

Article

Deep Learning for Sign Language Recognition: A Comparative Review

Shahad Thamear Abd Al-Latief ^{1*}, Salman Yussof ², Azhana Ahmad ³, Saif Khadim ⁴

¹ College of Graduate Studies (COGS), Universiti Tenaga Nasional (National Energy University), Malaysia.

² Institute of Informatics and Computing in Energy, Universiti Tenaga Nasional (National Energy University), Malaysia.

³ College of Computing and Informatics, Universiti Tenaga Nasional (National Energy University), Malaysia.

⁴ College of Graduate Studies (COGS), Universiti Tenaga Nasional (National Energy University), Malaysia.

* Corresponding Author: Shahad Thamear Abd Al-Latief, Email: PT21299@student.uniten.edu.my

Received 27 May 2024; Accepted 05 June 2024; Published 15 June 2024

Abstract: Sign language can be regarded as a unique form of communication method between human beings, which relies basically on visualized gestures of the individual body parts to transfer messages and obtains a substantial role in the life of impaired people having hearing and speaking disabilities deaf. There are various different signs in every sign language with differences in representation using hand shape, motion type, and location of the hand, face, and body portions participate in every sign. Understanding sign language by individuals without disabilities is a challenging operation. Therefore, automated sign language recognition has become a significant need to bridge the communication gap and facilitate the interaction between the deaf society, and the normal hearing majority. In this work, an extensive review of automated sign language recognition and translation of different languages around the world has been conducted. More than 140 research articles have been reviewed, and all of them are relying on deep learning techniques, which were published between 2018 and 2022, to recognize, and translate sign language. A brief review of concepts related to sign language is also presented including its types, and acquiring methods, as well as an introduction to deep learning, and the main challenges facing the recognition process. A description of the various types of public datasets of sign language in different languages is also presented and discussed.

Keywords: Sign language, Recognition, Deep Learning, Classification.

Copyright © 2023 Journal of Smart Internet of Things (JSIoT) published by Future Science for Digital Publishing and Sciendo . This is an open access article license CC BY (<https://creativecommons.org/licenses/by/4.0/>)

1. Introduction

Communication plays an essential role with enormous effects on individuals' lives, such as in gaining and exchanging knowledge, interacting, developing social relationships, and revealing

feelings and needs. While most humans communicate verbally, there are those with limited verbal abilities who need to communicate using Sign language (SL). Sign Language is a type of language that is visual, which is utilized by the deaf individuals and mainly relies on the various parts of an individual's body including fingers, hand, arm, head, body, and facial expression to transfer information rather than using the vocal tract [1]. According to the World Federation of the Deaf, there are more than seventy million deaf people around the world that use more than 300 types of sign language [2]. However, it is not popular among the individuals with typical hearing and communication abilities, and a few of them are able to understand, and learn sign languages. This reveals a genuine communication gap between deaf individuals and the rest of society. Automated recognition and translation of sign language by performing sign language recognition would help to break down these barriers by providing a comfortable communication platform between deaf, and hearing individuals and give the same opportunities for deaf individuals to obtain information as everyone else [3]. Machine translation demonstrates a remarkable capacity for overcoming linguistic barriers, particularly through the utilization of Deep Learning (DL), as a branch of the field. Deep learning exhibits outstanding and exceptional performance in diverse domains, including image classification, pattern recognition, and various other fields and applications [4]. The advancement of DL networks has witnessed a significant surge in their performance, particularly in the realm of video-related tasks, such as Human Action Recognition, Motion Capturing, Gesture Recognition [5-7]. Basically, DL techniques offer remarkable attributes that render them highly advantageous in Sign Language Recognition (SLR). This is primarily attributed to their hidden layers, which autonomously extract latent features, as well as their capacity to effectively handle the intricate nature of hand gestures in sign language. This is achieved by leveraging extensive datasets, enabling the generation of accurate outcomes without time-consuming processes, a characteristic often lacking in conventional translation methods [8]. This paper presents a review of various deep learning models used to recognize sign language in to spot the light the key challenges encountered in using deep learning for sign language recognition and determine the unresolved issues. Additionally, this paper provides some suggestions to overcome these challenges that are based on our knowledge have not been solved.

1.1. Motivation

The communication gap that exists between normal and deaf individuals is the most important motivation in designing and building an interpreter to facilitate communication between them. When embarking on the design of such a translator, a comprehensive set of objectives must be taken into account. These include ensuring accuracy, speed, efficiency, scalability, and other factors that contribute to delivering a satisfactory translation outcome for both parties involved. However, numerous challenges have been identified in the realm of sign language recognition, necessitating the development of an efficient and robust system to address various issues related to environmental conditions, movement speed, occlusions, and adherence to linguistic rules. Deep-learning-based sign language recognition models have gained significant interest in the last few years due to the quality of the recognition and translation that they provide and their ability in dealing with the various sign language recognition challenges

1.2. Contribution

The main contributions of this work are:

- Provide a description of important concepts related to sign language including acquiring methods, types of sign language, and a description of many public datasets in different languages around the world.
- Identify the various challenges and problems encountered in the implementation of sign

language recognition using DL.

- Review more than 140 related works for DL-based sign language recognition from the year 2018 to 2022.
- Classify these relevant works according to the specific problem addressed and the technique or method employed to overcome the specified challenge or problem.

1.3. Paper Organization

This paper is organized into eight main sections as described in Fig1. To facilitate a better and smoother reading for this review, a detailed description of each section is presented below:

1. Introduction: Provides a brief introduction about sign language, deep learning, describes the motivation behind this review, introduces the main contributions, and illustrates the main layout of this work.
2. Sign language Overview: Provides a comprehensive overview of sign language, encompassing its historical context and the fundamental principles employed in its construction and development. Additionally, it includes a description of the various forms used to represent letters, words, and sentences in sign language, as well as the acquisition methods employed for capturing sign language.
3. Deep Learning Background: Introduces the historical background of DL networks structures, properties, layers, and commonly utilized architectures.
4. Sign Language Recognition Challenges Using Deep Learning: Describes the main challenges and problems facing the recognition of sign language using DL.
5. Sign Language Public Datasets Description: Presents an overview of widely accessible sign language datasets, encompassing various languages and types (such as images and videos) and available in different formats (including RGB, depth, and skeleton data). Additionally, provide a description of public action datasets related to sign language.
6. Deep learning-based sign language-related works: Introduces a considerable number of promising related works for sign language using DL techniques from 2018 to 2022 that are organized based on the type of problem being addressed.
7. Discussion: Discusses the results and methods utilized by the presented related works.
8. Conclusion: Concludes the review conducted, and illustrates the conclusions reached by performing this review paper on sign language recognition using DL, in addition to a set of recommendations for future research in this area.

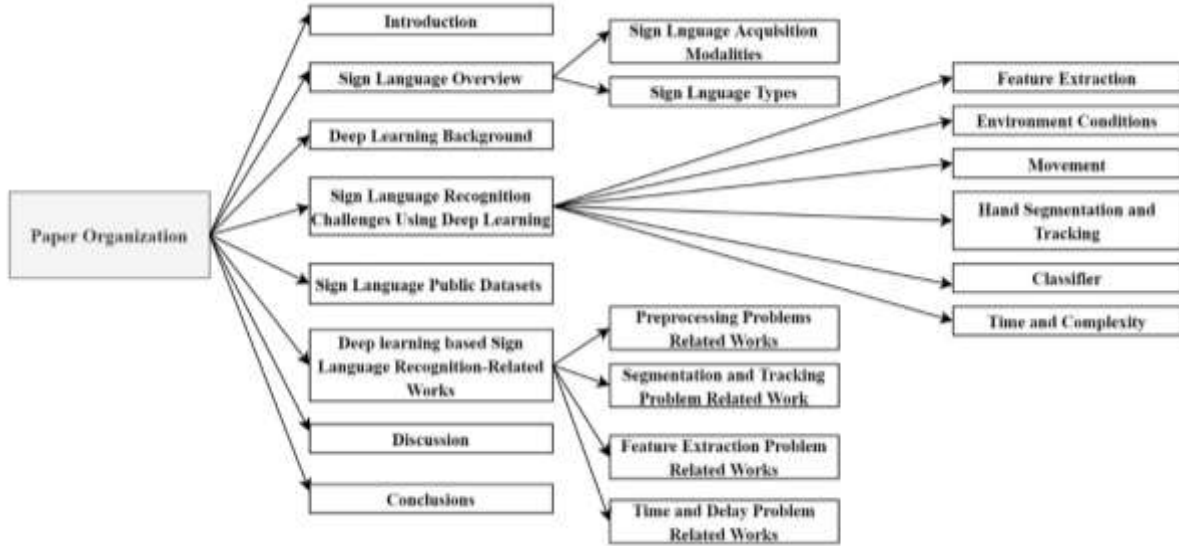


Figure 1: Paper Organization

2. Sign Language Overview

Sign language (SL) serves as a crucial means of communication for individuals who experience difficulties in speaking or hearing. Unlike spoken language, understanding sign language does not rely on auditory perception, nor does it involve vocalization. Instead, sign language is primarily conveyed through the simultaneous combination of hand shapes, orientations, movements, as well as facial expressions, making it a visual language. [9]. Historically, the linguistic studies of sign language started in the 1970s [10] and show that it is like spoken languages, in which it arranges elementary units called phonemes into meaningful units known as semantic units, which contain lingual information which include different symbols and letters. Sign language is not derived from spoken languages, instead, it has its own independent vocabulary and grammatical constructions [11]. However, the signs used by individuals who are deaf possess an internal structure similar to spoken words. Just as a limited number of sounds can generate hundreds of thousands of words in spoken languages, signs are formed by a finite set of gestural features. Thus, signs are not only gestures, but they are actually a group of linguistically significant features. There is a common misapprehension, that there is only a universal, and single sign language. Just like spoken languages, sign languages evolve, and grow inherently across time, and space [12]. Many countries have their own national sign languages. However, there is also regional variance and domestic dialects. Moreover, the signs do not have a one-to-one mapping to a specific word. Therefore, sign language recognition is a complex process that extends beyond a simple substitution of individual signs with their corresponding spoken language counterparts. This is attributed to the fact that sign languages possess distinct vocabularies and grammatical structures that are not confined to any particular spoken language. Furthermore, even within regions where the same spoken language is used, there can be significant variations in the sign languages employed [13].

2.1. Sign Language Acquisition Modalities

The signs of sign language must be captured and attained to provide input for the recognition system and there are various acquisition techniques that provide several types of input such as

image, video, and signals. Basically, the main acquisition methods for any sign language recognition system depend on one of these acquisition techniques.

1- Vision-Based: In this type of system signs are captured using single or multiple images capturing devices in the form of single images or video stream and in some cases uses an active and invasive device, to collect the depth information that represent an accurate information associated to the distance between the image plane and the relevant object in the intended image [14]. This category is easy and presents a low computational. There are many imaging devices for signs capturing images in the form of RGB and depth data including [15]:

- Single camera: Refers to a filming technique or production method that involves using only one camera such as Webcam, digital cam, video cam, and smartphone cam.
- Stereo-camera: Obtains many monocular cameras, or thermal ones to capture in-depth information.
- Active methods: Utilizes the projection of structured light using devices such as Kinect and Leap Motion Controller (LMC), which are 3D cameras that can gather movement and skeletal data.
- Other methods such as body markers in colored gloves, wrist bands, and LED lights.

Generally, the major advantages of vision methods are that it is not costly, convenient, and non-intrusive. The user simply needs to communicate using sign language naturally in front of an image capturing device. This makes it suitable for real-time applications [16]. However, the use of vision-based input suffers from a set of problems including [17]:

- Too much redundant information causes low recognition efficiency.
- Low recognition accuracy, due to occlusion and motion blur.
- The variances in sign language style between individuals resulted in poor generalization of algorithms.
- Small recognizable words vocabulary due to the large vocabulary datasets containing similar words.
- Have some challenging matters about time, speed, and overlapping.
- Need more feature extraction methods to operate correctly.

2- Hardware-Based: This type mainly depends on the use of some types of hardware devices that can capture or sense the signs performed by the user when attached to his/her arm, hand, or fingers, and convert these signs into signals, or images, or in some cases video. Motion sensors are the most widely utilized devices that can track the movements, position, shapes, and velocity of fingers and hands [18]. Electronic gloves serve as the predominant sensor technology employed for capturing hand pose and associated motion. They are affixed to both hands to acquire precise data on hand movements and gestures. The hand's position, orientation, and location are calculated precisely due to the hundreds of sensors supplemented in the gloves. The most significant advantage of this method is its fast reaction [19], so it is highly accurate. However, since it depends on costly sensors it cannot be considered an affordable method for the common deaf people. Moreover, it suffers from relatively low accuracy or complicated structures, and the insufficient amount of information provided by the wearable sensors often affects the overall performance of this system. Some popular examples of sensors are described below [20]:

- Inertial Measurement Unit (IMU): It is an electronic device employed to measure and report an object's specific force, position, angular rate, and sometimes orientation with respect to an inertial reference frame, and acceleration. It typically consists of a combination of

accelerometers, gyroscopes, and sometimes magnetometers.

- Electromyography (EMG): It is the device that uses electrodes placed on or inserted into the skin near the muscle of interest to measure human muscle's electrical pulses and employ the bio-signal to detect movements.
- Wi-Fi and Radar: These devices mainly depend on radio waves, broad beam radar, or spectrogram to detect in-air signal strength variation. They are employed to monitor the movements and positions of the deaf by capturing the reflections of radio waves off their body or hand movements. Radar systems can provide data on the dynamics and trajectories of sign language gestures. This information then can be used for analysis or recognition purposes.
- Others include flex sensors, ultrasonic, mechanical, electromagnetics, and haptic technologies.

In general, although these methods exhibit higher speed and accuracy, the necessity for individuals to wear sensors remains impractical for the following reasons [21]:

- (1) It may cause a burden on the users because they must take electronic devices with them when moving.
- (2) Portable electronic devices require a battery, which needs to be charged from time to time.
- (3) Specific equipment is required to process the signals acquired from the wearable devices.

3- Hybrid-based: In this type, the vision-based cameras together with other types of sensors, such as infrared depth sensors, are combined to acquire multi-mode information regarding the shapes of the hands [22]. This approach requires calibration between the hardware and vision-based modalities, which can be particularly challenging. The purpose of a hybrid system was to enhance data acquisition, and accuracy, and attempt to reduce the challenges and problems of both visions and hardware-based approaches [23].

2.2. Sign Language Types:

1. **Static:** A specific hand configuration and pose, depicted through a single image, is employed for the recognition of fingerspelled gestures of alphabets and digits. This recognition process relies solely on still images as input to predict and generate the corresponding output, without incorporating any movement. It is considered to be very inconvenient, due to the time required to perform prediction each time an input is given, and depends basically on handshapes, hand positions, and facial expressions to convey meaning [24].
2. **Dynamic:** Refers to a variant of sign language, in which signs are produced with movement. This form of communication encompasses not only handshapes and positions but also incorporates the movement of hands, arms, and other body parts to convey meaning. To capture and represent this type of sign language, video streams are required [25]. There are certain words in sign language, such as in American Sign Language, which necessitate hand movements for proper representation, making it a dynamic form. It plays a vital role in facilitating communication, as well as establishing linguistic and cultural identities within the deaf community. Dynamic signs find application in various contexts, including everyday conversations, education, storytelling, performances, and broadcasting. Broadly speaking, dynamic signs can be categorized into two main types based on what they represent, be it individual words or complete sentences. These are described below [26]:

- a) Isolated: The input dynamic signs are used to represent words by performing more than one sign each time and pauses only happen between words.
- b) Continues: The continuous dynamic entries are mainly employed to represent sentences because it incorporates more than one sign performed continuously without any pause between signs [27].

3. Deep Learning Background

The Deep Neural Network is basically a branch of Machine Learning (ML), that was originally inspired by and resembles the human nervous system, and the structure of the brain. It is composed of several layers and nodes, in which the layers are processing units systemized in input, output, and hidden layers. The nodes or units in every layer are linked to nodes in contiguous layers, and every connection owns its singular weight value. The inputs are multiplied by the intended weights and summed at every unit. The summation result undergoes a transformation depending on some type of activation function such as, Sigmoid function, Tan hyperbolic or Rectified Linear Unit (ReLU) [28]. Thus, DL includes stacking many learning layers to learn high-level abstractions in the data within approximate highly nonlinear functions giving the learning algorithm the ability to learn hierarchical features from the input data. This feature learning highly replaced the hand-engineered features and owes its regeneration to effective optimization methods, and powerful computational resources [29]. DL powerful properties give it the ability to taking the lead in achieving the desired results depending on a set of factors including: [30]

- Feature learning refers to the capacity to acquire descriptive features from data that have an impact on other correlated tasks. This implies that numerous relevant factors are disentangled within these features, in contrast to handcrafted features that are designed to remain constant with respect to the targeted factors.
- Hierarchical Impersonation: the features in this type of method are represented in a hierarchical format, in which the simple ones are represented in the lower layers, and the high layers learn increasingly complicated features. This will provide a successful encoding for properties of two types, including local and global in the last features representation.
- Distributed Impersonation: Signifies a many-to-many relationship where the representations are dispersed. This occurs because multiple neurons can demonstrate a single factor, while one neuron can account for multiple factors. Such an arrangement eradicates the potential for dimensionality and offers a compact and comprehensive representation.
- Large-Scale Datasets: the DL is able to deal with the datasets with a vast number of samples and gives outstanding performance in many domains.

In recent years, DL methods have demonstrated exceptional performance surpassing previous state-of-the-art ML techniques across various domains. One domain in which DL has emerged as a prominent methodology is computer vision, particularly in the context of sign language recognition. DL has provided novel solutions to challenges in sign language recognition and has become a leading approach in this field [31]. Many architectures of DL have been utilized for sign language recognition and in an accurate, fast, and efficient manner, due to their ability in dealing with most challenges and the complexity of sign language [32]. The most popular and utilized DL architectures are the Convolutional Neural Network (CNN), Deep Boltzmann Machine (RBM), Deep Belief Network (DBN), Auto Encoder (AE), Variational Auto Encoder (VAE), Generative Adversarial Network (GAN), and Recursive Neural Network (RNN) including Long Short-Term

Memory (LSTM) and Gated Recurrent Unit (GRU) [33].

4. Sign Language Recognition Challenges Using Deep Learning

Detection, tracking, pose estimation, gesture recognition, and pose recovery represent key sub-fields within sign language recognition. These sub-fields are extensively employed in human-computer interaction applications that utilize DL techniques. Nevertheless, the recognition or conversion of signs performed by deaf individuals using DL presents a range of challenges that can significantly impact the output results. These challenges include:

4.1. Feature Extraction

Feature extraction is the process used to select and /or combine variables into features, and effectively reduce the amount of data that must be processed, while still accurately and completely describing the original data. It has a role in addressing the problem of finding the most compact and informative set of features, to enhance the efficiency of data storage and processing. Defining feature vectors remains the most common and convenient means of data representation for classification and regression [34]. In the context of sign language recognition, the process of extracting pertinent features plays a vital, and decisive role. Irrelevant features, on the other hand, can result in misclassification and erroneous recognition [35]. Within the realm of DL techniques for data classification, automatic feature extraction holds paramount importance. Integrating various features extracted from both training and testing images without any data loss is a crucial step that greatly impacts the recognition accuracy of sign language. In general, two types of features are considered in sign language including manual, and non-manual features. Manual features encompass the movements of hands, fingers, and arms. On the other hand, nonmanual features represent the fundamental component in sign language [36] like facial expressions, eye gaze, head movements, upper body motion, and positioning. The combination of manual and non-manual features offers a comprehensive representation of sign language. In the domain of DL, features can be classified into two categories: spatial and temporal. Spatial features pertain to the geometric representation of shapes within a specific coordinate space, while temporal features account for time-related aspects during movement especially when dealing with a sequence of images as input. By employing feature fusion and combining multiple types of features in the process of sign language recognition using DL, one can achieve the desired outcomes effectively [37].

4.2. Environment Conditions

The variability in the environment during sign capture poses a significant technical challenge with a notable impact on sign language recognition. When capturing an image, numerous factors, such as lighting conditions (spectra, source distribution, and intensity) and camera characteristics (sensor response and lenses), exert their influence on the appearance of the captured sign. Additionally, skin reflectance properties and internal camera controls [38] further contribute to these effects. Moreover, noise originating from other elements present in the background and landmarks can also influence sign recognition outcomes.

4.3. Movement

The movements in sign language are dynamic acts, exhibiting trajectories with distinct beginnings and ends. The representation of dynamic sign language involves both isolated and continuous signing, wherein signs are performed consecutively without pauses. This introduces challenges related to similarity and occlusion, arising from variations in hand movements and

orientations, involving one or both hands in different angles and directions [39]. The determination of each sign's precise beginning and end presents a significant hurdle, resulting in what is termed Movement Epenthesis (ME) or transition segments. These ME segments act as connectors between sequential signs when transitioning from the final position of one sign to the initial position of the next. However, ME segments do not convey any specific sign information; instead, they contribute to the complexity of recognizing continuous sign sequences. The lack of well-defined rules for making such transitions poses a significant challenge [40], demanding careful attention and a demonstrable approach to address effectively.

4.4.Hand Segmentation, and Tracking

The segmentation process stands out as one of the most formidable challenges in computer vision, especially in the context of sign language recognition, where the extraction of hands from video frames or images holds particular significance due to their critical role in the recognition process. To address this, image segmentation is employed to isolate relevant hand data while eliminating undesired elements, such as background and other objects in the input, which might conflict with the classifier operations [41]. Image segmentation restricts the data region, enabling the classifier to focus solely on the Region of Interest (ROI). Segmentation methods can be categorized as contextual and non-contextual. Contextual segmentation takes spatial relationships between features into account, often using edge detection techniques. Conversely, non-contextual segmentation does not consider spatial relationships; rather, it gathers pixels based on global attributes. Hand tracking can be viewed as a subfield of segmentation, and it typically poses challenges, particularly when the hand moves swiftly, leading to significant appearance changes within a few frames [42].

4.5.Classifier

In the realm of sign language recognition, the classifier's selection and design require meticulous attention. It is essential to carefully determine the architecture of the classifier, encompassing its layers and parameters, in order to steer clear of potential problems like overfitting or underfitting. The primary objective is to achieve optimal performance in classifying sign language. Furthermore, the classifier's ability to generalize effectively across diverse data types, rather than being confined to specific subsets, is of paramount importance [43].

4.6.Time and Complexity

Real-time recognition of sign language is considered to be an important concern and one of the main problems that needs a real solution in order to give an efficient interpreter to fill up the communication gap between the deaf and public communities. The time problem arises from the need to process video data in real-time or with minimal delay. Computational complexity, both in hardware and software, can be quite demanding and may present challenges for the deaf community to effectively deal with [44].

5.Sign Language Public Datasets

The availability of sign language datasets is limited and can be considered as one of the main obstacles in designing an accurate recognition system, in which there are few datasets available for sign language compared to gesture databases. Several sign language datasets are created with many variations such as regional differences, type of images (RGB or Depth), type of acquiring methods (images, video), and so on. Sign language differs from one region to another just like spoken languages, and each type has its own properties and linguistic grammar. Most publicly available and utilized sign and gesture datasets in different languages are described in this section and categorized depending on the type of language as illustrated in Table 1, and Fig2.

Table 1: Public sign language datasets

Dataset	Language	Equipment	Modalities	Signers	Samples
ASL alphabets [45]	American	Webcam	RGB images	-	87,000
MNIST [46]	American	Webcam	Grey images	-	27,455
ASL Fingerspelling A [47]	American	Microsoft Kinect	RGB and depth images	5	48,000
NYU [48]	American	Kinect	RGB and depth images	36	81,009
ASL by Surrey [49]	American	Kinect	RGB and depth images	23	130,000
Jochen-Triesch [50]	American	Cam	Grey images with different background	24	720
MKLM [51]	American	Leap Motion device and a Kinect sensor	RGB and depth images	14	1400
NTU-HD [52]	American	Kinect sensor	RGB and depth images	10	1000
HUST [53]	American	Microsoft Kinect	RGB and depth images	10	10880
RVL-SLLL [54]	American	Cam	RGB video	14	
ChicagoFSWild [55]	American	Collected online from YouTube	RGB video	160	7,304
ASLG-PC12 [56]	American	Cam	RGB video	-	880
American Sign Language Lexicon Video (ASLLVD) [57]	American	Cam	RGB videos of different angles	6	3,300
MU [58]	American	Cam	RGB images with illumination variations in five different angles	5	2515
ASLID [59]	American	Web cam	RGB images	6	809
KSU-SSL [60]	Arabic	Cam and Kinect	RGB Videos with uncontrolled environment	40	16000
KArSL [61]	Arabic	Kinect V2	RGB video	3	75,300
ArSL by University of Sharjah [62]	Arabic	Analog camcorder	RGB images	3	3450
JTD [63]	Indian	Webcam	RGB images with 3 different backgrounds	24	720
IISL2020 [64]	Indian	Webcam	RGB video with uncontrolled environment	16	12100
RWTH-PHOENIX-Weather 2014 [65]	German	Webcam	RGB Video	9	8,257
SIGNUM [66]	German	Cam	RGB Video	25	33210
DEVISIGN-D [67]	Chinese	Cam	RGB videos	8	6000
DEVISIGN-L [67]	Chinese	Cam	RGB videos	8	24000
CSL-500 [68]	Chinese	Cam	RGB, depth and skeleton videos	50	25,000
Chinese Sign Language [69]	Chinese	Kinect	RGB, depth and skeleton videos	50	125000
38 BdSL [70]	Bengali	Cam	RGB images	320	12,160
Ishara-Lipi [71]	Bengali	Cam	Greyscale images	-	1800
ChaLearn14 [72]	Italian	Kinect	RGB and depth video	940	940
Montalbano II [73]	Italian	Kinect	RGB and depth video	20	940
UFOP-LIBRAS [74]	Brazilian	Kinect	RGB, depth and skeleton videos	5	2800
AUTSL [75]	Turkish	Kinect v2	RGB, depth and skeleton videos	43	38,336
RKS-PERSIANSIGN [76] in	Persian	Cam	RGB video	10	10,000
LSA64 [77]	Argentine	Cam	RGB video	10	3200
Polytropon (PGSL) [78]	Greek	Cam	RGB video	6	840
kETI [79]	Korean	Cam	RGB video	40	14,672



Figure 2: Samples of sign language datasets.

Several critical factors contribute to the evaluation of sign language datasets. One such factor is the number of signers involved in performing the signs, which significantly impacts the dataset's diversity and subsequently affects the evaluation of recognition systems' generalization rate. Additionally, the quantity of distinct signs within the datasets, particularly in isolated and continuous formats, holds considerable importance. Furthermore, the number of samples per sign plays a crucial role in training systems that require an ample representation of each sign. Adequate sample representation helps improve the robustness and accuracy of the recognition systems. Moreover, when dealing with continuous datasets, annotating them with temporal information for continuous sentence components is very important. This temporal information is vital for effectively processing and understanding this type of dataset [80]. Although sign language recognition is one of the gesture recognition applications, gesture datasets are seldom utilized for sign language recognition for many reasons. First, the classes count in the gesture recognition dataset has some degree of limitation. Secondly, sign language involves the simultaneous use of manual and non-manual gestures, posing challenges in annotating both types of gestures within a single gesture dataset. Moreover, sign language relies on hand gestures, while gesture datasets are broader and include gestures about full body movements. Additionally, gesture datasets lack the necessary details about hand fingers, which are essential for developing accurate sign language recognition systems [81]. Nevertheless, despite these limitations, gesture datasets can still play a role in training sign recognition systems. In this context, Table 2 presents a comprehensive overview of various gesture datasets, and Fig3 illustrates some representative examples.

Table 2: Gesture public datasets

Name	Modality	device	signers	samples
LMDHG [82]	RGB, and depth videos	Kinect and	21	608
SHREC Shape Retrieval Contest (SHREC) [83]	RGB, and depth videos	Intel RealSense short range depth camera	28	2800
UTD-MHAD [84]	RGB, depth and skeleton videos	Kinect and wearable inertial sensor	8	861
The Multicamera Human Action Video Data (MuHAVi) [85]	RGB video	8 camera views	14	1904
NUMA [86]	RGB, depth and skeleton videos	10 Kinect with three different views	10	1493

WEIZMANN [87]	Low resolution RGB video	Camera with 10 different viewpoints	9	90
NTU RGB [88]	RGB, depth and skeleton videos	Kinect	40	56 880
Cambridge hand gesture [89]	RGB video captured under five different illuminations	Cam	9	900
VIVA [90]	RGB, and depth videos	Kinect	8	885
MSR [91]	RGB, and depth videos	Kinect	10	320
CAD-60 [92]	RGB and depth video in different environments, such as a kitchen, a living room, and office	Kinect	4	48
HDM05MoCap (motion capture) [93]	RGB video	Cam	5	2337
CMU [94]	RGB images	CAM	25	204
isoGD [95]	RGB and depth videos	Kinect	21	47,933
NVIDIA [96]	RGB and depth video	Kinect	8	885
G3D [97]	RGB and depth video	Kinect	16	1280
UT Kinect [98]	RGB and depth video	Kinect	10	200
First-Person [99]	RGB and depth video	RealSense SR300 cam	6	1,175
Jester [100]	RGB	Cam	25	148,092
Ego Guster [101]	RGB and depth video	Kinect	50	2,081
NUS II [102]	RGB images with complex backgrounds, and various hand shapes and sizes	Cam	40	2000

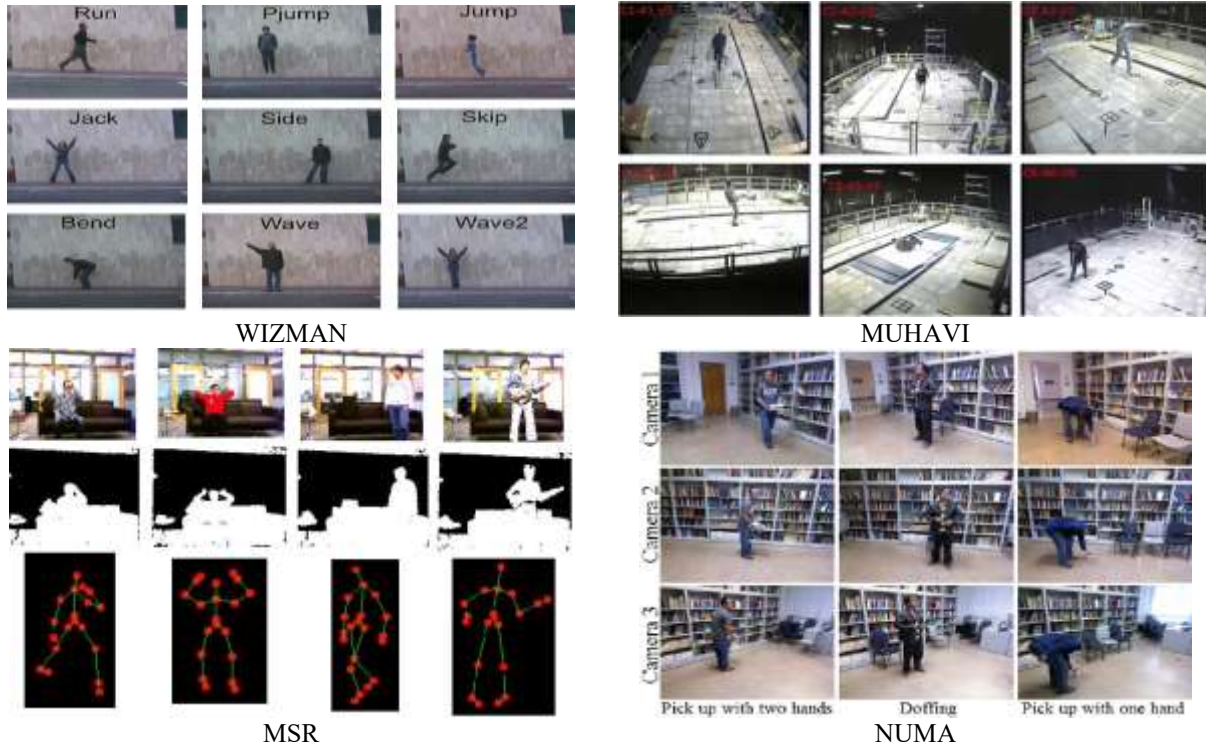


Figure 3: Samples of gesture datasets

6. Deep Learning based Sign Language Recognition-Related Works

Numerous research efforts have been dedicated to the recognition and translation of sign language across diverse languages worldwide, aiming to facilitate its conversion into other communication forms used by individuals, such as text or sound. This study categorizes the works of sign language recognition using DL according to the primary challenges encountered in recognition and the corresponding solutions proposed by each of the investigated works. Any sign language recognition system consists of key stages, which include signs acquisition, hand segmentation, tracking, preprocessing, feature extraction, and classification, as depicted in Fig4.

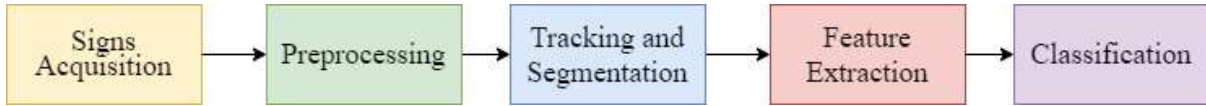


Figure 4: The procedural stages of sign language recognition

In sign Acquisition, the input modalities as mentioned earlier are either an image or a video stream using one type of vision-based capturing device or depth information using one of the hardware-based collecting equipment. The input modality may be in any format including RGB-colored, greyscale, and binary. In general, DL techniques need high quality data samples with sufficient amount for training to be conducted.

Accuracy is one of the most common performance measurements considered in any type of recognition system, in addition to the percentage of error that may be identified using the Equal Error Rate, Word Error Rate, and False Rate. Another evaluation metric named Bilingual Evaluation Understudy Score (BLEU), is used to measure the matching between the resulting sentences to the entered sign language. The perfect match results in a score of 1.0, while the worst score that represents mismatching is 0.0, so it is also considered as a measurement of accurate translation and widely used in machine learning systems [103]

The related sign language works using DL are categorized based on the type of problem solved in this work, and what is the technique utilized to get the desired result.

6.1. Related Works on Preprocessing Problem

The acquired signs may exhibit issues such as low quality, noise, varying degrees of orientation, or enormous size. Therefore, the preprocessing step becomes indispensable to rectify these issues in sign images and videos, effectively eliminating any environmental influences that might have affected them, such as variations in illumination and color. This phase involves the application of filters and other techniques to adjust the size, orientation, and color, ensuring improved data quality for subsequent analysis and recognition. The primary advantage of preprocessing is enhancing the image quality, which enables efficient hand segmentation from the scene for effective feature extraction. In the case of video streams, preprocessing serves to eliminate redundant and similar frames from the input video, thereby increasing the processing speed of the neural network without sacrificing essential information. Many sign language recognition using DL overcome the environmental condition problem using a variety of techniques. Table 3 illustrates the most important related work, the environmental condition being addressed, and the proposed technique. Fig5 shows a sample of the NUSII dataset to show the environmental condition problem.

Table 3: Related works on SLR using DL that address the various environmental conditions problem.

Author (s)	Year	Language	Modality	Type of condition	Deal with technique	results
[130]	2018	Bengali	RGB images	Variant background and skin colors	Modified VGG net	84.68%
[134]	2018	American	RGB images	noise and missing data	Augmentation	98.13%
[150]	2018	Indian	RGB video	Different viewing angles, background lighting, and distance	Novel CNN	92.88%
[158]	2019	American	Binary images	Noise	Erosion, closing, contour generation, and polygonal approximation,	96.83%
[159]	2019	American	Depth image	Variant illumination, and background	Attain depth images	88.7%
[164]	2019	chines	RGB, and depth video	Variant illumination, and background	Two-stream spatiotemporal network	96.7%

[173]	2019	Indian	RGB, and depth video	Variant illumination, background, and camera distance	Four stream CNN	86.87%
[178]	2020	Arabic	RGB images	Variant illumination, and skin color	DCNN	94.31%
[179]	2020	Arabic	RGB videos	Variant illumination, background, pose, scale, shape, position, and clothes	Bi-directional Long Short-Term Memory (BiLSTM)	89.59%
[180]	2020	Arabic	RGB Videos	Variant illumination, clothes, position, scale, and speed	3DCNN and SoftMax function	87.69%
[182]	2020	Arabic	RGB Videos	Variations in heights and distances from camera	Normalization	84.3%
[194]	2020	Arabic	RGB images	variant illumination, and background	VGG16 and the ResNet152 with enhanced softmax layer	99%
[201]	2020	American	Grayscale images	illumination, and skin color	Set the hand histogram	95%
[202]	2020	American	RGB images	Variant illumination, background	DCNN	99.96%
[206]	2021	Indian	RGB video	Variant illuminations, camera positions, and orientations	Google net+ BiLSTM	76.21%
[207]	2021	Indian	RGB images	Light and dark backgrounds	DCNN with few numbers of parameters	99.96%
[209]	2021	American	RGB video	Noise	Gaussian Blur	99.63%
[213]	2021	Korean	Depth Videos	Low resolution	Augmentation	91%
[224]	2021	Bengali	RGB images	Variant backgrounds, camera angle, light contrast, and skin tone	Conventional deep learning + Zero-shot learning ZSL	93.68%
[225]	2021	Arabic	RGB video	Variant illumination, background, and clothes	Inception-BiLSTM	84.2%
[227]	2021	American	Thermal images	Varying illumination	Adopt live images taken by a low-resolution thermal camera	99.52%
[229]	2021	Indian	RGB video	Varying illumination	3DCNN	88.24%
[230]	2021	American	RGB video	Noise, varying illumination	Median filtering + histogram equalization	96%
[236]	2021	Arabic	RGB images	Variant illumination, and background	Region-based Convolutional Neural Network (R-CNN)	93.4%
[239]	2022	Indian	RGB video	Variant illumination, and views	Grey scale conversion and histogram equalization	98.7%
[241]	2022	Arabic	RGB video	Variant illumination, and background	CNN+ RNN	98.8%
[249]	2022	Arabic	Greyscale images	Variant illumination, and background	Sobel filter	97%
[253]	2022	Arabic	RGB, and depth video	Variant Background	ResNet50-BiLSTM	99%
[259]	2022	American	RGB, and depth images	Noise and illumination variation	Median filtering and histogram equalization	91.4%
[261]	2022	American	Skeleton video	Noise in video frames	An innovative weighted least square (WLS) algorithm	97.98%
[270]	2022	English	Wi-Fi signal	Noise and uncleaned Wi-Fi signals.	Principal Component Analysis (PCA)	95.03%



Figure 5: Sample images (class 9) from NUS hand posture dataset-II (data subset A), showing the variations in hand posture sizes and appearances.

Another challenge arises when attempting to recognize signs, particularly in the dynamic type, where movement is considered one of the key phonological parameters in sign phonology.

This pertains to the variations in hand location, speed, orientation, and angles during the signing process [104]. A consensus on how to characterize and organize movement types and their associated features in a phonological representation has been lacking. Due to divergent approaches and perspectives, there remains uncertainty about the most suitable and standardized way to define and categorize movements in sign language. In general, there are three main types of movements in sign language [105,106]:

- Movement of the hands and arms: include waving, pointing, or tracing shapes in the air.
- Movement of the body: include twisting, turning, or leaning to indicate direction or location.
- Movement of the face and head: include nodding, shaking the head, or raising the eyebrows to convey different meanings or emotions.

The movement involved in demonstrating sign language also involves a significant challenge, which includes dealing with similar paths of movement (Trajectory), and Occlusion. The arm trajectory formation refers to the principles and laws that invariantly govern the selection, planning, and generation processes of multi-joint movements, as well as to the factors that dictate their kinematics, namely geometrical and temporal features [107]. The sign language movement trajectory swerves to some extent, due to the action speed, and arm length of the user; even for the same user, the psychological changes resulted in inconsistent implementation speed of sign language movement [108]. Movement trajectory recognition is the key link of sign language translation research, which influence directly on the accuracy of sign language translation, in which the same signs match with variant movement trajectory predominantly refer to two variant meanings, that is, illustrating different sign language [109]. On the other hand, occlusion means that some fingers or parts of the hand would be covered (not in view of the camera) or hidden by other parts of the scene, so the sign cannot be detected accurately [110]. The occlusion may appear in various parts including hand/hand, and hand/face depending on the movement and the captured scene. The occlusion has a great effect on the segmentation procedure especially skin segmentation techniques [111]. Table 4 summarizes the most important related DL works that handle these types of problems in sign language recognition.

Table 4: Related works on SLR using DL that address movement orientation, trajectory, occlusion problems.

Author(s)	Year	Type of variation	language	Signing mode	Model	Accuracy	Error Rate
[129]	2018	similarities, and occlusion	American	Static	DCNN	92.4%	
[135]	2018	Movement	Brazilian	Isolated	Long-term Recurrent Convolutional Networks	99%	-
[138]	2018	size, shape, and position of the fingers or hands	American	Static	CNN	82%	-
[140]	2018	Hand movement	American	Isolated	VGG 16	99%	-
[144]	2018	Movement	American	Isolated	Leap Motion Controller	88.79%	-
[145]	2018	3D motion	Indian	Isolated	Joint Angular Displacement Maps (JADMs)	92.14%	
[150]	2018	head and hand movements	Indian	Continues	CNN	92.88 %	-
[155]	2019	Hand movement	Indian	Continues	Wearable systems to measure muscle intensity, hand orientation, motion, and position	92.50%	-
[156]	2019	Variant hand orientations	Chines	Continues	Hierarchical Attention Network (HAN) and Latent Space	82.7%	-
[165]	2019	Similarity and trajectory	Chines	Isolated	Deep 3-d Residual ConvNet + BiLSTM	89.8%	-
[166]	2019	orientation of camera, hand position and movement, inter hand relation	Vietnam	Isolated	DCNN	95.83%	

[173]	2019	Movement, self-occlusions, orientation, and angles	Indian	Continues	Four stream CNN	86.87%	
[174]	2019	Movement in different distance from the camera	American	Static	Novel DNN	97.29%	-
[176]	2020	Angles, distance, object size, and rotations	Arabic	Static	Image Augmentation	90%	0.53
[180]	2020	fingers' configuration, hand's orientation, and its position to the body	Arabic	Isolated	Multilayer perceptron+ Autoencoder	87.69%	
[185]	2020	Hand Movement	Persian	Isolated	Single Shot Detector (SSD) +CNN+LSTM	98.42%	
[186]	2020	shape, orientation, and trajectory	Greek	Isolated	Fully convolutional attention-based encoder-decoder	95.31%	-
[192]	2020	Trajectory	Greek	Isolated	incorporate the depth dimension in the coordinates of the hand joints	93.56%	-
[195]	2020	finger angles and Multi finger movements	Taiwan	Continues	Wristband with ten modified barometric sensors+ dual DCNN	97.5%	
[196]	2020	movement of fingers and hands	Chinese	Isolated	Motion data from IMU sensors	99.81%	-
[197]	2020	finger movement	Chinese	Isolated	Trigno Wireless sEMG acquisition system used to collect multichannel sEMG signals of forearm muscles	93.33%	
[199]	2020	finger and arm motions, two-handed signs, and hand rotation	Chinees	Continues	Two armbands embedded with an IMU sensor and multi-channel sEMG sensors are attached on the forearms to capture both arm, and finger movements	-	10.8%
[76]	2020	Hand occlusion	Persian	Isolated	Skeleton detection	99.8%	
[204]	2020	Trajectory	Brazilian	Isolated	Convert the trajectory information into spherical coordinates	64.33%	
[210]	2021	Trajectory	Arabic	Isolated	Multi-Sign Language Ontology (MSLO)	94.5%	
[213]	2021	Movement	Korean	Isolated	3DCNN	91%	
[214]	2021	finger movement	Chines	Isolated	Design a low-cost data glove with simple hardware structure to capture finger movement and bending simultaneously	77.42%	
[218]	2021	Skewing, and angle rotation	Bengali	Static	DCNN	99.57	0.56
[219]	2021	Hand motion	American	Continues	Sensing Gloves	86.67%	
[223]	2021	spatial appearance and temporal motion	Chines	Continues	Lexical prediction network	91.72%	6.10
[226]	2021	finger self-occlusions, view invariance	Indian	Continues	Motion modelled deep attention network (M2DA-Net)	84.95%	
[228]	2021	Occlusions of hand/hand, hands/face, or hands/upper body postures.	American	Continues	Novel hyperparameter based optimized Generative Adversarial Networks (H-GANs) Deep Long Short-Term Memory (LSTM) as generator and LSTM with 3D Convolutional Neural Network (3D-CNN) as a discriminator	97%	1.4
[230]	2021	Variant view	American	Isolated	3-D CNN's cascaded	96%	
[233]	2021	Hand occlusion,	Italian	Isolated	LSTM+CNN	99.08%	
[237]	2021	Finger occlusion, motion blurring, variant signing styles.	Chines	Continues	Dual Network up on a Graph Convolutional Network (GCN).	98.08%	
[239]	2022	self-structural characteristics, and occlusion	Indian	Continues	Dynamic Time Warping (DTW)	98.7%	
[240]	2022	High similarity and complexity	American	Static	DCNN	99.67%	0.0016
[241]	2022	Movement	Arabic	Isolated	The difference function	98.8%	
[259]	2022	Hand Occlusion	American	Static	Re-formation layer in the CNN	91.40%	

[260]	2022	Trajectory, hand shapes, and orientation	American	Isolated	Media Pipe's Landmarks with GRU	99%	
[261]	2022	ambiguous and 3D double-hand motion trajectories	American	Isolated	3D extended Kalman filter (EKF) tracking, and approximation of a probability density function over a time frame.	97.98%	
[262]	2022	Movement	Turkish	Continues	Motion History Images (MHI) generated from RGB video frames	94.83%	
[264]	2022	Movement	Argentina	Continues	Propose an accumulative video motion (AVM) technique	91.8%	
[269]	2022	orientation angle, prosodic, and similarity	American	continues	Develop robust fast fisher vector (FFV) in in Deep Bi-LSTM	98.33%	
[270]	2022	variant length, sequential patterns,	English	Isolated	Novel Residual-Multi Head model	95.03%	

6.2. Related Works on Segmentation and Tracking Problem

Detecting the signer hand in a still image or tracking it in a video stream is challenging and affected by many factors discussed earlier in the preprocessing phase such as environment, movement, hand shape, and occlusion. Hence, the careful choice of an appropriate segmentation technique is of utmost importance, as it profoundly influences the recognition of sign language and the work of the subsequent phases (feature extraction and classification). The hand segmentation identifies the beginning and end of each sign. This is necessary for accurate recognition and understanding of the signer's message. Through the process of segmenting the sign language input, the recognition system can concentrate on discerning individual signs and their respective meanings, thereby avoiding the interpretation of the entire continuous signing stream as a single sign. In addition to enhancing recognition accuracy, segmentation contributes to system efficiency and speed. By dividing the input into distinct signs, the system can process each sign independently, reducing computational complexity and improving response time. Furthermore, segmentation facilitates advancements in sign language recognition technology by enabling the creation of sign language corpora annotated with information about individual signs. Such resources are valuable for training and evaluating sign language recognition systems and conducting linguistic research on sign language structure and syntax. Various segmentation techniques are employed, including Background subtraction [112], Skin color detection [113], Template matching [114], Optical flow [115], and Machine learning [116]. Table 5 presents DL for sign language recognition-related works that focus on addressing the segmentation and tracking challenges to achieve optimal system performance.

Table 5: Related works on SLR using DL that address segmentation problem.

Author(s)	Year	Input Modality	Segmentation method	Results
[131]	2018	RGB image	HSV color model	99.85%
[148]	2018	RGB image	Skin segmentation algorithm based on color information	94.7%
[149]	2018	RGB images	k-means-based algorithm	94.37%
[158]	2019	RGB images	Color segmentation by MLP network	96.83%
[159]	2019	Depth image	Wrist line localization by algorithm-based thresholding	88.7%
[164]	2019	RGB, and depth video	Aligned Random Sampling in Segments (ARSS)	96.7%
[168]	2019	RGB, and depth images	Depth based segmentation using data of Kinect RGB-D camera	97.71%
[171]	2019	RGB video	Design an adaptive temporal encoder to capture crucial RGB visemes and skeleton signees	94.7%
[179]	2020	RGB videos	Hand semantic Segmentation named as DeepLabv3+	89.59 %
[180]	2020	RGB Videos	Novel method based on open pose	87.69 %
[182]	2020	RGB Videos	Viola and Jones, and human body part ratios	84.3%
[183]	2020	RGB images	Robert edge detection method	99.3 %

[185]	2020	RGB video	SSD is a feed-forward convolutional network A Non-Maximum Suppression (NMS) step is used in the final step to estimate the final detection	98.42%
[187]	2020	RGB images	Sobel edge detector, and skin color by thresholding	98.89%
[188]	2020	RGB images	Open-CV with a Region of Interest (ROI) box in the driver program	93%
[189]	2020	RGB Videos	Frame stream density compression (FSDC) algorithm	10.73 error
[199]	2020	RGB Videos	Design an attention-based encoder-decoder model to realize end-to-end continuous SLR without segmentation	10.8% WER
[200]	2020	RGB images	Single Shot Multi Box Detection (SSD)	99.90%
[209]	2021	RGB Video	Canny	99.63%
[216]	2021	RGB images	Erosion, Dilation, and Watershed Segmentation	99.7 %
[219]	2021	RGB Video	Data sliding window	86.67%
[236]	2021	RGB images	R-CNN	93%
[239]	2022	RGB videos	Novel Adaptive Hough Transform (AHT)	98.7%
[246]	2022	RGB images, and video	Grad Cam and Cam shift algorithm	99.85%
[248]	2022	Grey images	YCbCr, HSV and watershed algorithm	99.60%,
[249]	2022	RGB images	Sobel operator method	97 %
[263]	2022	RGB images	Semantic	99.91%
[267]	2022	RGB images	R-CNN	99.7%
[268]	2022	RGB video	Mask is created by extracting the maximum connected region in the foreground assuming it to be the hand+ Canny method	99%

6.3.Related Works on Feature Extraction Problem

The feature extraction goal is to capture the most essential information about the sign language gestures while removing any redundant or irrelevant information that may be present in the input data. The process of feature extraction offers numerous advantages in sign language recognition. It enhances accuracy by effectively representing the distinctive characteristics of each sign and gesture, thereby facilitating the system's ability to differentiate between them. Moreover, feature extraction reduces both processing time and computational complexity, as the extracted features are typically represented in a more compact and informative manner compared to raw input data. Additionally, feature extraction confers robustness against noise and variability, as features can be designed to be invariant to specific types of variations, such as changes in lighting conditions or background clutter [117,118]. This enables the recognition system to maintain its performance even in challenging and diverse environments. Table 6 shows related DL works for sign language recognition that focus on solving the problem of features extraction.

Table 6: Related works on SLR using DL that address feature extraction problem.

Author(s)	Year	Dataset	Technique	Signing mode	Feature(s)	Result
[130]	2018	Collected	DCNN	static	Hand shape	84.6%
[135]	2018	Collected	3D CNN	Isolated	spatiotemporal	99%
[138]	2018	ASL Finger Spelling	CNN	Static	depth and intensity	82%
[141]	2018	RWTH-2014	3D Residual Convolutional Network (3D-ResNet)	Continues	Spatial information, and temporal connections across frames	37.3 WER
[143]	2018	Collected	3D-CNNs	Isolated	spatiotemporal	88.7%
[144]	2018	Collected	DCNN	Isolated	hand palm sphere radius, and position of hand palm and fingertip	88.79%
[149]	2018	ASL Finger Spelling	Histograms of oriented gradients, and Zernike moments	Static	Hand shape	94.37%
[150]	2018	Collected	CNN	Continues	Hand shape	92.88 %
[151]	2018	Collected	3DRCNN	Continues/ Isolated	motion, depth, and temporal	69.2%

[152]	2018	SHREC	Leap Motion Controller (LMC) sensor	Isolated, static	finger bones of hands.	96.4%
[153]	2018	Collected	Hybrid Discrete Wavelet Transform, Gabor filter, and histogram of distances from Centre of Mass	Static	Hand shape	76.25%
[154]	2018	Collected	DCNN	Static	Facial expressions	89%
[156]	2019	Collected	Two-stream 3-D CNN	Continues	Spatiotemporal	82.7%
[158]	2019	Collected	CNN	Static	Hand shape	96.83%
[79]	2019	Collected	Open Pose library	Continues	human key points (hand, face, body)	55.2%
[159]	2019	ASL fingerspelling	PCA Net	Static	hand shape (corners, edges, blobs, or ridges)	88.7%
[161]	2019	SIGNUM	Stacked temporal fusion layers in DCNN	Continues	spatiotemporal	2.80 WER
[162]	2019	Collected	Leap motion device	Continues Isolated	3D positions of the fingertips	72.3% 89%
[163]	2019	Collected	CNN	Static	Hand shape	95%
[164]	2019	CSL	D-shift Net	Continues	spatial features time features, and temporal.	96.7%
[165]	2019	DEVSIGN_D	B3D Res-Net	Isolated	spatiotemporal	89.8%
[166]	2019	Collected	Local and GIST Descriptor	Isolated	Spatial and scene-based features	95.83%
[169]	2019	Collected	Restricted Boltzmann Machine (RBM)	Isolated	Handshape, and network generated features	88.2%
[170]	2019	KSU-SSL	3D-CNN	Isolated	hand shape, position, orientation, and temporal dependence in consecutive frames	77.32%
[171]	2019	Collected	C3D, and Kinect device	Continues	Temporal, and Skeleton	94.7%
[175]	2019	Collected	Open Pose library with Kinect V2	Static	3D skeleton	98.9%.
[177]	2020	Ishara-Lipi	Mobile Net V1	Isolated	Two hands shape	95.71%
[178]	2020	Collected	DCNN	Static	Hand shape	94.31%.
[179]	2020	Collected	Single layer Convolutional Self-Organizing Map (CSOM)	Isolated	Hand shape	89.59%
[180]	2020	KSU-SSL	Enhanced C3D architecture	Isolated	Spatiotemporal of hand and body	87.69 %
[182]	2020	KSU-SSL	3DCNN	Isolated	Spatiotemporal	84.3%
[185]	2020	Collected	ResNet50 model	Isolated	Hand shape, Extra Spatial hand Relation (ESHR) features, and Hand Pose (HP), temporal.	98.42%
[186]	2020	Polytropon (PGSL)	ResNet-18	Isolated	Optical flow of skeletal, handshapes, and mouthing	95.31%
[187]	2020	Collected	Discrete cosines transform, Zernike moment, scale-invariant feature transform, and social ski driver optimization algorithm	Static	Hand shape	98.89%
[189]	2020	RWTH-2014	Temporal convolution unit and dynamic hierarchical bidirectional GRU unit	Continues	spatiotemporal	10.73% BLEU
[191]	2020	Collected	Standard score' normalization on the raw Channel State Information (CSI) acquired from the Wi-Fi device, and MIFS algorithm	Static, and continues	The cross-cumulant features (unbiased estimates of covariance, normalized skewness, normalized kurtosis)	99.9%
[192]	2020	GSL	Open Pose human joint detector	Isolated	3D hand skeletal, and region of hand, and mouth	93.56%
[197]	2020	Collected	Four channel surface electromyography (sEMG) signals	Isolated	time-frequency joint features	93.33%
[199]	2020	Collected	Euler angle, Quaternion from IMU signal	Continues	Hand Rotation	10.8% WER
[76]	2020	RKS-PERSIANSIGN	3DCNNs	Isolated	Spatiotemporal	99.8%
[202]	2020	ASL fingerspelling A	DCNN	Static	Hand Shape	99.96%

[203]	2020	Collected	Construct a color-coded topographical descriptor from joint distances and angles, to be used in 2 streams (CNN)	Isolated	distance and angular	93.01%
[204]	2020	Collected	Two CNN models and a descriptor based on Histogram of cumulative magnitudes	Isolated	Two hands, skeleton, and body	64.33%
[208]	2021	RWTH-2014T	Semantic Focus of Interest Network with Face Highlight Module (SFoI-Net-FHM)	Isolated	Body and facial expression	10.89 Bleu
[210]	2021	Collected	(ConvLSTM)	Isolated	Spatiotemporal	94.5%
[212]	2021	Collected	ResNet50	Static	hand area, the length of axis of first eigenvector, and hand position changes.	96.42%.
[214]	2021	Collected	f-CNN (fusion of 1-D CNN and 2-D CNN)	Isolated	Time and spatial-domain features of finger resistance movement	77.42%
[217]	2021	MU	Modified Alex Net and VGG16	Static	Hand edges and shape	99.82%
[222]	2021	Collected	VGG net of six convolutional layers	Static	Hand shape	97.62%
[224]	2021	38 BdSL	DenseNet201, and Linear Discriminant Analysis	Static	Hand shape	93.68%
[225]	2021	KSU-ArSL	Bi-LSTM	Isolated	spatiotemporal	84.2%
[226]	2021	Collected	Paired pooling network in view pair pooling net (VPPN)	Isolated	spatiotemporal	84.95%
[228]	2021	ASLLVD	Bayesian Parallel Hidden Markov Model (BPaHMM) + stacked denoising variational autoencoders (SD-VAE) + PCA	Continues	Shape of hand, palm, and face, along with their position, speed, and distance between them	97%
[230]	2021	ASLLVD	3-D CNN's cascaded	Isolated	spatiotemporal	96.0%
[231]	2021	Collected	leap motion controller	Static, and Isolated	sphere radius, angles between fingers their distance	91.82%
[232]	2021	RWTH- 2014	(3 C 2 C 1) D ResNet	Continues	height, motion of hand, and frame blurriness levels	23.30 WER
[233]	2021	Montalbano II	AlexNet + Optical Flow (OF) + Scene Flow (SF) methods	Isolated	Pixel level, and hand pose	99.08%
[234]	2021	RWTH- 2014	GAN	Continues	spatiotemporal	23.4 WER
[235]	2021	MNIST	DCNN	Static	Hand shape	98.58%
[236]	2021	Collected	R-CNN	Static	Hand shape	93%
[237]	2021	CSL-500	Multi-scale spatiotemporal attention network (MSSTA)	Isolated	Spatiotemporal	98.08%
[242]	2022	MNIST	modified CapsNet	Static	Spatial, and orientations	99.60%
[243]	2022	RKS-PERSIANSIGN	Singular value decomposition SVD	Isolated	3D hand key points between the segments of each finger, and their angles.	99.5%
[244]	2022	Collected	2DCRNN + 3DCRNN	Continues	Spatiotemporal out of small patches	99%
[246]	2022	Collected	Atrous convolution mechanism, and semantic spatial multi-cue model	Static Isolated	pose, face, and hand, and Spatial, full frame,	99.85%
[253]	2022	Collected	4 DNN models using 2D and 3D CNN	Isolated	Spatiotemporal	99%
[255]	2022	Collected	Scale-Invariant Feature Transformation (SIFT)	Static	Corner, edges, rotation, blurring, and illumination.	97.89%
[256]	2022	Collected	InceptionResNetV2	Isolated	Hand shape	97%
[257]	2022	Collected	Alex net	Static	Hand shape	94.81%
[258]	2022	Collected	Sensor + mathematical equations+ CNN	Continues	Mean, Magnitude of Mean, Variance, correlation, Covariance, and frequency domain features+ spatiotemporal	0.088 WER
[260]	2022	Collected	Media Pipe framework	Isolated	hands, body, and face	99%
[261]	2022	Collected	Bi-RNN network, maximal information correlation, and leap motion controller	Isolated	hand shape, orientation, position, and motion of 3D skeletal videos.	97.98%

[264]	2022	LSA64	dynamic motion network (DMN)+ Accumulative motion network (AMN)	Isolated	spatiotemporal	91.8%
[265]	2022	CSL-500	Spatial-temporal-channel attention (STCA) is proposed	isolated	spatiotemporal	97.45%
[268]	2022	Collected	SURF (Speeded Up Robust Features)	Isolated	distribution of the intensity material within the neighborhood of the interest point	99%
[269]	2022	Collected	Thresholding and Fast Fisher Vector Encoding (FFV)	Isolated	Hand, palm, finger shape, and position and 3D skeletal hand characteristics	98.33%

6.4. Related Works on Classification Problem

Classification is the final phase of any sign language recognition system and used before transferring the sign language into another form of data whether text or sound. In general, a particular sign is recognized by comparing it with the trained dataset, in which it categorizes the data into respective classes, depending on the feature vector obtained. Moreover, the system can calculate the probability associated with each class, allowing the data to be categorized under the respective class based on probability values. Overall, the classification conditions for sign language using DL involve selecting appropriate data representation, feature extraction techniques, classification algorithms, evaluation metrics, and ensuring sufficient and diverse training data. These factors collectively contribute to the accuracy and effectiveness of the sign language classification system. However, it may have some kinds of problems such as overfitting. In the realm of DL, overfitting occurs when a neural network model becomes too specialized in learning from the training data to the extent that it fails to generalize effectively to new, and unseen data. In other words, the model "memorizes" the training examples instead of learning the underlying patterns or relationships. When a DL model overfits, it performs very well on the training data but struggles to accurately predict or classify new instances that it has not encountered during training [119]. Various causes and indicators of overfitting exist, including a high model complexity with numerous parameters, insufficient training data, lack of regularization, excessive training epochs, and reliance on training data for evaluation [120]. To mitigate overfitting in deep models, several effective techniques can be employed. These include regularization methods [121], the incorporation of dropout layers [122], early stopping criteria [123], data augmentation strategies [124], and increasing training data [125]. These techniques can help to enhance model generalization and prevent the adverse effects of overfitting. Table 7 summarizes some related work of sign language recognition systems using DL that focuses on solving the problem of overfitting.

Table 7: Related works on SLR using DL that address overfitting problem.

Author(s)	Year	dataset	Model	technique	result
[129]	2018	NTU	DCNN	Augmentation	92.4%
[130]	2018	Collected	Modified VGG net	Dropout	84.68%
[132]	2018	Ishara-Lipi	DCNN	Dropout	94.88%
[133]	2018	Collected	DCNN	small convolutional filter sizes, Dropout, and learning strategy	85.3%
[136]	2018	HUST	Deep Attention Network (DAN)	data augmentation	73.4%
[142]	2018	ASL Finger Spelling A	DNN	Dense Net	90.3%
[143]	2018	Collected	3DCNN	SGD	88.7%
[146]	2018	SIGNUM	CNN-HMM hybrid	Augmentation	7.4 error
[157]	2019	Collected	DCNN	Augmentation	93.667%
[79]	2019	Collected	ResNet-152	batch size, Augmentation	55.28%
[163]	2019	Collected	VGG-16	Dropout	95%
[166]	2019	Collected	DCNN	Augmentation	95.83%
[167]	2019	Collected	DCNN	Dense Net	90.3%

[171]	2019	Collected	LSTM	Increase hidden state number	94.7%
[172]	2019	NVIDIA	Squeeze-net	Augmentation	83.29%
[173]	2019	G3D	Four stream CNN	Sharing of multi modal features with RGB spatial features during training and drop out	86.87%
[175]	2019	Collected	DCNN	Augmentation	98.9%.
[176]	2020	Collected	DCNN	Pooling Layer	90%
[181]	2020	Collected	DCNN	Reduce epochs to 30, and dropout added after each maxpooling	97.6%
[184]	2020	Collected	CNN with 8 layers	Augmentation	89.32 %
[188]	2020	MNIST	CNN	Dropout	93%
[190]	2020	Collected	Enhanced Alex Net	Augmentation	89.48%
[191]	2020	Collected	SVM	Augmentation, and k-fold cross validation	99.9%
[193]	2020	KETI	CNN+LSTM	New data augmentation	96.2%
[194]	2020	Collected	VGG16, and ResNet152 with enhanced softmax layer	Augmentation	99%
[196]	2020	Collected	RNN-LSTM	dropout layer (DR)	99.81%
[201]	2020	Collected	CNN	dropout layer, and augmentation	95%
[203]	2020	NTU	2 stream CNN	randomness in the features interlocking fusion with dropout	93.01%
[207]	2021	Jochen-Triesch's	DCNN	two dropouts	99.96%
[214]	2021	Collected	Generic temporal convolutional network (TCN)	Dropout	77.42%
[215]	2021	Collected	DCNN	Dropout	96.65%
[216]	2021	Collected	DCNN	Cyclical learning rate method	99.7%
[217]	2021	MU	Modified AlexNet and VGG16	Augmentation	99.82%
[222]	2021	Collected	CNN	Dropout	97.62%
[229]	2021	Collected	3DCNN	Dropout & Regularization	88.24%
[236]	2021	Collected	ResNet-18	Zero-patience stopping criteria	93.4%
[238]	2021	Collected	DCNN	Synthetic Minority Oversampling Technique (SMOTE)	97%
[240]	2022	Collected	DCNN	Augmentation	99.67%
[253]	2022	Collected	ResNet50-BiLSTM	Augmentation	99%
[256]	2022	Collected	LSTM, and GRU	Dropout	97%
[263]	2022	BdSL	CNN	Augmentation	99.91%

Another critical issue that must be considered when designing a deep model for sign language recognition is the generalization, which refers to the capability of a model to operate accurately on unseen data that is distinct from the training one. The model demonstrates a high degree of generalization ability by consistently achieving impressive performance across a wide range of diverse and distinct datasets [126]. Having consistent results across different datasets is an important characteristic for a model to be considered robust and reliable, which demonstrates that it can be applied effectively to various real-world scenarios. The datasets can have different characteristics, biases, or noise levels. Therefore, it is crucial to carefully evaluate and validate the model's performance on each specific dataset to ensure its reliability and generalization ability [127]. Table 8, presents relevant works in sign language recognition using DL, focusing on the model's generalization ability by evaluating its performance on diverse datasets.

Table 8: Related works on SLR using DL that aim to achieve generalization.

Author(s)	Year	Datasets	Technique	Result
[129]	2018	ASL finger spelling A NTU	DCNN	92.4% 99.7%
[134]	2018	NYU MU ASL Fingerspelling A ASL Surrey	Restricted Boltzmann Machine (RBM)	90.01% 99.31% 98.13% 97.56%
[136]	2018	NTU HUST	DAN	98.5% 73.4%
[143]	2018	Collected CSL ChaLearn14	3D-CNN	88.7% 95.3%

[145]	2018	Collected HMD05 CMU	JADM+CNN	88.59% 87.92% 87.27%
[146]	2018	RWTH 2012 RWTH 2014 SIGNUM	CNN-HMM hybrid	30.0 WER 32.5 7.4
[156]	2019	Collected RWTH- 2014	Hierarchical Attention Network (HAN) + Latent Space LS-HAN	82.7% 61.6%
[161]	2019	RWTH- 2014 SIGNUM	DCNN	22.86 WER 2.80
[164]	2019	CSL IsoGD	Proposed multimodal two-stream CNN	96.7% 63.78%
[165]	2019	DEVISIGN-D Collected	Deep 3-d Residual ConvNet + BiLSTM	89.8% 86.9%
[170]	2019	KSU-SSL ArSL RVL-SLLL	3D-CNN	77.32% 34.90% 70%
[173]	2019	Collected RGB-D MSR UT Kinect G3D	Four stream CNN	86.87% 86.98% 85.23% 88.68%
[174]	2019	Jochen-Triesch MKLM Novel SI-PSL	Novel DNN	97.29% 96.8% 51.88%
[182]	2020	KSU-SSL ArSL by University of Sharjah RVL-SLLL	3DCNN	84.38% 34.9% 70%
[186]	2020	PGSL ChicagoFSWild RWTH 2014T	DCNN	95.31% 92.63% 76.30%
[187]	2020	ASL MU	Deep Elman recurrent neural network	98.89% 97.5%
[192]	2020	GSL ChicagoFSWild	CNN	93.56% 91.38%
[76]	2020	NYU First-Person, RKS-PERSIANSIGN	CNN	4.64 error 91.12% 99.8%
[202]	2020	NUS American fingerspelling A	DCNN	94.7% 99.96%
[203]	2020	HDM05 CMU NTU Collected	2 stream CNN	93.42% 92.67% 94.42% 93.01%
[204]	2020	UTD-MHAD IsoGD Collected	linear SVM classifier	94.81% 67.36% 64.33%
[207]	2021	Collected RGB images. Jochen- Triesch's	DCNN	99.96% 100%
[210]	2021	LSA64 LSA Collected	3DCNN	98.5% 99.2 % 94.5%
[211]	2021	ASLG-PC12 RWTH- 2014	GRU and LSTM Bahdanau and Luong's attention mechanisms	66.59% 19.56% BLEU
[221]	2021	ASL alphabet, ASL MNIST MSL	Optimized CNN based on PSO	99.58% 99.58% 99.10%
[225]	2021	KSU-ArSL Jester NVIDIA	Inception-BiLSTM	84.2% 95.8% 86.6%
[226]	2021	Collected NTU MuHAVi, WEIZMANN NUMA	Motion modelled deep attention network (M2DA-Net)	84.95% 89.98% 85.12% 82.25% 88.25%
[228]	2021	RWTH- 2014 ASLLVD	Novel hyperparameter based optimized Generative. Adversarial Networks (H-GANs)	73.9% 97%

[232]	2021	RWTH- 2014 Collected	Bidirectional encoder representations from transformers (BERT) + ResNet	20.1 23.30 WER
[233]	2021	Montalbano II isoGD MSR CAD-60	LSTM+CNN	99.08% 86.10% 98.40% 95.50%
[234]	2021	RWTH2014 (CSL) (GSL)	GAN	23.4 2.1 2.26
[237]	2021	CSL-500 DEVISIGN-L	Dual Network up on a Graph Convolutional Network (GCN).	98.08% 64.57%
[242]	2022	SLDD MNIST	Modified Caps Net architecture (SLR-Caps Net)	99.52% 99.60%
[243]	2022	RKS-PERSIANSIGN First-Person ASVID isoGD	Single shot detector, 2D convolutional neural network, singular value decomposition (SVD), and LSTM	99.5% 91% 93% 86.1%
[247]	2022	Collected Collected ASL finger spelling	DCNN+ diffGrad optimizer	92.43% 88.01% 99.52%
[248]	2022	38 BdSL Collected Ishara-Lipi	BenSignNet	94.00% 99.60% 99.60%
[251]	2022	Collected Collected Collected	DCNN	99.41% 99.48% 99.38%
[254]	2022	Collected Cambridge hand gesture	Hybrid model based on VGG16-BiLSTM	83.36% 97%
[255]	2022	Collected MNIST, JTD NUS	Hybrid Fist CNN	97.89%, 95.68% 94.90% 95.87%
[256]	2022	ASL GSL AUTSL IISL2020	LSTM+GRU	95.3% 94% 95.1% 97.1%
[261]	2022	Collected SHREC LMDHG	DLSTM	97.98% 96.99% 97.99%
[262]	2022	AUTSL Collected	3D-CNN	93.53% 94.83%
[265]	2022	CSL-500 Jester Ego Gesture	deep R (2+1) D	97.45% 97.05% 94%
[266]	2022	MU HUST-ASL	end-to-end fine-tuning method of a pre-trained CNN model with score-level fusion technique	98.14% 64.55%
[269]	2022	SHREC Collected LMDHG	FFV-Bi-LSTM	92.99% 98.33% 93.08%

The choice of DL layers significantly influences the classification model's performance, as it determines the model's architecture and its ability to learn and represent intricate patterns in the input data. Selecting the right layers involves a comprehensive understanding of the data's characteristics, problem complexity, and available resources for training and inference. Often, it necessitates experimentation, tuning, and domain expertise to discover the optimal combination of layers that maximizes classification performance for a particular task [128]. In sign language recognition, numerous authors have designed and utilized deep models to achieve desired performance levels, as depicted in Table 9.

Table 9 Related works' Classifiers employed in SLR using DL.

Author	year	Input modality	Classifier	result
[129]	2018	Static	DCNN	92.4%
[131]	2018	Static	DCNN	99.85%
[133]	2018	Static	DCNN	85.3 %
[134]	2018	Static	restricted Boltzmann machine	98.13 %
[135]	2018	Isolated	LRCNs and 3D CNNs	99 %
[136]	2018	Static	DAN	73.4%
[137]	2018	Static	(CNNs) of variant depth sizes and stacked denoising autoencoders	92.83%
[139]	2018	Static	DCNN	82.5%
[142]	2018	Static	DCNN	90.3 %
[145]	2018	Isolated	DCNN	88.59%
[146]	2018	Continues	CNN-HMM hybrid	7.4 error
[147]	2018	Static	DCNN	98.05 %
[151]	2018	Isolated	3DCNN, and enhanced fully connected (FCRNN)	69.2 %
[155]	2019	Continues	Deep Capsule networks and game theory	92.50%
[156]	2019	Continues	Hierarchical Attention Network (HAN) and Latent Space	82.7 %
[157]	2019	Static	DCNN	93.667%
[160]	2019	Static	DCNN	97 %
[161]	2019	Continues	DCNN	2.80 WER
[162]	2019	Continues Isolated	Modified LSTM	72.3% 89%
[167]	2019	Isolated	DCNN based Dense NET	90.3 %
[168]	2019	Static	DCNN	97.71%
[176]	2020	Static	DCNN	90%
[181]	2020	Static	DCNN	97.6%
[184]	2020	Static	Eight CNN layers+ stochastic pooling, batch normalization and dropout	89.32 %
[185]	2020	Isolated	Cascaded model (SSD, CNN, LSTM)	98.42 %
[187]	2020	Static	Deep Elman recurrent neural network	98.89 %
[188]	2020	Static	DCNN	93%
[190]	2020	Static	Enhanced Alex Net	89.48%
[198]	2020	Static	Multimodality fine-tuned VGG16 CNN+ Leap Motion network	82.55%
[199]	2020	Continues	Multi-channel CNN	10.8 WER
[200]	2020	Static	Hybrid model based on the Inception v3+ SVM	99.90%
[201]	2020	Static	11 Layer CNN	95%
[205]	2021	Static	Three-layered CNN model	90.8%
[206]	2021	Isolated	Hybrid deep learning with convolutional (LSTM)+ and BiLSTM.	76.21%
[209]	2021	Isolated	DCNN+ Sentiment analysis	99.63%
[211]	2021	Continues	GRU+LSTM	19.56 error
[214]	2021	Isolated	Generic temporal convolutional network	77.42%
[215]	2021	Static	DCNN	96.65%
[216]	2021	Static	DCNN	99.7%
[220]	2021	Static	Pretrained InceptionV3+ Mini-batch gradient descent optimizer	85%
[221]	2021	Static	Apply the PSO algorithm to find the optimal parameters of the convolutional neural networks	99.58%
[223]	2021	Continues	Visual hierarchy to lexical sequence alignment network H2SNet	91.72%
[227]	2021	Static	Novel lightweight deep learning model based on bottleneck motivated from deep residual learning	99.52%
[228]	2021	Continues	Novel hyperparameter based optimized Generative Adversarial Networks (H-GANs)	97%
[229]	2021	Isolated	3DCNN	88.24%
[232]	2021	Continues	Bidirectional encoder representations from transformers (BERT) + ResNet	23.30 WER
[234]	2021	Continues	Generative Adversarial Network (SLRGAN)	23.4 WER
[238]	2021	Static	DCNN	97%
[239]	2022	Static	Optimized DCNN hybridization of Electric Fish Optimization (EFO), and Whale Optimization Algorithm (WOA) called Electric Fish based Whale Optimization Algorithm (E-WOA).	98.7%
[241]	2022	Isolated	CNN+ RNN	98.8%
[242]	2022	Static	Modified CapsNet architecture, (SLR-CapsNet)	99.60%
[245]	2022	Static	DCNN	99.52%
[247]	2022	Static	DCNN+ diffGrad optimizer	88.01%
[250]	2022	Static	DCNN	92%

[251]	2022	Static	DCNN	99.38%
[252]	2022	Static	Lightweight CNN	94.30%
[254]	2022	Isolated	Hybrid model based on VGG16-BiLSTM	83.36%

6.5.Related Works in Time and Delay Problem

In real-world classification scenarios using DL, time and delay is a principal factor to consider. It is important to strike a balance between achieving accurate classification results and minimizing the time required. The specific requirements and constraints of the application, such as the desired response time or the available computational resources, should be considered when designing and deploying DL models. As a result, one of the major requirements that make the recognition system of sign language efficient is the recognition time. Table 10 illustrates the related DL works for sign language recognition that focus on improving the recognition time.

Table 10: Related works on SLR using DL that aim to minimize the required time.

Discussion

Designing systems for recognizing sign language has become an emerging need in society and attracted the attention of academics and practitioners, due to its significant role in eliminating the communication barriers between the hearing and deaf communities. However, many challenges appeared when trying to design a sign language recognition system such as the dynamic gestures, environmental conditions, the availability of public datasets, and the multi-dimensional feature vectors. Still, many researchers are attempting to develop accurate, generalized, reliable, and robust sign language recognition models using deep learning. Deep learning technology is widely applied in many fields and research areas such as speech recognition, image processing, graphs, medicine, computer vision. With the emergence of DL approaches, sign language recognition has managed to significantly improve its accuracy. From the previous tables that illustrate some promising related works on sign language recognition using DL architectures, it is noticed that the most widely utilized deep architecture is CNN. Convolutional Neural Networks (CNNs) exhibit a remarkable capacity to extract discriminative features from raw data, enabling them to achieve impressive results in several types of sign language recognition tasks. They demonstrate robustness and flexibility, being employed either independently or in combination with other architectures, such as Long Short-Term Memory (LSTM), to enhance performance in sign language recognition. Moreover, CNNs prove to be highly advantageous in handling multi-modality data, such as RGB-D data, skeleton information, and finger points. These modalities provide rich information about the signer's actions, and their utilization has been instrumental in enhancing and addressing multiple challenges in sign language recognition. A set of related works focuses on solving only one type of problem facing the sign language recognition using DL such as in [132, 137, 139, 141, 147, 148, 152, 153, 154, 160, 169, 177, 195, 198, 205, 208, 212, 218, 220, 231, 235, 244, 247, 250, 252, 257, 258, 266], while others trying to solve multiple problems such as in [185, 199]. The most widely used feature is the spatiotemporal, that depends on the hand shape, and the location information of the hand [135, 143, 156, 161, 165, 180, 182, 189, 76, 210, 225, 226, 230, 234, 237, 244, 253, 264, 265]. However, there are works that make use of more than one type of features in addition to spatiotemporal such as facial expression, skeleton, orientation of hand and angles [138, 141, 144, 151, 152, 79, 159, 162, 164, 166, 170, 171, 175, 185, 186, 191, 192, 197, 199, 203, 204, 208, 212, 228, 231, 232, 233, 245, 246, 255, 258, 260, 261, 268, 269]. Some works apply separate feature extraction techniques rather than depending only on the DL extracted features and managed to

obtain recognition results [149, 152, 153, 79, 159, 162, 166, 169, 171, 175, 177, 179, 187, 189, 191, 192, 197, 199, 203, 204, 208, 228, 231, 233, 235, 237, 245, 246, 255, 258, 260, 261, 265, 268, 269]. Recent works especially from 2020 onwards focus on developing a recognition system for continuous sentences in sign language, which is still an open problem that gathers the most attention and is not completely solved or employed in any commercial application. Two factors that may contribute to an improved accuracy of continuous sign language recognition including feature extraction from frame sequences of the entered video and coordination between the features of every segment in the video and its corresponding sign label. Acquiring features from video frames that are more descriptive and discriminative resulted in better performance. While recent models in continuous sign language recognition have an uptrend in model performance using DL abilities in computer vision and Natural Language Processing (NLP), there is still much space for performance enhancement in this area. One of the main problems that many researchers deal with is the trajectory [186, 192, 204, 210, 260], and occlusion [129, 173, 76, 226, 228, 233, 237, 239, 259]. Furthermore, selecting or designing the appropriate deep model is one of the main challenges that have been addressed to deal with a particular type of challenges in sign language recognition by a variety of research in order to reach the desired accuracy goal. Others focus on solving some classification problems which is the overfitting that leads to the failure of the system. Applying any recognition system on more than one dataset with different properties is significant (high generalization), and one of the major factors that make the system highly effective. Thus, many researchers focus on implementing the recognition system of sign language on more than one dataset with a lot of variation and do not achieve the same results as in [129, 136, 143, 146, 156, 161, 164, 170, 182, 186, 204, 228, 234, 237, 254, 266]. Consequently, based on the information gathered from the preceding tables, deep learning stands out as a potent approach that has achieved the most impressive outcomes in sign language recognition. However, it's important to note that no existing research has successfully tackled all the associated challenges comprehensively. Some studies prioritize achieving high accuracy without considering time constraints, while others concentrate on addressing feature extraction issues and functioning in various environmental conditions. Yet, there's a lack of consideration for the complexity and overall applicability of the model. In addition, a significant aspect not extensively discussed in the related works pertains to hardware cost and complexity, both of which exert a substantial impact on the efficiency of the recognition system, particularly in real-world applications.

7. Conclusions and Future Work

Sign language recognition goes a long way from its first start in recognizing alphabets and digits reaching up to words and sentences. Recent systems of sign language recognition have a degree of success in dealing with dynamic signs that are based on hand and body motion, obtained from vision, or hardware devices. The use of DL for sign language recognition has improved the system performance into a higher level and confirmed its effectiveness in recognizing the signs of different forms including letters, word, and sentences, which are captured using different devices and convert them into another form such as text and sound. In this paper, related works on the use of DL for sign language recognition from the year of 2018 to 2022 have been reviewed, and a conclusion is possessed that DL reach to the desired performance, within high result in many aspects. Nevertheless, there remains room for further improvement to develop a comprehensive system capable of effectively handling all challenges encountered in sign language recognition. The goal is to achieve accurate and rapid results across various environmental conditions while utilizing diverse datasets. As future work, the primary objective is to address the issue of generalization and minimize the time needed for sign language recognition. Our objective is to present a deep learning model that can provide precise and highly accurate recognition outcomes for various types of sign language, encompassing both static and dynamic ones in different languages including English,

Arabic, Malay, and Chinese. Notably, this model aims to achieve these outcomes while minimizing hardware expenses and the required training time with high recognition accuracy.

Declaration of Competing Interest: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments: Not Applicable

Ethical approval: Not Applicable.

Consent to participate: The authors provide the appropriate consent to participate.

Consent for publication: The authors provide the consent to publish the images in the manuscript. The data used in the publication is publicly available. We provide respective citations for each of the data sources.

Code availability: Not Applicable.

Funding Statement: The authors received no specific funding for this study.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

- [1] Cheok, Ming Jin, Zaid Omar, and Mohamed Hisham Jaward. "A review of hand gesture and sign language recognition techniques." *International Journal of Machine Learning and Cybernetics* 10 (2019): 131-153.
- [2] World Federation of the deaf. Rome, Italy. Retrieved from <http://wfdeaf.org/our-work/>. (Accessed 18 January 2023).
- [3] Abd Al-Latief, Shahad Thamear, Salman Yussof, Azhana Ahmad, Saif Mohanad Khadim, and Raed Abdulkareem Abdulhasan. "Instant Sign Language Recognition by WAR Strategy Algorithm Based Tuned Machine Learning." *International Journal of Networked and Distributed Computing* (2024): 1-18.
- [4] Druzhkov, P. N., and V. D. Kustikova. "A survey of deep learning methods and software tools for image classification and object detection." *Pattern Recognition and Image Analysis* 26 (2016): 9-15.
- [5] Wu, Di, Nabin Sharma, and Michael Blumenstein. "Recent advances in video-based human action recognition using deep learning: A review." In *2017 International Joint Conference on Neural Networks (IJCNN)*, pp. 2865-2872. IEEE, 2017.
- [6] Hussain, Soeb, Rupal Saxena, Xie Han, Jameel Ahmed Khan, and Hyunchul Shin. "Hand gesture recognition using deep learning." In *2017 International SoC design conference (ISOCC)*, pp. 48-49. IEEE, 2017.
- [7] Alexiadis, Dimitrios S., Anargyros Chatzitofis, Nikolaos Zioulis, Olga Zoidi, Georgios Louizis, Dimitrios Zarpalas, and Petros Daras. "An integrated platform for live 3D human reconstruction and motion capturing." *IEEE Transactions on Circuits and Systems for Video Technology* 27, no. 4 (2016): 798-813.
- [8] Adaloglou, Nikolas, Theocharis Chatzis, Ilias Papastratis, Andreas Stergioulas, Georgios Th Papadopoulos, Vassia Zacharopoulou, George J. Xydopoulos, Klimnis Atzakas, Dimitris Papazachariou, and Petros Daras. "A comprehensive study on deep learning-based methods for sign language recognition." *IEEE Transactions on Multimedia* 24 (2021): 1750-1762.
- [9] Subburaj, S., and S. Murugavalli. "Survey on sign language recognition in context of vision-based and deep learning." *Measurement: Sensors* 23 (2022): 100385.
- [10] Mandel, M. (1977). *Iconic devices in American sign language. On the other hand, New perspectives on American sign language.*
- [11] Sandler, Wendy, and Diane Lillo-Martin. *Sign language and linguistic universals.* Cambridge University Press, 2006.
- [12] Goldin-Meadow, Susan, and Diane Brentari. *"Gesture, sign, and language: The coming of age of sign*

- language and gesture studies." *Behavioral and brain sciences* 40 (2017): e46.
- [13] Ong, Sylvie CW, and Surendra Ranganath. "Automatic sign language analysis: A survey and the future beyond lexical meaning." *IEEE Transactions on Pattern Analysis & Machine Intelligence* 27, no. 06 (2005): 873-891.
 - [14] Joudaki, Saba, Dzulkifli bin Mohamad, Tanzila Saba, Amjad Rehman, Mznah Al-Rodhaan, and Abdullah Al-Dhelaan. "Vision-based sign language classification: a directional review." *IETE Technical Review* 31, no. 5 (2014): 383-391.
 - [15] Sharma, Sakshi, and Sukhwinder Singh. "Vision-based sign language recognition system: A Comprehensive Review." In *2020 international conference on inventive computation technologies (ICICT)*, pp. 140-144. IEEE, 2020.
 - [16] Pansare, Jayshree R., and Maya Ingle. "Vision-based approach for American sign language recognition using edge orientation histogram." In *2016 international conference on image, vision and computing (ICIVC)*, pp. 86-90. IEEE, 2016.
 - [17] Aran, Oya. "Vision based sign language recognition: modeling and recognizing isolated signs with manual and non-manual components." Bogazi» ci University (2008).
 - [18] Al-Qurishi, Muhammad, Thariq Khalid, and Riad Souissi. "Deep learning for sign language recognition: Current techniques, benchmarks, and open issues." *IEEE Access* 9 (2021): 126917-126951.
 - [19] Li, Kehuang, Zhengyu Zhou, and Chin-Hui Lee. "Sign transition modeling and a scalable solution to continuous sign language recognition for real-world applications." *ACM Transactions on Accessible Computing (TACCESS)* 8, no. 2 (2016): 1-23.
 - [20] Rosero-Montalvo, Paul D., Pamela Godoy-Trujillo, Edison Flores-Bosmediano, Jorge Carrascal-Garcia, Santiago Otero-Potosi, Henry Benitez-Pereira, and Diego H. Peluffo-Ordenez. "Sign language recognition based on intelligent glove using machine learning techniques." In *2018 IEEE Third Ecuador Technical Chapters Meeting (ETCM)*, pp. 1-5. IEEE, 2018.
 - [21] Kudrinko, Karly, Emile Flavin, Xiaodan Zhu, and Qingguo Li. "Wearable sensor-based sign language recognition: A comprehensive review." *IEEE Reviews in Biomedical Engineering* 14 (2020): 82-97.
 - [22] Li, Shao-Zi, Bin Yu, Wei Wu, Song-Zhi Su, and Rong-Rong Ji. "Feature learning based on SAE-PCA network for human gesture recognition in RGBD images." *Neurocomputing* 151 (2015): 565-573.
 - [23] Amin, Muhammad Saad, Syed Tahir Hussain Rizvi, and Md Murad Hossain. "A Comparative Review on Applications of Different Sensors for Sign Language Recognition." *Journal of Imaging* 8, no. 4 (2022): 98.
 - [24] Theodorakis, Stavros, Vassilis Pitsikalis, and Petros Maragos. "Dynamic-static unsupervised sequentiality, statistical subunits and lexicon for sign language recognition." *Image and Vision Computing* 32, no. 8 (2014): 533-549.
 - [25] Plouffe, Guillaume, and Ana-Maria Cretu. "Static and dynamic hand gesture recognition in depth data using dynamic time warping." *IEEE transactions on instrumentation and measurement* 65, no. 2 (2015): 305-316.
 - [26] Agrawal, Subhash Chand, Anand Singh Jalal, and Rajesh Kumar Tripathi. "A survey on manual and non-manual sign language recognition for isolated and continuous sign." *International Journal of Applied Pattern Recognition* 3, no. 2 (2016): 99-134.
 - [27] El-Alfy, El-Sayed M., and Hamzah Luqman. "A comprehensive survey and taxonomy of sign language research." *Engineering Applications of Artificial Intelligence* 114 (2022): 105198.
 - [28] Dong, Shi, Ping Wang, and Khushnood Abbas. "A survey on deep learning and its applications." *Computer Science Review* 40 (2021): 100379.
 - [29] Najafabadi, Maryam M., Flavio Villanustre, Taghi M. Khoshgoftaar, Naeem Seliya, Randall Wald, and Edin Muharemagic. "Deep learning applications and challenges in big data analytics." *Journal of big data* 2, no. 1 (2015): 1-21.
 - [30] LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton. "Deep learning." *nature* 521, no. 7553 (2015): 436-444.
 - [31] Sarker, Iqbal H. "Deep learning: a comprehensive overview on techniques, taxonomy, applications and research directions." *SN Computer Science* 2, no. 6 (2021): 420.
 - [32] Rastgoo, Raziieh, Kourosh Kiani, Sergio Escalera, and Mohammad Sabokrou. "Sign language production: A review." In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 3451-3461. 2021.
 - [33] Yadav, Ashima, and Dinesh Kumar Vishwakarma. "Sentiment analysis using deep learning architectures: a review." *Artificial Intelligence Review* 53, no. 6 (2020): 4335-4385.
 - [34] Abdulhasan, Raed Abdulkareem, Shahad Thamear Abd Al-latief, and Saif Mohanad Kadhim. "Instant learning based on deep neural network with linear discriminant analysis features extraction for accurate iris recognition system." *Multimedia Tools and Applications* 83, no. 11 (2024): 32099-32122.

- [35] Madhiarasan, Dr M., Prof Roy, and Partha Pratim. "A Comprehensive Review of Sign Language Recognition: Different Types, Modalities, and Datasets." arXiv preprint arXiv:2204.03328 (2022).
- [36] Yang, Hee-Deok, and Seong-Whan Lee. "Robust sign language recognition by combining manual and non-manual features based on conditional random field and support vector machine." *Pattern Recognition Letters* 34, no. 16 (2013): 2051-2056.
- [37] Chen, Feng-Sheng, Chih-Ming Fu, and Chung-Lin Huang. "Hand gesture recognition using a real-time tracking method and hidden Markov models." *Image and vision computing* 21, no. 8 (2003): 745-758.
- [38] Ibrahim, Nada B., Hala H. Zayed, and Mazen M. Selim. "Advances, challenges and opportunities in continuous sign language recognition." *J. Eng. Appl. Sci* 15, no. 5 (2020): 1205-1227.
- [39] Smith, Paul, Niels da Vitoria Lobo, and Mubarak Shah. "Resolving hand over face occlusion." *Image and Vision Computing* 25, no. 9 (2007): 1432-1448.
- [40] Yang, Ruiduo, Sudeep Sarkar, and Barbara Loeding. "Handling movement epenthesis and hand segmentation ambiguities in continuous sign language recognition using nested dynamic programming." *IEEE transactions on pattern analysis and machine intelligence* 32, no. 3 (2009): 462-477.
- [41] Zhang, Hui, Jason E. Fritts, and Sally A. Goldman. "Image segmentation evaluation: A survey of unsupervised methods." *computer vision and image understanding* 110, no. 2 (2008): 260-280.
- [42] Cai, Shanshan, and Desheng Liu. "A comparison of object-based and contextual pixel-based classifications using high and medium spatial resolution images." *Remote sensing letters* 4, no. 10 (2013): 998-1007.
- [43] Kausar, Sumaira, and M. Younus Javed. "A survey on sign language recognition." In *2011 Frontiers of Information Technology*, pp. 95-98. IEEE, 2011.
- [44] Aloysius, Neena, and M. Geetha. "Understanding vision-based continuous sign language recognition." *Multimedia Tools and Applications* 79, no. 31-32 (2020): 22177-22209.
- [45] <https://www.kaggle.com/grassknotted/asl-alphabet>
- [46] <https://www.kaggle.com/datasets/datamunge/sign-language-mnist>
- [47] Pugeault, Nicolas, and Richard Bowden. "Spelling it out: Real-time ASL fingerspelling recognition." In *2011 IEEE International conference on computer vision workshops (ICCV workshops)*, pp. 1114-1119. IEEE, 2011.
- [48] Tompson, Jonathan, Murphy Stein, Yann Lecun, and Ken Perlin. "Real-time continuous pose recovery of human hands using convolutional networks." *ACM Transactions on Graphics (ToG)* 33, no. 5 (2014): 1-10.
- [49] Ong, Eng-Jon, Helen Cooper, Nicolas Pugeault, and Richard Bowden. "Sign language recognition using sequential pattern trees." In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2200-2207. IEEE, 2012.
- [50] Triesch, Jochen, and Christoph Von Der Malsburg. "Robust classification of hand postures against complex backgrounds." In *Proceedings of the second international conference on automatic face and gesture recognition*, pp. 170-175. IEEE, 1996.
- [51] Marin, Giulio, Fabio Dominio, and Pietro Zanuttigh. "Hand gesture recognition with leap motion and kinect devices." In *2014 IEEE International conference on image processing (ICIP)*, pp. 1565-1569. IEEE, 2014.
- [52] Ren, Zhou, Junsong Yuan, and Zhengyou Zhang. "Robust hand gesture recognition based on finger-earth mover's distance with a commodity depth camera." In *Proceedings of the 19th ACM international conference on Multimedia*, pp. 1093-1096. 2011.
- [53] Feng, Bin, Fangzi He, Xinggang Wang, Yongjiang Wu, Hao Wang, Sihua Yi, and Wenyu Liu. "Depth-projection-map-based bag of contour fragments for robust hand gesture recognition." *IEEE Transactions on Human-Machine Systems* 47, no. 4 (2016): 511-523.
- [54] Wilbur, Ronnie, and Avinash C. Kak. "Purdue RVL-SLLL American sign language database." (2006).
- [55] Shi, Bowen, Aurora Martinez Del Rio, Jonathan Keane, Jonathan Michaux, Diane Brentari, Greg Shakhnarovich, and Karen Livescu. "American sign language fingerspelling recognition in the wild." In *2018 IEEE Spoken Language Technology Workshop (SLT)*, pp. 145-152. IEEE, 2018.
- [56] Othman, Achraf, Zouhour Tmar, and Mohamed Jemni. "Toward developing a very big sign language parallel corpus." In *Computers Helping People with Special Needs: 13th International Conference, ICCHP 2012, Linz, Austria, July 11-13, 2012, Proceedings, Part II* 13, pp. 192-199. Springer Berlin Heidelberg, 2012.
- [57] Neidle, Carol, and Augustine Opoku. *A User's Guide to the American Sign Language Linguistic Research Project (ASLLRP) Data Access Interface (DAI) 2—Version 2*. American Sign Language Linguistic Research Project Report No. 18, Boston University. No. 18. Linguistic Research Project Report, 2020.
- [58] Barczak, A. L. C., N. H. Reyes, M. Abastillas, A. Piccio, and Teo Susnjak. "A new 2D static hand gesture colour image dataset for ASL gestures." (2011).
- [59] http://vlm1.uta.edu/~srujana/ASLID/ASL_Image_Dataset.html

- [60] <https://ieee-dataport.org/documents/ksu-arasl-arabic-sign-language>
- [61] Sidig, Ala Addin I., Hamzah Luqman, Sabri Mahmoud, and Mohamed Mohandes. "KArSL: Arabic sign language database." *ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP)* 20, no. 1 (2021): 1-19.
- [62] Shanableh, Tamer, Khaled Assaleh, and Mohammad Al-Rousan. "Spatio-temporal feature-extraction techniques for isolated gesture recognition in Arabic sign language." *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)* 37, no. 3 (2007): 641-650.
- [63] <https://www.idiap.ch/webarchives/sites/www.idiap.ch/resource/gestures/>
- [64] https://github.com/DeepKothadiya/Custom_ISLDataset/tree/main
- [65] Forster, Jens, Christoph Schmidt, Oscar Koller, Martin Bellgardt, and Hermann Ney. "Extensions of the Sign Language Recognition and Translation Corpus RWTH-PHOENIX-Weather." In *LREC*, pp. 1911-1916. 2014.
- [66] Agris, Ulrich von, and Karl-Friedrich Kraiss. "Signum database: Video corpus for signer-independent continuous sign language recognition." In *sign-lang@ LREC 2010*, pp. 243-246. European Language Resources Association (ELRA), 2010.
- [67] Chai, Xiujuan, Guang Li, Yushun Lin, Zhihao Xu, Yili Tang, Xilin Chen, and Ming Zhou. "Sign language recognition and translation with kinect." In *IEEE conf. on AFGR*, vol. 655, p. 4. 2013.
- [68] <https://paperswithcode.com/dataset/csl-daily>
- [69] <http://home.ustc.edu.cn/~pjh/openresources/cslr-dataset-2015/index.html>
- [70] Rafi, A.M.; Nawal, N.; Bayev, N.S.; Nima, L.; Shahnaz, C.; Fattah, S.A. Image-based bengali sign language alphabet recognition for deaf and dumb community. In *Proceedings of the 2019 IEEE Global Humanitarian Technology Conference (GHTC)*, Seattle, WA, USA, 17–20 October 2019; pp. 1–7
- [71] Islam, Md Sanzidul, Sadia Sultana Sharmin Mousumi, Nazmul A. Jessan, AKM Shahariar Azad Rabby, and Sayed Akhter Hossain. "Ishara-lipi: The first complete multipurposeopen access dataset of isolated characters for bangla sign language." In *2018 International Conference on Bangla Speech and Language Processing (ICBSLP)*, pp. 1-4. IEEE, 2018.
- [72] Asadi-Aghbolaghi, Maryam, Hugo Bertiche, Vicent Roig, Shohreh Kasaei, and Sergio Escalera. "Action recognition from RGB-D data: Comparison and fusion of spatio-temporal handcrafted features and deep strategies." In *Proceedings of the IEEE International conference on computer vision workshops*, pp. 3179-3188. 2017.
- [73] Escalera S, Gonzalez J, Baro X, Reyes M, Lopes O, Guyon I, Athitsos V, Escalante H (2013) Multi-modal gesture recognition challenge 2013: dataset and results, In *Proceedings of the 15th ACM on International conference on multimodal interaction*, 445–452
- [74] Cerna, Lourdes Ramirez, Edwin Escobedo Cardenas, Dayse Garcia Miranda, David Menotti, and Guillermo Camara-Chavez. "A multimodal LIBRAS-UFOP Brazilian sign language dataset of minimal pairs using a microsoft Kinect sensor." *Expert Systems with Applications* 167 (2021): 114179.]
- [75] Sincan, Ozge Mercanoglu, and Hacer Yalim Keles. "Autsl: A large scale multi-modal turkish sign language dataset and baseline methods." *IEEE Access* 8 (2020): 181340-181355.
- [76] Rastgoo, Razieh, Kourosh Kiani, and Sergio Escalera. "Hand sign language recognition using multi-view hand skeleton." *Expert Systems with Applications* 150 (2020): 113336.
- [77] Ronchetti, Franco, Facundo Quiroga, César Armando Estrebow, Laura Cristina Lanzarini, and Alejandro Rosete. "LSA64: an Argentinian sign language dataset." In *XXII Congreso Argentino de Ciencias de la Computación (CACIC 2016)*. 2016.
- [78] Efthimiou, Eleni, Kiki Vasilaki, Stavroula-Evita Fotinea, Anna Vacalopoulou, Theodoros Goulas, and Athanasia-Lida Dimou. "The POLYTROPON parallel corpus." In *sign-lang@ LREC 2018*, pp. 39-44. European Language Resources Association (ELRA), 2018.
- [79] Ko, Sang-Ki, Chang Jo Kim, Hyedong Jung, and Choongsang Cho. "Neural sign language translation based on human keypoint estimation." *Applied sciences* 9, no. 13 (2019): 2683.
- [80] Luqman, Hamzah, and Sabri A. Mahmoud. "A machine translation system from Arabic sign language to Arabic." *Universal Access in the Information Society* 19, no. 4 (2020): 891-904.
- [81] Ruffieux, Simon, Denis Lalanne, Elena Mugellini, and Omar Abou Khaled. "A survey of datasets for human gesture recognition." In *Human-Computer Interaction. Advanced Interaction Modalities and Techniques: 16th International Conference, HCI International 2014, Heraklion, Crete, Greece, June 22-27, 2014, Proceedings, Part II* 16, pp. 337-348. Springer International Publishing, 2014.
- [82] Boulahia, Said Yacine, Eric Anquetil, Franck Multon, and Richard Kulpa. "Dynamic hand gesture recognition based on 3D pattern assembled trajectories." In *2017 seventh international conference on image*

- processing theory, tools and applications (IPTA), pp. 1-6. IEEE, 2017.
- [83] Avola, Danilo, Marco Bernardi, Luigi Cinque, Gian Luca Foresti, and C. Massaroni. "Exploiting recurrent neural networks and leap motion controller for sign language and semaphoric gesture recognition." arXiv preprint arXiv:1803.10435.
 - [84] Chen, Chen, Roozbeh Jafari, and Nasser Kehtarnavaz. "UTD-MHAD: A multimodal dataset for human action recognition utilizing a depth camera and a wearable inertial sensor." In 2015 IEEE International conference on image processing (ICIP), pp. 168-172. IEEE, 2015.
 - [85] S. Singh, S.A. Velastin, H. Ragheb, Muhavi: A multicamera human action video dataset for the evaluation of action recognition methods, in 2010 7th IEEE International Conference on Advanced Video and Signal Based Surveillance, IEEE, 2010, pp. 48–55
 - [86] Zheng, Jingjing, Zhuolin Jiang, P. Jonathon Phillips, and Rama Chellappa. "Cross-View Action Recognition via a Transferable Dictionary Pair." In *bmvc*, vol. 1, no. 2, p. 7. 2012.
 - [87] L. Gorelick, M. Blank, E. Shechtman, M. Irani, R. Basri, Actions as space-time shapes, *IEEE Trans. Pattern Anal. Mach. Intell.* 29
 - [88] A. Shahroudy, J. Liu, T.-T. Ng, G. Wang, Ntu rgb+ d: A large scale dataset for 3d human activity analysis, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 1010–1019
 - [89] Kim, T-K.; Wong, S-F.; Cipolla, R.: Tensor canonical correlation analysis for action classification. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Minneapolis, MN (2007)
 - [90] Zhang, Yi, Chong Wang, Ye Zheng, Jieyu Zhao, Yuqi Li, and Xijiong Xie. "Short-term temporal convolutional networks for dynamic hand gesture recognition." arXiv preprint arXiv:2001.05833 (2019).
 - [91] Wang J, Liu Z, Wu Y, Yuan J (2012) Mining actionlet ensemble for action recognition with depth cameras, In *Computer Vision and Pattern Recognition (CVPR)*, 2012 IEEE Conference on, pp. 1290–1297
 - [92] Koppula, Hema Swetha, Rudhir Gupta, and Ashutosh Saxena. "Learning human activities and object affordances from rgb-d videos." *The International journal of robotics research* 32, no. 8 (2013): 951-970.
 - [93] Müller, Meinard, Tido Röder, Michael Clausen, Bernhard Eberhardt, Björn Krüger, and Andreas Weber. "Mocap database hdm05." *Institut für Informatik II, Universität Bonn* 2, no. 7 (2007).]
 - [94] Gross, Ralph, and Jianbo Shi. "The cmu motion of body (mobo) database. Robotics Institute." Pittsburgh, PA (2001).
 - [95] Wan J et al. (2016) ChaLearn looking at people RGB-D isolated and continuous datasets for gesture recognition, *IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Las Vegas, NV, USA
 - [96] Gupta, P. M. X. Y. S., and K. K. S. T. J. Kautz. "Online detection and classification of dynamic hand gestures with recurrent 3d convolutional neural networks." In *CVPR*, vol. 1, no. 2, p. 3. 2016.
 - [97] Bloom, Victoria, Dimitrios Makris, and Vasileios Argyriou. "G3D: A gaming action dataset and real time action recognition evaluation framework." In 2012 IEEE Computer society conference on computer vision and pattern recognition workshops, pp. 7-12. IEEE, 2012.
 - [98] Xia, Lu, Chia-Chih Chen, and Jake K. Aggarwal. "View invariant human action recognition using histograms of 3d joints." In 2012 IEEE computer society conference on computer vision and pattern recognition workshops, pp. 20-27. IEEE, 2012.
 - [99] Garcia-Hernando, Guillermo, Shanxin Yuan, Seungryul Baek, and Tae-Kyun Kim. "First-person hand action benchmark with rgb-d videos and 3d hand pose annotations." In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 409-419. 2018.
 - [100] Materzynska, Joanna, Guillaume Berger, Ingo Bax, and Roland Memisevic. "The jester dataset: A large-scale video dataset of human gestures." In *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*, pp. 0-0. 2019.
 - [101] Zhang, Yifan, Congqi Cao, Jian Cheng, and Hanqing Lu. "Egogesture: a new dataset and benchmark for egocentric hand gesture recognition." *IEEE Transactions on Multimedia* 20, no. 5 (2018): 1038-1050.
 - [102] Pisharady, Pramod Kumar, Prahlad Vadakkepat, and Ai Poh Loh. "Attention based detection and recognition of hand postures against complex backgrounds." *International Journal of Computer Vision* 101 (2013): 403-419
 - [103] Papineni, Kishore, Salim Roukos, Todd Ward, and Wei-Jing Zhu. "Bleu: a method for automatic evaluation of machine translation." In *Proceedings of the 40th annual meeting of the Association for Computational Linguistics*, pp. 311-318. 2002.
 - [104] Mann, Wolfgang, Chloe R. Marshall, Kathryn Mason, and Gary Morgan. "The acquisition of sign language: The impact of phonetic complexity on phonology." *Language Learning and Development* 6, no. 1 (2010): 60-86.

- [105] Padden, Carol, Irit Meir, Mark Aronoff, and Wendy Sandler. The grammar of space in two new sign languages. na, 2010.
- [106] Lillo-Martin, Diane, and Richard P. Meier. "On the linguistic status of 'agreement' in sign languages." (2011): 95-141.
- [107] Binder, Marc D., Nobutaka Hirokawa, and Uwe Windhorst, eds. Encyclopedia of neuroscience. Vol. 3166. Berlin, Germany: Springer, 2009.
- [108] Chen, Xiang, Xu Zhang, Zhang-Yan Zhao, Ji-Hai Yang, Vuokko Lantz, and Kong-Qiao Wang. "Hand gesture recognition research based on surface EMG sensors and 2D-accelerometers." In 2007 11th IEEE International Symposium on Wearable Computers, pp. 11-14. IEEE, 2007.
- [109] Li, Wenguo, Zhizeng Luo, and Xugang Xi. "Movement trajectory recognition of sign language based on optimized dynamic time warping." Electronics 9, no. 9 (2020): 1400.
- [110] Mino, Ajkel, Mirela Popa, and Alexia Briassouli. "The Effect of Spatial and Temporal Occlusion on Word Level Sign Language Recognition." In 2022 IEEE International Conference on Image Processing (ICIP), pp. 2686-2690. IEEE, 2022.
- [111] Aran, Oya. "Vision based sign language recognition: modeling and recognizing isolated signs with manual and non-manual components." Bogaziçi University (2008).
- [112] KaewTraKulPong, Pakorn, and Richard Bowden. "An improved adaptive background mixture model for real-time tracking with shadow detection." Video-based surveillance systems: Computer vision and distributed processing (2002): 135-144.
- [113] Kakumanu, Praveen, Sokratis Makrogiannis, and Nikolaos Bourbakis. "A survey of skin-color modeling and detection methods." Pattern recognition 40, no. 3 (2007): 1106-1122.
- [114] Yun, Liu, Zhang Lifeng, and Zhang Shujun. "A hand gesture recognition method based on multi-feature fusion and template matching." Procedia Engineering 29 (2012): 1678-1684.
- [115] Kartika, Dyah Rahma, and Riyanto Sigit. "Sign language interpreter hand using optical flow." In 2016 International Seminar on Application for Technology of Information and Communication (ISemantic), pp. 197-201. IEEE, 2016.
- [116] Neverova, Natalia, Christian Wolf, Graham W. Taylor, and Florian Nebout. "Hand segmentation with structured convolutional learning." In Computer Vision--ACCV 2014: 12th Asian Conference on Computer Vision, Singapore, Singapore, November 1-5, 2014, Revised Selected Papers, Part III 12, pp. 687-702. Springer International Publishing, 2015.
- [117] Tyagi, Akansha, and Sandhya Bansal. "Feature extraction technique for vision-based indian sign language recognition system: A review." Computational Methods and Data Engineering: Proceedings of ICMDE 2020, Volume 1 (2020): 39-53.
- [118] Shanableh, Tamer, Khaled Assaleh, and Mohammad Al-Rousan. "Spatio-temporal feature-extraction techniques for isolated gesture recognition in Arabic sign language." IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics) 37, no. 3 (2007): 641-650.
- [119] Rice, Leslie, Eric Wong, and Zico Kolter. "Overfitting in adversarially robust deep learning." In International Conference on Machine Learning, pp. 8093-8104. PMLR, 2020.
- [120] Ying, Xue. "An overview of overfitting and its solutions." In Journal of physics: Conference series, vol. 1168, p. 022022. IOP Publishing, 2019.
- [121] Bisong, Ekaba, and Ekaba Bisong. "Regularization for deep learning." Building Machine Learning and Deep Learning Models on Google Cloud Platform: A Comprehensive Guide for Beginners (2019): 415-421.
- [122] Srivastava, Nitish, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. "Dropout: a simple way to prevent neural networks from overfitting." The journal of machine learning research 15, no. 1 (2014): 1929-1958.
- [123] Caruana, Rich, Steve Lawrence, and C. Giles. "Overfitting in neural nets: Backpropagation, conjugate gradient, and early stopping." Advances in neural information processing systems 13 (2000).
- [124] Khosla, Cherry, and Baljit Singh Saini. "Enhancing performance of deep learning models with different data augmentation techniques: A survey." In 2020 International Conference on Intelligent Engineering and Management (ICIEM), pp. 79-85. IEEE, 2020.
- [125] Zhang, Chiyuan, Oriol Vinyals, Remi Munos, and Samy Bengio. "A study on overfitting in deep reinforcement learning." arXiv preprint arXiv:1804.06893 (2018).
- [126] Neyshabur, Behnam, Srinadh Bhojanapalli, David McAllester, and Nati Srebro. "Exploring generalization in deep learning." Advances in neural information processing systems 30 (2017).
- [127] Kawaguchi, Kenji, Leslie Pack Kaelbling, and Yoshua Bengio. "Generalization in deep learning." arXiv preprint arXiv:1710.05468 (2017).

- [128] Hu, Xia, Lingyang Chu, Jian Pei, Weiqing Liu, and Jiang Bian. "Model complexity of deep learning: A survey." *Knowledge and Information Systems* 63 (2021): 2585-2619.
- [129] Tao, Wenjin, Ming C. Leu, and Zhaozheng Yin. "American Sign Language alphabet recognition using Convolutional Neural Networks with multiview augmentation and inference fusion." *Engineering Applications of Artificial Intelligence* 76 (2018): 202-213.
- [130] Hossen, M. A., Arun Govindaiah, Sadia Sultana, and Alauddin Bhuiyan. "Bengali sign language recognition using deep convolutional neural network." In *2018 joint 7th international conference on informatics, electronics & vision (iciev) and 2018 2nd international conference on imaging, vision & pattern recognition (icIVPR)*, pp. 369-373. IEEE, 2018.
- [131] Lazo, Cristian, Zaid Sanchez, and Christian del Carpio. "A Static Hand Gesture Recognition for Peruvian Sign Language Using Digital Image Processing and Deep Learning." In *Brazilian Technology Symposium*, pp. 281-290. Springer, Cham, 2018.
- [132] Islam, Sanzidul, Sadia Sultana Sharmin Mousumi, AKM Shahariar Azad Rabby, Sayed Akhter Hossain, and Sheikh Abujar. "A potent model to recognize bangla sign language digits using convolutional neural network." *Procedia computer science* 143 (2018): 611-618.
- [133] Bao, Peijun, Ana I. Maqueda, Carlos R. del-Blanco, and Narciso García. "Tiny hand gesture recognition without localization via a deep convolutional network." *IEEE Transactions on Consumer Electronics* 63, no. 3 (2017): 251-257.
- [134] Rastgoo, Razieh, Kourosh Kiani, and Sergio Escalera. "Multi-modal deep hand sign language recognition in still images using restricted Boltzmann machine." *Entropy* 20, no. 11 (2018): 809.
- [135] Amaral, Lucas, Givanildo LN Júnior, Tiago Vieira, and Thales Vieira. "Evaluating deep models for dynamic brazilian sign language recognition." In *Iberoamerican congress on pattern recognition*, pp. 930-937. Springer, Cham, 2018.
- [136] Li, Yuan, Xinggang Wang, Wenyu Liu, and Bin Feng. "Deep attention network for joint hand gesture localization and recognition using static RGB-D images." *Information Sciences* 441 (2018): 66-78.
- [137] Oyedotun, Oyebade K., and Adnan Khashman. "Deep learning in vision-based static hand gesture recognition." *Neural Computing and Applications* 28, no. 12 (2017): 3941-3951.
- [138] Ameen, Salem, and Sunil Vadera. "A convolutional neural network to classify American Sign Language fingerspelling from depth and colour images." *Expert Systems* 34, no. 3 (2017): e12197.
- [139] Bheda, Vivek, and Dianna Radpour. "Using deep convolutional networks for gesture recognition in american sign language." *arXiv preprint arXiv:1710.06836* (2017).
- [140] Ji, Yangho, Sunmok Kim, Young-Joo Kim, and Ki-Baek Lee. "Human-like sign-language learning method using deep learning." *ETRI Journal* 40, no. 4 (2018): 435-445.
- [141] Pu, Junfu, Wengang Zhou, and Houqiang Li. "Dilated convolutional network with iterative optimization for continuous sign language recognition." In *IJCAI*, vol. 3, p. 7. 2018.
- [142] Daroya, Rangel, Daryl Peralta, and Prospero Naval. "Alphabet sign language image classification using deep learning." In *TENCON 2018-2018 IEEE Region 10 Conference*, pp. 0646-0650. IEEE, 2018.
- [143] Huang, Jie, Wengang Zhou, Houqiang Li, and Weiping Li. "Attention-based 3D-CNNs for large-vocabulary sign language recognition." *IEEE Transactions on Circuits and Systems for Video Technology* 29, no. 9 (2018): 2822-2832.
- [144] Chong, Teak-Wei, and Boon-Giin Lee. "American sign language recognition using leap motion controller with machine learning approach." *Sensors* 18, no. 10 (2018): 3554.
- [145] Kumar, E. Kiran, P. V. V. Kishore, A. S. C. S. Sastry, M. Teja Kiran Kumar, and D. Anil Kumar. "Training CNNs for 3-D sign language recognition with color texture coded joint angular displacement maps." *IEEE Signal Processing Letters* 25, no. 5 (2018): 645-649.
- [146] Koller, Oscar, Sepehr Zargaran, Hermann Ney, and Richard Bowden. "Deep sign: Enabling robust statistical continuous sign language recognition via hybrid CNN-HMMs." *International Journal of Computer Vision* 126, no. 12 (2018): 1311-1325.
- [147] Taskiran, Murat, Mehmet Killioglu, and Nihan Kahraman. "A real-time system for recognition of American sign language by using deep learning." In *2018 41st international conference on telecommunications and signal processing (TSP)*, pp. 1-5. IEEE, 2018.
- [148] Shahriar, Shadman, Ashraf Siddiquee, Tanveerul Islam, Abesh Ghosh, Rajat Chakraborty, Asir Intisar Khan, Celia Shahnaz, and Shaikh Anowarul Fattah. "Real-time american sign language recognition using skin segmentation and image category classification with convolutional neural network and deep learning." In *TENCON 2018-2018 IEEE Region 10 Conference*, pp. 1168-1171. IEEE, 2018.
- [149] Hu, Yong, Hai-Feng Zhao, and Zhi-Gang Wang. "Sign language fingerspelling recognition using depth

- information and deep belief networks." *International Journal of Pattern Recognition and Artificial Intelligence* 32, no. 06 (2018): 1850018.
- [150] Kishore, P. V. V., G. Anantha Rao, E. Kiran Kumar, M. Teja Kiran Kumar, and D. Anil Kumar. "Selfie sign language recognition with convolutional neural networks." *International Journal of Intelligent Systems and Applications* 10, no. 10 (2018): 63.
 - [151] Ye, Yuancheng, Yingli Tian, Matt Huenerfauth, and Jingya Liu. "Recognizing american sign language gestures from within continuous videos." In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 2064-2073. 2018.
 - [152] Avola, Danilo, Marco Bernardi, Luigi Cinque, Gian Luca Foresti, and Cristiano Massaroni. "Exploiting recurrent neural networks and leap motion controller for the recognition of sign language and semaphoric hand gestures." *IEEE Transactions on Multimedia* 21, no. 1 (2018): 234-245.
 - [153] Ranga, Virender, Nikita Yadav, and Pulkit Garg. "American sign language fingerspelling using hybrid discrete wavelet transform-gabor filter and convolutional neural network." *Journal of Engineering Science and Technology* 13, no. 9 (2018): 2655-2669.
 - [154] Vega, AM Rincon, A. Vasquez, W. Amador, and A. Rojas. "Deep learning for the recognition of facial expression in the Colombian sign language." *Annals of Physical and Rehabilitation Medicine* 61 (2018): e96.
 - [155] Suri, Karush, and Rinki Gupta. "Continuous sign language recognition from wearable IMUs using deep capsule networks and game theory." *Computers & Electrical Engineering* 78 (2019): 493-503.
 - [156] Huang, Jie, Wengang Zhou, Qilin Zhang, Houqiang Li, and Weiping Li. "Video-based sign language recognition without temporal segmentation." In *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 32, no. 1. 2018.
 - [157] Tolentino, Lean Karlo S., RO Serfa Juan, August C. Thio-ac, Maria Abigail B. Pamahoy, Joni Rose R. Forteza, and Xavier Jet O. Garcia. "Static sign language recognition using deep learning." *Int. J. Mach. Learn. Comput* 9, no. 6 (2019): 821-827.
 - [158] Pinto, Raimundo F., Carlos DB Borges, Antônio Almeida, and Iális C. Paula. "Static hand gesture recognition based on convolutional neural networks." *Journal of Electrical and Computer Engineering* 2019 (2019).
 - [159] Aly, Walaa, Saleh Aly, and Sultan Almotairi. "User-independent American sign language alphabet recognition based on depth image and PCANet features." *IEEE Access* 7 (2019): 123138-123150.
 - [160] Joy, Jestin, Kannan Balakrishnan, and M. Sreeraj. "SignQuiz: a quiz-based tool for learning fingerspelled signs in indian sign language using ASLR." *IEEE Access* 7 (2019): 28363-28371.
 - [161] Cui, Runpeng, Hu Liu, and Changshui Zhang. "A deep neural framework for continuous sign language recognition by iterative training." *IEEE Transactions on Multimedia* 21, no. 7 (2019): 1880-1891.
 - [162] Mittal, Anshul, Pradeep Kumar, Partha Pratim Roy, Raman Balasubramanian, and Bidyut B. Chaudhuri. "A modified LSTM model for continuous sign language recognition using leap motion." *IEEE Sensors Journal* 19, no. 16 (2019): 7056-7063.
 - [163] Kulhandjian, Hovannes, Prakshi Sharma, Michel Kulhandjian, and Claude D'Amours. "Sign language gesture recognition using doppler radar and deep learning." In *2019 IEEE Globecom Workshops (GC Wkshps)*, pp. 1-6. IEEE, 2019.
 - [164] Zhang, Shujun, Weijia Meng, Hui Li, and Xuehong Cui. "Multimodal spatiotemporal networks for sign language recognition." *IEEE Access* 7 (2019): 180270-180280.
 - [165] Liao, Yanqiu, Pengwen Xiong, Weidong Min, Weiqiong Min, and Jiahao Lu. "Dynamic sign language recognition based on video sequence with BLSTM-3D residual networks." *IEEE Access* 7 (2019): 38044-38054.
 - [166] Vo, Anh H., Van-Huy Pham, and Bao T. Nguyen. "Deep learning for vietnamese sign language recognition in video sequence." *International Journal of Machine Learning and Computing* 9, no. 4 (2019): 440-445.
 - [167] Liang, Zhi-jie, Sheng-bin Liao, and Bing-zhang Hu. "3D convolutional neural networks for dynamic sign language recognition." *The Computer Journal* 61, no. 11 (2018): 1724-1736.
 - [168] Bhagat, Neel Kamal, Y. Vishnusai, and G. N. Rathna. "Indian sign language gesture recognition using image processing and deep learning." In *2019 Digital Image Computing: Techniques and Applications (DICTA)*, pp. 1-8. IEEE, 2019.
 - [169] Yu, Yi, Xiang Chen, Shuai Cao, Xu Zhang, and Xun Chen. "Exploration of Chinese sign language recognition using wearable sensors based on deep belief net." *IEEE journal of biomedical and health informatics* 24, no. 5 (2019): 1310-1320.
 - [170] Al-Hammadi, Muneer, Ghulam Muhammad, Wadood Abdul, Mansour Alsulaiman, and M. Shamim Hossain. "Hand gesture recognition using 3D-CNN model." *IEEE Consumer Electronics Magazine* 9, no. 1 (2019): 95-101.

- [171] Guo, Dan, Wengang Zhou, Anyang Li, Houqiang Li, and Meng Wang. "Hierarchical recurrent deep fusion using adaptive clip summarization for sign language translation." *IEEE Transactions on Image Processing* 29 (2019): 1575-1590.
- [172] Kasukurthi, Nikhil, Brij Rokad, Shiv Bidani, and Dr Dennisan. "American Sign Language Alphabet Recognition using Deep Learning." *arXiv preprint arXiv:1905.05487* (2019).
- [173] Ravi, Sunitha, Maloji Suman, P. V. V. Kishore, Kiran Kumar, and Anil Kumar. "Multi modal spatio temporal co-trained CNNs with single modal testing on RGB-D based sign language gesture recognition." *Journal of Computer Languages* 52 (2019): 88-102.
- [174] Ferreira, Pedro M., Diogo Pernes, Ana Rebelo, and Jaime S. Cardoso. "Desire: Deep signer-invariant representations for sign language recognition." *IEEE Transactions on Systems, Man, and Cybernetics: Systems* 51, no. 9 (2019): 5830-5845.
- [175] Mazhar, Osama, Benjamin Navarro, Sofiane Ramdani, Robin Passama, and Andrea Cherubini. "A real-time human-robot interaction framework with robust background invariant hand gesture detection." *Robotics and Computer-Integrated Manufacturing* 60 (2019): 34-48.
- [176] Kamruzzaman, M. M. "Arabic sign language recognition and generating Arabic speech using convolutional neural network." *Wireless Communications and Mobile Computing* 2020 (2020).
- [177] Angona, Tazkia Mim, ASM Siamuzzaman Shaon, Kazi Tahmid Rashad Niloy, Tajbia Karim, Zarin Tasnim, SM Salim Reza, and Tasmima Noushiba Mahbub. "Automated Bangla sign language translation system for alphabets by means of MobileNet." *TELKOMNIKA (Telecommunication Computing Electronics and Control)* 18, no. 3 (2020): 1292-1301.
- [178] Elsayed, Eman K., and Doaa R. Fathy. "Sign language semantic translation system using ontology and deep learning." *International Journal of Advanced Computer Science and Applications* 11, no. 1 (2020).
- [179] Aly, Saleh, and Walaa Aly. "DeepArSLR: A novel signer-independent deep learning framework for isolated arabic sign language gestures recognition." *IEEE Access* 8 (2020): 83199-83212.
- [180] Al-Hammadi, Muneer, Ghulam Muhammad, Wadood Abdul, Mansour Alsulaiman, Mohammed A. Bencherif, Tareq S. Alrayes, Hassan Mathkour, and Mohamed Amine Mekhtiche. "Deep learning-based approach for sign language gesture recognition with efficient hand gesture representation." *IEEE Access* 8 (2020): 192527-192542.
- [181] Latif, Ghazanfar, Nazeeruddin Mohammad, Roaa AlKhalaf, Rawan AlKhalaf, Jaafar Alghazo, and Majid Khan. "An automatic Arabic sign language recognition system based on deep CNN: an assistive system for the deaf and hard of hearing." *International Journal of Computing and Digital Systems* 9, no. 4 (2020): 715-724.
- [182] Al-Hammadi, Muneer, Ghulam Muhammad, Wadood Abdul, Mansour Alsulaiman, Mohamed A. Bencherif, and Mohamed Amine Mekhtiche. "Hand gesture recognition for sign language using 3DCNN." *IEEE Access* 8 (2020): 79491-79509.
- [183] Abdulhussein, Abdulwahab A., and Firas A. Raheem. "Hand gesture recognition of static letters american sign language (ASL) using deep learning." *Engineering and Technology Journal* 38, no. 6 (2020): 926-937.
- [184] Jiang, Xianwei, Mingzhou Lu, and Shui-Hua Wang. "An eight-layer convolutional neural network with stochastic pooling, batch normalization and dropout for fingerspelling recognition of Chinese sign language." *Multimedia Tools and Applications* 79, no. 21 (2020): 15697-15715.
- [185] Rastgoo, Razieh, Kourosh Kiani, and Sergio Escalera. "Video-based isolated hand sign language recognition using a deep cascaded model." *Multimedia Tools and Applications* 79, no. 31 (2020): 22965-22987.
- [186] Papadimitriou, Katerina, and Gerasimos Potamianos. "Multimodal Sign Language Recognition via Temporal Deformable Convolutional Sequence Learning." In *INTERSPEECH*, pp. 2752-2756. 2020.
- [187] Arun, C., and R. Gopikakumari. "Optimisation of both classifier and fusion based feature set for static American sign language recognition." *IET Image Processing* 14, no. 10 (2020): 2101-2109.
- [188] Sabeenian, R. S., S. Sai Bharathwaj, and M. Mohamed Aadhil. "Sign language recognition using deep learning and computer vision." *J. Adv. Res. Dyn. Contr. Syst* 12 (2020): 964-968.
- [189] Zheng, Jiangbin, Zheng Zhao, Min Chen, Jing Chen, Chong Wu, Yidong Chen, Xiaodong Shi, and Yiqi Tong. "An improved sign language translation model with explainable adaptations for processing long sign sentences." *Computational Intelligence and Neuroscience* 2020 (2020).
- [190] Jiang, Xianwei, Bo Hu, Suresh Chandra Satapathy, Shui-Hua Wang, and Yu-Dong Zhang. "Fingerspelling identification for Chinese sign language via AlexNet-based transfer learning and Adam optimizer." *Scientific Programming* 2020 (2020).
- [191] Ahmed, Hasmath Farhana Thariq, Hafisoh Ahmad, Kulasekharan Narasingamurthi, Houda Harkat, and Swee King Phang. "DF-WiSLR: Device-free Wi-Fi-based sign language recognition." *Pervasive and Mobile*

- Computing 69 (2020): 101289.
- [192] Parelli, Maria, Katerina Papadimitriou, Gerasimos Potamianos, Georgios Pavlakos, and Petros Maragos. "Exploiting 3d hand pose estimation in deep learning-based sign language recognition from rgb videos." In European Conference on Computer Vision, pp. 249-263. Springer, Cham, 2020.
 - [193] Park, Chan-Il, and Chae-Bong Sohn. "Data augmentation for human keypoint estimation deep learning based sign language translation." *Electronics* 9, no. 8 (2020): 1257.
 - [194] Saleh, Yaser, and Ghassan Issa. "Arabic sign language recognition through deep neural networks fine-tuning." (2020): 71-83.
 - [195] Gao, Qinghua, Shuo Jiang, and Peter B. Shull. "Simultaneous hand gesture classification and finger angle estimation via a novel dual-output deep learning model." *Sensors* 20, no. 10 (2020): 2972.
 - [196] Lee, Boon Giin, Teak-Wei Chong, and Wan-Young Chung. "Sensor fusion of motion-based sign language interpretation with deep learning." *Sensors* 20, no. 21 (2020): 6256.
 - [197] Li, Wenguo, Zhizeng Luo, Yan Jin, and Xugang Xi. "Gesture recognition based on multiscale singular value entropy and deep belief network." *Sensors* 21, no. 1 (2020): 119.
 - [198] Bird, Jordan J., Anikó Ekárt, and Diego R. Faria. "British sign language recognition via late fusion of computer vision and leap motion with transfer learning to american sign language." *Sensors* 20, no. 18 (2020): 5151.
 - [199] Wang, Zhibo, Tengda Zhao, Jinxin Ma, Hongkai Chen, Kaixin Liu, Huajie Shao, Qian Wang, and Ju Ren. "Hear sign language: A real-time end-to-end sign language recognition system." *IEEE Transactions on Mobile Computing* (2020).
 - [200] Abiyev, Rahib H., Murat Arslan, and John Bush Idoko. "Sign language translation using deep convolutional neural networks." *KSII Transactions on Internet and Information Systems (TIIS)* 14, no. 2 (2020): 631-653.
 - [201] Ojha, Ankit, Ayush Pandey, Shubham Maurya, Abhishek Thakur, and P. Dayananda. "Sign language to text and speech translation in real time using convolutional neural network." *International Journal of Engineering Research & Technology (IJERT)* 8, no. 15 (2020).
 - [202] Adithya, V., and Reghunadhan Rajesh. "A deep convolutional neural network approach for static hand gesture recognition." *Procedia Computer Science* 171 (2020): 2353-2361.
 - [203] Kumar, E. Kiran, P. V. V. Kishore, M. Teja Kiran Kumar, and D. Anil Kumar. "3D sign language recognition with joint distance and angular coded color topographical descriptor on a 2-stream CNN." *Neurocomputing* 372 (2020): 40-54.
 - [204] Cardenas, Edwin Jonathan Escobedo, and Guillermo Camara Chavez. "Multimodal hand gesture recognition combining temporal and pose information based on CNN descriptors and histogram of cumulative magnitudes." *Journal of Visual Communication and Image Representation* 71 (2020): 102772.
 - [205] Sharma, Prachi, and Radhey Shyam Anand. "A comprehensive evaluation of deep models and optimizers for Indian sign language recognition." *Graphics and Visual Computing* 5 (2021): 200032.
 - [206] Venugopalan, Adithya, and Rajesh Reghunadhan. "Applying deep neural networks for the automatic recognition of sign language words: A communication aid to deaf agriculturists." *Expert Systems with Applications* 185 (2021): 115601.
 - [207] Sharma, Sakshi, and Sukhwinder Singh. "Vision-based hand gesture recognition using deep learning for the interpretation of sign language." *Expert Systems with Applications* 182 (2021): 115657.
 - [208] Zheng, Jiangbin, Yidong Chen, Chong Wu, Xiaodong Shi, and Suhail Muhammad Kamal. "Enhancing Neural Sign Language Translation by highlighting the facial expression information." *Neurocomputing* 464 (2021): 462-472.
 - [209] Kulkarni, Aishwarya. "Dynamic sign language translating system using deep learning and natural language processing." *Turkish Journal of Computer and Mathematics Education (TURCOMAT)* 12, no. 10 (2021): 129-137.
 - [210] Elsayed, Eman K., and Doaa R. Fathy. "Semantic deep learning to translate dynamic sign language." *Int. J. Intell. Eng. Syst* 14 (2021).
 - [211] Amin, Mohamed, Hesahm Hefny, and Mohammed Ammar. "Sign language gloss translation using deep learning models." *International Journal of Advanced Computer Science and Applications* 12, no. 11 (2021).
 - [212] Martinez-Martin, Ester, and Francisco Morillas-Espejo. "Deep learning techniques for Spanish sign language interpretation." *Computational Intelligence and Neuroscience* 2021 (2021).
 - [213] Park, HyeonJung, Youngki Lee, and JeongGil Ko. "Enabling real-time sign language translation on mobile platforms with on-board depth cameras." *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 5, no. 2 (2021): 1-30.
 - [214] Dong, Yongfeng, Jielong Liu, and Wenjie Yan. "Dynamic hand gesture recognition based on signals from

- specialized data glove and deep learning algorithms." *IEEE Transactions on Instrumentation and Measurement* 70 (2021): 1-14.
- [215] Gauni, Sabitha, Ankit Bastia, B. Sohan Kumar, Prakhar Soni, and Vineeth Pydi. "Translation of Gesture-Based Static Sign Language to Text and Speech." In *Journal of Physics: Conference Series*, vol. 1964, no. 6, p. 062074. IOP Publishing, 2021.
 - [216] Aksoy, Bekir, Osamah Khaled Musleh Salman, and Özge Ekrem. "Detection of Turkish Sign Language Using Deep Learning and Image Processing Methods." *Applied Artificial Intelligence* 35, no. 12 (2021): 952-981.
 - [217] Barbhuiya, Abul Abbas, Ram Kumar Karsh, and Rahul Jain. "CNN based feature extraction and classification for sign language." *Multimedia Tools and Applications* 80, no. 2 (2021): 3051-3069.
 - [218] Alam, Md, Mahib Tanvir, Dip Kumar Saha, and Sajal K. Das. "Two-Dimensional Convolutional Neural Network Approach for Real-Time Bangla Sign Language Characters Recognition and Translation." *SN Computer Science* 2, no. 5 (2021): 1-13.
 - [219] Wen, Feng, Zixuan Zhang, Tianyi He, and Chengkuo Lee. "AI enabled sign language recognition and VR space bidirectional communication using triboelectric smart glove." *Nature communications* 12, no. 1 (2021): 1-13.
 - [220] Halvardsson, Gustaf, Johanna Peterson, César Soto-Valero, and Benoit Baudry. "Interpretation of swedish sign language using convolutional neural networks and transfer learning." *SN Computer Science* 2, no. 3 (2021): 1-15.
 - [221] Fregoso, Jonathan, Claudia I. Gonzalez, and Gabriela E. Martinez. "Optimization of convolutional neural networks architectures using pso for sign language recognition." *Axioms* 10, no. 3 (2021): 139.
 - [222] Wangchuk, Karma, Panomkhawn Riyamongkol, and Rattapoom Waranusast. "Real-time Bhutanese sign language digits recognition system using convolutional neural network." *Ict Express* 7, no. 2 (2021): 215-220.
 - [223] Gao, Liqing, Haibo Li, Zhijian Liu, Zekang Liu, Liang Wan, and Wei Feng. "RNN-transducer based Chinese sign language recognition." *Neurocomputing* 434 (2021): 45-54.
 - [224] Nihal, Ragib Amin, Sejuti Rahman, Nawara Mahmood Broti, and Shamim Ahmed Deowan. "Bangla sign alphabet recognition with zero-shot and transfer learning." *Pattern Recognition Letters* 150 (2021): 84-93.
 - [225] Abdul, Wadood, Mansour Alsulaiman, Syed Umar Amin, Mohammed Faisal, Ghulam Muhammad, Fahad R. Albogamy, Mohamed A. Bencherif, and Hamid Ghaleb. "Intelligent real-time Arabic sign language classification using attention-based inception and BiLSTM." *Computers and Electrical Engineering* 95 (2021): 107395.
 - [226] Suneetha, M., M. V. D. Prasad, and P. V. V. Kishore. "Multi-view motion modelled deep attention networks (M2DA-Net) for video-based sign language recognition." *Journal of Visual Communication and Image Representation* 78 (2021): 103161.
 - [227] Breland, Daniel S., Simen B. Skriubakken, Aveen Dayal, Ajit Jha, Phaneendra K. Yalavarthy, and Linga Reddy Cenkeramaddi. "Deep learning-based sign language digits recognition from thermal images with edge computing system." *IEEE Sensors Journal* 21, no. 9 (2021): 10445-10453.
 - [228] Elakkiya, R., Pandi Vijayakumar, and Neeraj Kumar. "An optimized Generative Adversarial Network based continuous sign language classification." *Expert Systems with Applications* 182 (2021): 115276.
 - [229] Singh, Dushyant Kumar. "3D-CNN based Dynamic Gesture Recognition for Indian Sign Language Modeling." *Procedia Computer Science* 189 (2021): 76-83.
 - [230] Sharma, Shikhar, and Krishan Kumar. "ASL-3DCNN: American sign language recognition technique using 3-D convolutional neural networks." *Multimedia Tools and Applications* 80, no. 17 (2021): 26319-26331.
 - [231] Lee, Carman KM, Kam KH Ng, Chun-Hsien Chen, Henry CW Lau, S. Y. Chung, and Tiffany Tsoi. "American sign language recognition and training method with recurrent neural network." *Expert Systems with Applications* 167 (2021): 114403.
 - [232] Zhou, Zhenxing, Vincent WL Tam, and Edmund Y. Lam. "SignBERT: A BERT-Based Deep Learning Framework for Continuous Sign Language Recognition." *IEEE Access* 9 (2021): 161669-161682.
 - [233] Rastgoo, Razieh, Kourosh Kiani, and Sergio Escalera. "Hand pose aware multimodal isolated sign language recognition." *Multimedia Tools and Applications* 80, no. 1 (2021): 127-163.
 - [234] Papastratis, Ilias, Kosmas Dimitropoulos, and Petros Daras. "Continuous sign language recognition through a context-aware generative adversarial network." *Sensors* 21, no. 7 (2021): 2437.
 - [235] Jain, Vanita, Achin Jain, Abhinav Chauhan, Srinivasu Soma Kotla, and Ashish Gautam. "American sign language recognition using support vector machine and convolutional neural network." *International Journal of Information Technology* 13, no. 3 (2021): 1193-1200.

- [236] Alawwad, Rahaf Abdulaziz, Ouiem Bchir, and Mohamed Maher Ben Ismail. "Arabic Sign Language Recognition using Faster R-CNN." *International Journal of Advanced Computer Science and Applications* 12, no. 3 (2021).
- [237] Meng, Lu, and Ronghui Li. "An attention-enhanced multi-scale and dual sign language recognition network based on a graph convolution network." *Sensors* 21, no. 4 (2021): 1120.
- [238] Alani, Ali A., and Georgina Cosma. "ArSL-CNN: a convolutional neural network for Arabic sign language gesture recognition." *Indonesian journal of electrical engineering and computer science* 22 (2021).
- [239] Kowdiki, Manisha, and Arti Khaparde. "Adaptive hough transform with optimized deep learning followed by dynamic time warping for hand gesture recognition." *Multimedia Tools and Applications* 81, no. 2 (2022): 2095-2126.
- [240] Mannan, Abdul, Ahmed Abbasi, Abdul Rehman Javed, Anam Ahsan, Thippa Reddy Gadekallu, and Qin Xin. "Hypertuned deep convolutional neural network for sign language recognition." *Computational Intelligence and Neuroscience* 2022 (2022).
- [241] Balaha, Mostafa Magdy, Sara El-Kady, Hossam Magdy Balaha, Mohamed Salama, Eslam Emad, Muhammed Hassan, and Mahmoud M. Saafan. "A vision-based deep learning approach for independent-users Arabic sign language interpretation." *Multimedia Tools and Applications* (2022): 1-20.
- [242] Xiao, Hongwang, Yun Yang, Ke Yu, Jiao Tian, Xinyi Cai, Usman Muhammad, and Jinjun Chen. "Sign language digits and alphabets recognition by capsule networks." *Journal of Ambient Intelligence and Humanized Computing* 13, no. 4 (2022): 2131-2141.
- [243] Rastgoo, Razieh, Kourosh Kiani, and Sergio Escalera. "Real-time isolated hand sign language recognition using deep networks and SVD." *Journal of Ambient Intelligence and Humanized Computing* 13, no. 1 (2022): 591-611.
- [244] Boukdir, Abdelbasset, Mohamed Benaddy, Ayoub Ellahyani, Othmane El Meslouhi, and Mustapha Kardouchi. "Isolated Video-Based Arabic Sign Language Recognition Using Convolutional and Recursive Neural Networks." *Arabian Journal for Science and Engineering* 47, no. 2 (2022): 2187-2199.
- [245] Sharma, Sakshi, and Sukhwinder Singh. "Recognition of Indian sign language (ISL) using deep learning model." *Wireless Personal Communications* 123, no. 1 (2022): 671-692.
- [246] Rajalakshmi, E., R. Elakkiya, Alexey L. Prikhodko, M. G. Grif, Maxim A. Bakaev, Jatinderkumar R. Saini, Ketan Kotecha, and V. Subramaniaswamy. "Static and Dynamic Isolated Indian and Russian Sign Language Recognition with Spatial and Temporal Feature Detection Using Hybrid Neural Network." *ACM Transactions on Asian and Low-Resource Language Information Processing* 22, no. 1 (2022): 1-23.
- [247] Nandi, Utpal, Anudyuti Ghorai, Moirangthem Marjit Singh, Chiranjit Changdar, Shubhankar Bhakta, and Rajat Kumar Pal. "Indian sign language alphabet recognition system using CNN with diffGrad optimizer and stochastic pooling." *Multimedia Tools and Applications* (2022): 1-22.
- [248] Miah, Abu Saleh Musa, Jungpil Shin, Md Al Mehedi Hasan, and Md Abdur Rahim. "BenSignNet: Bengali Sign Language Alphabet Recognition Using Concatenated Segmentation and Convolutional Neural Network." *Applied Sciences* 12, no. 8 (2022): 3933.
- [249] Duwairi, Rehab Mustafa, and Zain Abdullah Halloush. "Automatic recognition of Arabic alphabets sign language using deep learning." *International Journal of Electrical & Computer Engineering* (2088-8708) 12, no. 3 (2022).
- [250] Musthafa, Najla, and C. G. Raji. "Real time Indian sign language recognition system." *Materials Today: Proceedings* 58 (2022): 504-508.
- [251] Kasapbaşı, Ahmed, Ahmed Eltayeb AHMED ELBUSHRA, AL-HARDANEE Omar, and Arif Yilmaz. "DeepASLR: A CNN based human computer interface for American Sign Language recognition for hearing-impaired individuals." *Computer Methods and Programs in Biomedicine Update* 2 (2022): 100048.
- [252] AlKhuraym, Batool Yahya, Mohamed Maher Ben Ismail, and Ouiem Bchir. "Arabic Sign Language Recognition using Lightweight CNN-based Architecture." *International Journal of Advanced Computer Science and Applications* 13, no. 4 (2022).
- [253] Ismail, Mohammad H., Shefa A. Dawwd, and Fakhrulddin H. Ali. "Dynamic hand gesture recognition of Arabic sign language by using deep convolutional neural networks." *Indonesian Journal of Electrical Engineering and Computer Science* 25, no. 2 (2022): 952-962.
- [254] Venugopalan, Adithya, and Rajesh Reghunadhan. "Applying Hybrid Deep Neural Network for the Recognition of Sign Language Words Used by the Deaf COVID-19 Patients." *Arabian Journal for Science and Engineering* (2022): 1-14.
- [255] Tyagi, Akansha, and Sandhya Bansal. "Hybrid FiST_CNN approach for feature extraction for vision-based indian sign language recognition." *Int. Arab J. Inf. Technol.* 19, no. 3 (2022): 403-411.

- [256] Kothadiya, Deep, Chintan Bhatt, Krenil Sapariya, Kevin Patel, Ana-Belén Gil-González, and Juan M. Corchado. "Deepsign: Sign Language Detection and Recognition Using Deep Learning." *Electronics* 11, no. 11 (2022): 1780.
- [257] Alsaadi, Zaran, Easa Alshamani, Mohammed Alrehaili, Abdulmajeed Ayesh D. Alrashdi, Saleh Albelwi, and Abdelrahman Osman Elfaki. "A Real Time Arabic Sign Language Alphabets (ArSLA) Recognition Model Using Deep Learning Architecture." *Computers* 11, no. 5 (2022): 78.
- [258] Zhou, Zhenxing, Vincent WL Tam, and Edmund Y. Lam. "A Portable Sign Language Collection and Translation Platform with Smart Watches Using a BLSTM-Based Multi-Feature Framework." *Micromachines* 13, no. 2 (2022): 333.
- [259] Sharma, Shikhar, Krishan Kumar, and Navjot Singh. "Deep eigen space based ASL recognition system." *IETE Journal of Research* 68, no. 5 (2022): 3798-3808.
- [260] Samaan, Gerges H., Abanoub R. Wadie, Abanoub K. Attia, Abanoub M. Asaad, Andrew E. Kamel, Salwa O. Slim, Mohamed S. Abdallah, and Young-Im Cho. "MediaPipe's Landmarks with RNN for Dynamic Sign Language Recognition." *Electronics* 11, no. 19 (2022): 3228.
- [261] Abdullahi, Sunusi Bala, and Kosin Chamnongthai. "American Sign Language Words Recognition of Skeletal Videos Using Processed Video Driven Multi-Stacked Deep LSTM." *Sensors* 22, no. 4 (2022): 1406.
- [262] Sincan, Ozge Mercanoglu, and Hacer Yalim Keles. "Using Motion History Images with 3D Convolutional Networks in Isolated Sign Language Recognition." *IEEE Access* 10 (2022): 18608-18618.
- [263] Podder, Kanchon Kanti, Muhammad EH Chowdhury, Anas M. Tahir, Zaid Bin Mahbub, Amith Khandakar, Md Shafayet Hossain, and Muhammad Abdul Kadir. "Bangla sign language (bdsl) alphabets and numerals classification using a deep learning model." *Sensors* 22, no. 2 (2022): 574.
- [264] Luqman, Hamzah. "An Efficient Two-Stream Network for Isolated Sign Language Recognition Using Accumulative Video Motion." *IEEE Access* 10 (2022): 93785-93798.
- [265] Han, Xiangzu, Fei Lu, Jianqin Yin, Guohui Tian, and Jun Liu. "Sign Language Recognition Based on R (2+ 1) D With Spatial–Temporal–Channel Attention." *IEEE Transactions on Human-Machine Systems* (2022).
- [266] Sahoo, Jaya Prakash, Allam Jaya Prakash, Paweł Pławiak, and Saunak Samantray. "Real-Time Hand Gesture Recognition Using Fine-Tuned Convolutional Neural Network." *Sensors* 22, no. 3 (2022): 706.
- [267] Yirtici, Tolga, and Kamil Yurtkan. "Regional-CNN-based enhanced Turkish sign language recognition." *Signal, Image and Video Processing* (2022): 1-7.
- [268] Katoch, Shagun, Varsha Singh, and Uma Shanker Tiwary. "Indian Sign Language recognition system using SURF with SVM and CNN." *Array* 14 (2022): 100141.
- [269] Abdullahi, Sunusi Bala, and Kosin Chamnongthai. "American Sign Language Words Recognition using Spatio-Temporal Prosodic and Angle Features: A sequential learning approach." *IEEE Access* 10 (2022): 15911-15923.
- [270] Zhang, Nengbo, Jin Zhang, Yao Ying, Chengwen Luo, and Jianqiang Li. "Wi-Phrase: Deep Residual-MultiHead Model for WiFi Sign Language Phrase Recognition." *IEEE Internet of Things Journal* (2022).