science and Engineering
National Institute of Technology Puducherry
Karaikal, India
harshithachandrasekhar@gmail.com

Sendil Vadivu D
Dept. of Computer Science and Engineering
National Institute of Technology Puducherry
Karaikal, India
ashasendilvadivu@gmail.com

Narendran Rajagopalan
Dept. of Computer Science and Engineering
National Institute of Technology Puducherry
Karaikal, India
narendran@nitpy.ac.in

***Abstract*— Sign language is an essential communication medium for individuals with hearing impairments. It enables them to convey messages, disseminate knowledge, and transfer ideas within the deaf community. However, not everyone understands sign language, making it challenging for individuals with hearing impairments to communicate effectively with other common people. Sign Language Recognition (SLR) is vital for bridging the communication gap between persons with hearing loss and non-sign language speakers. This survey paper aims to explore Machine Learning (ML) and Deep Learning (DL) techniques for recognition of sign languages. Investigate further research scope in this field.**

***Keywords*—Sign Language Recognition, Indian Sign Language, American Sign Language, ML, DL.**

I. INTRODUCTION

Sign language is a visual-gestural communication used by hearing impairments and people having hard time to hear to express themselves. Sign language recognition (SLR) stands at the intersection of technology and accessibility, aiming to bridge communication gaps for individuals with hearing impairments. Just as spoken languages convey meaning through vocalizations, sign languages utilize hand movements, gestures, and facial expressions to communicate complex concepts and emotions.

Sign Language Recognition (SLR) breaks down the communication barrier faced by them, enhance education, employment, healthcare, and public services accessibility. Deaf individuals often encounter communication barriers when interacting with healthcare providers, which can lead to misunderstandings, misdiagnoses, and inadequate care. SLR systems can enhance educational experiences by providing real-time translation of classroom lectures, instructional videos, and digital learning materials into sign language, ensuring equitable access to educational content for all students. Deaf individuals also face challenges in accessing employment opportunities and advancing in their careers due to communication barriers around their workspace. SLR provides facilitates independent life and social interaction without any intermediators.

This potential makes SLR a pivotal area of research promoting effective communication for the hard-of-hearing community across various aspects of life. However, despite the richness and expressiveness of sign languages, the technological tools for their recognition and interpretation have historically lagged behind those for spoken languages. Traditional communication barriers often limit their ability to fully participate in various aspects of life. Despite the progress made in SLR research, significant challenges remain. The complexity and variability of sign language gestures, coupled with the need for real-time performance and robustness to environmental factors, pose formidable obstacles to the development of effective SLR systems. Moreover, the diversity of sign languages across different regions and communities necessitates the creation of adaptable and scalable solutions capable of accommodating linguistic variations and cultural nuances.

In recent years, we can observe a surge in the area of advancements in SLR technologies, propelled by the growing recognition of the need for inclusive communication. The introduction sets the stage by highlighting the societal impact of improved SLR systems and the potential they hold for enhancing the lives of individuals who rely on sign languages for communication. As we explore the landscape of SLR, a one-size-fits-all strategy is shown to be insufficient, requiring a sophisticated comprehension of the unique difficulties presented by various sign languages.

This survey paper aims to provide an overview of the current SLR techniques, with a specific focus on the application of machine learning (ML) and deep learning (DL) techniques [1]. The study delves into the unique challenges and opportunities posed by Indian Sign Language (ISL) and American Sign Language (ASL), recognizing the importance of SLR systems. ISL and ASL, representing the rich diversity of sign languages, are the primary focus of this survey. Considering the cultural and linguistic intricacies that distinguish ISL and ASL [2], we concentrate on these two languages to present a complete survey of the current state of the SLR.

979-8-3315-4068-5/24/$31.00 ©2024 IEEE

Fig. 1. Hand gestures of Indian Sign Language (ISL) [3]

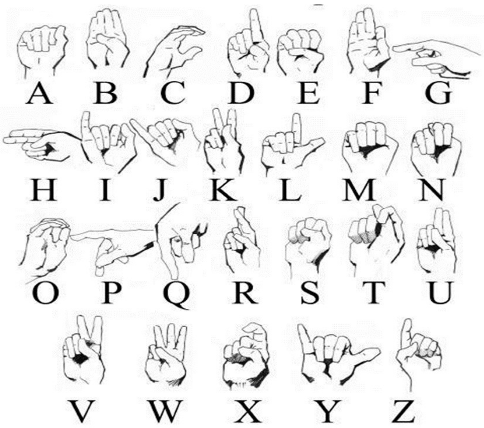

Fig. 2. Hand gestures of American Sign Language (ASL) [4]

The figure 1 and 2 represent the hand gestures of alphabets from A to Z in ISL and ASL respectively. There is existence of research conducted on a variety of sign languages, such as Korean, Arabic, Malayalam, Bangla, and Iranian sign languages. However, this survey has intentionally opted to limit its scope to ISL and ASL literatures. The choice to focus exclusively on these two sign languages is based on their recognition as standard languages within their respective regions. By specifying that ASL and ISL are considered standard languages, the survey indicates that these sign languages have achieved a level of formalization, widespread use, and cultural acceptance. Standardization often involves the establishment of linguistic norms, rules, and conventions, making the languages suitable for formal communication, education, and research.

The rest of the paper is arranged as follows. Section II, briefly elaborates the technologies used in the papers surveyed. Section III, elaborates the literature survey. Section IV presents the conclusion and Section V describes the scope of future enhancements in SLR.

II. Brief About Technologies Used

A. Image Processing

Image processing techniques are crucial in sign language recognition projects, aiding in the analysis and interpretation of sign gestures captured in images or videos. They are utilized throughout the recognition pipeline for preprocessing, feature extraction, and classification tasks. Initially, these techniques preprocess images or videos to enhance quality, including noise reduction and contrast enhancement, standardizing input data. Hand detection and tracking algorithms are then employed to identify and monitor the signer's hands, enabling the extraction of spatial-temporal information essential for sign recognition. Subsequently, features like hand shape and motion trajectories are extracted using image processing techniques, facilitating descriptive representations suitable for classification.

B. OpenCV

Over the years, OpenCV has continually pushed the boundaries of sign language recognition systems, not only in its functionalities but also in how it's implemented. Initially serving basic functions like image acquisition and preprocessing, its utilization has matured significantly. One common approach involves using OpenCV for image acquisition from cameras or video streams. Researchers then preprocess these images using OpenCV's tools for noise reduction, edge detection, and contour extraction to enhance the quality of the captured data. These preprocessed images serve as input for subsequent analysis. Furthermore, OpenCV's integration with machine learning frameworks facilitates the development of robust recognition models. Researchers employ OpenCV in conjunction with libraries like TensorFlow or scikit-learn to train and deploy machine learning algorithms on extracted features. These models learn to recognize and classify sign language gestures with high accuracy, thanks to the rich feature representations provided by OpenCV.

In summary, OpenCV's implementation in sign language recognition involves a combination of image acquisition, preprocessing, feature extraction, machine learning integration, and real-time processing. By harnessing its versatile capabilities, researchers are able to develop robust and efficient sign language recognition systems that contribute to communication accessibility for individuals who suffer from hard of hearing.

C. TensorFlow

In sign language recognition projects, TensorFlow serves as a versatile and powerful tool for model development and deployment. Leveraging TensorFlow's high-level APIs like Keras, researchers and developers can design intricate neural network architectures tailored to the nuances of sign language gestures. TensorFlow's extensive library of pre-implemented layers and functions streamlines the process of building convolutional, recurrent, or hybrid neural networks optimized for sign language recognition tasks. Furthermore, TensorFlow's efficient implementation of optimization algorithms and automatic differentiation facilitates the training of these models on large datasets, ensuring robust performance. Once trained, models can be deployed seamlessly using TensorFlow Serving or TensorFlow Lite for real-time applications on various platforms, including mobile devices and embedded systems. This integration of TensorFlow across the entire development pipeline empowers researchers and developers to create an accurate and efficient sign language recognition systems, ultimately enhancing communication and accessibility for individuals within the deaf and hard-of-hearing communities.

D. Machine Learning

In sign language recognition projects, machine learning techniques are integral for interpreting sign gestures without relying on neural networks. These projects often begin with feature extraction from sign language images or videos, where essential characteristics such as hand shapes and motion trajectories are captured. For better understanding of traditional machine learning algorithms and its types Figure 3 is created which is applied to classify these extracted features based on learned patterns. Dimensionality reduction techniques may also be employed to reduce computational complexity and enhance model efficiency.

Feature engineering plays a vital role in SLR projects, where domain-specific knowledge is utilized to design effective features that capture the distinctive aspects of sign gestures. These features may include hand shape descriptors, motion direction histograms, or spatial-temporal patterns. Once features are extracted, a classifier is trained on labeled datasets of sign language to learn the relationship between the extracted features and corresponding sign language symbols or phrases.

Evaluation of the trained classifier includes assessing its performance using various metrics such as accuracy, precision, recall, and F1-score on a separate validation dataset. Additionally, techniques like cross-validation is employed to ensure the efficiency of the model. Real-time recognition is another key consideration, requiring optimization of the classifier for speed and efficiency to enable seamless integration into applications for communication aids or assistive technologies.

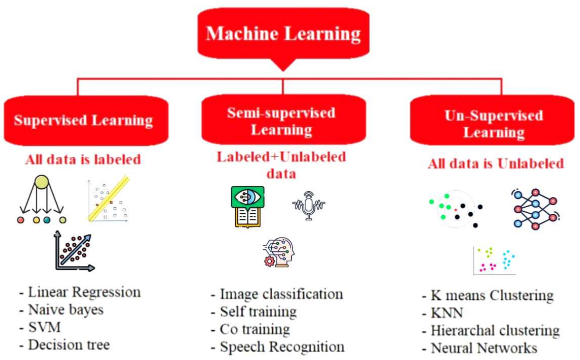

Fig. 3. Machine Learning and its types

E. Deep Learning

In sign language recognition projects, deep learning models are instrumental in achieving accurate and efficient interpretation of sign gestures. Researchers leverage various types of deep learning architectures as represented in Figure 4 which are tailored to the specific requirements of SLR tasks. Convolutional Neural Networks (CNNs) are widely used for capturing spatial features from images, allowing effective recognition of static sign gestures. These models excel at learning intricate patterns and variations in hand shapes and spatial configurations, contributing to high recognition accuracy. Recurrent Neural Networks (RNNs), on the other hand, are utilized to model temporal dependencies in sign language videos, enabling accurate recognition of dynamic gestures that involve motion and temporal changes. Hybrid architectures that combine CNNs and RNNs are also employed to leverage both spatial and temporal information for improved recognition performance.

Additionally, transfer learning techniques are often applied to adapt pre-trained models to the specific characteristics of sign language datasets, facilitating faster convergence and better generalization performance. Attention mechanisms further enhance the interpretability and robustness of deep learning models by enabling them to focus on relevant parts of the input data, leading to improved recognition accuracy in challenging scenarios. Overall, deep learning models have a pivotal role in advancing SLR technology, contributing to enhanced communication for individuals within the deaf and hard-of-hearing communities in the society.

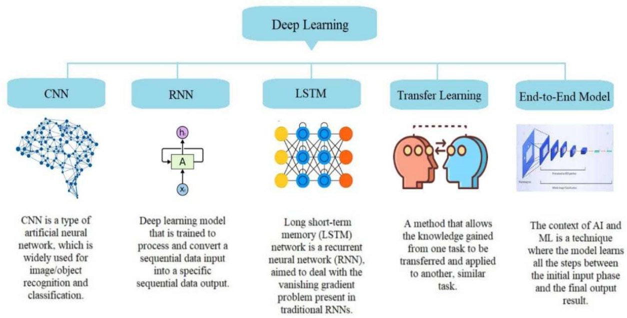

Fig. 4 Deep Learning and its models

III. Literature Survey

The survey of sign language recognition projects showcases a diverse range of approaches and techniques utilized by various researchers which are discussed below.

Yulius Obi's work introduces a unique Sign Language Recognition System designed specifically to improve communication with individuals with disabilities. Employing a CNN architecture and the Kaggle "ASL Hand Sign Dataset," the study achieved an impressive accuracy of 96.3%. In classification process, the image undergoes filtering, followed by classification to predict the class of the hand gestures. Therefore, they have formulated a plan to develop a desktop application capable of real-time sign language recognition and conversion to text. [5].

Sanaa Mohsin's investigation delves into American Sign Language Recognition, employing Transfer Learning Algorithms. Through the utilization of deep learning models and transfer learning techniques with the MNIST dataset, the study attains an impressive accuracy of 96%. A range of deep learning architectures, such as VGG16, ResNet50, MobileNetV2, InceptionV3, and CNN, are assessed using the MNIST dataset. InceptionV3 emerges as the most successful model, underscoring the efficacy of transfer learning in augmenting ASL recognition accuracy. [6].

Harshitha C.'s study focuses on Sign Language Recognition using Machine Learning techniques, primarily employing OpenCV. A simple model was built using teachable machine which was directly implemented in the project. With a dataset created using 100-150 images for each letter, the study achieved an accuracy of 85.45% in recognizing sign language gestures. The confidence rate for each alphabet has been computed and tabulated [7].

Soumen Das advances the field with a specialized deep sign language recognition system customized for Indian Sign Language. Utilizing the VGG-19 deep learning model on an extensive dataset curated by Sridhar, the study achieves an accuracy of 87.67%. Furthermore, a vision-based HCBSLR (Hand-Centric BiLSTM Sign Language Recognition) system is proposed, incorporating a Histogram Difference (HD) based key-frame extraction method to streamline preprocessing and eliminate redundant frames. The HCBSLR system integrates VGG-19 for spatial feature extraction and BiLSTM for temporal feature extraction, presenting a holistic approach to sign language recognition. [8].

Anudyuti Ghorai explores specialized network architectures for Indian Sign Language recognition, employing deconvolution and spatial transformer network techniques. Utilizing the VUCS_ISL_I and VUCS_ISL_II datasets, the study achieved an impressive accuracy of 96.83%. Most contemporary ISL recognition systems rely on CNNs, which can inadvertently capture redundant data because of image correlations. To address this issue, this study suggests an ISL recognition system that utilizes network deconvolution to minimize redundancy. Furthermore, it incorporates a spatial transformer network to improve spatial invariance during convolution operations in the face of transformations [9].

R. Sreemathy delves into the realm of sign language recognition using artificial intelligence. Employing a Backpropagation Neural Network (BPN) trained with HOG features, the system achieves an accuracy of 89.5%, leveraging 5184 input features and 50 hidden neurons. During real-time gesture testing, deep learning models such as AlexNet, GoogleNet, VGG-16, and VGG-19 attain accuracies of 99.11%, 95.84%, 98.42%, and 99.11% respectively, with MATLAB serving as the simulation platform. This innovative technology, functioning as a teaching assistant for individuals with special needs, showcases a notable 60–70% enhancement in children's cognitive abilities. [10].

Utpal Nandi's research delves into Indian Sign Language alphabet recognition employing Convolutional Neural Networks (CNNs), coupled with a diffGrad optimizer and stochastic pooling. Utilizing an extensive dataset featuring 2400 images for each ISL alphabet, the study devised a finger-spelling recognition system for the Indian sign language alphabet using CNNs. Techniques such as data augmentation, batch normalization, dropout, stochastic pooling, and the diffGrad optimizer, the method achieved training and validation accuracies of 99.76% and 99.64%, respectively, surpassing other systems examined. [11].

Soumen Das introduces the Automated Indian Sign Language Recognition System (AISLRSEW), which integrates deep learning with handcrafted features. By leveraging words from the ISL dataset by Aditya V, the AISLRSEW's performance is assessed using a two-fold cross-validation method and contrasted with existing models, achieving an accuracy of 94.42%. [12].

Shashank Kumar Singh presents a robust and effective machine learning pipeline tailored for the recognition of American Sign Language gestures using surface electromyography (sEMG) sensors. The study, which encompasses a dataset encompassing 24 alphabets and 0-9 digits, attains an accuracy of 99.91%. By leveraging sEMG signals, which remain unaffected by lighting conditions, the system records two sEMG datasets for ASL gestures, resulting in approximately 450 features per channel post-preprocessing. Moreover, endeavors to optimize sensor and feature selection are undertaken without compromising accuracy, offering valuable insights for the future development of Sign Language Recognition Systems [13].

Abul Abbas Barbhuiya's research concentrates on the classification and localization of American Sign Language (ASL) hand gestures. They utilize a Deep Ensemble Neural Network coupled with VGG-16 architecture. The incorporation of a self-attention module aids in refining image features crucial for distinguishing between various gesture categories. Furthermore, a weighted ensemble model is introduced for consequently improving the overall performance of the network. By utilizing datasets such as Kinect Leap and HGR Dataset, their study demonstrates remarkable accuracies of 99.76% and 95.10% respectively. [14].

Vaishnaw Kale's research involves the development of a Sign Language Platform using CNNs. The research introduces a sign language learning platform built on Next.js, capable of recognizing and translating signs through a Convolutional Neural Network (CNN). With datasets encompassing A to Z alphabet and 1 to 9 numerals, the study achieved an accuracy of 98.67% [15].

Pranati Rakshit introduces a Sign Language Detection system employing convolutional neural networks to recognize 26 alphabetical letter signs. The proposed models utilize a blend of CNN architecture and SoftMax activation function. A bespoke dataset comprising 78,000 RGB images of ASL alphabets sized 200*200 has been created and is accessible on Kaggle. The study attains an accuracy of 98.44% and an F1 score of 98.41%. [16].

These studies collectively demonstrate the diverse range of techniques and methodologies employed in SLR, contributing to the advancement of communication accessibility for individuals with disabilities. The summary of the reviewed papers in the survey is presented in the Table 1.

TABLE I. Summary of the related works.

S.no	Author	Technique(s) used	Accuracy obtained	Dataset
1	Yulius Obi 2022 [5]	CNN	96.3%	Kaggle- “ASL Hand Sign Dataset”
2	Sanaa Mohsin 2023 [6]	Deep learning & transfer learning	96%	MNIST
3	Harshitha C 2023 [7]	Model using Teachable machine	85.45%	Dataset containing A -Z alphabets
4	Soumen Das 2023 [8]	VGG-19	87.67%	A large-scale dataset by Sridhar.
5	Anudyuti Ghorai 2023 [9]	Network deconvolution technique	96.83%	VUCS_ISL_I & VUCS_ISL_II
6	R. Sreemat hy 2023 [10]	AlexNet, VGG-16, GoogleNet and VGG-19	99.11, 98.42, 95.84, & 99.11%	R2019b
7	Utpal Nand i 2023 [11]	CNN	99.76%	ISL alphabets each having 2400 images
8	Soumen Das 2023 [12]	Combination of CNN and local handcrafted features (AISLRSW)	94.42%	Words from ISL dataset by Aditya V
9	Shashank Kumar Singh 2023 [13]	surface elec-tromyography (sEMG) using ML	99.91%	EMG dataset was created for 24 alphabets & 0-9 digits
10	Abul Abbas Barbhuiya 2023 [14]	VGG-16	99.76% and 95.10%	Kinect Leap Dataset & HGR Dataset
11	Vaishnaw Kale 2023 [15]	CNN	98.67%	A to Z alphabet & 1 to 9 numerals
12	Pranati Rakshit 2024 [16]	CNN	98.44%	ASL from Kaggle

In the dynamic landscape of real-time SLR, cutting-edge systems are engineered to swiftly interpret hand gestures spanning from A to Z. Leveraging continuous video input, these systems meticulously capture and analyze intricate hand movements. Through sophisticated computer vision techniques, such as convolutional neural networks (sCNNs) [17], the visual data is processed and dissected, allowing the system to discern the unique patterns and shapes formed by the gestures.

Utilizing advanced machine learning algorithms, the system intelligently categorizes these hand movements, mapping them to their respective alphabets. apping of two alphabets A and C in real-time SLR is depicted in the Figure 5 and Figure 6 . This intricate process enables the system to swiftly and accurately identify the intended letters in real time. As each gesture is recognized, the system promptly generates and displays the corresponding alphabet letter, providing instantaneous feedback to both the signer and the observer.

By incorporating additional image recognition algorithms and neural network architectures, the system can swiftly identify and interpret a broader range of hand gestures. As depicted in the accompanying images, the real-time system seamlessly captures and analyzes the hand movements corresponding to the letters from "A" to "Z", demonstrating its versatility and accuracy in recognizing various sign language gestures in real time. This advanced technology not only enhances communication accessibility but also fosters inclusivity and understanding among diverse communities.

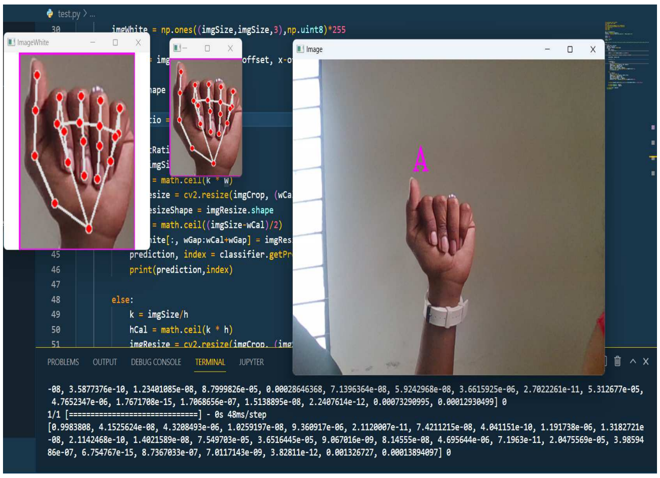

Fig. 5 Alphabet “A” recognized in real-time

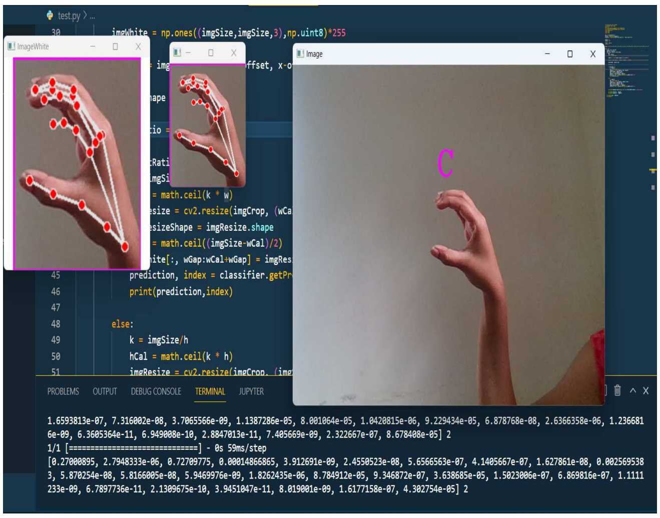

Fig. 6 Alphabet “C” recognized in real-time

IV. Conclusion

The papers reviewed in the table showcase a diverse range of techniques employed for sign language recognition systems. Authors have utilized various methods, including Convolutional Neural Networks (CNNs), Transfer Learning Algorithms, OpenCV, network deconvolution, and hybrid models combining deep learning techniques with handcrafted features. The choice of datasets varies among the authors, with some using general datasets like Kaggle's "ASL Hand Sign Dataset" and MNIST, while others focus on region-specific sign languages such as Korean, Iranian, ISL or ASL. This highlights the importance of tailoring models to specific sign languages to improve accuracy and applicability.

Certain authors have proposed hybrid models by combining DL techniques with handcrafted features. This approach recognizes the complementary strengths of automated feature extraction and human-designed features for effective sign language recognition. Some papers delve into specialized applications incorporating surface electromyography (sEMG) using ML techniques for ASL recognition. This trend indicates a growing exploration of diverse input modalities to enhance sign language.

A collaborative effort to enhance the efficiency and precision of sign language recognition systems is seen in the use of cutting-edge architectures such as VGG-16, VGG-19, and ensemble neural networks in a number of publications. The chronological progression of papers shows continuous research in this domain, with the latest paper from 2024 introducing a six-layered Convolutional Neural Network (CNN), suggesting ongoing refinement and evolution of models in this field.

V. Future Enhancements

The development of larger and more diverse datasets specific to various sign languages remains crucial. This would allow for more comprehensive training of models, making them capable of recognizing a broader range of gestures and expressions, and facilitating the creation of more inclusive and globally applicable systems. Additionally, there is room for further research in the fusion of multiple modalities. Integrating not only visual data but also incorporating other sensory inputs, such as depth information or even incorporating natural language processing for sign language interpretation, could enhance the overall robustness and accuracy of recognition systems.

Looking ahead, one key direction involves the exploration of more advanced DL architectures and techniques. Researchers could delve more into emerging models beyond VGG, AlexNet, and GoogleNet, exploring architectures that may offer improved performance, efficiency, or adaptability to diverse sign languages. In terms of practical uses, scientists should concentrate on improving sign language recognition system's applicability in various settings. This could entail incorporating these systems into assistive technologies for the disabled or optimizing models for real-time applications, like in wearable technology. Additionally, there is room to investigate human-in-the-loop methods, in which user input is actively included into the process of learning. This could make the systems more flexible and individualized by addressing issues with individual variances in signing preferences and styles

In summary, the future of sign language recognition systems holds promise for advancements in DL architectures, multimodal integration, dataset diversity, real-world applications, user feedback incorporation, and ethical considerations, contributing to more effective and inclusive communication technologies.

References

[1] Wadhawan, A., & Kumar, P. (2021). "Sign language recognition systems: A decade systematic literature review." Archives of Computational Methods in Engineering, 28, 785-813.

[2] Ardiansyah, A., Hitoyoshi, B., Halim, M., Hanafiah, N., & Wibisurya, A. (2021). "Systematic literature review: American Sign Language translator." Procedia Computer Science, 179, 541-549.

[3] https://www.1specialplace.com/2021/02/11/all-about-indian-sign-language/

[4] https://www.researchgate.net/figure/American-Sign-Language-alphabet_fig1_340721626

[5] Obi, Y., Claudio, K. S., Budiman, V. M., Achmad, S., & Kurniawan, A. (2023). "Sign language recognition system for communicating to people with disabilities." Procedia Computer Science, 216, 13-20.

[6] Mohsin, S., Salim, B. W., Mohamedsaeed, A. K., Ibrahim, B. F., & Zeebaree, S. R. (2023). "American Sign Language Recognition Based on Transfer Learning Algorithms." International Journal of Intelligent Systems and Applications in Engineering, 12(5s), 390-399.

[7] Harshitha, C., & Girish, L. (2023). "Sign Language Recognition using Machine Learning." International Journal of Advanced Scientific Innovation, 5(2).

[8] Das, S., Biswas, S. K., & Purkayastha, B. (2023). "A deep sign language recognition system for Indian sign language." Neural Computing and Applications, 35(2), 1469-1481.

[9] Nandi, U., Ghorai, A., Singh, M. M., Changdar, C., Bhakta, S., & Kumar Pal, R. (2023). "Indian sign language alphabet recognition system using CNN with diffGrad optimizer and stochastic pooling." Multimedia Tools and Applications, 82(7), 9627-9648.

[10] Sreemathy, R., Turuk, M., Jagdale, J., Agarwal, A., & Kumar, V. (2023, March). "Indian Sign Language Interpretation using Convolutional Neural Networks." In 2023 10th International Conference on Signal Processing and Integrated Networks (SPIN) (pp. 789-794). IEEE.

[11] Nandi, U., Ghorai, A., Singh, M. M., Changdar, C., Bhakta, S., & Kumar Pal, R. (2023). "Indian sign language alphabet recognition system using CNN with diffGrad optimizer and stochastic pooling." Multimedia Tools and Applications, 82(7), 9627-9648.

[12] Das, S., Biswas, S. K., & Purkayastha, B. (2023). "Automated Indian sign language recognition system by fusing deep and handcrafted feature." Multimedia Tools and Applications, 82(11), 16905-16927.

[13] Singh, S. K., & Chaturvedi, A. (2023). "A reliable and efficient machine learning pipeline for American Sign Language gesture recognition using EMG sensors." Multimedia Tools and Applications, 82(15), 23833-23871.

[14] Barbhuiya, A. A., Karsh, R. K., & Jain, R. (2023). "ASL hand gesture classification and localization using deep ensemble neural network." Arabian Journal for Science and Engineering, 48(5), 6689-6702.

[15] Kale, V. G., Rafeeq, S. M., Ehsanullah, A. M., & Vishnu, R. (2023). "Development of Sign Language Platform using CNN." International Journal of Computing and Digital Systems, 14(1), 1.

[16] Rakshit, P., Paul, S., & Dey, S. (2024). Sign language detection using convolutional neural network. *Journal of Ambient Intelligence and Humanized Computing*, 1-26

[17] Renjith, S., and Rashmi Manazhy. "Sign language: a systematic review on classification and recognition." *Multimedia Tools and Applications* (2024):1-51.