Kerbala Journal for Engineering Science
https://kjes.uokerbala.edu.iq/

Sign Language Recognition and Hand Gestures Review

Atyaf Hekmat Mohammed Ali [a]*, Hawraa H. Abbas [a], Haider Ismael Shahadi [a]

[a] Electrical and Electronic Department, Collage of Engineering, University of Kerbala, Karbala, Iraq.
* Corresponding author, Email: atyaf.h@s.uokerbala.edu.iq

Received: 07 July 2022; Revised: 12 October 2022; Accepted: 29 December 2022; Published: 31 December 2022

Abstract

Deaf people use movements and physical expressions to reveal their ideas and feelings to their world. These expressions are called 'sign language', and like natural languages, there are many forms of signs worldwide. Deaf use one or two hands and sometimes use other body parts like the head, lips or eyes. Their gestures are by either static or dynamic hands, and they are a bit of complicated language. Therefore, other people need to understand the meaning of each of these signs and gestures to communicate with the Deaf community successfully. Human-computer interaction is an effective tool and an excellent trend to facilitate the communication and comprehension of the different sign languages used worldwide. The research community has tried to review the most important techniques and models used in deciphering and understanding sign languages. Every new research effort is directed towards improving these ways of communication. Some proposed models dealt with isolated signs, and others focused on continuous signs. This article represents a summary of multiple comprehensive reviews that studied different literature conducted on sign language recognition. The discussion in this review focuses on the systems and approaches that only deal with static hand gesture recognition. This work aims to provide a guide for researchers and practitioners to relate their work to existing research and gain insights into what their work can contribute to the field.

Keywords: Sign language review; feature extraction; feature selection; deep learning; static hand recognition

1. Introduction

There are about 90 million deaf people around the world who together use more than 300 different types of sign languages [1]. Sign language represents a unique personal communication between deaf people and their community. Visual gestures, hand movements, signs and facial expressions convey meanings and information. Single and combinations of these bodily movements can express letters, numbers, words and sentences. The use of sign language is not limited only to personal interaction between people, but it can also be employed in smart interactive environments. For instance, a hand gesture can be used instead of a vocal term to refer to a function or an option in smart home applications. Technically, according to their temporal nature, there are two categories of hand gestures used in sign languages: static hand gestures and dynamic hand gestures. The importance of studying sign language linguistics and classifying their gestures lies in how deaf people use the language elements and how they can be interpreted and comprehended by receivers from other people to whom the sign language messages are directed.

Human-computer interaction (HCI) has become an indispensable part of daily life and can be found in many applications. Gesture recognition for sign language interpretation is one of these applications in which images of hand gestures of a sign language are input to a computer, processed and converted into audible or written output signals understandable by users, including deaf people. Therefore, such HCI algorithms must be reliable, fast, adaptable and precise, especially when they involve deaf people in emergencies. Thus, based on these requirements, there can be different methods and techniques of gesture input, gesture processing and output interpretation.

The remainder of this paper is organized as follows; Section (2) describes the types of hand gestures in sign language. Section (3) presents an overview of sign language recognition approaches. Section (4) discusses the different methods of sign identification systems. Section (5) presents a comprehensive review of related literature. Section (6) focuses on static hand sign language. Finally, section (7) recaps the paper by discussing the conclusions and insights for future work.

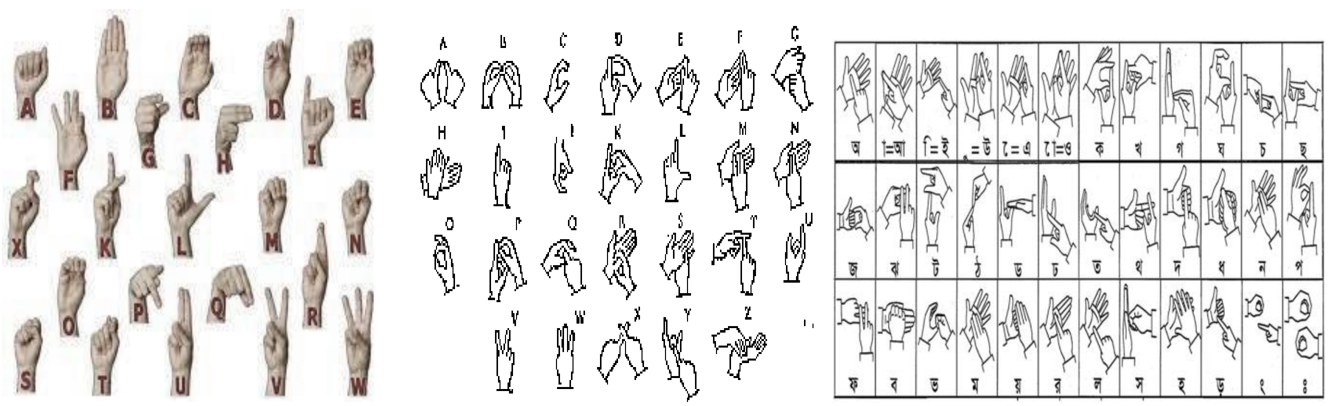
2. Hand Gesture Types in Sign Language

In any language, there are dozens of letters, thousands of words, and an infinite number of sentence formations used by speakers to communicate with each other in that language. Similarly, deaf people express their feelings and ideas using groupings of hand signs and gestures that can be classified according to their temporal relation into static and dynamic signs [2].

2.1 Static Hand Gesture

In this category of gestures, the position of the hand does not affect the gesture's meaning and the gestures are time-independent [3]. n American sign language (ASL), static one-hand gestures refer to the English alphabet, as shown in Figure 1 (a) Therefore, these signs are very important because signers use them to spell out names such as their names, names of places they want to go to, names of food they want to eat or to name anything that does not have special signs [4] .

In Indian sign language (ISL), the letters of the alphabet are represented by both 2-static and 1-static hands [2] as shown in Figure 1 (b) In Bangla sign language (BSL), which is the common sign language used in Bangladesh, the alphabet is represented by 2-static hands [5] as shown in Figure 1 (c) This type of gesture can be delivered to the recognition system individually to be deciphered by one. It can also be input to the system continuously in the form of a video that will be divided into frames to be recognized by the recognition algorithm [6].



(a) (b) (c)

Figure 1 Hand gesture alphabet (a) [ASL] American [7], (b) [ISL] Indian [8], (c) [BdSL] Bangla [5].

2.2 Dynamic Hand Gestures

Dynamic gestures are time-dependent, and any change in the position of the hand gesture means that the signer wants to express a new idea [3]. Therefore, dynamic gestures can convey a wide range of meanings. For instance, to sign “mom” in ASL, the signer places the tip of the thumb of their open hand against the chin. The same hand sign can mean “dad” by placing the thumb against the forehead. Figure 2 shows some dynamic gestures and their meanings in ASL, where the arrow represents the direction of the movement. Typically, dynamic hand gestures are converted into a sequence of image frames, and their recognition depends on the time that every frame enters the recognition system [9].

Figure 2 Some dynamic gestures in ASL [10].

3. Overview of Sign Language Recognition Approaches

According to the technique of inputting the gesture into the computing system, there are three approaches to sign language recognition (SLR): sensor-based, vision-based and hybrid-based [11] as shown in Figure 3.

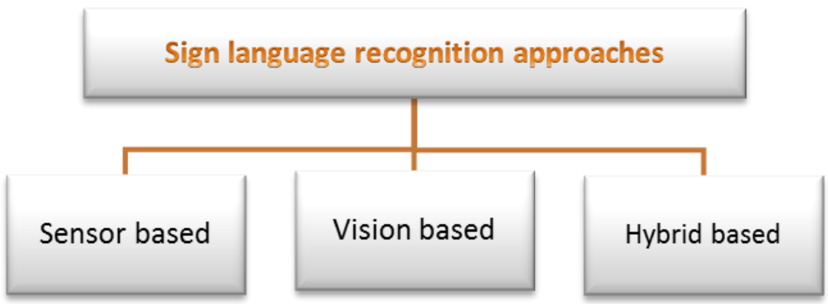


Figure 3 [SLR] Approaches [11].

3.1 A Sensor-Based Strategy

This approach uses sensors and computation electronic circuits to acquire and condition the input data, as shown in Figure (4) in the sensory gloves system. The advantages of this technique include simplicity because no complex data processing is required, flexibility because there are no restrictions on movements such as sitting behind a desk or chair, reliability as hand shape recognition is unaffected by background conditions, and its lightweight as well as mobility that enables the device to be carried easily and comfortably [11].

The sensory glove contains different sensors such as an accelerometer and gyroscope for obtaining the orientation of the hand and fingers and also angle and acceleration information. Some gloves also contain flex sensors that provide the system with finger-bending information [12].

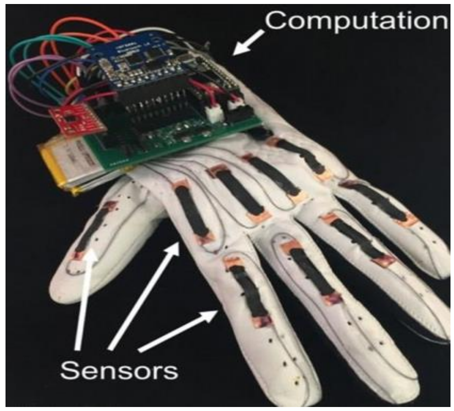


Figure 4 Sensory glove system [13].

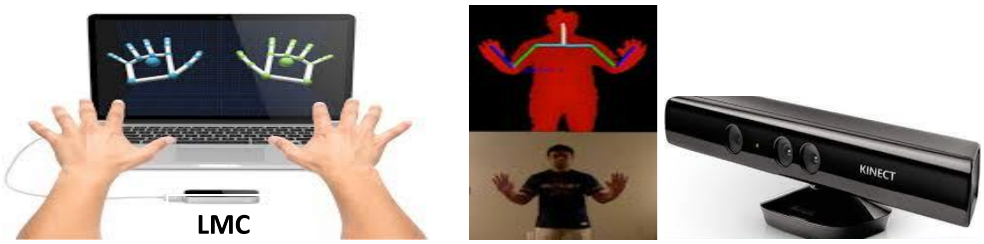
3.2 A Vision-Based Strategy

This approach depends on the acquisition of images or many frames of one video taken by a camera such as a webcam or a smartphone camera. It is considered flexible, simple and cost-effective as it does not use expensive advanced sensors. Compared to the sensory glove system, this approach does not pose personal safety concerns such as skin damage, burns and the spread of infectious diseases [14]. The vision-based technique was first used in 1988 for recognizing Japanese SL using a normal analogue camera which caused distortion in the images so filters were required to reduce the noise [1].

In this approach, there is a need for image preprocessing and processing to get the feature vector. It also requires an effective image-capturing environment that takes into consideration brightness levels, shading, skin tone variations, etc. However, this approach is not without its disadvantages which are represented by real-time computational delay, the need for high computational power and a low recognition rate [15].

3.3 A Hybrid Based Strategy

As the name suggests, this technique combines sensors and image-capturing cameras to acquire the input data, such as combining gloves of sensors, leap motion, and Microsoft Kinect. It enables designers to benefit from the advantages of spatial sensors that support the performance of computer vision methods [16]. The leap motion controller {LMC} in Figure 5 (a) is a sensor sensitive to position and motion information. It contains 3 infrared LEDs and 2-IR cameras [17] while Microsoft Kinect {MSK} sensors in Figure 5 (b) are sensitive to depth and skeleton shape. This device provides the RGB information obtained from the captured images [18]. The main difference between smart gloves and {LMC, MSK} devices is that the gloves are wearable. Using the three aforementioned devices, different combinations can be formed. For example, [16] used {camera+ gloves} while [19] used {webcam+ MSK}. [20] used {2-digital cameras+ LMC}. In these three contributions, researchers discussed dynamic gestures and their datasets. However, dynamic gestures are out of the scope of this study and only static gestures are considered.



(a) (b)

Figure 5 (a) Leap motion controller (LMC), (b) Microsoft Kinect (MSK) [21].

4. Methods for Sign Identification Systems

The other aspect that was discussed by researchers in the field of gesture recognition is the methods of processing the information of acquired sign gestures. This topic was studied in two research directions:

4.1 Feature Extraction Technique

Feature extraction can be defined as the process of identifying the information that must be extracted from the input image after applying some preprocessing functions. The extracted features are either local or global based on what image traits they describe. For instance, local features describe a segment of the image, i.e., a single or a certain group of the image pixels, such as the colour or brightness of these pixels. On the other hand, global features are related to the entire image, such as the image shape, texture and contours, etc. Therefore, various extraction techniques were developed with different performance levels and capabilities, such as (HOG, LBP, SURF, SIFT, DWT, GLCM,..). Researchers also investigated blended techniques where two or more of these methods were integrated [22].

After the dataset is collected from any camera, whether it is a webcam or obtained from a website and preprocessing is performed, features of the image are extracted by one of the methods above then the system must convert the information of features into a vector form for every image of the dataset. The vector's size depends on the feature extraction method used. Then, the vectors for all images of the dataset must be transformed into a matrix, and this matrix is carried over to the next step in the process [23].

4.1.1 Feature Selection Technique

Feature selection refers to the process of selecting the most important or dominant features from the complete set of extracted features. Feature selection algorithms can be divided into three types:

- Filter methods examine the data intrinsic features while ignoring the classifier.
- Embedded methods include the process of selection within the classifier learning step.
- Wrapper methods utilize classifiers for scoring a suitable subset of features.

In most of these methods, two operations are performed concurrently: subset selection and feature ranking. In other methods, these two operations are performed sequentially, with ranking done first, followed by the selection function. In this regard, programmers have proposed many algorithms, including (CFS, ILFS, DGUFS, Relief, LASSO,..). After extracting features, the feature selection technique loads the obtained information to one of these algorithms to select the most relevant features important for the used model in the subsequent step. This ranking and minimizing step will provide efficient characteristics like speed and little memory usage, which make the classification process more productive [24].

4.1.2 Classification Techniques

Many traditional methods like (SVM [25], KNN [26], Random Forest [27], Linear Discriminant Analysis LDA[28],..) are used by researchers to provide the machine with the ability to learn from the input data how to distinguish between the multi classes in the dataset of hand gestures. For a certain type of input information, some of these methods are suitable and effective while other methods might give undesirable results. Also, there is a differentiation between the algorithm application fundamentals. In a support vector machine (SVM), there are (linear, polynomial, and radial function) kernels, while in K-Nearest Neighbor (KNN) the number of the neighbours (K) that are used in the calculation of the distance significantly impacts the results.

4.2 Deep Learning Technique

In this technique, most of the tasks are done by the machine. It involves multi-layers of filters, multi-layers of comparisons, convolutional layers, hidden feature extraction layers, and classification. Deep learning using convolution neural networks [CNN] has recently been a common tool used for hand gesture recognition. CNN can be applied in many algorithms such as (Alexnet, VGG, Googlenet, squeeze net, YOLOnet,..). However, despite the fact that this technique offers high accuracy, it needs more computing time [29].

5. Related Work

Sign language recognition has received a great deal of attention from many researchers who dedicated their work to discussing the techniques and concepts highlighted in the previous sections. In this section, a comprehensive review of a considerable body of recent literature is presented. Table 1 lists the reviewed work, its scope, currency (year of publication) and the number of Certified references. The purpose of this review is to function as a guide for interested researchers to relate their work to existing research and to gain insights into what their work can contribute to the field. Many attempts of systematic review and analysis of SLR techniques have been identified in the literature. Some of these studies covered the work published over a limited period such as in [11], which covered a period of 10 years from 2007 to 2017, [30] and [31] surveyed the work in the last 5 and 10 years, respectively. However, [32] went beyond this limit and extended their review period to 20 years, from 2001 to 2021. Other contributions targeted a specific sign language in their review, such as in [2][22][8] which limited their discussion to Indian SL only, and [33] and [34] focused on Chinese and Arabic SL, respectively.

On the other hand, most of the authors structured their review around the SLR approaches. For example, [11] has collected and analyzed research that talked about **different sensor gloves** with different numbers and types of sensors (proximity, accelerometers (ACCs), flexion, and abduction) sensors. The purpose of these sensors was either for finger bending detection or for detecting hand movement and orientation. Hardware aspects such as the microcontroller and its processing unit, which represents the mind of the system, were also considered. Some of the reviewed systems were compared according to the architecture and type of their microcontrollers, such as microchips with 8-bit AVR or central processor module, ARM7.

The **vision-based approach** has also been widely investigated by researchers for its outstanding performance and results [2][14][22][30] [35]. Authors in [14] discussed different strategies of the vision-based approach as follows:

- Colour-based recognition with two sides: using (glove marker& skin colour)
- Appearance-based recognition with 2D and 3D images by extracted features to get a visual appearance model.
- Motion- based recognition uses a series of image frames and then extracts the features.
- Skeleton-based recognition by making the geometric attributes as a description and constraints for getting the skeletal joints characteristics of the hand gesture.
- Depth-based recognition by using different types of cameras to get 3D geometric information.
- Deep learning-based recognition was a modern application because of the use of a learning role principle and multilayers for learning data. The biggest challenge facing this technique is the requirement of the dataset to train algorithms which may affect the processing time.

Authors in [2] have focused their work on a different aspect which is the elements of the **vision-based system** which include hand tracking and segmentation, feature extraction, and classification and recognition. Also, a discussion was made on the difficulties faced by researchers in the vision-based system such as in the segmentation step. These challenges are caused by skin tone variations, lighting conditions, or how to detect hand gestures in a complex background. For skin color detection, YCbCr and HSV colour models were effective due to their ability to separate chrominance and luminance components. Bayesian classifier and K-mean clustering were also used for this purpose.

Besides, information about the hand position such as motion, blob, and colour tracking can be obtained from 2D Tracking algorithms such as camshaft, mean-shift, and viola jones. These techniques achieve a successful segmentation, especially with a background that is complex. For features extraction, the author offers two strategies: (i) Contour-based shape description and representation methods like Fourier Descriptor and Wavelet Descriptor, and (ii) Region-based shape description and representation methods like Zernike Moments, and Geometric Moments.

Some studies had potential limitations such as the work in [22] in which the reviewed feature extraction methods were limited to ISL only focusing on the vision-based approach only. Furthermore, the work reviewed the literature from 2010 to 2019 only. The work has highlighted several observations about the feature extraction techniques:

- There was a statistical category for techniques that make use of statistical parameters (Zernike moments and Contour moments)
- Shape transform-based technique based on shape extraction phenomenon which was invariant to a translation like (Fourier Descriptor and Discrete Wavelet Transform)
- The two techniques above are categorized as Content-based image retrieval (CBIR) because they can make feature extraction based on image content like color, texture and shape.
- Soft computing techniques were invariant to illumination and outplace the use of image preprocessing (SIFT & SURF& HOG).
- Hybrid techniques which combine (CBIR) and soft computing were also identified.

Deep learning with vision-based was proposed in [30]. The study discussed the most significant deep learning models used for hand gesture recognition including Convolutional Neural Network (CNN), Deep Boltzmann Machine (DBM), Generative Adversarial Networks (GAN), and Recursive Neural Networks (RNN) which consists of Long Short Term Memory (LSTM) and Gated Recurrent Unit (GRU).

CNN was mostly used with static hand gestures; it was also used with dynamic gestures. However, researchers preferred the use of RNN-LSTM for dynamic gestures because of its effectiveness with time. In [35], the author has displayed some proposed systems and made a comparison between their results.

In [32] the authors presented a critical review and analysis of sign language recognition methods. The study emphasized the importance of the topic and showed that it has been extensively researched. The authors stated that using the keyword “sign language recognition” in the search engine of the Scopus database returned 1321 results, i.e., scholarly papers. At last, the study reviewed different techniques of image segmentation such as thresholding, edge detection, region-based, clustering-based segmentation, or ANN-based segmentation.

Table 1 Different studies on sign language recognition.

S.	Ref.	Determinant of the study	Publisher \Year	No. of Certified references
1	[11]	Based Sensory Gloves for(2007 ,2017)	Sensors \2018	(98)research
2	[12]	For all	Springer \2019	(150) research
3	[14]	computer vision	journal of Imaging\2020	(100) research
4	[1]	the recent trends only	IEEE \2020	(22) research
5	[15]	Machine learning based	Springer \ 2021	(150) research
6	[2]	vision based,(ISL)	International Journal on smart sensing and intelligent systems\2014	(58)research
7	[22]	vision-based, Feature extraction, (ISL)	Springer\2021	(81) research
8	[30]	Vision-based, deeplearning, last 5-years	ELSEIVER \ 2021	(128) research
9	[38]	For all	ARPN Journal \ 2014	(55) research
10	[23]	Feature extraction	IEEE \2018	(34) research
11	[32]	Machinelearning methods for(2001-2021)	ELSEIVER \2021	(300)research
12	[36]	feature extraction , image-based	(IJCSE)\ 2017	(60) research
13	[8]	For all: Indian Sign Language	IRJET\ 2016	(15) research
14	[34]	Arabic(SL) ,advanced machine learning	Springer\2021	(85) research
15	[21]	image processing	(IJERT) \2016	(20) research
16	[31]	For 10-years	Springer\ 2021	(134) research
17	[39]	For all	IEEE \2021	(16)research
18	[35]	Vision-based	IEEE \2020	(32)research
19	[37]	Real-Time, (EMG), Machine Learning	Sensors\2020	(132)research
20	[33]	Artificial Intelligence, Chinese (SL)	Springer\ 2020	(158)research

Feature extraction techniques were the main focus of the studies [23], [36] which presented a comparison between some features such as (PCA and SIFT) and the classifiers used like (SVM and MDC). In [37], real timelessness and surface electromyography with machine learning were studied. Surface electromyography represents a non-medical procedure with which the activity levels of skeletal muscles are measured. The measurement is implemented by measuring the electrical potential produced by the muscles using surface sensors. Then, the behaviour of the muscles is studied by analyzing the acquired electrical signals.

The electrical activity is produced from 2-states of a skeletal muscle: i. when a skeletal muscle is at rest, an electric potential of (–80 mV) is produced by each muscular cell, ii. when a skeletal muscle is contracted to produce the electric potential in a motor unit. Static and dynamic muscular contractions are the two types of muscle contractions considered in the test. When someone holds his or her hand motionless or makes the peace sign, the lengths of the muscle fibers do not alter and the joints do not move, but the muscle fibers still contract. When someone waves their hand to say hello, there are variations in the lengths of the muscle fibers and the joints are in motion during a dynamic contraction.

Different techniques of sign language recognition were also studied in [12] [38][39].

6. A Static-Hand Based Sign Language Review

This section reviews of the different recognition techniques that deal with static gestures of sign languages.

Starting with the **sensor-based approach**, [17] used a **leap motion controller** (LMC) with 28 static one-hand gestures that represented the (ArSL) alphabet. For every letter, 100 frames were used, therefore the total images of the whole alphabet were (2800). The (LMC) offered (23) features for each frame of the gestures. Still only the 12 most relevant features were selected to be sent to the classifier step, which was implemented by two types of classifiers: (i) Nave Bayes Classifier (NBC), (ii) Multilayer Perceptron Neural networks(MLP). The overall accuracy achieved with these two classifiers was 98.3% and 99.1%, respectively. However, the LMC system faces a challenge in detecting the hands and fingers moving from one side. Some of the fingers may be occluded by others and consequently reduce the reparability of some signs.

In [40] the authors presented improved results using an algorithm of 3D hand posture to extract 24 features and by XGBoost model and 10-fold cross-validation to select only the top 6 features. Then, Gaussian Naïve Bayes (GNB) classifier was used to classify the features. The proposed technique was validated with data from 10 postures in Chinese SL and 50 training samples. The results proved the accuracy and the stability of the algorithm to get close to 100%.

In [41], the authors utilized a **smart glove** with multi-sensors to get the orientation of the hand and fingers. To measure the finger orientations, 5-Flex sensors were used, and to measure (hand movements and orientation), an accelerometer and gyroscope were used. The measured data was transmitted to a smartphone. An application was developed in the smartphone to process the received data and speak out the corresponding alphabet. The Decision Trees classifier classified the features. The result was true for all (ASL) except the 2 letters u and v. The study also discussed dynamic gesture, but dynamic gestures recognition is out of the scope of this review.

The system in [42] also used gloves with 5 flex sensors, gyroscopes, and accelerometers embedded in them. The gestures were captured in real-time by the sensors and collected by Arduino Nano microcontroller to send them to a PC via Bluetooth for processing, and then classified by SVM classifier to get (100%) accuracy. This system was programmed to recognize sign gestures of two sign languages: (ASL) and (ISL).

In [43], another sensor approach was proposed based on a **magnetic positioning** system (MPS) which was used to track and recognize the static gestures of (ASL) after removing the 2-dynamic letters (j,z). 6 magnetic nodes were tracked to get the position and the orientation, Tracking nodes of the MPS were mounted on a glove. (SVM) was used as a classifier to achieve 97% accuracy.

Microsoft Kinect in real-time was the approach studied in [18], but with the (ISL) gestures which were preprocessed by using a median filter for noise reduction, then segmented by K-clustering. Features that were extracted here were (HOG, SURF, and LBP) which were fed to (SVM) classifier to get about 71.85% accuracy.

This result was less than that in [27], although the same device was used. The author here tried to improve his system by adding a color glove. The preprocessing step was implemented by segmenting the data of (ASL) gestures using a per-pixel classification algorithm to obtain: palm, 5 lower finger sections, and 5 fingertips. Then the feature extraction step was performed by extracting the joint positions using a dimension-adaptive mean-shift mode-seeking algorithm. At last, Random Forest (RF) classifier was used, and the accuracy was 90%. This result was better than that achieved in

[25], where Microsoft Kinect was utilized to get 89% from the (SVM) classifier and after extracting HOG features that are the depth information, and fingertip positions for 4-gestures: (Open hand, Peace, Ok, and Like) poses as input data.

Authors in [44] used another technique, the **Media Pipe**, a machine learning platform invented by Google. The proposed method was used with (ASL) alphabet to obtain 3D coordinates for 21 joints to get distance and angle features from RGB images captured by a webcam. Then, two classification methods, the (light gradient boosting machine GBM) and, (SVM) were applied. Different multi-type data were used at the input which are the Massey dataset, finger spelling A, and ASL alphabets containing images. With the first two data types, the achieved accuracy was 97.80% and 99.39% with GBM and SVM, classifiers, respectively which represents the highest results obtained. However, with the third data type used, the system performance degraded, and the accuracy dropped to 86.12% and 87.60% with the used classifiers, respectively.

In [45], the same technique was used but with one type of complex background, and the KNN algorithm was used for prediction. The accuracy achieved by this configuration was improved and increased from 86% to 91% compared to the previous work. For a smart home application, authors in [46] used MediaPipe with the (CNN) technique for the prediction step to get an accuracy of 99%. The machine learning model consisted of three convolutional 2D layers in Tensor Flow with a rectified linear unit activation function on each. The input video was captured by the integrated video camera as individual frames by Open CV. Each frame was stored during runtime as a two-dimensional array of (480 x 60 pixels x3 RGB values), then the video frames were transmitted to a server for processing.

The system in [26] depended on **vision-based** and used (BdSL) as input gestures in real-time. A segmentation of the hand area by the Haar-like feature-based cascaded classifier was done, and then the human skin color-based hand shape segmentation approach was used. After that, noise was reduced by Gaussian smoothing and converted to grayscale to obtain the binary image. Then, hand shape features such as finger position and fingertip were extracted. Finally, the KNN classifier was used to recognize the gestures with an accuracy of 98.17% for Vowels, and 94.75% for Consonants.

Authors in [28] tried to convert (ISL) gestures to text by a system that used the Otsu algorithm for the segmentation process. Then the image components were obtained by morphological filtering tools, and some of their features such as Eigen values and Eigen vectors were extracted and classified by the Linear Discriminant Analysis (LDA) algorithm to recognize the gesture as a text.

With Chinese sign language in [47], hand gestures were segmented from the background, resized to (256×256) pixels, and converted into a grey-level image to make it suitable for the feature extraction step which was done by grey-level co-occurrence matrix (GLCM) algorithm. Then, the extracted features were classified by SVM to get (85.3%) accuracy.

In [48], a speeded-up robust feature (SURF) detector was used as a feature extraction method for Pakistanis SL (PSL) which is based on one static hand gesture. The preprocessing step included resizing and converting the RGB images into grayscale. The segmentation technique was achieved by applying a threshold to the grey-scale images. Some letters in the PSL were represented by dynamic gestures so they were provided to the system as videos and processed frame by frame to act as static images. The (SVM) classifier was used to get an accuracy of (97.80%).

Authors in [49] applied segmentation of the background depending on a fixed image. The information of the image was subtracted from the images with gestures. After that, the resulting image was filtered for noise reduction and converted into grayscale. The feature extraction (FE) step was achieved by Discrete Wavelet Transformation (DWT) and Singular Value Decomposition (SVD) to get 37 different features. For more efficient results, the feature selection (FS) step was implemented by the genetic algorithm (GA). At last, (SVM) classifier was used to get an accuracy of [FE]:61.15% [FS]: 77.55%. It is obvious from the results that although the strong features were selected, the achieved accuracy was low compared with other works.

Bag-of-features by (SIFT) was the feature choice for authors in [50] to recognize 7 gestures from ASL with a white background. It was noticed that this approach was better than SURF features because the latter was not stable with image rotation changes. The system was proposed with offline training and online testing to achieve 96.23% accuracy by K-Means Clustering and SVM classifier.

(ArSL) alphabet with a uniform colored background was the input dataset for the system in [51] which extracted multi-features to make a comparison between the accuracies of different features using the (SVM) classifier. The features that were extracted were (HOG, EHD, LBP, DWT, and GLCM) descriptors and their accuracies were (63.56 %, 42%, 9.78%, 8%, and 2.89%), respectively.

Satisfying results were achieved with the system in [52] after choosing the (HOG) features to extract from (ASL) gesture data. A preprocessing step was implemented by a Logarithmic Transformation and Histogram Equalization for reducing the noise from the images of the dataset. The background was segmented using L*a*b color space. Canny edges are detected to prepare the image for the feature extraction step, then ended with the KNN classifier to get an accuracy of about (94.23%).

A considerable amount of research has focused on using deep learning algorithms for sign language recognition such as [53] which used Artificial Neural Network (ANN) with feedforward, backpropagation algorithm. 30 feature vectors were used to recognize 37 signs of ASL with black background from a mobile video camera. For extracting these features the system employed an algorithm called the “K convex hull” method. Before feature extraction, some preprocessing was done for the input image including resizing the image to (260×260), converting RGB to binary, and applying a median filter to rotated images. The accuracy was (94.32%) in real-time.

Another technique of deep learning was used in [54] to recognize (ISL) gestures, that is the Convolution Neural Network (CNN). It was also used only to classify the images. So there were: (RGB to gray conversion, Background Segmentation, and Gaussian blur filtering) as a preprocessing step and (Canny edge detection) as a feature extraction step. The accuracy of that system was 86%.

Authors in [55] used 2-types of (CNNs): (GoogLeNet) and (AlexNet) they did not require any external preprocessing or feature extraction because these types of networks have multi-hidden layers for doing all tasks needed to recognize the gestures. There was only one step that must be done which was resizing the image into (224x224x3) for GoogLeNet and (227x227x3) for AlexNet. The system achieved favorable results although the input data was from complex backgrounds and different environments. Accuracy results were (95.52%) by GoogLeNet, and (99.39%) by Alexnet.

With the Static hand of the German alphabet, the authors in [56] proposed a system that contained 3-deep networks (HandSegNet, PoseNet, PosePrior network) to Estimate 3D hand pose from the input regular RGB images. The first net (segmentation network) was for localizing the hand within the image. Then, the image was cropped and resized to be suitable for the next net (PoseNet) which was localizing a set of hand key points represented as score maps. (PosePrior network) was used for classification to get results with a 33.2% error rate.

Two CNNs, {Yolov3 and DarkNet-53} were proposed in [57] for finger-pointing positions of numbers: (1, 2, 3, 4, 5) as a data input in real-time. YOLO annotation was used to label the training data and then fed to the DarkNet-53. The accuracy of the studied system was 97.68%.

Furthermore, authors in [58], [6], and [59] discussed the use of CNN's algorithms with different types of datasets which were ASL, RGB, and 2D (complex background, MNIST dataset, Surrey dataset). They achieved an accuracy of (95%, 99%, and 83.29%), respectively. This variation in the results is attributed to the difference in data input and the analysis methods used for the pre-trained networks.

A detailed summary of all the work reviewed above is tabulated in table 2 comparing the input dataset, the approach used, the preprocessing needed, the feature extraction method, the classifier type, and the achieved accuracy.

Table 2 A detailed summery of the reviewed techniques for sign language recognition with static hand gestures.

Ref.	Dataset	Approaches	Preprocessing	Feature Extraction	Classifier	Accuracy
[17]	(ArSL)Arabic alphabet\1-hand	LMC sensor	-	Finger:length,width Hand pitch palm position	(NBC) (MLP)	98.3% 99.1%
[40]	10-postures in Chinese(SL)	LMC sensor	-	[FE]:3D key points collected/[FS]: XGBoost model & 10-fold cross	(GNB)Gaussian Naive Bayes	100% nearly
[41]	take (ASL) alpha. lef words	Smart glove 0f sensors	-	Orientation of hand and fingers	Decision Trees	All give true except u-v
[42]	(ASL) &(ISL)	Smart glove 0f sensors	-	(Normalize, Sample, Rescale, Linearize)the data	(SVM)Support Vector Machine	100%
[43]	ASL alphabet remove (j,z) only	magnetic positioning Sensor	-	The positions and the orientations of the fingers	(SVM)Support Vector Machine	97%
[18]	(ISL) Microsoft Kinect/real time	Microsoft Kinect/real time	*Use.medianfilter *Segmented by K- clustering	(HOG), (SURF), (LBP)	SVM	up to 71.85%.
[27]	(ASL) depth and RG images	Microsoft Kinect+ color glove	hand segmented into: palm, 5 lower finger sections, 5 fingertips	*depth features *hierarchical mode-seeking to localize hand joint positions	(RF) Random Forest	90%
[25]	4-gestures of: (Open hand, Peace, Ok , and Like) pose	Microsoft Kinect: RGB-D	hand segmentation, conv.tograyscale	HOG feature vectors, depth information, fingertip positions	(SVM)	up to 89%
[44]	ASL(multitype, high accuracy with: Massey	media-pipe	-	Obtained3Dcoordinates for 21 joints to get (distance,angle)Features	(SVM), light (GBM)	99.39% 97.80%

[45]	ASL	media-pipe	-	Find: 21 hand points 3D Landmarks of Palm	KNN algorithm	86 to 91%.
[46]	ASL\20 sign	media-pipe	-	21 hand points 3D Landmarks of Palm	CNN	99%
[26]	(BdSL)\2-hand Bengali language	vision-based	skin color segm. ,reduce noise by Gaussian conv. gray-BW	extract hand shape features: like (finger position, fingertip)	(KNN)	98.17% for Vowels 94.75%for Consonants
[28]	(ISL)\1&2hand RGB	vision-based	Segmentation, morphological filtering	Eigen values and Eigen vectors	Linear Discriminant Analysis(LDA)	Not specified
[47]	Chinese sign language	vision-based	*hand segmented from background *resize to 256×256 *converted into gray-level image	gray-level co-occurrence matrix (GLCM) as a fetures	SVM with mediumGaussian	85.3%
[48]	(PSL) Pakistanis One static Hand UniformBackground	vision-based	RGB into grayscale Segmentation	features extracted (SURF)	(SVM)	97.80%
[49]	10-numeral signs	vision-based	subtractBackground Conv.into grayscale	[FE]: (DWT), (SVD) [FS]: GA genetic	(SVM)	[FE]:61.15% [FS]: 77.55%
[50]	7- gestures from ASL whiteBackground	vision-based by camera	Skin detection for postures alone	bag-of-features by (SIFT)	(SVM)	96.23%
[51]	(ArSL) alphabets uniform colored background	vision-based by Smartphon	Conv. into gray	HOG descriptor EHD descriptor LBP descriptor DWT descriptor GLCM descriptor	(SVM)	63.56 % 42% 9.78% 8% 2.89%
[52]	ASL\ 572 images as a dataset	vision-based	Noise redaction, background segmented, edges detected	HOG	(KNN)	94.23%
[53]	ASL /black background	vision-based by mobile camera\real time	resized to 260×260 RGB to binary median filter\rotated images	fingertip finder, eccentricity, elongatedness	ANN	94.32 %
[54]	(ISL)	vision-based mobile camera	RGB to gray Background Segmentation Gaussian blur	Canny edge detection	2D-CNN	86%
[55]	ASL\ RGB with complex background	vision-based	resized to: 224x224x3for GoogLeNet *227x227x3 for AlexNet	Hidden features extraction by:GoogLeNet & AlexNet	Hidden classifiers in the 2-CNNs algorithms	95:52% by GoogLeNet, 99:39% by Alexnet
[56]	Static German alphabet RGB	vision-based	resized to 320*320/ randomCropped to(256*256) in1st network	Estimates 3D hand pose by 3deep networks (HandSegNet, PoseNet, Viewpoint+PosePrior)	Classifier by fully connected 3-layer-networks	33.2 % Error rate
[57]	finger-pointing positions of numbers: 1, 2, 3, 4, and 5.	vision-based by camera in real-time	UsingYOLO with200images increasing 2-fold, each image was duplicated	Obtaining boundary box (x-axis, y-axis, height, and width) in hidden layers: YOLOv3& DarkNet-53	Hidden classifiers in YOLOv3 and DarkNet-53 CNNnet	97.68%

[58]	ASL\ RGB complex background	vision-based by smart phoneCamera	J,Z dynamic: excluded resized to 32 x 32.	Hidden features extraction by:CNN	Hidden classifiers CNN	95%
[6]	ASL\ RGB MNIST dataset	vision-based by camera	image segmented conv.to grayscale	Hidden features extraction by:CNN	CNN hidden classifier	99%
[59]	ASL\ RGB\ 2D Surrey dataset	vision-based for mobile	Resized images	Hidden features by CNN Squeezenet	CNN Squeezenet	Training: 87.47% Validation: 83.29%

7. Discussion and Comments

With the **sensor-based** systems that were used **(MLP)**, there was a difference in the results between the two research this might be because of the number of gestures that were used as a dataset (28&10), this means, When the number of the classes is less, the accuracy will be better. Another cause was the number of features that were selected which means that this type of device needs only about six features for giving good results.

With the **gloves** systems, the results failed with two gestures in a system that was designed to be a smartphone accessory, but when it was used to feed the data of gestures to a PC device the accuracy was 100% which means that the system is more effective and compatible with a PC because of its better data processing ability. However, this technology is not recommended by researchers because it requires a wearable device.

Microsoft Kinect is a sensing input. It is designed with a 3D sensor camera. Compared with the other two systems, it did not provide the best accuracy despite all the preprocessing done, the sensitive properties, and its independence of lighting conditions.

MediaPipe which delivered 3D skeletal joint points from a 2D image is an open-source framework from Google so these systems need to always be connected to the web in order to be used.

Many types of features can be extracted in **vision-based** systems. It was very important to decide what feature was the most suitable to extract. The accuracy was different according to the method of image preprocessing, the feature that was selected, the type of classifier, and the data that was used in the training step which was either with complex background or simple form. Sometimes the real-time condition was the cause of the low accuracy.

From the above, there was an excellent result with the extraction of hand-shape features like (finger position, and fingertip) although the system was used in real-time conditions. The system still had more complexity than systems that depended on deep learning techniques which did not need preprocessing steps, or manual feature extraction steps, and offered excellent accuracy. These characteristics make a choice fall on this technique in modern research, despite the required storage space and the time consumed to complete the machine learning process.

8. Conclusion

Sign language can be recognized using different methods based on sensors and vision. Sensors-based methods were outperformed in real-time while vision-based suffered from computation delay. Vision-based methods are a suitable choice for those who are looking for affordable methods because they do not require expensive sensors. Sensor-based methods do not require complex data processing while most vision-based methods need image preprocessing and processing to get feature vectors. Sensor-based methods extract the measurements of the hand such as the hands (speed, position, and joint orientation), and because of the skeletal data, it offers a higher recognition rate. Hybrid-based methods, in which different techniques are combined, are needed for multi-devices and more computation steps and that makes them suitable for dynamic gestures.

In conclusion, there are many aspects to consider in every stage of the recognition process, like the data acquisition technique, static or dynamic signs, single or double-handed signs, feature extraction, or deep learning methods. Also, the feature extraction technique and the need for feature selection are two important factors to study in designing any gesture recognition system. In feature extraction, there were more manual steps than in deep learning, but deep learning needs more memory and time computation. The accuracy was different according to the data type and if the recognition was in real-time or not. Feature selection is another crucial step in the process because it provides the most relevant and decisive data for the classification phase. The classification method depends on the data form provided to the system.

References:

[1] K. P. Nimisha and A. Jacob, "A Brief Review of the Recent Trends in Sign Language Recognition," *Proc. 2020 IEEE Int. Conf. Commun. Signal Process. ICCSP 2020*, pp. 186–190,

2020, doi: 10.1109/ICCSP48568.2020.9182351.

[2] A. Ghotkar, "Study of Vision Based Hand Gesture Recognition Using," vol. 7, no. 1, pp. 96–115, 2014.

[3] M. A. Almasre and H. Al-Nuaim, "A comparison of Arabic sign language dynamic gesture recognition models," *Heliyon*, vol. 6, no. 3, p. e03554, 2020, doi: 10.1016/j.heliyon.2020.e03554.

[4] V. Bheda and D. Radpour, "Using Deep Convolutional Networks for Gesture Recognition in American Sign Language," 2017, [Online]. Available: http://arxiv.org/abs/1710.06836.

[5] M. Sanzidul Islam, S. Sultana Sharmin Mousumi, N. A. Jessan, A. Shahariar Azad Rabby, and S. Akhter Hossain, "Ishara-Lipi: The First Complete MultipurposeOpen Access Dataset of Isolated Characters for Bangla Sign Language," *2018 Int. Conf. Bangla Speech Lang. Process. ICBSLP 2018*, no. September, pp. 1–4, 2018, doi: 10.1109/ICBSLP.2018.8554466.

[6] R. S. Sabeenian, S. Sai Bharathwaj, and M. Mohamed Aadhil, "Sign language recognition using deep learning and computer vision," *J. Adv. Res. Dyn. Control Syst.*, vol. 12, no. 5 Special Issue, pp. 964–968, 2020, doi: 10.5373/JARDCS/V12SP5/20201842.

[7] U. M. Butt *et al.*, "Feature based algorithmic analysis on American sign language dataset," *Int. J. Adv. Comput. Sci. Appl.*, vol. 10, no. 5, pp. 583–589, 2019, doi: 10.14569/ijacsa.2019.0100575.

[8] J. Ekbote and M. Joshi, "A Survey on Hand Gesture Recognition for Indian Sign Language," pp. 1039–1044, 2016.

[9] H. Mahmud, M. M. Morshed, and M. K. Hasan, "A Deep Learning-based Multimodal Depth-Aware Dynamic Hand Gesture Recognition System," 2021, [Online]. Available: http://arxiv.org/abs/2107.02543.

[10] S. Nadgeri, "An Analytical study of signs used in Baby Sign Language using MobileNet Framework."

[11] M. A. Ahmed, B. B. Zaidan, A. A. Zaidan, M. M. Salih, and M. M. Bin Lakulu, "A review on systems-based sensory gloves for sign language recognition state of the art between 2007 and 2017," *Sensors (Switzerland)*, vol. 18, no. 7, 2018, doi: 10.3390/s18072208.

[12] M. J. Cheok, Z. Omar, and M. H. Jaward, "A review of hand gesture and sign language recognition techniques," *Int. J. Mach. Learn. Cybern.*, vol. 10, no. 1, pp. 131–153, 2019, doi: 10.1007/s13042-017-0705-5.

[13] "Low-Cost Glove Wirelessly Tran [IMAGE] | EurekAlert! Science News Releases." [Online].

Available: https://www.eurekalert.org/multimedia/566594.

[14] M. Oudah, A. Al-Naji, and J. Chahl, “Hand Gesture Recognition Based on Computer Vision: A Review of Techniques,” *J. Imaging*, vol. 6, no. 8, 2020, doi: 10.3390/JIMAGING6080073.

[15] R. Elakkiya, “Machine learning based sign language recognition: a review and its research frontier,” *J. Ambient Intell. Humaniz. Comput.*, no. 0123456789, 2020, doi: 10.1007/s12652-020-02396-y.

[16] R. Van Culver, “A hybrid sign language recognition system,” *Proc. - Int. Symp. Wearable Comput. ISWC*, pp. 30–33, 2004, doi: 10.1109/iswc.2004.2.

[17] M. Mohandes, S. Aliyu, and M. Deriche, “Arabic sign language recognition using the leap motion controller,” *IEEE Int. Symp. Ind. Electron.*, no. November 2015, pp. 960–965, 2014, doi: 10.1109/ISIE.2014.6864742.

[18] T. Raghuveera, R. Deepthi, R. Mangalashri, and R. Akshaya, “A depth-based Indian Sign Language recognition using Microsoft Kinect,” *Sadhana - Acad. Proc. Eng. Sci.*, vol. 45, no. 1, 2020, doi: 10.1007/s12046-019-1250-6.

[19] J. L. Raheja, A. Mishra, and A. Chaudhary, “Indian sign language recognition using SVM,” *Pattern Recognit. Image Anal.*, vol. 26, no. 2, pp. 434–441, 2016, doi: 10.1134/S1054661816020164.

[20] M. El Badawy, A. Samir Elons, H. Sheded, and M. F. Tolba, “A proposed hybrid sensor architecture for Arabic sign language recognition,” *Adv. Intell. Syst. Comput.*, vol. 323, pp. 721–730, 2015, doi: 10.1007/978-3-319-11310-4_63.

[21] Manisha U. Kakde, Mahender G. Nakrani, and Amit M. Rawate, “A Review Paper on Sign Language Recognition System For Deaf And Dumb People using Image Processing,” *Int. J. Eng. Res.*, vol. V5, no. 03, 2016, doi: 10.17577/ijertv5is031036.

[22] A. Tyagi and S. Bansal, “Feature extraction technique for vision-based indian sign language recognition system: A review,” *Adv. Intell. Syst. Comput.*, vol. 1227, pp. 39–53, 2021, doi: 10.1007/978-981-15-6876-3_4.

[23] Suharjito, F. Wiryana, G. P. Kusuma, and A. Zahra, “Feature Extraction Methods in Sign Language Recognition System: A Literature Review,” *1st 2018 Indones. Assoc. Pattern Recognit. Int. Conf. Ina. 2018 - Proc.*, no. September, pp. 11–15, 2019, doi: 10.1109/INAPR.2018.8626857.

[24] G. Roffo, “Feature Selection Library (MATLAB Toolbox),” 2016, [Online]. Available: http://arxiv.org/abs/1607.01327.

[25] C. R. Mihalache and B. Apostol, "Hand pose estimation using HOG features from RGB-D data," *2013 17th Int. Conf. Syst. Theory, Control Comput. ICSTCC 2013; Jt. Conf. SINTES 2013, SACCS 2013, SIMSIS 2013 - Proc.*, pp. 356–361, 2013, doi: 10.1109/ICSTCC.2013.6688985.

[26] M. A. Rahaman, M. Jasim, M. H. Ali, and M. Hasanuzzaman, "Real-time computer vision-based Bengali sign language recognition," *2014 17th Int. Conf. Comput. Inf. Technol. ICCIT 2014*, no. May 2019, pp. 192–197, 2014, doi: 10.1109/ICCITechn.2014.7073150.

[27] C. Dong, M. C. Leu, and Z. Yin, "American Sign Language alphabet recognition using Microsoft Kinect," *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Work.*, vol. 2015-Octob, pp. 44–52, 2015, doi: 10.1109/CVPRW.2015.7301347.

[28] M. Kumar, "Conversion of Sign Language into Text," *Int. J. Appl. Eng. Res.*, vol. 13, no. 9, pp. 7154–7161, 2018, [Online]. Available: http://www.ripublication.com.

[29] A. A. Barbhuiya, R. K. Karsh, and R. Jain, "CNN based feature extraction and classification for sign language," *Multimed. Tools Appl.*, vol. 80, no. 2, pp. 3051–3069, 2021, doi: 10.1007/s11042-020-09829-y.

[30] R. Rastgoo, K. Kiani, and S. Escalera, "Sign Language Recognition: A Deep Survey," *Expert Syst. Appl.*, vol. 164, no. February 2020, p. 113794, 2021, doi: 10.1016/j.eswa.2020.113794.

[31] A. Wadhawan and P. Kumar, "Sign Language Recognition Systems: A Decade Systematic Literature Review," *Arch. Comput. Methods Eng.*, vol. 28, no. 3, pp. 785–813, 2021, doi: 10.1007/s11831-019-09384-2.

[32] I. A. Adeyanju, O. O. Bello, and M. A. Adegboye, "Machine learning methods for sign language recognition: A critical review and analysis," *Intell. Syst. with Appl.*, vol. 12, p. 200056, 2021, doi: 10.1016/j.iswa.2021.200056.

[33] X. Jiang, S. C. Satapathy, L. Yang, S. H. Wang, and Y. D. Zhang, *A Survey on Artificial Intelligence in Chinese Sign Language Recognition*, vol. 45, no. 12. Springer Berlin Heidelberg, 2020.

[34] M. Mustafa, "A study on Arabic sign language recognition for differently abled using advanced machine learning classifiers," *J. Ambient Intell. Humaniz. Comput.*, vol. 12, no. 3, pp. 4101–4115, 2021, doi: 10.1007/s12652-020-01790-w.

[35] S. Sharma and S. Singh, "Vision-based sign language recognition system: A Comprehensive Review," *Proc. 5th Int. Conf. Inven. Comput. Technol. ICICT 2020*, no. February 2020, pp. 140–144, 2020, doi: 10.1109/ICICT48043.2020.9112409.

[36] H. Bhavsar and J. Trivedi, “Review on Feature Extraction methods of Image based Sign Language Recognition system,” *Indian J. Comput. Sci. Eng.*, vol. 8, no. 3, pp. 249–259, 2017.

[37] A. Jaramillo-Yánez, M. E. Benalcázar, and E. Mena-Maldonado, “Real-time hand gesture recognition using surface electromyography and machine learning: A systematic literature review,” *Sensors (Switzerland)*, vol. 20, no. 9, pp. 1–36, 2020, doi: 10.3390/s20092467.

[38] A. K. Sahoo, G. S. Mishra, and K. K. Ravulakollu, “Sign language recognition: State of the art,” *ARPN J. Eng. Appl. Sci.*, vol. 9, no. 2, pp. 116–134, 2014.

[39] R. A. Yuvraj Grover, “Sign Language Translation System for Hearing/Speech Impaired People: A Review,” *Int. Conf. Inov. Pract. Technol. Manag.*, pp. 10–14, 2021.

[40] yu zhang, tian zhang, and ruimei li, “Algorithm of 3D hand posture recognition with space coordinates based on optimal feature selection,” no. August 2018, p. 87, 2018, doi: 10.1117/12.2502939.

[41] S. C. Bodda, P. Gupta, G. Joshi, and A. Chaturvedi, “A new architecture for hand-worn Sign language to Speech translator,” 2020, [Online]. Available: http://arxiv.org/abs/2009.03988.

[42] M. M. Chandra, S. Rajkumar, and L. S. Kumar, “Sign Languages to Speech Conversion Prototype using the SVM Classifier,” *IEEE Reg. 10 Annu. Int. Conf. Proceedings/TENCON*, vol. 2019-Octob, pp. 1803–1807, 2019, doi: 10.1109/TENCON.2019.8929356.

[43] M. Rinalduzzi *et al.*, “Gesture recognition of sign language alphabet using a magnetic positioning system,” *Appl. Sci.*, vol. 11, no. 12, 2021, doi: 10.3390/app11125594.

[44] J. Shin, A. Matsuoka, M. A. M. Hasan, and A. Y. Srizon, “American sign language alphabet recognition by extracting feature from hand pose estimation,” *Sensors*, vol. 21, no. 17, pp. 1–19, 2021, doi: 10.3390/s21175856.

[45] K. Gomase, A. Dhanawade, P. Gurav, and S. Lokare, “Sign Language Recognition using Mediapipe,” *Int. Res. J. Eng. Technol.*, vol. 9, no. 1, pp. 744–746, 2022, [Online]. Available: https://www.irjet.net/archives/V9/i1/IRJET-V9I1133.pdf.

[46] B. Bagby, D. Gray, R. Hughes, Z. Langford, and R. Stonner, “Simplifying Sign Language Detection for Smart Home Devices using Google MediaPipe,” 2021, [Online]. Available: https://bradenbagby.com/Portfolio/Resources/PDFs/ResearchPaper.pdf.

[47] X. Jiang, “Isolated Chinese Sign Language Recognition Using Gray-Level Co-occurrence Matrix and Parameter-Optimized Medium Gaussian Support Vector Machine,” *Adv. Intell. Syst. Comput.*, vol. 1014, no. January, pp. 182–193, 2020, doi: 10.1007/978-981-13-9920-6_19.

[48] M. Shaheer Mirza, S. Muhammad Munaf, S. Ali, F. Azim, and S. Jawaid Khan, “Vision-based

Pakistani Sign Language Recognition Using Bag-of-Words and Support Vector Machines," 2022, [Online]. Available: https://doi.org/10.21203/rs.3.rs-1204236/v1.

[49] R. A. Bhuiyan, A. K. Tushar, A. Ashiquzzaman, J. Shin, and M. R. Islam, "Reduction of gesture feature dimension for improving the hand gesture recognition performance of numerical sign language," *20th Int. Conf. Comput. Inf. Technol. ICCIT 2017*, vol. 2018-Janua, no. December 2019, pp. 1–6, 2018, doi: 10.1109/ICCITECHN.2017.8281833.

[50] N. H. Dardas and N. D. Georganas, "Real-Time Hand Gesture Detection and Recognition.pdf," vol. 60, no. 11, pp. 3592–3607, 2011.

[51] R. Alzohairi, R. Alghonaim, W. Alshehri, S. Aloqeely, M. Alzaidan, and O. Bchir, "Image based Arabic Sign Language recognition system," *Int. J. Adv. Comput. Sci. Appl.*, vol. 9, no. 3, pp. 185–194, 2018, doi: 10.14569/IJACSA.2018.090327.

[52] I. Mahmud, T. Tabassum, M. P. Uddin, E. Ali, A. M. Nitu, and M. I. Afjal, "Efficient Noise Reduction and HOG Feature Extraction for Sign Language Recognition," *2018 Int. Conf. Adv. Electr. Electron. Eng. ICAEEE 2018*, no. November, pp. 1–4, 2019, doi: 10.1109/ICAEEE.2018.8642983.

[53] M. M. Islam, S. Siddiqua, and J. Afnan, "Real time Hand Gesture Recognition using different algorithms based on American Sign Language," *2017 IEEE Int. Conf. Imaging, Vis. Pattern Recognition, icIVPR 2017*, 2017, doi: 10.1109/ICIVPR.2017.7890854.

[54] and S. B. Shubham Kr. Mishra, Sheona Sinha, Sourabh Sinha, "Recognition of Hand Gestures and Conversion of Voice for Betterment of Deaf and Mute People," *Springer Nat. Singapore*, vol. 1046, no. July, pp. 46–57, 2019, doi: 10.1007/978-981-13-9942-8.

[55] M. E. M. Cayamcela and W. Lim, "Fine-tuning a pre-trained Convolutional Neural Network Model to translate American Sign Language in Real-time," *2019 Int. Conf. Comput. Netw. Commun. ICNC 2019*, pp. 100–104, 2019, doi: 10.1109/ICCNC.2019.8685536.

[56] C. Zimmermann and T. Brox, "Learning to Estimate 3D Hand Pose from Single RGB Images," *Proc. IEEE Int. Conf. Comput. Vis.*, vol. 2017-Octob, pp. 4913–4921, 2017, doi: 10.1109/ICCV.2017.525.

[57] A. Mujahid *et al.*, "Real-time hand gesture recognition based on deep learning YOLOv3 model," *Appl. Sci.*, vol. 11, no. 9, 2021, doi: 10.3390/app11094164.

[58] L. Y. Bin, G. Y. Huann, and L. K. Yun, "Study of Convolutional Neural Network in Recognizing Static American Sign Language," *Proc. 2019 IEEE Int. Conf. Signal Image Process. Appl. ICSIPA 2019*, pp. 41–45, 2019, doi: 10.1109/ICSIPA45851.2019.8977767.

[59] N. Kasukurthi, B. Rokad, S. Bidani, and D. A. Dennisan, “American Sign Language Alphabet Recognition using Deep Learning,” 2019, [Online]. Available: http://arxiv.org/abs/1905.05487.

التعرف على لغة الإشارة ومراجعة لإيماءات اليد

الخلاصة: *يستخدم الصم الحركات والتعبيرات الجسدية للكشف عن أفكارهم ومشاعرهم لعالمهم. تسمى هذه التعبيرات "لغة الإشارة" ومثل اللغات الطبيعية ، هناك العديد من أشكال الإشارات حول العالم فهناك لغة اشارة امريكية ولغة اشارة عربية واخرى هندية وهكذا.*

ان مجتمع الصم يقوم اما باستخدام يد واحدة أو يديه الاثنتين وأحيانًا يستخدم أجزاء أخرى من الجسم مثل الرأس أو الشفتين أو العينين اضافة الى ايماءة اليد. إيماءاتهم هذه قد تكون إما بأيدي ثابتة أو متحركة، كذلك قد يشير كل منها الى حرف او كلمة او جملة كاملة. وعليه يمكن اعتبار انها لغات معقدة بعض الشيء، لذلك من الضروري أن يفهم الأشخاص الآخرون العاديون (اي أولئك الذين لا يعانون من الصمم أبدا) معنى كل من هذه الإشارات والإيماءات حتى يتمكنوا من التواصل مع مجتمع الصم بنجاح. يعد التفاعل بين الإنسان والحاسوب أداة فعالة واتجاهًا ممتازًا لاستخدامه في تسهيل الاتصال وفهم لغات الإشارة المختلفة المستخدمة في جميع أنحاء العالم. لقد بذلت محاولات عديدة من قبل الباحثين لغرض مراجعة أهم التقنيات والنماذج المستخدمة في مجال فك رموز لغات الإشارة وفهمها. يتم توجيه كل جهد بحثي جديد نحو تحسين طرق الاتصال هذه. ان بعض النماذج المقترحة تعاملت مع علامات منعزلة والبعض الآخر ركز على العلامات المستمرة. تمثل هذه المقالة ملخصًا للعديد من المراجعات الشاملة التي درست الأدبيات المختلفة التي أجريت حول التعرف على لغة الإشارة. تركز المناقشة في هذه المراجعة على الأنظمة والأساليب التي تتعامل مع التعرف على إيماءات اليد الثابتة فقط. يهدف هذا العمل إلى توفير دليل للباحثين والممارسين لربط عملهم بالبحوث الحالية واكتساب رؤى حول ما يمكن لعملهم أن يساهم به في هذا المجال.