Review Article

Ahmed Sultan, Walied Makram, Mohammed Kayed*, Abdelmaged Amin Ali

Sign language identification and recognition: A comparative study

https://doi.org/10.1515/comp-2022-0240
received April 29, 2021; accepted April 12, 2022

Abstract: Sign Language (SL) is the main language for handicapped and disabled people. Each country has its own SL that is different from other countries. Each sign in a language is represented with variant hand gestures, body movements, and facial expressions. Researchers in this field aim to remove any obstacles that prevent the communication with deaf people by replacing all device-based techniques with vision-based techniques using Artificial Intelligence (AI) and Deep Learning. This article highlights two main SL processing tasks: Sign Language Recognition (SLR) and Sign Language Identification (SLID). The latter task is targeted to identify the signer language, while the former is aimed to translate the signer conversation into tokens (signs). The article addresses the most common datasets used in the literature for the two tasks (static and dynamic datasets that are collected from different corpora) with different contents including numerical, alphabets, words, and sentences from different SLs. It also discusses the devices required to build these datasets, as well as the different preprocessing steps applied before training and testing. The article compares the different approaches and techniques applied on these datasets. It discusses both the vision-based and the data-gloves-based approaches, aiming to analyze and focus on main methods used in vision-based approaches such as hybrid methods and deep learning algorithms. Furthermore, the article presents a graphical depiction and a tabular representation of various SLR approaches.

*** Corresponding author: Mohammed Kayed,** Faculty of Computers and Artificial Intelligence, Computer Science Department, Beni-Suef University, Egypt, e-mail: mskayed@gmail.com
Ahmed Sultan: Faculty of Computers and Artificial Intelligence, Computer Science Department, Beni-Suef University, Egypt, e-mail: ahmed.soltan@fcis.bsu.edu.eg
Walied Makram: Faculty of Computers and Information, Information System Department, Minia University, Egypt, e-mail: waleedmakram@minia.edu.eg
Abdelmaged Amin Ali: Faculty of Computers and Information, Computer Science Department, Minia University, Egypt, e-mail: abdelmgeid@yahoo.com

Keywords: sign language, sign language recognition, artificial intelligence, convolutional neural network, deep learning

1 Introduction

Based on the World Health Organization (WHO) statistics, there are over 360 million people with hearing loss disability (WHO 2015 [1,2]). This number has increased to 466 million by 2020, and it is estimated that by 2050 over 900 million people will have hearing loss disability. According to the world federation of deaf people, there are about 300 sign languages (SLs) used around the world. SL is the bridge for communication between deaf and normal people. It is defined as a mode of interaction for the hard of hearing people through a collection of hand gestures, postures, movements, and facial expressions or movements which correspond to letters and words in our real life. To communicate with deaf people, an interpreter is needed to translate real-world words and sentences. So, deaf people can understand us or *vice versa*. Unfortunately, deaf people do not have a written form and have a huge lack of electronic resources. The most common SLs are American Sign Language (ASL) [3], Spanish Sign Language (SSL) [4], Australian Sign Language (AUSLAN) [5], and Arabic Sign Language (ArSL) [6]. Some of these societies use only one hand for sign languages such as USA, France, and Russia, while others use two-hands like UK, Turkey, and Czech Republic.

The need for an organized and unified SL was first discussed in World Sign Congress in 1951. The British Deaf Association (BDA) Published a book named Gestuno [7]. Gestuno is an International SL for the Deaf which contains a vocabulary list of about 1,500 signs. The name “Gestuno” was chosen referencing gesture and oneness. This language arises in the Western and Middle Eastern languages. Gestuno is considered a pidgin of SLs with limited lexicons. It was established in different countries such as US, Denmark, Italy, Russia, and Great Britain, in order to cover

the international meetings of deaf people. Although, Gestuno cannot be considered as a language due to several reasons. First, no children or ordinary people grow up using this global language. Second, it has no unified grammar (their book contains only a collection of signs without any grammar). Third, there are a fewer number of specialized people who are fluent or professional in practicing this language. Last, it is not used daily in any single country and it is not likely that people replace their national SL with this international one [8].

ASL has many linguistics that is difficult to be understood by researchers who are interested in technology, so experts of SLs are needed to facilitate these difficulties. SL has many building blocks that are known as phonological features. These features are represented as hand gestures, facial expressions, and body movements. Each one of these three phonological features has its own shape which differs and varies from one sign to another one. A word/an expression may have similar phonological features in different SLs. For example, the word "drink" could be represented similarly in the three languages ASL, ArSL, and SSL [16]. On the other hand, a word/an expression may have different phonological features in different SLs. For example, the word "Stand" in American and the word "يقف" (stand) in Arabic are represented differently in the two SLs. The process of understanding a SL by a machine is called Sign Language Processing (SLP) [9]. Many research problems are suggested in this domain such as Sign Language Recognition (SLR), Sign Language Identification (SLID), Sign Language Synthesis, and Sign Language Translation [10]. This article covers the first two tasks: SLR and SLID.

SLR basically depends on what is the translation of any hand gesture and posture included in SL, and continues/deals from sign gesture until the step of text generation to the ordinary people to understand deaf people. To detect any sign, a feature extraction step is a crucial phase in the recognition system. It plays the most important role in sign recognition. They must be unique, normalized, and preprocessed. Many algorithms have been suggested to solve sign recognition ranging from traditional machine learning (ML) algorithms to deep learning algorithms as we shall discuss in the upcoming sections. On the other hand, few researchers have focused on SLID [11]. SLID is the task of assigning a language when given a collection of hand gestures, postures, movements, and facial expressions or movements. The term "SLID" raised in the last decade as a result of many attempts to globalize and identify a global SL. The identification process is considered as a multiclass classification problem. There are many contributions in SLR with prior surveys. The latest survey was a workshop [12,13]. To the best of our knowledge, no prior works have surveyed SLID in previous Literature. This shortage was due to the need for experts who can explain and illustrate many different SLs to researchers. Also, this shortage due to the distinction between any SL and its spoken language [8,14] (i.e., ASL is not a manual form of English and does not have a unified written form).

Although many SLR models have been developed, to the best of our knowledge, none of them can be used to recognize multiple SLs. At the same time, in recent decades, the need for a reliable system that could interact and communicate with people from different nations with different SLs is of great necessity [15]. COVID-19 Coronavirus is a global pandemic that forced a huge percentage of employees to work and contact remotely. Deaf people need to contact and attend online meetings using different platforms such as Zoom, Microsoft Team, and Google Meeting rooms. So, we need to identify and globalize a unique SL as excluding deaf people and discarding their attendance will affect the whole work progress and damage their psyche which emphasizes the principle of "nothing about us without us." Also, SL occupies a big space of all daily life activities such as TV sign translators, local conferences sign translators, and international sign translators which is a big issue to translate all conference's points to all deaf people from different nations, as every deaf person requires a translator of their own SL to translate and communicate with him. In Deaflympics 2010, many deaf athletics were invited for this international Olympics. They need to interact and communicate with each other or even with anybody in their residence [16]. Building an interactive unified recognizer system is a challenge [11] as there are many words/expressions with the same sign in different languages, other words/expressions with different signs in the different languages, and other words/expressions could be expressed using the hands beside the movements of the eyebrows, mouth, head, shoulders, and eye gaze. For example, in ASL, raised eyebrows indicate an open-ended question, and furrowed eyebrows indicate a yes/no question. SLs could also be modified by mouth movements. For example, expressing the sign CUP with different mouth positions may indicate cup size, also body movements which may be included while expressing any SL provides different meanings. **SLID** will help in breaking down all these barriers for SL across the world.

Traditional machine and deep learning algorithms were applied to different SLs to recognize and detect signs. Most proposed systems achieved promising results

and indicated significant improvements in SL recognition accuracy. According to higher results in SLR on different SLs, a new task of SLID arises to achieve more stability and facility in deaf and ordinary people communication. SLID has many subtasks starting from image preprocessing, segmentation, feature extraction, and image classification. Most proposed models for recognition were applied to a single dataset, whereas the proposed SLID was applied to more than one SL dataset [11]. SLID inherits all SLR challenges, such as background and illumination variance [17], also skin detection and hands segmentation using both static and dynamic gestures. Challenges are doubled and maximized in SLID as many characters and words in different signs share the same hand's gestures, body movements, and so on, but may differ by considering facial expressions. For example, in ASL, raised eyebrows indicate an open-ended question, and furrowed eyebrows indicate a yes/no question, SL could also be modified by mouth movements.

Despite the need for interdisciplinary learning and knowledge of sign linguistics, most existing research does not go in depth but tackles the most important topics and separate portions. In this survey, we will introduce three important questions – (1) Why SLID is important? (2) What are the challenges to solve SLID? (3) what is the most used sign language for identifications and why? A contact inequality of SLs arises from this communication, whether it is in an informal personal context or in a formal international context. Deaf people have therefore used a kind of auxiliary gestural system for international communication at sporting or cultural events since the early 19th century [18]. Spoken languages like English are the most used language between all countries and many people thought it is a globally spoken language. Unfortunately, it is like all local languages. On the other side, for deaf people, many of them thought that ASL is the universal SL.

Furthermore, this article compares the different machine and deep learning models applied on different datasets, identifies best deep learning parameters such as, neural network, activation function, number of Epochs, best optimization functions, and so on, and highlights the main state of the art contributions in SLID. It also covers the preprocessing steps required for sign recognition, the devices used for this task, and the used techniques. The article tries to answer the questions: Which algorithms and datasets had achieved high accuracy? What are the main sub-tasks that every paper seeks to achieve? Are they successful in achieving the main goal or not?

This survey will also be helpful in our next research which will be about SLID, which requires deep understanding of more trending techniques and procedures used in SLR and SLID. Also, the survey compares the strengths and weaknesses of different algorithms and preprocessing steps to recognize signs in different SLs. Furthermore, it will be helpful to other researchers to be more aware of SL techniques.

The upcoming sections are arranged as follows: Datasets of different SLs are described in Section 2. The preprocessing steps for these datasets, that are prerequisite for all SL aspects, and the required devices will be discussed in Section 3. Section 4 includes the applied techniques for SL. Section 5 comprehensively compares the results and the main contributions of these addressed models. Finally, the conclusion and the future work will be discussed in Section 6.

2 Datasets

In this section, we discuss many datasets that had been used in different SL aspects such as skin and body detection, image segmentation, feature extraction, gesture recognition, and sign identification for more advanced approaches. For each dataset, we try to explore the structure of the dataset, the attributes with significant effects in the training and testing processes, the advantages and disadvantages of the dataset, and the content of the dataset (images, videos, or gloves). Also, we try to compare the accuracies of the dataset when applying different techniques to it [19]. Table 1 summarizes the results of these comparisons.

CopyCate Game [20]: a dataset that was collected from deaf children for educational adventure game purpose. These games facilitate interaction with computers using gesture recognition technology. Children wear two colored gloves (red and purple), one glove on each hand. They had collected about 5,829 phrases over 4 phases, with a total number of 9 deployments, each phrase has about 3, 4, or 5 signs taken from a vocabulary token of about 22 signs which is a list of adjectives, objects, prepositions, and subjects. The phrases have the following format:

[*adjective1*] *subject preposition* [*adjective2*] *object*.

Some **disadvantages** of this dataset are library continuity, sensor changes, varied environments, data integrity, and

Table 1: A comparison between different datasets

Name	Devices	No. of participants	Supported language	Size	Content (image/video/gloves)
RWTH German fingerspelling [21]	Webcam	20	GSL	1,400 (images)	**Image**
RWTH-BOSTON-104 [22,23]	Digital camera [black – white]	Not mentioned	ASL	201 (videos)	Video
CORPUS-NGT [24]	Digital camera	92	NSL	72 h	Video
CopyCat games [20]	Gloves with accelerometer	30	ASL	5,829 (phrases)	Gloves
Multiple dataset [25]	DG5-VHand data gloves and Polhemus G4	Not mentioned	ArSL	40 (phrases)	Gloves
ArSL dataset [26]	Digital camera	Not mentioned	ArSL	20 (words)	Video
SIGNUM [27]	Digital camera	25	DGS	455 (signs)/19k (sentences)	Video
ISL-HS [28]	Apple iPhone-7	6	ISL	468 videos [58,114 images]	Image/video
Camgoz et al. [29]	Microsoft Kinect v2 sensor	10	TSL	855 (signs)	Video
SMILE [30]	Microsoft Kinect, 2 video cameras and 3 webcams	60 learners and signers	SGSL	100 (signs)	Video
Gebre et al. [11]	Digital camera	9 (British) 10 (Greek)	BSL/GSL	~16 h	Video
Adaloglou et al. [31]	Intel RealSense D435 RGB + D camera	7 (Native Greek signers)	GSL	6.44 h	Video
Sahoo [32]	Sony digital camera	100 users	ISL	5,000 images	Image
RKS-PERSIANSIGN [33]	Digital camera	10	PSL	10,000 RGB videos	Video
MS-ASL [34]	—	100	ASL	25,000 videos	Video
WASL [35]	—	119	ASL	21,083 videos	Video
AUTSL [36]	Microsoft Kinect v2	43	TSL	38,336 videos	Video
KArSL [37]	Multi-modal Microsoft Kinect V2	3	ArSL	75,300 videos	Video
Breland [38]	Raspberry pi with a thermal camera	—	—	3,200 images	Image
Mittal et al. [39]	Leap motion sensor	6	ISL	3,150 signs	Video

ASL: American Sign Language, DGS: German Sign Language, NSL: Netherlands Sign Language, TSL: Turkish Sign Language, LSF: French Sign Language, ISL: Irish Sign Language, SGSL: Swiss German Sign Language, ISL: Indian Sign Language, PSL: Persian Sign Language.

sign variation. Another **disadvantage** and disability are wearing gloves because users must interact with systems using gloves. On the other hand, it has the advantage of integrating new data they gathered from other deployments into its libraries.

Multiple Dataset [25]: Collected two datasets of ArSL, consisting of 40 phrases with 80-word lexicon, each phrase was repeated 10 times, using DG5-Vhand data glove with five sensors on each finger with an embedded accelerometer. It was collected using two Polhemus G4 motion trackers providing six different measurements. Dataset (Number.2) was collected using a digital camera without wearing gloves for capturing signs.

ArSL Dataset [26]: Digital cameras are used to capture signer's gestures, then videos are stored as a AVI video format to be analyzed later. Data were captured from deaf volunteers to generate samples for training and testing the model. It consists of 20 lexicons, with 45 repetitions for every word, 20 for training and 18 for testing. All signer's hands are bare, and no wearable gloves are required. Twenty-five frames are captured per second with a resolution of 640 × 480.

Weather Dataset [40]: A continuous SL composed of three state-of-the-art datasets for the SL recognition purpose: RWTH-PHOENIX-Weather 2012, RWTH-PHOENIX-Weather 2014, and SIGNUM.

SIGNUM [27]: It is used for pattern recognition and SL recognition tasks. A video database is collected from daily life activities and sentences like going to the cinema, ride a bus, and so on. Signer's gestures were captured by digital cameras.

CORPUS-NGT [24]: Great and huge efforts are done to collect and record videos of the SL of Netherlands (Nederlandes Gebarentaal: NGT), providing global access to this corpus for all researchers and sign language studies. About 100 native signers of different ages participated in collecting and recording signs for about 72 h. It provided annotation or translation of some of these signs.

RWTH German fingerspelling [21]: A German SL dataset is collected from 20 participants, producing about 1,400 image sequences. Each participant was asked to record every sign twice on different days by different cameras (one webcam and one camcorder) without any background limitations or restrictions on wearing clothes while gesturing. Dataset contains about 35 gestures with video sequences of alphabets and first 5 numbers (1–5).

RWTH-BOSTON-104: A dataset published by the national center of SL and gesture resources by Boston University. Four cameras were used to capture signs, three of them are white/black cameras and one is a color camera. Two white/black cameras in front of the signers to form stereo, another camera on the side of the signer, and the colored camera was focused between the stereo cameras. It consists of 201 annotated videos, about 161 videos are used for training and about 40 for testing. Captured movies of sentences consist of 30 fps [312 × 242] using only the upper center [195 × 165].

Oliveira et al. [28]: An Irish SL dataset captured human subjects using handshapes and movements, producing 468 videos. It is represented as two datasets which were employed for static (each sign language is transferred by a single frame) and dynamic (each sign is expressed by different frames and dimensions) Irish Sign Language recognition.

ISL-HS consists of 486 videos that captured 6 persons performing Irish SL with a rotating hand while signing each letter. Only arms and hands are considered in the frames. Also, videos whose background was removed by thresholding were provided. Further, 23 labels are considered, excluding j, x, and z letters, because they require hand motion which is out of the research area of the framework [41].

Camgoz et al. [29]: It presented the Turkish sign language. It was recorded using the state-of-the-art Microsoft Kinect v2 sensor. This dataset contains about 855 signs from everyday life domains from different fields such as finance and health. It has about 496 samples in health domain, about 171 samples in finance domain, and the remaining signs (about 181) are commonly used signs in everyday life. Each sign was captured by 10 users and was repeated 6 times, each user was asked to perform about 30–70 signs.

SMILE [30]: It prepared an assessment system for lexical signs of Swiss German Sign Language that relies on SLR. The aim of this assessment system is to give adult L2 (learners of the sign language) ofDSGS feedback on the correctness of the manual parameters such as hand position, shape, movement, and location. As an initial step, the system will have feedback for a subset of DSGS vocabulary production of 100 lexical words, to provide the SLR as a component of the assessment systems a huge dataset of 100 items was recorded with the aid of 11 adult L1 signers and 19 adult L2 learners of DSGS.

Most SLR techniques begin with extracting the upper body pose information, which is a challenge task due to the difference between signer and background color, another challenge is motion blur. To overcome all these challenges, they used a diverse set of visual sensors including high-speed and high-resolution GoPro video cameras, and a Microsoft Kinect V2 depth sensor.

Gebre et al. [11]: Dataset includes two languages British and Greek sign language which are available on Dicta-Sign corpus. The corpus has recordings for 4 sign languages with at least 14 signers per language and approximately 2 h using the same material across languages, from this selection, two languages BL and GL were selected. The priority of the signer's selection was based on their skin's color difference from background color. About 95% F1 score of accuracy was achieved.

Sahoo [32]: About 5,000 images of digital numbers (0,1,2...9) from 100 users (31 were female and 69 were male) were collected. Each signer was asked to repeat each character 5 times. Sony digital camera with resolution up to 16.1MP is used. Image format is JPEG with a resolution of 4,608 × 3,456 of the captured images. Image resolution was resized to 200 × 300. Finally, the dataset was divided into two groups for training and testing.

RKS-PERSIANSIGN [33] include a large dataset of 10 contributors with different backgrounds to produce 10,000 videos of Persian sign language (PSL), containing 100 videos for each PSL word, using the most commonly used words in daily communication of people.

Joze and Koller [34] proposed a large-scale dataset for understanding ASL including about 1,000 signs registered over 200 signers, comprising over 25,000 videos.

WASL [35] constructed a wide scale ASL dataset from authorized websites such as ASLU and ASL_LEX. Also, data were collected from YouTube based on clear titles that describe signs. About 21,083 videos were accessed from 20 different websites. Dataset was performed by 119 signers, producing only one video for every sign.

AUTSL [36] presented a new large-scale multi-modal dataset for Turkish Sign Language dataset (AUTSL). 226 signs were captured by 43 different signers, producing 38,336 isolated sign videos. Some samples of videos containing a wide variety of background recorded in indoor and outdoor environments.

KArSL [37], a comprehensive benchmark for ArSL containing 502 signs recorded by 3 different signers. Each sign was repeated 50 times by each signer, using Microsoft Kinect V2 for sign recording.

Daniel [38] used Raspberry pi with a thermal camera to produce 3,200 images with low resolution of 32 × 32 pixel. Each sign has 320 thermal images, so we conclude capturing images of about 10 signs.

Mittal et al. [39] created an ISL dataset recorded by six participants. The dataset contains 35 sign words, each word was repeated at least 15 times by each participant, so the size of the dataset is 3,150 (35 × 15 × 6).

3 Preprocessing steps

Two main preprocessing steps are required for different sign language processing tasks: segmentation and filtration. These tasks include the following subtasks: skin detection, handshape detection, feature extraction, image/video segmentation, gesture recognition, and so on. In this section, we shall briefly discuss all these subtasks. Figure 1 shows the sequence of different preprocessing steps that are almost required for different SL models. Each model usually starts with a signer's image, applying color space conversion. Non-skin images are rejected, while other images continue the processing by applying image morphology (erosion and dilation) for noise reduction. Each image is validated by checking whether it has a hand or not. If yes, then the Region-of-Interest (ROI) is detected using hand mask images and segment fingers using defined algorithms. Image enhancement such as image filtering, data augmentation, and some other algorithms could be used to detect edges.

Skin-detection: It is the process of separating the skin color from the non-skin color. Ref. [42] approved that it is not possible to provide a uniform method for detection and segmentation of human skin as it varies from one person to another. RGB is a widely used color mode, but it is not preferred in skin detection because of its chrominance and luminance and its non-uniform characteristics. Skin detection is applied on HSV (HUE, and Saturation Values) images and YCbCr.

ROI [42]: It is focused on detecting [43] hand gestures and extracting the most interesting points. The hand region is detected using skin-detection from the original image using some defined masks and filters as shown in Figure 2.

Image resize: It is the process of resizing images by either expanding or decreasing image size. Ref. [44] applied an interpolation algorithm that changes the image accuracy from one to another. Bicubic, a new pixel B (r^', c^') is formed by interpolating the nearest 4 × 4.

Ref. [45] proposed a promising skin-color detection algorithm, giving the best results even with complex backgrounds. Starting with acquiring an image from the video input stream, then adjusting image size, converting an image from RGB color space to YCbCr space (also denoting that YCbCr space is the most suitable one for skin color detection), and finally identifying color based on different values of threshold [46,47] and marking the detected skin with white color, otherwise with black color. Figure 1 includes a sub-flowchart that illustrates this algorithm.

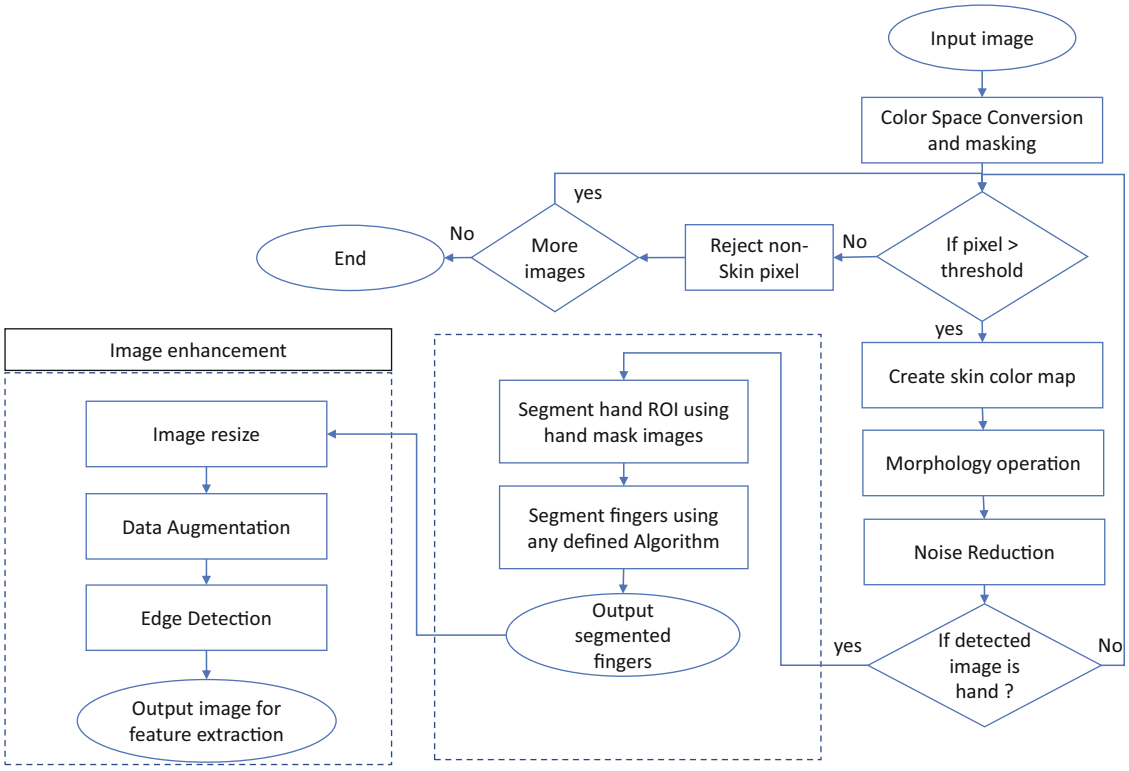


Figure 1: A flowchart that demonstrates the different image preprocessing steps.

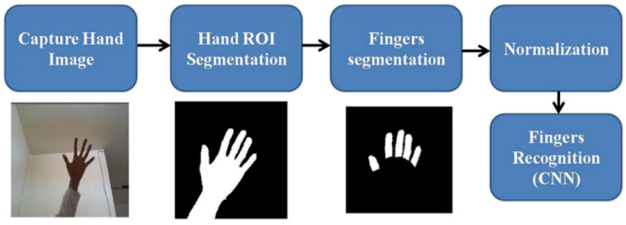


Figure 2: A proposed hand gesture recognition system using ROI [64].

In Ref. [48], binarization of images from RGB color mode to black and white color mode using Ostu's algorithm of global thresholding was performed, images captured were then resized to 260 × 260 pixels for width and height and then Ostu's method was used to convert the image. Ref. [49] applied new technique for feature extraction known as 7Hu moments invariant, which are used as a feature vector of algebraic functions. Their values are invariant because of the change in size, rotation, and translation. 7Hu moments were developed by Mark Hu in 1961. Structural shape descriptors [23] are proposed in five terms, aspect ratio, solidity, elongation, speediness, and orientation.

Ref. [50] used a face region skin detector which includes eyes and mouth which are non-skin non-smooth regions, which affect and decrease the accuracy.

A window of 10 × 10 around centered pixel of signer's face is used to detected skin, but it is not accurate because in most cases it detects nose as it suffers from high illumination conditions [51].

Image segmentation: It refers to the extraction of hands from video frames or images. Either background technique or skin detector algorithm is applied first to detect skin of signer and then segmentation algorithm is applied [52]. Ref. [53] applied skin color segmentation using Artificial Neural Network (ANN), features extracted from left and right hands are used for neural network model with average recognition of 92.85%.

Feature extraction: It is the process of getting most important data-items or most interesting points of segmented image or gesture. Ref. [25] applied two techniques for feature extraction including window-based statistical feature and 2D discrete cosine transform (DCT) transformation. Ref. [48] applied five types of feature extraction including fingertip finder, elongatedness, eccentricity, pixel segmentation, and rotation. Figures 3 and 4 depict a promising accuracy in percentage of different feature extraction algorithms. Figure 3 illustrates the strength

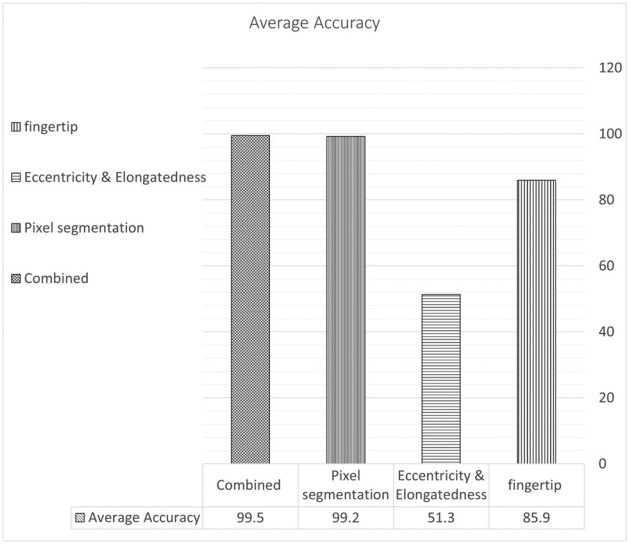


Figure 3: Average recognition rate (%) for each feature extraction algorithm and percentage of combining them.

combining different three feature extraction algorithms: pixel segmentation, eccentricity, and elongatedness and fingertip and applying each one individually. Combined algorithms have largest accuracy with 99.5%.

Tracking: Tracking body parts facilitate the SLR process. How important are accurate tracking of body parts and its movements? How accurate are its contribution to SLR? And how does the comparison and differences occur to just use the tracked image for feature extraction. **Hand Tracking:** hands of the signer convey most of the recognition of signs in most SL. Ref. [27] employed a free tracking system that is based on dynamic programming tracking (DPT). **Tracking facial landmarks:** introduced Active Appearance Models (AAMs) which was then reformulated by Matthews et al. [54].

4 Required devices

In the last decade, researchers depended on electronic devices to detect and recognize hand position and its gestures, because of many reasons [55]. One of them is SLR using signer independent or signer dependent. Signer dependent is the main core of any SLR system, as the signer performs both training and testing phases. So, this type affects the recognition rate positively. On the other side, signer independence is a challenging phase, as signers perform only the training phase, not admitted in the testing phase. This discarding is a challenge in adapting the system to accept another signer. The target of SL systems can be achieved by (I) image-based

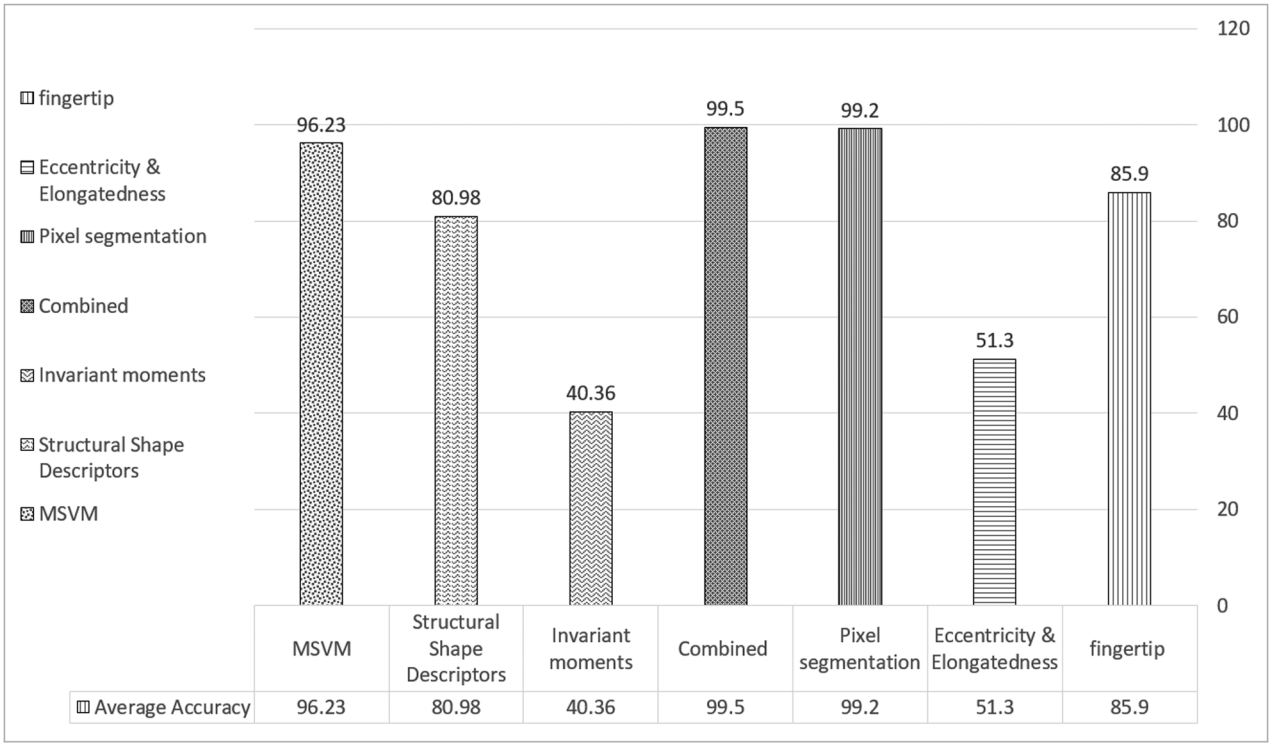


Figure 4: Comparison of different feature algorithms.

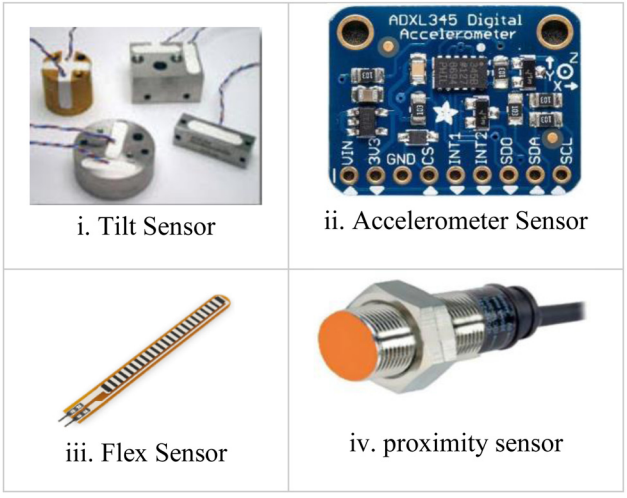
approach [56] or (II) glove-based approach based on sensors as shown in Figure 5 or (III) a new method for gesture recognition called virtual button [57].

One of the disadvantages of data-gloves or electronic devices mainly are, data-gloves gives accurate information but with little information, the more advanced technology of sensors used, the more the cost, finally the data-gloves must be on-off-on each time of hand gesture recognition, which adds more obstacles with people who do not or are not aware of communication with this technology especially when they are in public places. Below is a short description of most used devices for SLR.

Glove-based approach with built-in sensors: It is an electromechanical input device used for human computer interactions, widely used in haptic applications, and works like a glove. It is used to capture many physical data such as hand gestures and postures, body movements, and motion tracker, all these movements are interpreted by software that accompanies the glove. In 1983, this approach was used for SLR of signers from recorded videos and to circumvent many approaches of computer vision, especially, recognition of signs from videos. There are many types of sensors attached to this glove to facilitate motion and movement capture as shown in Figure 6.

- **Tilt sensor:** It is a device that produces an electrical signal that varies with an angular movement, used to measure slope and tilt with a limited range of motion.
- **Accelerometer sensor:** It measures 3-axis acceleration caused by gravity and motion; in another word it is used to measure the rate of change of velocity.
- **Flex sensor:** It is a very thin and lightweight electric device, used for the measurement of bending or deflection. Usually is stocked to the surface of fingers and the resistance of the sensor varied by bending the surface.

Figure 5: Cyber-glove.



i. Tilt Sensor ii. Accelerometer Sensor

iii. Flex Sensor iv. proximity sensor

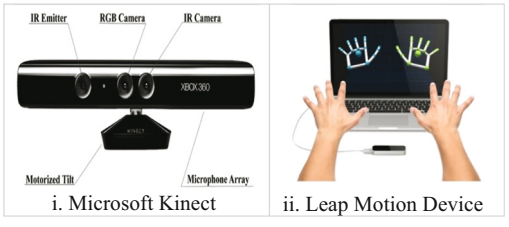
Figure 6: Different sensor types attached to hand gloves.

- **Motion (proximity) sensor:** It is an electrical device which utilizes a sensor to capture motion, or it is used to detect the presence of objects without any physical contact.

As previously illustrated all these kinds of sensors used to measure the bend angles of fingers, the orientation or direction of the rest, abduction, and adduction between fingers. These sensors give an advantage over vison-based systems. Sensors can directly report with required data without any preprocessing steps for feature extraction (pending degree, orientation, etc.) in terms of voltage values to the system, but on the other side, vision-based systems require to apply tracking and feature extraction algorithms. But ref. [58] mentioned that using data-gloves and sensors do not provide the naturalness of HCI systems.

As a part of electronic devices which may have built-in sensors, there are two devices widely used in many fields, infra-red sensors such as Microsoft Kinect and Leap-Motion devices as shown in Figure 7 (Table 2).

Vision-based approach: The great development in computer techniques and ML algorithms motivate many researchers to depend on vision-based methodology. A camera is used to capture images and then process



i. Microsoft Kinect ii. Leap Motion Device

Figure 7: Digital devices with built-in sensors used to capture dynamic gestures of human expressions.

Table 2: Most widely used electronic devices for hand gesture recognition

Author (year)	Devices	Techniques	Dataset			No. of hands	Accuracy (%)
			Number	Alphabet	Word/Phrases		
Hussain [59], 1999	Colored glove [red and blue]	ANN	—	✓(ASL)	—	Two	95.57
Oz and Leu [60], 2011	CyberGlove [18 sensors]	ANN	—	—	✓(ASL)	Right hand	98
Kadous [61], 2014	Glove [Flex sensors]	Matching	—	✓(ASL)	—	Right hand	—
Kadous [61], 1996	PowerGlove	IBL and DTL	✓	✓	✓(AUSLAN)	—	80
Tubaiz et al. [62], 2015	DG5-VHand data gloves	—	—	—	✓ (ArSL)	—	98.9%
Daniel [38], 2021	Raspberry PI and Omron D6T thermal camera	CNN	✓	—	—	—	**99.5**
Mittal et al. [39], 2019	Leap motion sensor	LSTM	—	—	✓(ISL)	Two	72.3 (sentences) and 89.5% (words)
Rosero-Montalvo et al. [63], 2018	Gloves with flex sensors	KNN	✓	—	—	Right hand	85
Chen et al. [64], 2020	Myo armband	CNN	—	—	—	Right hand	98.81

ANFIS: Adaptive Neuro-Fuzzy Inference system, MSL: Malaysian Sign Language, IBL: instance-based learning, DTL: decision-tree learning, ISL: Indian Sign Language.
Bold indicated the highest accuracy of using electronic devices for hand gesture recognition.

to detect the most important features for recognition purposes. Most researchers prefer vision-based method because of its framework's adaptability, the involvement of facial expression, body movements, and lips perusing. So, this approach required only a **Camera** to capture a person's movements with a clear background without any gadgets. Previous gloves required an accompanying camera to register the gesture but does not work well in lightning conditions.

Virtual Button approach [57]: Depends on a virtual button generated by the system and receives hand's motion and gesture by holding and discharging individually. This approach is not effective for recognizing SL because every sign language required utilization of all hand's fingers and it also cannot be practical for real life communication.

4.1 Methodology and applied techniques

Many datasets were used in SL recognition, some of these datasets are based on the approaches of vision and some are based on the approach of soft computing like ANN, Fuzzy Logic, Genetic Algorithms, and others like Principal Component Analysis (PCA) and deep learning like Convolutional Neural Network (CNN).

Also, many algorithms and techniques were applied to recognize SLs and identify different languages with variance accuracy. Some of these techniques are classical algorithms and others are deep learning which has become the heading technique for most AI problems and overshadowing classical ML. A clear reason for depending on deep learning is that it had repeatedly demonstrated high quality results on a wide variety of tasks, especially those with big datasets.

Traditional ML algorithms are a set of algorithms that use training data to learn, then apply what they had learned to make informed decisions. Among traditional algorithms, there are classification and clustering algorithms used for SLR and SLID. Deep learning is considered as an evolution of ML, it uses programmable neural networks which make decisions without human intervention.

5 Methodology and applied techniques

K-Nearest Neighbors (KNN): It is one of the traditional ML algorithms used for classification and regression problems. Many researchers applied KNN on different SL

datasets, but accuracy was lower than expected. Ref. [65] achieved results of 28.6% accuracy when applying KNN with PCA for dimensionality reduction. Other researchers merged some preprocessing steps for better accuracies. Although KNN indicate lower accuracy for image classification problems, some researchers recommended using KNN because of its ease of use, implementation, and fewer steps. Table 3 discusses some of the KNN algorithms applied on different datasets.

Dewinta and Heryadi [65] classified ASL dataset using KNN classifier, varying the value of $K = 3, 5, 7, 9$, and 11. The highest accuracy was 99.8% using $K = 3$, while the worst accuracy was achieved by setting $K = 5$ using PCA for dimensionality reduction.

Fitri [66] proposed a framework using Simple Multi Attribute Rating Technique (SMART) weighting and KNN classifier. SMART was used to optimize and enhance accuracy of KNN classifier. The accuracy varied from 94 to 96% according to some lightening conditions. The accuracy decreases when lighting decreases and *vice versa* (Figure 8).

According to Figure 9 it is clear that different algorithms preferred that *K*-values should not be static to its default value which equals "1", but varying *K*-values to 1, 3, 5 or any odd number will result in good results. With $K = 3$, most researchers get the best accuracy.

Jadhav et al. [67] proposed a framework based on KNN for recognizing sign languages. The unique importance of this framework allowed users to define their own sign language. In his system, users must store their signs first in database, after that he can use these signs while communicating with others. While communicating with another person using those stored signs, the opposite person can see the signs and its meaning. This framework suggested using real time sign recognition. The framework is based on three main steps:

Skin detection, he created a "skin detector" method for converting images from BGR format to HSV format. He used an interpolation algorithm for shadow detection from the images and fills it with continuous dots using "FillHoles" method. Another method called "**DetectAndRecognize**" takes the fill hole image as input for detection and recognition and calculates the contours which detect the edges of the signs. **Title Blob detection** – have two methods "FillHoles" and "DetectAndRecognize" methods.

Umang [68] applied KNN and PNN as a classification technique to recognize ISL alphabets. 7Hu moments were used for feature extraction. 82% is the approximate accuracy they achieved, using KNN built-in function in MATLAB with default $K = 1$.

Table 3: KNN classification comparison on different datasets

Author (year)	Technique	Gesture type	*K*-Value	Accuracy (%)	Notes
Jadhav et al. [67], 2017	KNN	HGR	—	**100**	Template matching technique was applied with best recognition of hand's gestures, then used KNN for time reduction
Dewinta and Heryadi [65], 2015	KNN	ASL	3	99.8	KNN classifier used to detect and recognize ASL, giving promising accuracy with $K = 3$ and lower accuracy with $K = 5$, which is 28.6%
Utaminingrum [66], 2018	KNN	ASL	1	94:95	Applied KNN with SMART technique used to improve weights and get best accuracy
Tubaiz et al. [62], 2015	MKNN	ArSL	3	98.9	Recognize Arabic sign language based on two DG5-VHand data gloves electronic device. About 80-words lexicon were used to build up 40 sentences
Patel [68], 2017	KNN	ISL	1	82	Used MATLAB functions to convert captured hand gestures into text and speech based on classification algorithms (PNN and KNN) to recognize ISL alphabets
Saggio et al. [69], 2010	KNN	ISL	—	96.6	Proposed a wearable electronic device to recognize 10 gestures of Italian sign language. KNN and CNN were applied with accuracy of 96.6 and 98%, respectively
	CNN			98	
Sahoo [70], 2021	KNN	ISL	1	98.36	Performed classification of a standard ISL dataset (containing 10 static digit numbers) using two different classifiers (KNN and Naïve Bayes). Accuracy produced by KNN is higher than Naïve Bayes
	Naïve Bayes			97.79	
Saggio [69], 2021	KNN	—	—	97.12	Proposed a KNN classifier to analyze input video and extract the vocabulary of 20 gestures. Using a hand gesture dataset for training and testing, they got an overall accuracy of 97%

Bold indicated the highest accuracy using KNN algorithm using different K-values.

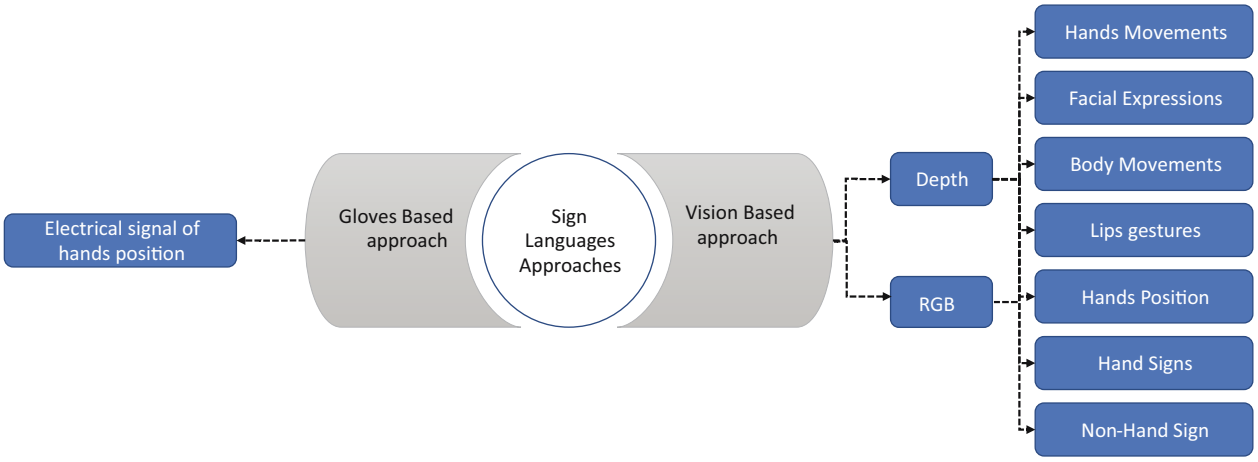


Figure 8: Sign languages-based approaches.

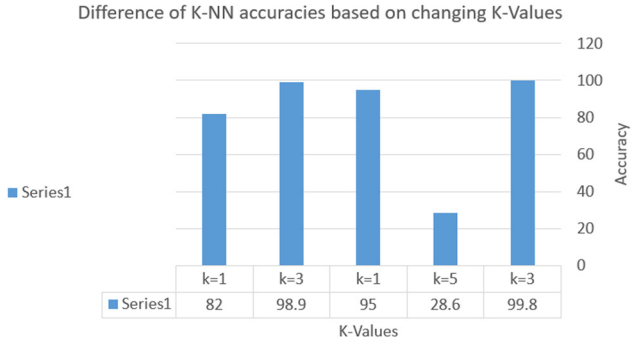


Figure 9: Different KNN results based on change in *K*-values.

Hidden Markov Model (HMM): Based on our review, HMM was one of the strongly recommended approaches for SL problems. Hybridization of HMM with CNN provided high accuracy with huge datasets. HMM is the most widely used technique for speech recognition and SL problems for both vision-based and data-gloves-based approach. Table 4 discusses some of the HMMs models applied on different datasets.

Parcheta and Martínez-Hinarejos [71] used an optimized sensor called "leap motion" that we presented previously. This leap device was used to capture 3D information of hands gestures. He applied one of the two available types of HMM which is discrete and continuous. Continuous HMM was used for gesture recognition. Hidden Markov Model Toolkit (*HTK*) was used to interact and interpret HMMs. He tried to recognize about 91 gestures collected using the aforementioned device by partitioning data into four parts, training HMM topologies through some defined models and producing accuracy of 87.4% for gesture recognition.

Starner [73] proposed a real-time HMM-based system to recognize sentences of ASL consisting of 40-word lexicon and capturing users' gestures using cameras mounted to a desk, producing 92% accuracy. The second camera was mounted to a cap of the user, producing 98% accuracy. This paper proved that vision-based approach is more useful than glove-based approaches.

Ref. [34] provides systems based on HMM to recognize real-time ArSL. The system is a signer-independent, removing all barriers to communicate with deaf people. They built their own dataset to recognize 20 Arabic isolated words. They used 6 HMM models with different number of states and different Gaussian mixtures per state. The best accuracy was 82.2%.

Oliveira et al. [28] built a framework for static and dynamic sign language.

Hand Segmentation used OpenPose [73] detector trained on the dataset, getting high results for hand segmentation among all evaluated detectors (HandSegNet [54] and hand detector [74]). The right hand is detected by applying a forward feed neural network based on VGG-19, and the left image is detected by flipping the image and applying the previous steps again.

Static Signs consists of 2D convolution which contains the features, first layer tends to know more about the basic feature's pixels like lines and corners. Each input frame is convolved with more than 32 filters to cover the network's scope which is narrower at the beginning. The model is interested in fewer features.

Dynamic Sign Language Model: It is concerned with two key-points. First, it considered three dimensions of the layer for temporal dimension. It extends over the temporal dimension; this is useful in sign language

Table 4: HMM comparison on different datasets

Author (year)	Technique	Gesture type	Accuracy (%)	Notes
Parcheta and Martínez-Hinarejos [71], 2017	HMM	SSL	87.4	Compared two results of using KNN + DTW which produced accuracy of 88.4% but recognition process speed was very high of 9,383 compared to HMM which was 519 to recognize Spanish sign language
Starner et al. [72], 1998	HMM	ASL	98	Recognized ASL based on HMM using two techniques based on camera mounted on a desk with result of 98% accuracy, and the second used camera stuck to a user's cap producing 92% accuracy
Youssif et al. [26], 2011	HMM	ArSL	82.22	Used their own captured dataset from deaf people to represent 20 Arabic words. Applying HMM to recognize and detect each word with 6 models and different Gaussian mixtures. Average accuracy is 82.22%
Roy et al. [79], 2021	HMM	ASL	77.75	Proposed HMM model to track hand motion in videos. Using camshift for hand tracking, accuracy of 77.75 % was achieved using 91 ASL hand gestures. Although this result is low, the author considered it very high as it contains more than double gestures with respect to existing approaches
Ghanbari Azar and Seyedarabi [80], 2019	HMM	PeSL	97.48	Dataset of 1200 videos captured based on 20 dynamic signs using 12 participants. HMM was used as a classifier with Gaussian density function as for observations. An average accuracy of 97.48% was obtained using both signer dependent and independent

recognition because it helps to model the local variations that describe the trajectory of gesture during its movement. Briefly, a dynamic model is implemented on a single frame followed by the gesture of each sequence.

For static results of ISL recognition, this paper achieves an accuracy of 0.9998 for cropped frames whereas it achieves an accuracy of 0.9979 for original frame, while for dynamic results, categorical accuracy is considered for each class. Classifier model was trained and tested on 8 classes, its accuracy was not high as it ranges between 0.66 and 0.76 for different streams.

Binyam Gebrekidan Gebre [11] proposed a method that gathers two methods of Stokoe's and H-M model as they assumed that features extracted from frames are independent of each other, But Gebre assumes that sign's features will be extracted from two frames. The next and previous one to get a hand or any movement. He Proposed an ideal SLID, the system subcomponents are: (1) skin detection, (2) feature extraction, (3) modeling, and (4) identification. For a modeling step, it used a random forest algorithm which generates many decision tree classifiers and aggregates their results. Extracted features include high performance, flexibility, and stability. He achieved about 95% F1 score of accuracy.

Invariant features [75] consists of three stages, namely, a training phase, a testing phase, and a recognition phase. The parameters of 7Hu invariant moment and structural shape descriptors which are created to form a new feature vector to recognize the sign are combined and then MSVM is applies for training the recognized signs of ISL. The effectiveness of the proposed method is validated on a dataset of 720 images with a recognition rate of 96%.

CNN [76] Kang et al. used CNN, specifically caffe implementation network (CaffeNet), consisting of 5 convolution layers, 3 max-pooling layers, and 3 fully connected layers.

FFANN [17,77] was used in ref. [48] achieving an average accuracy of 94.32% using convex hull eccentricity, elongatedness, pixel segmentation, and rotation for American number, and alphabets recognition of about 37 signs, whereas ref. [49] applied FFANN on facial and hand gestures of 11 signs, with an average accuracy of 93.6% depending on automatic gesture area segmentation and orientation normalization. Ref. [78] also used FFANN for Bengali alphabet with 46 signs achieving an accuracy of 88.69% for testing result depending on Fingertip finder algorithm with multilayered feedforward, back propagation training.

Effective ML algorithms were used to achieve high accuracy, but deep learning algorithms indicate more

accurate results. Deep learning types vary between unsupervised pre-trained networks, CNN, recurrent neural network, and recursive neural network which encourage more people to do more research, share, and compare their results. We will compare between types of deep learning algorithms and used parameters, to determine which activation function is the best? How to test and train the model?

Ref. [44] applied two CNN models on 24 letters of ASL with 10 images per letter, image size is 227 × 227 which is resized using the Bicubic interpolation method. The images were trained using 4 CNNs with 20 layers in each CNN. Each model had a different activation function and a different optimization algorithm. PReLU and ReLU were used in model 1 and model 2, respectively. Accuracy for model 1 is 99.3% as it was able to recognize all 24 letters, but the accuracy of model 2 was 83.33% as it recognizes only 20 letters of all the 24 letters.

Ref. [81] used deep learning algorithms to identify SL using three publicly available datasets. Also introduced a new public large-scale dataset for Greek sign language RGB + D, providing two CTC variations that were mostly used in other application fields EnCTC and StimCTC. Each frame was resized from 256 × 256 to 224 × 224. The models are trained using Adam optimizer, and initial learning rate of 0.0001 was reduced to 0.00001.

Ref. [82] proposed a deep learning model consisting of CNN (inception model) and RNN to capture images of ASL, this dataset consists of 2,400 images, divided into 1,800 images for training and the remaining for testing. CNN extracts feature from the frames, using two major approaches for classification as SoftMax layer and the pool layer. After retraining the model using the inception model, the extracted features were passed to RNN using LSTM Model.

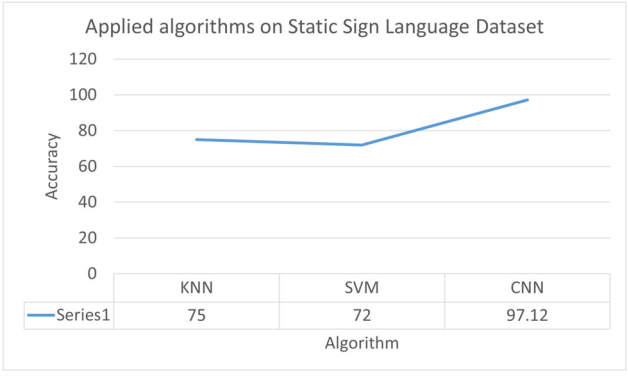


Figure 10: Traditional and deep learning algorithm results applied on the same dataset.

Table 5: Comparison of different machine learning algorithms based on different datasets

Author (year)	Technique	Gesture type	Accuracy (%)	Notes
Tu et al. [49], 2013	FFANN	HGR and face recognition	93.6	Automatic gesture segmentation and orientation normalization of hand were used
Islam et al. [48], 2017	FFANN	ASL and numeric numbers	94.32	real time recognition system using of 37 signs of numeric and alphabetic American characters
Dixit and Jalal [76], 2013	MSVM	ISL	**96**	Recognition of Indian sign language using 7Hu invariant moment and structural face descriptors, combining them to for new feature for sign recognition
Kang [76], 2015	CNN	Finger spelling	—	Real time sign language finger spelling recognition system using CNN
Utaminingrum [83], 2019	KNN	Alphabets	96	KNN classifier with $k = 1$ applied to a dataset of captured alphabets which were preprocessed to enhance images and detect hands using skin color algorithms, producing different accuracies of 94, 95, and 96% for dark, normal, and light images, respectively
Kamruzzaman [84], 2020	CNN	ArSL	90	Used CNN as a deep learning model to classify ArSL alphabets of 31 letters. Also, the proposed model produces a speech for the recognized letter
Varsha and Nair [85], 2021	CNN	ISL	93	Applied CNN model (inception model V3) on ISL, which receives its input as an image, achieving an average accuracy of 93%

Table 6: Comparison of deep learning of different sign language datasets focusing on technical parameters such as activation and optimization function, learning rate, and so on

Author	Deep learning algorithm	Dataset	Activation function	Learning rate (LR)	Epochs	Optimization function	Loss error	Accuracy (%)
Raheem and Abdulwahhab [44]	CNN	240 images of ASL	PReLU	0.01	100	SGDM	0.3274	99.3
			ReLU	0.01	100	RMSProp	0.485	83.33
Adaloglou et al. [82]	GoogLeNet + TConvs	—	—	0.0001:0.00001	10:25	ADAM optimizer	—	—
Bantupalli and Xie [83]	CNN + LSTM	2,400 images of ASL	SoftMax and pool layer	—	10	ADAM optimizer	0.3	91
Islam et al. [86]	CNN	—	ReLU & pool layer	0.001	60	SGD optimizer	—	97.12
Neethu et al. [43]	CNN	1,600 images	Average and max pool layer	—	—	—	—	96.2
Tolentino et al. [87]	CNN	ASL letters and [1–10] digits	SoftMax and max pool layer	0.01	50	SGD optimizer	—	93.67
Wangchuk et al. [88]	CNN	20,000 images of BSL digits [0–9]	VGGNet [ReLU & max pool]	—	34	ADAM optimizer	0.0021	97.62
Tolentino et al. [89]	CNN	35,000 images of ISL	ReLU	—	100	ADAM optimizer	—	**99.72**
Ferreira et al. [90]	CNN	1,400 images of ASL	SoftMax	0.001	—	SGD optimizer	—	93.17
Elboushaki et al. [91]	MultiD-CNN	isoGD	ReLU	0.001	—	ADAM optimizer	—	72.53
		SKIG						**99.72**
		NATOPS						95.87
		SBU						97.51
Kopuklu et al. [92]	CNN	EgoGesture	Average pooling and SoftMax	0.001	—	SGD optimizer	—	94.03
Yuxiao et al. [93]	DG-STA	DHG-14/28	Softmax	0.001	—	ADAM optimizer	—	91.9
Breland et al. [32]	CNN	Thermal dataset	ReLU	0.005	20	ADAM optimizer	—	**99.52**

ISL: Indian Sign Language.
Bold indicates highest results of applying different CNN models on various SL datasets.

Table 7: Distinction between CNN and SVM on different measurements

Performance analysis parameters (%)	CNN/SVM classification approach	Connected component analysis? CNN/SVM classification approach	AHE? connected component analysis? CNN/SVM classification approach
Sensitivity	91.5/89.1	96.8/90.5	98.1/92.1
Specificity	82.7/78.7	89.2/87.5	93.4/89.9
Accuracy	91.6/87.5	94.8/91.6	96.2/93.5
Recognition rate	90.7/88.2	96.2/90.5	98.7/91.5

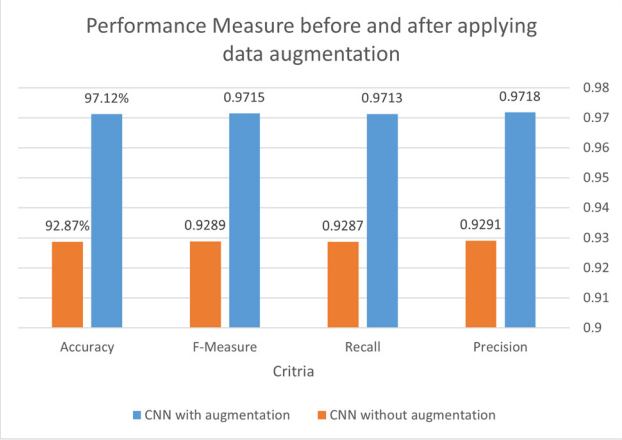


Figure 11: Performance measure before and after applying data augmentation.

Ref. [87] studied the effect of data-augmentation on deep learning algorithms, achieving an accuracy of 97.12% which is higher than that of the model before applying data augmentation by about 4%. Dataset consists of 10 static gestures to recognize, each class has 800 images for training and 160 for testing, resulting in 8,000 images for training and 1,600 for testing. This algorithm overcomes both SVM and KNN as shown in Figure 10, while being applied on the same dataset (Tables 5 and 6).

Ref. [43] uses CNN for hand gesture classification. First, the author used the algorithm of connected components analysis to select and segment hands from the image dataset using masks and filters, finger cropping, and segmentation. The author also used Adaptive Histogram equalization (AHE) for image enhancement to improve image contrast. CNN algorithm's accuracy was 96.2% which is higher than SVM classification algorithm applied by the author to achieve an accuracy of 93.5%. The following table illustrates this difference. Also, recognition time using CNN (0.356 s) is lower than SVM (0.647 s). As shown in Table 7, CNN exceeds SVM in different measurements like sensitivity, specificity, and accuracy.

Ref. [88] implemented training and testing using CNN by Keras and TensorFlow using SGD algorithm as its optimizer, having a learning rate of 0.01. The number of epochs is equal to 50 with a batch size of 500. Dataset has a set of static signs of letters, digits, and some words then resize of words to 50×50. Each class contains 1,200 images. The overall average accuracy of the system was 93.67%, of which 90.04, 93.44, and 97.52% for ASL alphabets, number recognition, and static word recognition, respectively. Tests were applied on 6 persons who were signer's interpreters and 24 students without any knowledge of using sign language (Figure 11).

Ref. [89] applied CNN algorithm on Bhutanese Sign Language digits recognition, collected dataset of 20,000 images of digits [0–9] from 21 students, each student was asked to capture 10 images per class. Images and videos were captured from different angles, directions, different backgrounds, and lighting conditions. Images were scaled to 64×64. TensorFlow was used as a deep learning library. Comparison with traditional ML was done and approved the superiority of deep learning CNN to SVM and KNN algorithms with average accuracy of 97.62% for CNN, 78.95% for KNN, and 70.25% for SVM, with lower testing time for CNN (Figure 11).

Ref. [90] applied CNN algorithm on ISL dataset which consists of distinct 100 images, generating 35,000 images of both colored and grayscale image types. The dataset includes digits [0–10] and 23 alphabets and about 67 most common words. Original image size of $126 \times 126 \times 16$ was reduced to $63 \times 63 \times 16$ using kernel filter of size 2. Many optimizers were applied such as ADAM, SGD, Adagrad, AdaDelta, RMSprop, and SGD. Using ADAM optimizer he achieved the best result of 99.17% and 98.8% for training and validation, respectively. Also, the proposed model accuracy exceeds other classifiers such as KNN (95.95%), SVM (97.9%), and ANN (98%).

6 Conclusion and future work

The variety of sign language datasets, which includes different gestures, leads to different accuracies as we

had discussed based on review of previous literature. This survey showed that different datasets have been used in the training and testing of SLR systems. It compared between vison-based approach and glove-based approach, showed the advantages and the disadvantages of both, illustrated the difference between signer dependent and signer independent, and addressed the basic preprocessing steps such as skin detector, image segmentation, hand tracking, feature extraction, and hand's gesture classification.

The survey also compares some ML techniques with the most used deep learning algorithm (CNN), showing that deep learning results exceed traditional ML. Some glove-based systems outperform deep learning algorithms due to the accurate signals that researchers get while feature extraction, while using deep learning their features get during model training which is not accurate as the gloves-based systems. According to this previous issue, we need to get rid of any obstacles (gloves, sensors, and leap devices) or any electronic device that may restrict user interaction with the system. Many trials had been done but with less accuracy.

Few researchers are working to solve SLID, although it is important for having a comprehensive SLR system. Including ArSL in our future work will be a challenging task. Also, trying to wear-off any gloves or any electric based systems will give user more comfort while communicating with others.

Conflict of interest: Authors state no conflict of interest.

Data availability statement: Data sharing is not applicable to this article as no new data were created or analyzed in this study.

References

[1] R. Kushalnagar, "Deafness and Hearing Loss," *Web Accessibility. Human–Computer Interaction Series*, Y. Yesilada, S. Harper, eds, London, Springer, 2019.

[2] World Federation of the Deaf. Our Work, 2018. http://wfdeaf.org/our-work/Accessed 2019–03–26.

[3] S. Wilcox and J. Peyton, "American Sign Language as a foreign language," *CAL. Dig.*, pp. 159–160, 1999.

[4] M. del Carmen Cabeza-Pereiro, J. M. Garcia-Miguel, C. G. Mateo, and J. L. A. Castro, "CORILSE: a Spanish sign language repository for linguistic analysis," *Proceedings of the Tenth International Conference on Language Resources and Evaluation* (*LREC'16*), 2016, May, pp. 1402–1407.

[5] T. Johnston and A. Schembri, *Australian Sign Language* (*Auslan*)*: An Introduction to Sign Language Linguistics*, Cambridge, UK, Cambridge University Press, 2007. ISBN 9780521540568. doi: 10.1017/CBO9780511607479.

[6] M. Abdel-Fattah, "Arabic Sign Language: A Perspective," *J. Deaf. Stud. Deaf. Educ.*, vol. 10, no. 2, 2005, pp. 212–221. doi: 10. 212-21. 10.1093/deafed/eni007.

[7] J. V. Van Cleve, *Gallaudet Encyclopedia of Deaf People and Deafness*, Vol 3, New York, New York, McGraw-Hill Company, Inc., 1987, pp. 344–346.

[8] D. Cokely, *Charlotte Baker-Shenk, American Sign Language*, Washington, Gallaudet University Press, 1981.

[9] U. Shrawankar and S. Dixit, Framing Sentences from Sign Language Symbols using NLP, *In IEEE conference*, 2016, pp. 5260–5262.

[10] N. El-Bendary, H. M. Zawbaa, M. S. Daoud, A. E. Hassanien, K. Nakamatsu, "ArSLAT: Arabic Sign Language Alphabets Translator," *2010 International Conference on Computer Information Systems and Industrial Management Applications* (*CISIM*)*, Krackow*, 2010, pp. 590–595.

[11] B. G. Gebre, P. Wittenburg, and T. Heskes, "Automatic sign language identification," *2013 IEEE International Conference on Image Processing*, Melbourne, VIC, 2013, pp. 2626–2630.

[12] D. Bragg, O. Koller, M. Bellard, L. Berke, P. Boudreault, A. Braffort, et al., "Sign language recognition, generation, and translation: an interdisciplinary perspective," *The 21st International ACM SIGACCESS Conference on Computers and Accessibility* (*ASSETS '19*), New York, NY, USA, Association for Computing Machinery, 2019, pp. 16–31.

[13] R. Rastgoo, K. Kiani, and S. Escalera, "Sign language recognition: A deep survey," *Expert. Syst. Appl.*, vol. 164, 113794, 2020.

[14] A. Sahoo, G. Mishra, and K. Ravulakollu, "Sign language recognition: State of the art," *ARPN J. Eng. Appl. Sci.*, vol. 9, pp. 116–134, 2014.

[15] A. Karpov, I. Kipyatkova, and M. Železný, "Automatic technologies for processing spoken sign languages," *Proc. Computer Sci.*, vol. 81, pp. 201–207, 2016. doi: 10.1016/j.procs.2016.04.050.

[16] F. Chou and Y. Su, "An encoding and identification approach for the static sign language recognition," *2012 IEEE/ASME International Conference on Advanced Intelligent Mechatronics* (*AIM*)*, Kachsiung*, 2012, pp. 885–889.

[17] https://en.wikipedia.org/wiki/Feedforward_neural_network.

[18] https://www.deafwebsites.com/sign-language/sign-language-other-cultures.html.

[19] D. Santiago, I. Benderitter, and C. García-Mateo, *Experimental Framework Design for Sign Language Automatic Recognition*, 2018, pp. 72–76. doi: 10.21437/IberSPEECH.2018-16.

[20] Z. Zafrulla, H. Brashear, P. Yin, P. Presti, T. Starner, and H. Hamilton, "American sign language phrase verification in an educational game for deaf children," IEEE, pp. 3846–3849, 2010, doi: 10.1109/ICPR.2010.937.

[21] K. B. Shaik, P. Ganesan, V. Kalist, B. S. Sathish, and J. M. M. Jenitha, "Comparative study of skin color detection and segmentation in HSV and YCbCr color space," *Proc. Computer Sci.*, vol. 57, pp. 41–48, 2015. doi: 10.1016/j.procs.2015.07.362.

[22] P. Dreuw, D. Rybach, T. Deselaers, M. Zahedi, and H. Ney, "Speech Recognition Techniques for a Sign Language Recognition System," *ICSLP, Antwerp, Belgium, August. Best Paper Award*, 2007a.

[23] K. Dixit and A. S. Jalal, "Automatic Indian Sign Language recognition system," *2013 3rd IEEE International Advance*

Computing Conference (*IACC*), *Ghaziabad*, 2013, pp. 883–887. doi: 10.1109/IAdCC.2013.6514343.

[24] I. Z. Onno Crasborn and J. Ros, "Corpus-NGT. An open access digital corpus of movies with annotations of Sign Language of the Netherlands," *Technical Report, Centre for Language Studies*, Radboud University Nijmegen, 2008. http://www.corpusngt.nl.

[25] M. Hassan, K. Assaleh, and T. Shanableh, "Multiple proposals for continuous arabic sign language recognition," *Sensing Imaging*, vol. 20, no. 1. pp. 1–23, 2019.

[26] A. Youssif, A. Aboutabl, and H. Ali, "Arabic sign language (ArSL) recognition system using HMM," *Int. J. Adv. Computer Sci. Appl.*, vol. 2, 2011. doi: 10.14569/IJACSA.2011.021108.

[27] O. Koller, J. Forster, and H. Ney, "Continuous sign language recognition: Towards large vocabulary statistical recognition systems handling multiple signers," *Computer Vis. Image Underst.*, vol. 141, pp. 108–125, 2015. doi: 10.1016/j.cviu.2015.09.013.

[28] M. Oliveira, H. Chatbri, Y. Ferstl, M. Farouk, S. Little, N. OConnor, et al., "A dataset for Irish sign language recognition," *Proceedings of the Irish Machine Vision and Image Processing Conference* (*IMVIP*), vol. 8, 2017.

[29] N. C. Camgoz, A. A. Kindiroğlu, S. Karabüklü, M. Kelepir, A. S. Ozsoy, and L. Akarun, BosphorusSign: a Turkish sign language recognition corpus in health and finance domains. In Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16), 2016, pp. 1383–1388.

[30] S. Ebling, N. C. Camgöz, P. B. Braem, K. Tissi, S. Sidler-Miserez, S. Stoll, and M. Magimai-Doss, "SMILE Swiss German sign language dataset," *Proceedings of the 11th International Conference on Language Resources and Evaluation* (*LREC*) *2018*, University of Surrey, 2018.

[31] N. M. Adaloglou, T. Chatzis, I. Papastratis, A. Stergioulas, G. T. Papadopoulos, V. Zacharopoulou, and P. Daras none, "A comprehensive study on deep learning-based methods for sign language recognition," *IEEE Trans. Multimedia*, pp. 1, 2021. doi: 10.1109/tmm.2021.3070438.

[32] A. Sahoo, "Indian sign language recognition using neural networks and kNN classifiers," *J. Eng. Appl. Sci.*, vol. 9, pp. 1255–1259, 2014.

[33] R. Rastgoo, K. Kiani, and S. Escalera, "Hand sign language recognition using multi-view hand skeleton," *Expert. Syst. Appl.*, vol. 150, p. 113336, 2020a.

[34] H. R. V. Joze and O. Koller, "MS-ASL: A large-scale dataset and benchmark for understanding American sign language. arXiv preprint arXiv:1812.01053," arXiv 2018, arXiv:1812.01053.

[35] D. Li, C. Rodriguez, X. Yu, and H. Li, "Word-level deep sign language recognition from video: A new large-scale dataset and methods comparison," *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, Snowmass Village, CO, USA, 1–5 March 2020, pp. 1459–1469.

[36] O. M. Sincan and H. Y. Keles, "AUTSL: A large-scale multi-modal Turkish sign language dataset and baseline methods," *IEEE Access*, vol. 8, pp. 181340–181355, 2020.

[37] A. A. I. Sidig, H. Luqman, S. Mahmoud, and M. Mohandes, "KArSL: Arabic sign language database," *ACM Trans. Asian Low-Resour. Lang. Inf. Process*, vol. 20, pp. 1–19, 2021.

[38] D. S. Breland, S. B. Skriubakken, A. Dayal, A. Jha, P. K. Yalavarthy, and L. R. Cenkeramaddi, "Deep learning-based sign language digits recognition from thermal images with edge computing system," *IEEE Sens. J.*, vol. 21, no. 9. pp. 10445–10453, 2021.

[39] A. Mittal, P. Kumar, P. P. Roy, R. Balasubramanian, and B. B. Chaudhuri, "A modified LSTM model for continuous sign language recognition using leap motion," *IEEE Sens. J.*, vol. 19, no. 16. pp. 7056–7063, 2019. doi: 10.1109/jsen.2019.2909837.

[40] O. Koller, S. Zargaran, H. Ney, and R. Bowden, "Deep sign: enabling robust statistical continuous sign language recognition via hybrid CNN-HMMs," *Int. J. Comput. Vis.*, vol. 126, pp. 1311–1325, 2018.

[41] I. Hernández, *Automatic Irish sign language recognition*, Trinity College, Diss. Thesis of Master of Science in Computer Science (Augmented and Virtual Reality), University of Dublin, 2018.

[42] P. S. Neethu, R. Suguna, and D. Sathish, "An efficient method for human hand gesture detection and recognition using deep learning convolutional neural networks," *Soft Comput.*, vol. 24, pp. 15239–15248, 2020. doi: 10.1007/s00500-020-04860-5.

[43] C. D. D. Monteiro, C. M. Mathew, R. Gutierrez-Osuna, F. Shipman, Detecting and identifying sign languages through visual features, *2016 IEEE International Symposium on Multimedia* (*ISM*), 2016. doi: 10.1109/ism.2016.0063.

[44] F. Raheem and A. A. Abdulwahhab, "Deep learning convolution neural networks analysis and comparative study for static alphabet ASL hand gesture recognition," *Xi'an Dianzi Keji Daxue Xuebao/J. Xidian Univ.*, vol. 14, pp. 1871–1881, 2020. doi: 10.37896/jxu14.4/212.

[45] A. Kumar and S. Malhotra, *Real-Time Human Skin Color Detection Algorithm Using Skin Color Map*, 2015.

[46] Y. R. Wang, W. H. Li and L. Yang, "A Novel real time hand detection based on skin color," *17th IEEE International Symposium on Consumer Electronics* (*ISCE*), 2013, pp. 141–142.

[47] K. Sheth, N. Gadgil, and P. R. Futane, "A Hybrid hand detection algorithm for human computer interaction using skin color and motion cues," *Inter. J. Computer Appl.*, vol. 84, no. 2. pp. 14–18, December 2013.

[48] M. M. Islam, S. Siddiqua, and J. Afnan, "Real time hand gesture recognition using different algorithms based on American sign language," *2017 IEEE International Conference on Imaging, Vision & Pattern Recognition* (*icIVPR*), 2017. doi: 10.1109/icivpr.2017.7890854.

[49] Y.-J. Tu, C.-C. Kao, and H.-Y. Lin, "Human computer interaction using face and gesture recognition," *2013 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference*, 2013. doi: 10.1109/apsipa.2013.6694276.

[50] M. Kawulok, "Dynamic skin detection in color images for sign language recognition," *Image Signal. Process*, vol. 5099, pp. 112–119, 2008.

[51] S. Bilal, R. Akmeliawati, M. J. E. Salami, and A. A. Shafie, "Dynamic approach for real-time skin detection," *J. Real-Time Image Proc.*, vol. 10, no. 2. pp. 371–385, 2015.

[52] N. Ibrahim, H. Zayed, and M. Selim, "An automatic arabic sign language recognition system (ArSLRS)," *J. King Saud.*

Univ. – Computer Inf. Sci., Vol. 30, no. 4, October 2018, Pages 470–477. doi: 10.1016/j.jksuci.2017.09.007.

[53] M. P. Paulraj, S. Yaacob, Z. Azalan, M. Shuhanaz, and R. Palaniappan, *A Phoneme-based Sign Language Recognition System Using Skin Color Segmentation*, 2010, pp. 1–5. doi: 10.1109/CSPA.2010.5545253.

[54] T. Simon, H. Joo, I. Matthews, and Y. Sheikh, "Hand Keypoint Detection in Single Images Using Multiview Bootstrapping" *2017 IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), 2017, pp. 4645–4653. doi: 10.1109/CVPR.2017.494.

[55] R. Akmeliawati, "Real-time Malaysian sign language translation using colour segmentation and neural network", *Proc. of the IEEE International Conference on Instrumentation and Measurement Technology 2007, Warsaw*, 2007, pp. 1–6.

[56] J. Lim, D. Lee, and B. Kim, "Recognizing hand gesture using wrist shapes," *2010 Digest of Technical Papers of the International Conference on Consumer Electronics* (*ICCE*)*, Las Vegas*, 2010, pp. 197–198.

[57] O.Al-Jarrah and A. Halawani, "Recognition of gestures in Arabic sign language using neuro-fuzzy systems," *Artif. Intell.*, vol. 133, pp. 117–138, 2001. doi: 10.1016/S0004-3702(01)00141-2.

[58] M. A. Hussain, *Automatic recognition of sign language gestures*, Master's Thesis. Jordan University of Science and Technology, Irbid, 1999.

[59] C. Oz and M. C. Leu, "American sign language word recognition with a sensory glove using artifcial neural networks," *Eng. Appl. Artifcial Intell.*, vol. 24, no. 7. pp. 1204–1213, Oct. 2011.

[60] M. W. Kadous, "Machine recognition of Auslan signs using PowerGloves: Towards large-lexicon recognition of sign language," *Proceedings of the Workshop on the Integration of Gesture in Language and Speech*, Wilmington, DE, USA, 1996, pp. 165–174.

[61] N. Tubaiz, T. Shanableh, and K. Assaleh, "Glove-based continuous Arabic sign language recognition in user-dependent mode," *IEEE Trans. Human-Mach. Syst.*, vol. 45, no. 4. pp. 526–533, 2015.

[62] P. D. Rosero-Montalvo, P. Godoy-Trujillo, E. Flores-Bosmediano, J. Carrascal-Garcia, S. Otero-Potosi, H. Benitez-Pereira, et al., "Sign language recognition based on intelligent glove using machine learning techniques," *2018 IEEE Third Ecuador Technical Chapters Meeting* (*ETCM*), 2018. doi: 10.1109/etcm.2018.8580268.

[63] L. Chen, J. Fu, Y. Wu, H. Li, and B. Zheng, "Hand gesture recognition using compact CNN via surface electromyography signals," *Sensors*, vol. 20, no. 3. p. 672, 2020. doi: 10.3390/s20030672.

[64] D. Aryanie and Y. Heryadi, "American sign language-based finger-spelling recognition using k-Nearest Neighbors classifier." *2015 3rd International Conference on Information and Communication Technology* (*ICoICT*), 2015, pp. 533–536.

[65] F. Utaminingrum, I. Komang Somawirata, and G. D. Naviri, "Alphabet sign language recognition using K-nearest neighbor optimization," *JCP*, vol. 14, no. 1. pp. 63–70, 2019.

[66] A. Jadhav, G. Tatkar, G. Hanwate, and R. Patwardhan, "Sign language recognition," *Int. J. Adv. Res. Comput. Sci. Softw. Eng.*, vol. 7, pp. 109–115, no. 3, 2017.

[67] U. Patel and A. G. Ambekar, "Moment Based Sign Language Recognition for Indian Languages," *2017 International Conference on Computing, Communication, Control and Automation* (*ICCUBEA*), 2017, pp. 1–6. doi: 10.1109/ICCUBEA.2017.8463901.

[68] G. Saggio, P. Cavallo, M. Ricci, V. Errico, J. Zea, and M. E. Benalcázar, "Sign language recognition using wearable electronics: implementing k-Nearest Neighbors with dynamic time warping and convolutional neural network algorithms," *Sensors*, vol. 20, no. 14. p. 3879, 2020. doi: 10.3390/s20143879.

[69] A. K. Sahoo, "Indian sign language recognition using machine learning techniques," *Macromol. Symp.*, vol. 397, no. 1. p. 2000241, 2021. doi: 10.1002/masy.202000241.

[70] Z. Parcheta and C.-D. Martínez-Hinarejos, "Sign language gesture recognition using HMM," in *Pattern Recognition and Image Analysis. Lecture Notes in Computer Science 2017*. L. Alexandre, J. Salvador Sánchez, J. Rodrigues, (Eds), IbPRIA, vol. 10255, Cham: Springer, pp. 419–426, 2017. doi: 10.1007/978-3-319-58838-4_46.

[71] T. Starner, J. Weaver, and A. Pentland, "Real-time American sign language recognition using desk and wearable computer-based video," *IEEE Trans. Pattern Anal. Mach. Intellig.*, vol. 20, no. 12. pp. 1371–1375, 1998.

[72] C. Zimmermann and T. Brox, "Learning to estimate 3D hand pose from single RGB images," *2017 IEEE International Conference on Computer Vision* (*ICCV*), 2017, pp. 4913–4921.

[73] D. Victor, *Real-Time Hand Tracking Using SSD on TensorFlow*, GitHub Repository, 2017.

[74] K. Dixit and A. S. Jalal, "Automatic Indian sign language recognition system," *2013 3rd IEEE International Advance Computing Conference* (*IACC*), 2013. doi: 10.1109/iadcc.2013.6514343.

[75] B. Kang, S. Tripathi, and T. Nguyen, "Real-time sign language fingerspelling recognition using convolutional neural networks from depth map," *3rd IAPR Asian Conference on Pattern Recognition*, Kuala Lumpur, Malaysia, 2015. doi: 10.1109/acpr.2015.7486481.

[76] https://en.wikipedia.org/wiki/Backpropagation.

[77] A. M. Jarman, S. Arshad, N. Alam, and M. J. Islam, "An automated bengali sign language recognition system based on fingertip finder algorithm," *Int. J. Electron. Inform.*, vol. 4, no. 1. pp. 1–10, 2015.

[78] P. P. Roy, P. Kumar, and B. -G. Kim, "An efficient sign language recognition (SLR) system using camshift tracker and hidden markov model (HMM)," *SN Computer Sci.*, vol. 2, 79, no. 2, 2021. doi: 10.1007/s42979-021-00485-z.

[79] S. Ghanbari Azar and H. Seyedarabi, "Trajectory-based recognition of dynamic persian sign language using hidden Markov Model," *arXiv e-prints*, p. arXiv-1912, 2019.

[80] N. M. Adaloglou, T. Chatzis, I. Papastratis, A. Stergioulas, G. T. Papadopoulos, V. Zacharopoulou, and P. Daras, "A Comprehensive Study on Deep Learning-based Methods for Sign Language Recognition," IEEE Transactions on Multimedia, p. 1, 2021. doi: 10.1109/tmm.2021.3070438.

[81] K. Bantupalli and Y. Xie, "American sign language recognition using deep learning and computer vision," *2018 IEEE International Conference on Big Data* (*Big Data*)*, Seattle, WA,*

USA, 2018, pp. 4896–4899. doi: 10.1109/BigData.2018.8622141.

[82] F. Utaminingrum, I. Komang Somawirata, and G. D. Naviri, "Alphabet sign language recognition using K-nearest neighbor optimization," *J. Comput.*, vol. 14, no. 1. pp. 63–70, 2019.

[83] M. M. Kamruzzaman, "Arabic sign language recognition and generating Arabic speech using convolutional neural network," *Wirel. Commun. Mob. Comput.*, vol. 2020, pp. 1–9, 2020. doi: 10.1155/2020/3685614.

[84] M. Varsha and C. S. Nair, "Indian sign language gesture recognition using deep convolutional neural network," *2021 8th International Conference on Smart Computing and Communications* (*ICSCC*), IEEE, 2021.

[85] M. Z. Islam, M. S. Hossain, R. ul Islam, and K. Andersson, "Static hand gesture recognition using convolutional neural network with data augmentation," *2019 Joint 8th International Conference on Informatics, Electronics & Vision* (*ICIEV*) *and 2019 3rd International Conference on Imaging, Vision & Pattern Recognition* (*icIVPR*)*, Spokane, WA, USA*, 2019, pp. 324–329. doi: 10.1109/ICIEV.2019.8858563.

[86] L. K. S. Tolentino, R. O. Serfa Juan, A. C. Thio-ac, M. A. B. Pamahoy, J. R. R. Forteza, and X. J. O. Garcia, "Static sign language recognition using deep learning," *Int. J. Mach. Learn. Comput.*, vol. 9, no. 6. pp. 821–827, 2019.

[87] K. Wangchuk, P. Riyamongkol, and R. Waranusast, "Real-time Bhutanese sign language digits recognition system using convolutional neural network," *ICT Exp.*, vol. 7, no. 2, pp. 215–220, 2020. doi: 10.1016/j.icte.2020.08.002.

[88] L. K. Tolentino, R. Serfa Juan, A. Thio-ac, M. Pamahoy, J. Forteza, and X. Garcia, "Static sign language recognition using deep learning," *Int. J. Mach. Learn. Comput.*, vol. 9, pp. 821–827, 2019. doi: 10.18178/ijmlc.2019.9.6.879.

[89] P. M. Ferreira, J. S. Cardoso, and A. Rebelo, "Multimodal Learning for Sign Language Recognition," *Pattern Recognition and Image Analysis. IbPRIA 2017. Lecture Notes in Computer Science*(), L. Alexandre, J. Salvador Sánchez, and J. Rodrigues, (eds), vol. 10255, Cham, Springer, 2017. doi: 10.1007/978-3-319-58838-4_35.

[90] A. Elboushaki, R. Hannane, A. Karim, and L. Koutti, "MultiD-CNN: a multi-dimensional feature learning approach based on deep convolutional networks for gesture recognition in RGB-D image sequences," *Expert. Syst. Appl.*, vol. 139, p. 112829, 2019. doi: 10.1016/j.eswa.2019.112829.

[91] O. Kopuklu, A. Gunduz, N. Kose, and G. Rigoll, "Real-time hand gesture detection and classification using convolutional neural networks," *2019 14th IEEE International Conference on Automatic Face & Gesture Recognition* (*FG 2019*), 2019. doi: 10.1109/fg.2019.8756576.

[92] Ch. Yuxiao, L. Zhao, X. Peng, J. Yuan, and D. Metaxas, *Construct Dynamic Graphs for Hand Gesture Recognition Via Spatial-temporal Attention, UK*, 2019, pp. 1–13. https://bmvc2019.org/wp-content/uploads/papers/0281-paper.pdf.

[93] A. Z. Shukor, M. F. Miskon, M. H. Jamaluddin, F. Bin Ali, M. F. Asyraf, and M. B. Bin Bahar., "A new data glove approach for malaysian sign language detection," *Procedia Computer Science*, vol. 76, pp. 60–67, 2015, doi: 10.1016/j.procs.2015.12.276.