Dean&Francis

ISSN 2959-6157

A Deep Neural Framework for Continuous Sign Language Recognition via Iterative Training

YITONG CAI

Faculty of Sciences, Engineering and Technology of B.Computer Science, University of Adelaide, Adelaide, South Australia, Australia
*Corresponding author: a1901006@adelaide.edu.au

Abstract:

Sign language is a visual language that uses a variety of hand gesture combinations to express distinct ideas. It acts as a bridge to communicate with the outside world and is a vital communication tool for the deaf and mute community. Recent years have seen a fast increase in computer technology, opening up new avenues for research on sign language recognition through developments in computer graphics, computer vision, neural networks, and associated hardware. However, many sign language movements are still hard to tell apart because of the intrinsic constraints of visual component combinations. Natural language can be incorporated to help with the recognition process in order to overcome this difficulty. This review study highlights the benefits and drawbacks of modern approaches to sign language recognition in terms of recording, identifying, translating, and depicting motions. The primary difficulties in the field of sign language technology are also covered, along with the introduction of a number of useful applications. In order to encourage and support additional achievements in this field, future research directions and possible developments are finally suggested.

Keywords: Gesture Recognition; Similarity Fusion; Relevant Feature Recognition; Cross-language gesture; Continuous Sign Language Recognition.

1. Introduction

Nowadays, various auxiliary technologies driven by the advancement of artificial intelligence are gradually emerging. These transformative technologies aim to enhance the quality of life for individuals who are hearing-impaired or deaf-mute, enabling them to better adapt to the rapidly evolving digital society. However, significant challenges remain, as these technologies still face problems that require resolution and breakthroughs across multiple fields [1].

Sign language recognition (SLR) is a scientific research and development field that utilizes computer vision and artificial intelligence technologies to capture and translate sign language through dynamic slicing techniques. Over the past decade, this research

ISSN 2959-6157

has made significant progress by introducing sensors such as RealSense™, Kinect, and LeapMotion controllers [2].

This review will focus on the research of sign language recognition, an emerging field with vast exploration and development potential. An important researcher in this field, William Stokoe, who is known as "the man who gave voice to sign language," proved that sign language has the same grammar, syntax, and complexity as any spoken language. His pioneering work has led to significant advancements in the fast evolving field of sign language studies. Sign language recognition (SLR), sign language synthesis and visualization, and sign language interpretation and linguistics are the three main research fields in sign language studies. This paper offers a scope-defined evaluation of the most recent technologies in the field of sign language recognition [3], given the significance of sign language for communication among the hundreds of millions of deaf and hard-of-hearing persons worldwide and the speed at which science and technology are developing.

The World Health Organization (WHO) estimates that 430 million people, or more than 5% of the global population, need rehabilitation treatment for debilitating hearing loss. There are over 34 million children among them. Some of these people were born with hearing loss, while others developed it as a result of illnesses or accidents. By 2050, this number is predicted to double to almost 700 million, or a tenth of the world's present population [4].

The hearing threshold in both ears of a person with normal hearing is 20 dB or above. People who have poorer hearing are said to suffer hearing loss. Hearing loss can be slight, moderate, severe, or profound in severity. It can impair one or both ears, making it challenging to hear loud noises or conversations. A hearing loss in the better-hearing ear of greater than 35 dB is considered disability-related hearing loss. About 80% of those who suffer from hearing loss due to a handicap live in low- and middle-income nations. Over 25% of adults over 60 suffer from disability-related hearing loss, and the prevalence of hearing loss rises with age. One of the most prevalent long-term deficits, "presbycusis," is defined by the degradation of sensory cells.

Those who are "hard of hearing" have modest to severe hearing loss. In addition to using subtitles, cochlear implants, hearing aids, and other assistive technology, people who are hard of hearing may typically communicate through speech. Most "deaf" people have severe hearing loss and are barely able to hear at all. They can benefit from cochlear implants. Some of them communicate through sign language. However, when they communicate with people who do not have hearing impairments in daily life, they often face difficulties and challenges, which seriously affect their daily lives and can also have a certain impact on their psychology, causing them to feel lonely, have their self-esteem hurt, or be depressed for a long time. How to better integrate people with hearing impairments into "normal society" is a very urgent problem to be solved.

2. Sign Language Recognition Methods

The basic characteristics of SLR, including datasets, modalities, features, classification, computational resources, and applications, are introduced in this article along with a classification of each property. Performance in recognition can be impacted by large datasets and modalities. Figure 1 [5] illustrates how performance can be enhanced by effective feature extraction techniques, classification models, and computational resources.

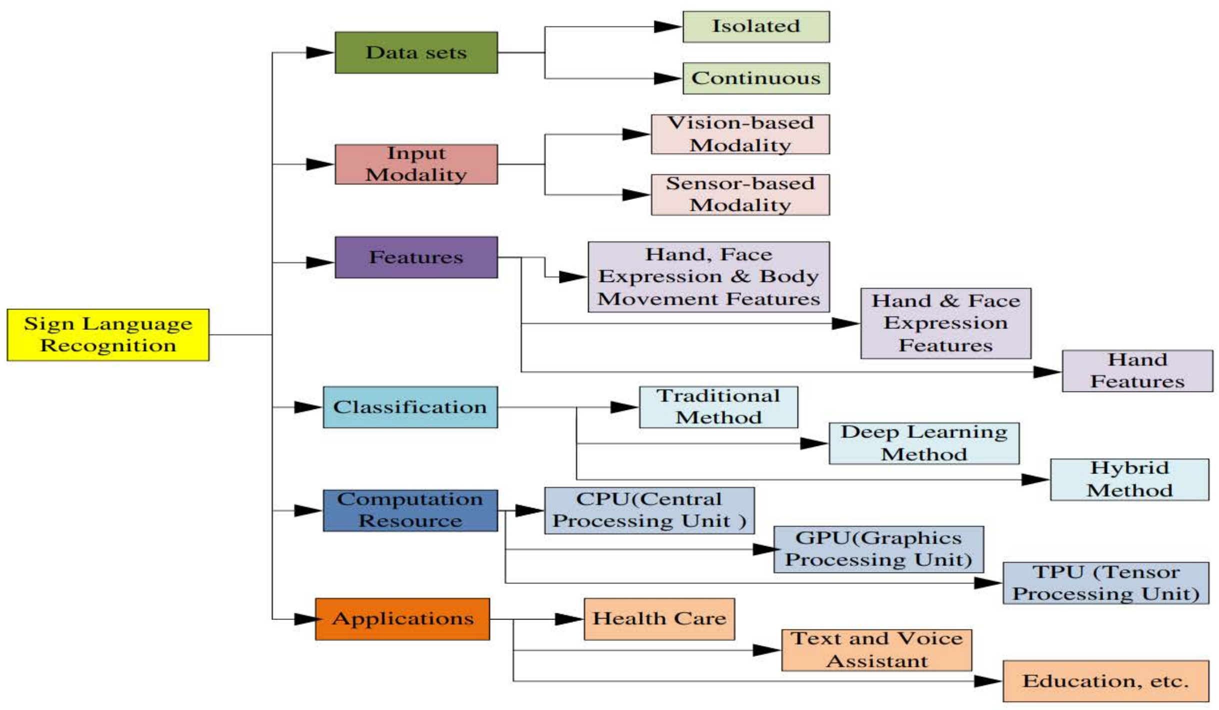

Fig. 1: SLR Taxonomy: the fundamental attributes of SLR like datasets, modality, features, classification, computation resources, and application, along with each attribute's categorization are shown here. Large-scale datasets and modalities affect the recognition performance. The efficient features extraction method and classification model with efficient power computation resources lead to high performance.

Fig. 1 SLR Taxonomy [5]

2.1 Classification of Sign Language Research

The use of sign language involves many concepts. Sign language is not international, and many countries may use different sign languages as shown in Table 1. For instance, Arabic sign language is considered the preferred communication and learning method for many deaf people. Some studies focus on experiments with Chinese and Chinese sign language [6].

Table 1. Distribution of Sign Languages in Different Regions

ABSL	Al-Sayyid Bedouin Sign Language
ASL	American Sign Language
Auslan	Australian Sign Language
BSL	ritish Sign Language
KSL	Kuwaiti Sign Language
LIU	Jordanian Sign Language
LSF	French Sign Language
LSL	Libyan Sign Language
LSM	Mexican Sign Language
LSQ	Quebec Sign Language
MVSL	Martha's Vineyard Sign Language
PSL	Palestinian Sign Language

ISSN 2959-6157

2.2 Current Status of Sign Language Recognition Research

The creation of efficient local sign language recognition (SLR) techniques is essential given the sharp rise in the population of individuals with hearing impairments [7].

Table 2. SLR Barriers and Problems [5]

Barrier	Problem
Computation speed and time	Crable complexiy lo the system and lake a lot of computation time.
Scaling and image orientation proble	The distance of input data capturing various signers.
Illumination of light	Performance varies with different illumination scenarios because most models use the RGB model. It is highly illumination sensitive.
Dynamic and non-uniform background environment	The noise, improper detection of hand, and face lead to affect the performance and mislead the sign recognition system.

Many design characteristics of other natural languages are also present in sign language, which is categorized as a natural language. Users of sign language communicate through body language, facial expressions, and gestures; nevertheless, as Figure 2 illustrates, the composition of these elements is restricted and has traits of ambiguity and difficulty in classification. Table 2 illustrates how environmental elements like illumination also have an effect. This significantly limits and diminishes the visual neural networks‘ capacity for recognition. Concept classification is necessary for current approaches to accurately evaluate all available data. The similarities and differences between sign language recognition techniques are outlined in this paper along with a general framework. Input patterns are very important in this sector, according to current research; recognition based on several data sources (such as visual and sensing channels) appears to be better than single-modal analysis. Furthermore, recent advancements have made it possible for researchers to translate continuous sign language communication with little latency, moving beyond basic letter and word identification. In a number of tasks, some proposed models have demonstrated comparatively good performance; nevertheless, none of them presently possesses the generalization capacity needed for commercial implementation. Nonetheless, the research is making significant strides, and if certain particular problems can be resolved, more innovations should be anticipated. For big vocabularies, it is currently difficult to create an SLR tool with 100% accuracy due to the numerous challenges this research task faces. As seen in Figures 3, 4, and 5. As a result, additional techniques must be created to assess and compile the benefits and drawbacks in order to progressively identify more dependable and superior alternatives. There is disagreement about the optimal network structure framework, which is a contentious topic, despite the fact that many academics think deep learning models are the best approach. Various competing structure framework approaches have had positive outcomes. Most study is based on the sign language grammar used by the majority of local sign language users because sign languages vary from one country to another [1].

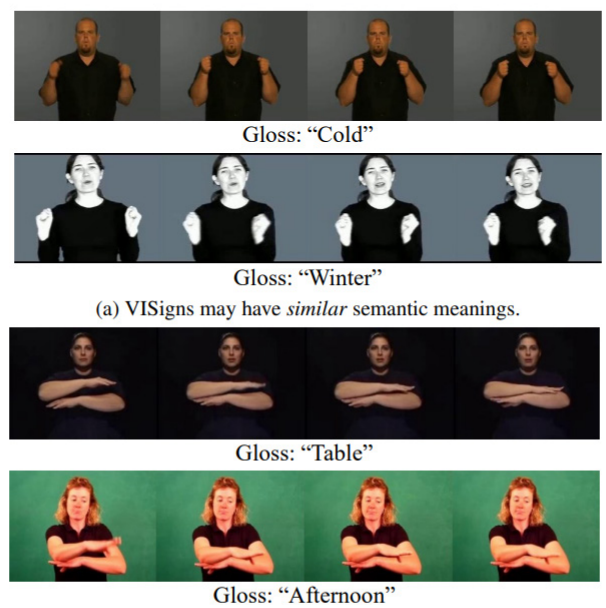

Fig. 2 Examples of Similar Gestures [8]

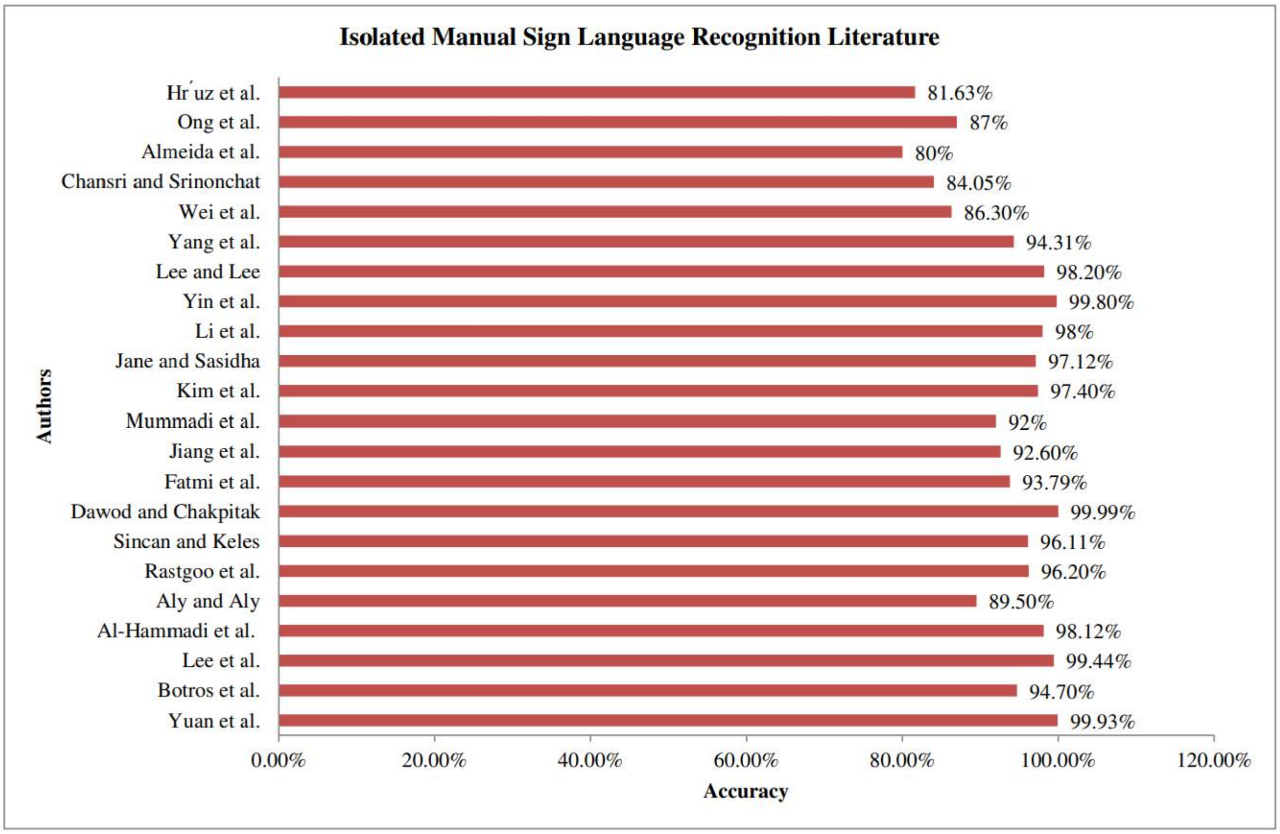

Fig. 3 Isolated Manual SLR Literature Performance [5]

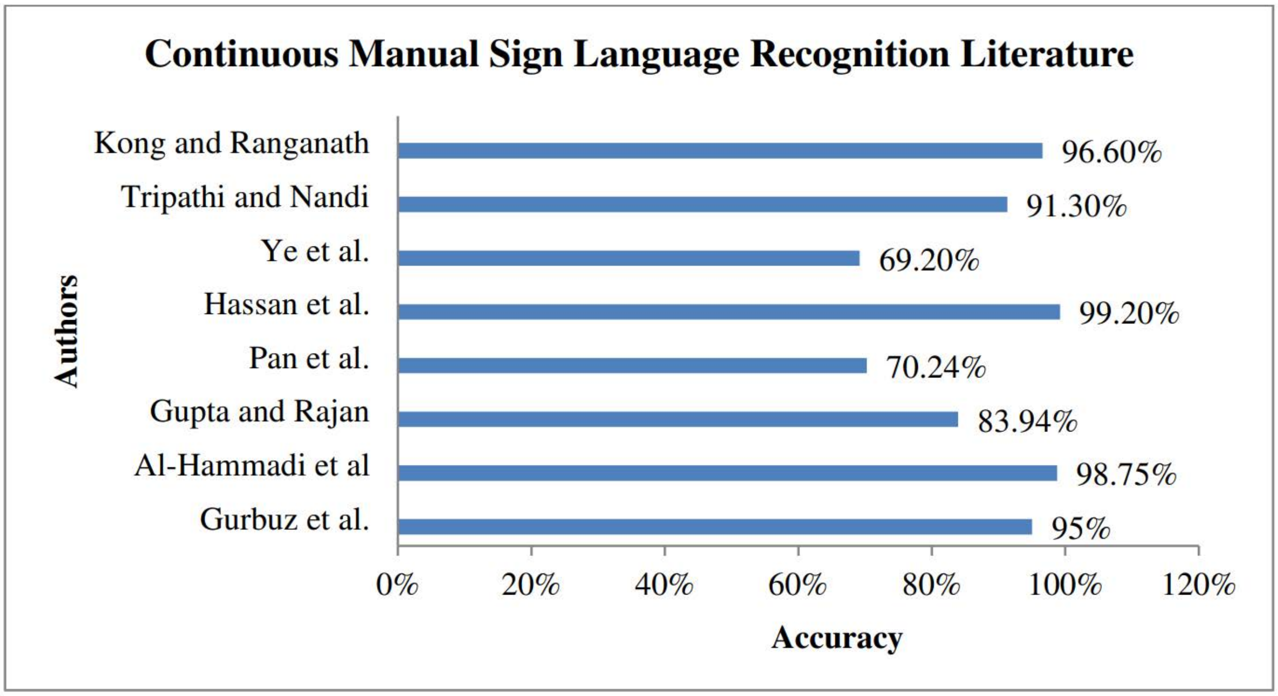

Fig. 4 Continuous Manual SLR Literature Performance [5]

ISSN 2959-6157

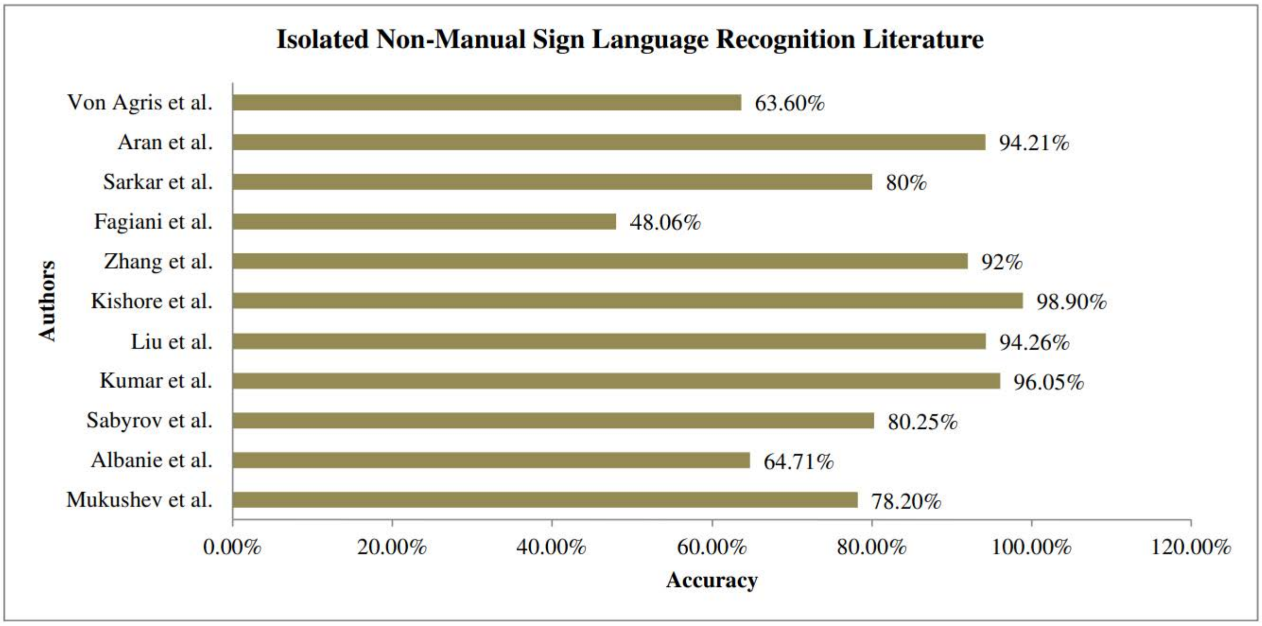

Fig. 5 Isolated Non-Manual SLR Literature Performance [5]

2.3 Sign Language Recognition Research Methods

2.3.1 Sign Language Recognition Research Methods Based on NLA-SLR

Some studies have proposed a natural language-assisted sign language recognition framework (NLA-SLR) that utilizes the semantic information present in sign language annotations (sign language labels). For difficult-to-distinguish gestures with identical semantics (VISigns), matching soft labels can be generated for each training gesture using a language-aware label smoothing technique. The normalized semantic similarity between annotations is used to calculate the smoothing weights, which facilitates training. Second, a cross-modal fusion technique is proposed for VISigns with distinct semantic meanings under mixed label supervision, using annotation and visual information to further maximize the separability of different motions [3].

2.3.2 Sign Language Recognition Research Methods Based on Machine Learning Techniques

Table 3. Comparison of Vision-based and Sensor-based Methods [5]

Method	Capturing Device	Obstacle	Efficiency	Cost	Limitation	Advantage
Vision-based Method	Video Camera	Environment, disturbance, and noise.	Low (depends on the resolution).	Low	Possess challenging concerns for time, speed, and overlapping. More Feature extraction techniques are required.	Fast Speed.
Sensor-based or gloves-based Method	Sensors and gloves	Environment, disturbance, and noise.	Better than vision-based method (depends on sensor performance).	High	Not suitable for real-time applica-tion.	Better performance. Require minimal feature extraction.

Certain advancements and triumphs in aiding sign language identification have also been made possible by machine learning techniques, as illustrated in Figure 6. In order to achieve high-accuracy target gesture identification, this study first used data input from wearable sensors that could translate users‘ movements directly, as indicated in Table 3. The data was then filtered using techniques such as support vector machines (SVM). Nevertheless, this approach mostly examined static content or distinct sign language gestures. A dynamic model that can comprehend continuous parts of sign language is currently required. While models that can recognize and interpret sign language continuously usually demand more sophisticated architectures and computational power, general random models are only appropriate for simple sign language recognition (SLR) applications. Naturally, more variables and models are used in more complicated SLR applications. Simpler machine learning techniques are also useful tools that can be used as comparison benchmarks, and simple fundamental models continue to be competitive and appealing [9][3].

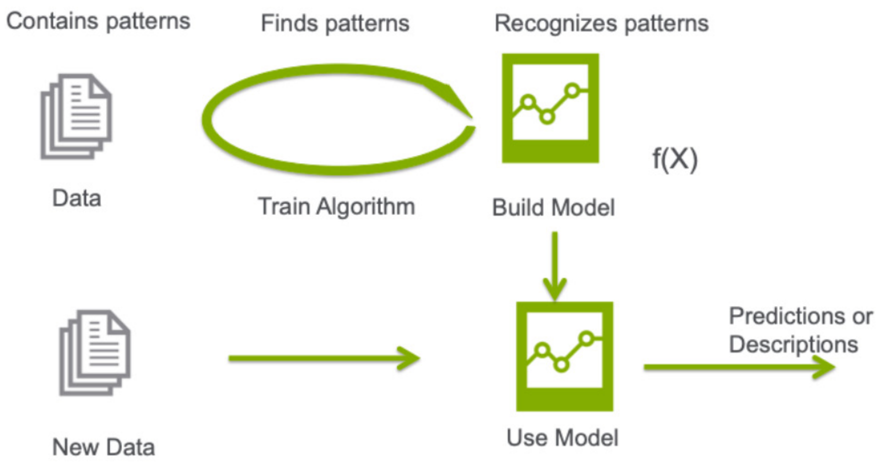

Fig. 6 Machine Learning (ML) Model [10]

2.3.3 Research Methods of Sign Language Recognition Based on Deep Neural Networks

However, more hierarchical architectures that use multi-layer structures and transmit information in vector form between layers—gradually optimizing until positive recognition is achieved—have supplanted fundamental machine learning techniques in recent years. These approaches, which are more complicated than machine learning techniques, are collectively referred to as deep neural networks. Recurrent neural networks (RNNs) with at least one recurrent layer, as seen in Figure 8, and convolutional neural networks (CNNs) with at least one convolutional layer, as illustrated in Figure 7, are two examples of distinct network architectural frameworks that are commonly employed for distinct tasks. These network structures can display various properties to accommodate various task types, depending on the number of layers and types. [11] The performance of these algorithms is significantly impacted by phased training. More stable training can be achieved with larger and more focused datasets, and further model improvements can be made by modifying the data and defining the training procedure [6].

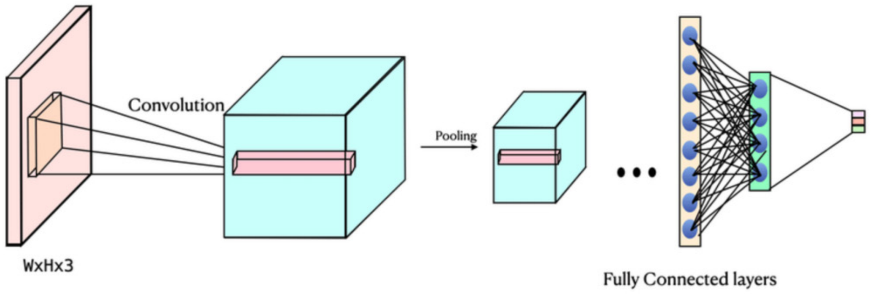

Fig. 7 CNN Model [9]

ISSN 2959-6157

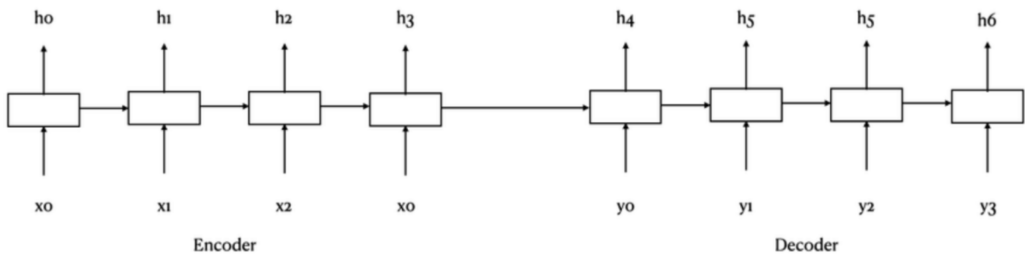

Fig. 8 Basic RNN Encoder-Decoder Model [9]

The majority of SLR studies aim to enhance recognition accuracy by more precisely interpreting body motions and sign language gestures. As seen in Figures 3, 4, and 5. Figure 9 illustrates how several studies use comparable techniques even when their procedures are not precisely the same. As seen in Figure 10 [12], the primary data gathering techniques that SLR has embraced are listed below, along with the common models utilized in the majority of research methodologies.

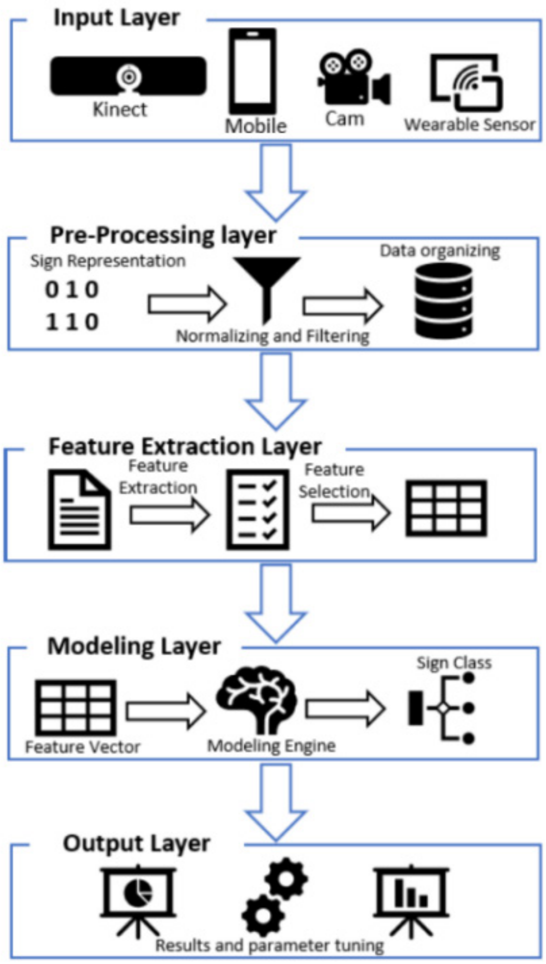

Fig. 9 A model that is generally applicable to most research methods [9]

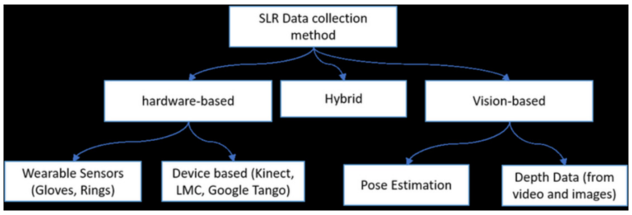

Fig. 10 The main data collection methods adopted by SLR [9]

2.3.4 Research Methods of Sign Language Recognition Based on Human-Computer Interface Technology

As human-computer interface (HCI) technology has advanced, as seen in Figures 11 and 12, research findings in a few areas that are closely related to sign language have been published one after the other [5]. Motion responses—particularly facial expressions and gestures—provide extra non-auditory information in human-computer interfaces that must be taken into account when engaging with computers. The creation of the new field of sign language recognition has been made possible by research on gesture recognition. While recognition techniques are classified as pattern recognition-based, neural network-based, or statistical classification-based, the architectures of gesture recognition systems are compared and summarized, and they may be separated into two categories: data-based and graphics-based [13][14].

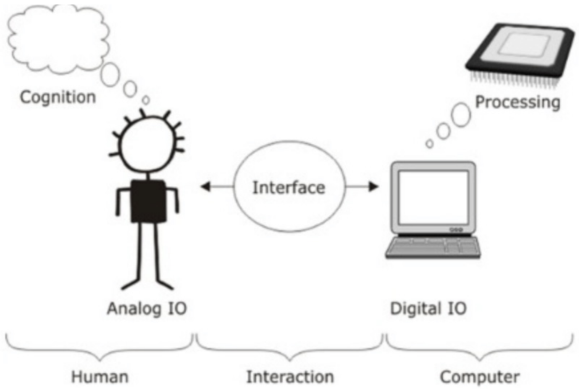

Fig. 11 HCI Model [15]

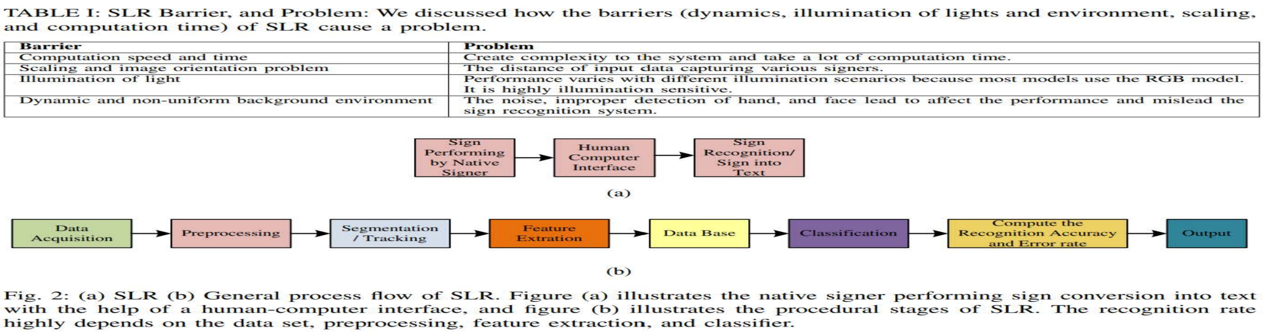

TABLE I: SLR Barrier, and Problem: We discussed how the barriers (dynamics, illumination of lights and environment, scaling, and computation time) of SLR cause a problem.

Barrier	Problem
Computation speed and time	Create complexity to the system and take a lot of computation time.
Scaling and image orientation problem	The distance of input data capturing various signers.
Illumination of light	Performance varies with different illumination scenarios because most models use the RGB model. It is highly illumination sensitive.
Dynamic and non-uniform background environment	The noise, improper detection of hand, and face lead to affect the performance and mislead the sign recognition system.

Fig. 2: (a) SLR (b) General process flow of SLR. Figure (a) illustrates the native signer performing sign conversion into text with the help of a human-computer interface, and figure (b) illustrates the procedural stages of SLR. The recognition rate highly depends on the data set, preprocessing, feature extraction, and classifier.

Fig. 12 (a) SLR Example (b) SLR Process Flow [5]

Iterative optimization is used in an interactive alignment network for continuous, under-supervised sign language recognition. It uses sequence models for feature learning and classification and is comprised of a 3D convolutional resnet. But just two decoders—Long Short-Term Memory (LSTM) and Connectionist Temporal Classification (CTC)—are used by this network. The outcomes demonstrate how successful the decoders that were used were [16][8].

This deep learning-based pipeline design uses Long Short-Term Memory (LSTM). Sign language is automatically recognized from RGB input movies using a 2D Convolutional Neural Network (2DCNN), a 3D Convolutional Neural Network (3DCNN), and a Single Shot Detector (SSD). This model outperforms the most advanced models in terms of performance. Because it creates an automatic method for recognizing sign language, this research is extremely important to this review [17].

Deaf-mute people and those with normal hearing can communicate more easily thanks to real-time sign language translation systems that can translate continuous sign language action sequences into text or speech. These systems fall into two categories based on the kind of input they receive: sensor-based and visual-based [18]. The sensor-based technique is currently used in the majority of commercial goods, however, it is an expensive and unintuitive option. On the other hand, there is a pressing need to build visual-based sign language translation systems; nevertheless, numerous technical obstacles must be overcome before they can fully work. According to preliminary studies, traditional continuous sign language recognition (CSLR) techniques like Conditional Random Field (CRF), Hidden Markov Model (HMM), and Dynamic Time Warping (DTW) typically begin with the isolated sign language recognition (ISLR) problem and then extend the solution to CSLR, which can result in performance degradation to some degree. The primary difficulty of motion insertion (ME) segments in continuous sign language is addressed clearly at the heart of all approaches. Research in this area has advanced as a result of the proposal of more practical visual-based CSLR solutions brought about by the advent of new technologies like deep learning [6][19].

3. Conclusion

The machine learning techniques that are frequently employed in other domains, such as pattern recognition, statistical methods, and neural network-based approaches, have been incorporated into the sign language recognition system. Due to its benefits in processing time series data, HMM fared well in the beginning. However, in recent years, neural networks and deep learning techniques have demonstrated more notable outcomes in this field. Thanks to their achievements in image/video recognition, time series analysis, and natural language processing, 3D-CNN, Transformer, and LSTM architectures have emerged as the industry standard for continuous sign language recognition. Because of its effectiveness in processing time series data, LSTM in particular has gained widespread adoption.

The performance of several sign languages in recognition, visualization, and synthesis still requires investigation. While cross-language comparison studies are rare, mostly because there aren't many cross-language sign language translation datasets, the majority of current studies often compare novel approaches with conventional ones for the same sign language. However, the literature study suggests that the state-of-the-art techniques for recognizing sign language function similarly across languages, meaning they can be slightly modified and used for other sign languages with comparable outcomes.

It is anticipated that sign language identification and visualization technologies will advance dramatically with the ongoing development of hardware and machine learning technologies, particularly neural networks and deep learning. Lack of adequate data and training samples covering all languages is currently the biggest obstacle, however, transfer learning advancements will eventually make this issue less severe. Public and commercial products that can translate multiple sign languages in real time are antici-

ISSN 2959-6157

pated to appear within the next five to ten years. This will facilitate smooth communication between speech and sign language and encourage the adoption of cross-sign language real-time translation systems.

References

[1] Joksimoski B, Zdravevski E, Lameski P, et al. Technological solutions for sign language recognition: a scoping review of research trends, challenges, and opportunities. IEEE Access, 2022, 10: 40979-40998.
[2] Lee C K M, Ng K K H, Chen C H, et al. American sign language recognition and training method with recurrent neural network. Expert Systems with Applications, 2021, 167: 114403.
[3] Aloysius N, Geetha M. Understanding vision-based continuous sign language recognition. Multimedia Tools and Applications, 2020, 79(31): 22177-22209.
[4] WHO, 26 February 2025, Deafness and hearing loss. https://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss
[5] Madhiarasan M, Roy P P. A comprehensive review of sign language recognition: Different types, modalities, and datasets. arXiv preprint arXiv:2204.03328, 2022.
[6] Wadhawan A, Kumar P. Sign language recognition systems: A decade systematic literature review. Archives of Computational Methods in Engineering, 2021, 28(3): 785-813.
[7] Adaloglou N, Chatzis T, Papastratis I, et al. A comprehensive study on deep learning-based methods for sign language recognition. IEEE Transactions on Multimedia, 2021, 24: 1750-1762.
[8] Al-Qurishi M, Khalid T, Souissi R. Deep learning for sign language recognition: Current techniques, benchmarks, and open issues. IEEE Access, 2021, 9: 126917-126951.
[9] Hao A, Min Y, Chen X. Self-mutual distillation learning for continuous sign language recognition. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021: 11303-11312.
[10] NVIDIA, August 16, 2021, Machine Learning – What Is It and Why Does It Matter? https://www.nvidia.com/en-au/glossary/machine-learning/
[11] Selvaraj P, Nc G, Kumar P, et al. OpenHands: Making sign language recognition accessible with pose-based pretrained models across languages. arXiv preprint arXiv:2110.05877, 2021.
[12] Papastratis I, Chatzikonstantinou C, Konstantinidis D, et al. Artificial intelligence technologies for sign language. Sensors, 2021, 21(17): 5843.
[13] Abduljabbar M, et al. A cloud-based 3D digital twin for Arabic sign language alphabet using machine learning object detection model. 2023 15th International Conference on Innovations in Information Technology (IIT). IEEE, 2023.
[14] Patil A, Kulkarni A, Yesane H, et al. Literature survey: sign language recognition using gesture recognition and natural language processing. Data Management, Analytics and Innovation: Proceedings of ICDMAI 2021, Volume 1, 2021: 197-210.
[15] Okelo B K, Mbuguah S M. Discoverability: a new learnability principle for children's application software. 2016.
[16] Zuo R, Wei F, Mak B. Natural language-assisted sign language recognition. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 14890-14900.
[17] Hu L, Gao L, Liu Z, et al. Continuous sign language recognition with correlation network. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 2529-2539.
[18] Wei F, Chen Y. Improving continuous sign language recognition with cross-lingual signs. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2023: 23612-23621.
[19] Min Y, Hao A, Chai X, et al. Visual alignment constraint for continuous sign language recognition. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021: 11542-11551.