# ARABIC SIGN LANGUAGE RECOGNITION SYSTEMS: A SYSTEMATIC REVIEW

Ahmad M. J. AL Moustafa

Faculty of Computing, Universiti Teknologi Malaysia, 81310 Johor Bahru, Johor, Malaysia
College of Computer Science, King Khalid University, Abha, Saudi Arabia
ahmad200000@hotmail.com

Mohd Shafry Mohd Rahim

Faculty of Computing, Universiti Teknologi Malaysia, 81310 Johor Bahru, Johor, Malaysia
Media and Game Innovation Centre of Excellence, Institute of Human Centered Engineering, Universiti Teknologi Malaysia, 81310 Johor Bahru, Johor, Malaysia
shafry@utm.my

Mahmoud M. Khattab

Faculty of Information and Communication Technology, International Islamic University Malaysia, Kuala Lumpur, Malaysia
mmkhattab2000@gmail.com

Akram M. Zeki

Faculty of Information and Communication Technology, International Islamic University Malaysia, Kuala Lumpur, Malaysia
akramzeki@iium.edu.my

Safaa S. Matter

Applied College, King Khalid University, Abha, Saudi Arabia
safaamatter2010@gmail.com

Amr Mohmed Soliman

Special Education Dept, College of Education, King Khalid University, Abha, Saudi Arabia
psy_amro@hotmail.com

Abdelmoty M. Ahmed

College of Computer Science, Nahda University, Beni Suef, Egypt
abd2005moty@yahoo.com

## Abstract

**Deaf individuals use sign language to interact with each other and with people in general. In most cases, an interpreter is needed in order for a hearing person to have a conversation with a deaf person. The study of recognizing visual signs is becoming more prevalent in the field of computer vision. A deaf person may be able to communicate with hearing people without an interpreter if reliable sign language recognition technology is made available to them. It can be used to generate voice or text, giving the mute greater independence. The study of Arabic sign language recognition is crucial. It facilitates better communication between the deaf and others in the community. This paper attempts to investigate the current Arabic sign language (ArSL) recognition approaches which are commonly and efficiently used for building the automatic translation and recognition of ArSL systems. Furthermore, the focus is given on highlighting the challenges and future work directions to explore the unsolved problems related to ArSL recognition systems.**

***Keywords*****: Sign language recognition; deep learning; vision-based; sensor-based.**

DOI : 10.21817/indjcse/2024/v15i1/241501008

## 1. Introduction

There is a growing demand for human-computer interaction (HCI) as a result of developments in sensor technology. These developments include advancements in camera hardware and the growth of the smartphone application industry. Gesture-based computer interaction is a relatively recent area of study compared to speech-based interaction, which has been studied for much longer. The use of hand gestures in communication may be seen as either a stand-alone modality or as a supplement to spoken exchanges. The everyday medium of communication for the hearing-impaired is sign languages, which are based on the gestures used in every aspect of human communication, both on purpose and by accident [1].

Sign languages (SL), similar to vocal languages, emerge and grow spontaneously among communities of deaf people, just as they did with vocal languages. The evolution of sign languages occurs apart from the linguistic traditions of each country or location in which they are used, meaning that they do so wherever there are deaf or hearing-impaired people. Every sign language has its own syntax and set of rules, and the only thing they have in common is that they can all be visually perceived [2]. Additionally, comprehension of a sign language hardly outspreads outside deaf communities and for a deaf individual the communication with hearing individuals often means knowledge of a different language entirely such as the capability of reading, writing, and lip-reading. Most of deaf kids are born to hearing parents consequently a language gap occurs even within a family. Furthermore, there is no standard procedure for sign languages, which makes tutoring a challenge for leading a deaf individual to be bilingual in order to read and write [3].

Pattern recognition, image processing, natural language processing, and linguistics are just a few of the numerous fields that contribute to the study of sign language recognition (SLR). The challenge has several dimensions, including the difficulty of learning to recognize hand movements visually and the wide variety of sign languages. Sign languages, like spoken languages, have a syntax and grammar, but they differ from spoken languages because the sign language structure uses several body motions simultaneously, rather than sequentially, to convey meaning [4]. Sign language's linguistic characteristics result from the use of not just hand gestures but also facial expressions and head movements to convey meaning [5].

Studies of hand gesture recognition began with gloves connected to multiple sensors and trackers [3]. While these gloves provided precise data for hand location and finger movement, they necessitate the use of cumbersome devices on the individual's hand. In comparison to vision-based systems that provide a natural situation for them. However, it also presented various difficulties, such as detecting occlusion or the detection and segmentation of the hand and fingers. Vision-based systems employ indications such as colored gloves for each hand or coloured indicators on fingers to address these concerns. Despite the various findings in the literature, the problem of indicator-free detection and tracking, in free surroundings, is still an interesting problem [6].

Other projects make use of depth cameras, such as the ones found in the Microsoft Kinect [7]. The main advantage of using a depth camera is the elimination of relying on a wired glove. The stationary camera allows the person to gesture at a greater distance and removes any concern about meddlesome wires. Researchers using the depth camera utilize computer vision algorithms to find vectors and segment the hand accordingly for reading. The disadvantage of this method is the constant computations the algorithms place on the computer. Not only does the program have to segment the hand constantly, but it also has to constantly compare those readings to the gestures it is programmed with [8].

Deep learning is a recent development in the field of artificial intelligence. It is a specialized method of machine learning that utilizes hundreds of neural networks to independently produce an output based on its input [2; 8]. The recent boom of deep learning research and methodologies makes it a prime candidate for this project, mainly for its area of image classification [9].

Unfortunately, there is a lack of modern technologies which assist deaf and dumb people in their communication with other persons. A deaf person can only enjoy his full human rights by using and recognizing sign languages to get bilingual education using sign language in addition to the interpretation of sign language for others. Access to all walks of life and education in the deaf community depend on the availability of explanations of the sign language [10]. There is a big lack of Arabic research in the field of serving disabled people by modern technical means. Also, there is a weakness in the Arabic computer applications that assist disabled people in education, training, and communication in general. Also, there is a lack of ArSL automation. In general, there is little Arabic research published in this field. As a result of the above, the automatic SLR systems have a serious shortage in the size of the dataset which represents the sign language, and that affects the quality of translation and interpretation [11].

This research paper aims to provide an extensive review of the main approaches which are commonly and efficiently used for building the automatic translation and recognition of ArSL systems, identify the main factors that affect the accuracy of automatic translation of ArSL systems, and investigate the significance of building and improving the automatic translation and recognition of ArSL systems. Most of the published literature have been summarized. In order to assist researchers in investigating unsolved challenges that are associated with ArSL recognition systems, certain recommendations and potential areas for future research have been identified.

The remaining sections of this paper are structured as follows. Section 2 and 3 illustrate ArSL recognition framework and challenges respectively. Section 4 describes the classification of ArSL recognition. Different ArSL recognition approaches and datasets are represented in Section 5 and 6 respectively. Section 7 provides a detailed discussion, while Section 8 concludes the paper.

## 2. ArSL Recognition Framework

An ArSL recognition framework is a system that can identify and interpret hand gestures produced in ArSL by employing computer vision and machine learning methods [12]. The framework typically consists of several components, including data acquisition stage, a preprocessing stage, a feature extraction stage, and a classification stage as shown in Fig. 1.

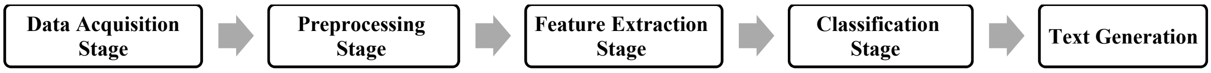

Fig. 1. Arabic Sign Language Recognition Framework

The data acquisition stage captures image or video of the signer performing the sign language gesture. The preprocessing stage is responsible for cleaning and enhancing the image or video data [13; 14], such as removing noise and correcting lighting [15; 16]. The feature extraction stage extracts relevant features from the image or video data, such as hand shape, movement, and orientation. The classification stage uses machine learning models to classify the extracted features and recognize the sign language gesture.

## 3. ArSL Recognition Challenges

There is no standard sign language. Even the ArSL has regional and dialectal variations that make it distinct from other spoken languages. This study focuses on the ArSL, which is understood by the vast majority of Arabs [17; 18]. In Arab society, fostering understanding and integration of the deaf community requires efforts to learn their language and develop deep learning models for ArSL. These models aim to teach ArSL to various groups, including family members, friends, and neighbors, enabling a broader segment of society to communicate effectively with the deaf and mute individuals. The primary goal of designing these deep models is to partially replace human interpreters and create a system capable of translating sign language captured in video frames or images into Arabic text. This system also aims to enhance communication between the deaf community and others, facilitating smoother interactions and inclusivity.

One of the key challenges in developing an ArSL recognition framework is the variability and sophisticated nature of sign language gestures. Sign languages are complex visual languages that involve a range of hand shapes, movements, facial expressions, and body postures. ArSL, in particular, has its own unique grammar and syntax, which makes it challenging to model and recognize using traditional machine learning techniques [11].

Researchers have implemented numerous approaches for ArSL recognition to overcome these challenges. These approaches include deep learning, computer vision, and natural language processing. In particular, deep learning techniques, namely convolutional neural networks (CNNs) and recurrent neural networks (RNNs), have been successful in recognizing sign language gestures from video data [19].

Another important consideration when developing an ArSL recognition framework is the need for real-time processing and low-latency response times. This is especially important for applications that require immediate feedback, such as sign language translation systems or assistive technologies for people with hearing impairments. To achieve real-time processing, researchers have explored the possibility of utilizing specialist hardware including graphics processing units (GPUs) and field-programmable gate arrays (FPGAs), as well as optimized software implementations that can run efficiently on standard hardware [19].

## 4. Classification of ArSL Recognition

To build an ArSL recognition framework, a large dataset of sign language images or videos needs to be collected and annotated with corresponding sign language labels. In the classification phase of machine learning, this dataset may be utilized for both training and testing the models. The data for ArSL recognition may be gathered in two distinct ways [20]: 1) using a vision-based approach, and 2) through a sensor-based approach as shown in Fig. 2.

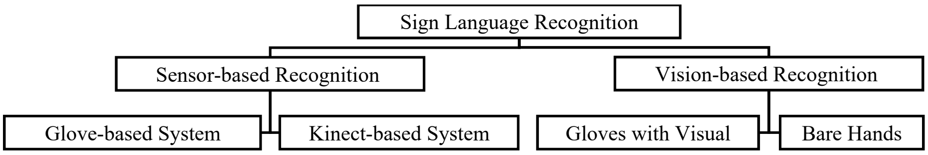

Fig. 2. Classification of ArSL Recognition [21]

A vision-based approach for ArSL recognition is the process of analysing and interpreting sign language gestures from image or video data with the use of computer vision techniques. This approach typically involves several steps, including gesture detection, feature extraction, and classification. Gesture detection involves locating and tracking the hands and other relevant body parts in the image or video data. This can be done using techniques such as colored skin segmentation, hand shape classification, or optical flow evaluation. Once the hands and other relevant body parts have been detected, feature extraction techniques are used to extract relevant information such as hand shape, movements, and orientations from the video data. This can be done using a range of techniques, including approaches based on deep learning such as CNNs, while others like histogram of oriented gradients (HOG) and local binary patterns (LBP) are more traditional approaches. Lastly, the retrieved information is employed by machine learning techniques such as support vector machines (SVMs), random forests (RFs), decision trees (DTs), and artificial neural networks (ANNs) to categorize the sign language gesture [20].

One challenge in utilizing a vision-based approaches for ArSL recognition is the management of variability and complexity of sign language gestures, which can vary significantly between different signers and in different contexts. To address this challenge, researchers have explored techniques such as data augmentation, multi-modal fusion, and transfer learning, which can help improve the robustness and generalization of the recognition system [20].

A sensor-based approach for ArSL recognition involves using sensors to detect and capture data related to sign language gestures. This approach can be used to complement or replace vision-based approaches and has several advantages and challenges. One advantage of using sensors is that they can provide more accurate and reliable data than vision-based approaches, especially in situations with poor lighting or occlusions. Sensors can also capture additional information that may not be visible to the naked eye, which include electromyography (EMG) signals, pressure, and force. Furthermore, there exist two main categories of systems: one is predicated upon the utilization of gloves [22], whilst the other is founded with the Microsoft Kinect [23]. Electronic and mechanical components in gloves [22] enable their usage in systems that interpret hand gestures. Wearing a glove connected to certain sensors that capture information might be uncomfortable for signers who are hearing or speech impaired, despite the fact that the approach may produce good results [24]. Kinect sensors are utilized to recognize signs made in the second category. Microsoft first created these sensor devices for use with the Xbox game to eliminate the need for traditional controllers [7]. Recent years have seen an increase in the use of this technology, which has led to the incorporation of recognition technologies such as sign language recognition.

In contrast, vision-based systems are able to interpret sign language through the use of visual information, such as images or videos [25], [26], [27], as well as image processing and machine learning methods [24]. There are two main types of these systems. One strategy makes use of colorful gloves or other visual cues to identify hand gestures [28]. Nevertheless, this approach limits the naturalness of sign language recognition systems, which is desirable in comparable HCI systems [28]. The second kind is dependent on images capturing sign language hand gestures [24]. These image-based recognition methods improve the usability of the system by removing the requirement for sensors or gloves with visual indicators for the hearing and speech impaired [28].

## 5. ArSL Recognition Approaches

In this section, we categorize the previous studies of ArSL recognition approaches into three classes, namely: alphabets recognition, isolated-words recognition, and continuous sentences recognition. Therefore, the following subsections describe the reviewed techniques as shown in Fig. 3. Also, summary of researcher’s efforts in designing ArSL recognition system for alphabets, isolated-words, and continuous sentences are outlined in Table 1, Table 2, and Table 3 respectively.

### 5.1. *Alphabet ArSL Recognition*

Recognizing the ArSL alphabet involves recognizing a set of gestures that correspond to the letters of the Arabic alphabet. This can be done using either a vision-based or/and a sensor-based approaches. One approach to recognizing the ArSL alphabet is to treat each letter as a distinct gesture and utilize machine learning techniques to classify the gestures based on visual or sensor features. In this section, a number of different approaches to the image-based recognition of Arabic alphabet signs are discussed.

In [29], Mohandes implements a system that allows for the automated recognition of the ArSL letters. For classification, SVMs are utilized, and moment invariants are utilized for feature selection. It is possible to get a recognition rate of 87%. Using a neuro-fuzzy system, AlJarrah and Halawani [28] analyze images of hand gestures. A PC camera reads the sign without the need for gloves. A PC camera captured the sign image without gloves. The model was stable over gesture position, size, and direction changes in the picture. They are successful in having a 93.55% recognition rate.

Al-Rousan and Hussain [30] develop a flexible neuro-fuzzy interference system to interpret letter. The hands area may be easily divided using a colorful glove to aid in the segmention procedure. A 9.55% success rate in recognition is obtained. Polynomial classifiers are used by Assaleh and Al-Rousan [23] to carry out the alphabet recognition. In the training phase, they build a new 2nd-order polynomial classifier for each class and use it to construct feature vectors. Both training and testing sets have identical error rates: 1.6% for training data and 6.59% for test data.

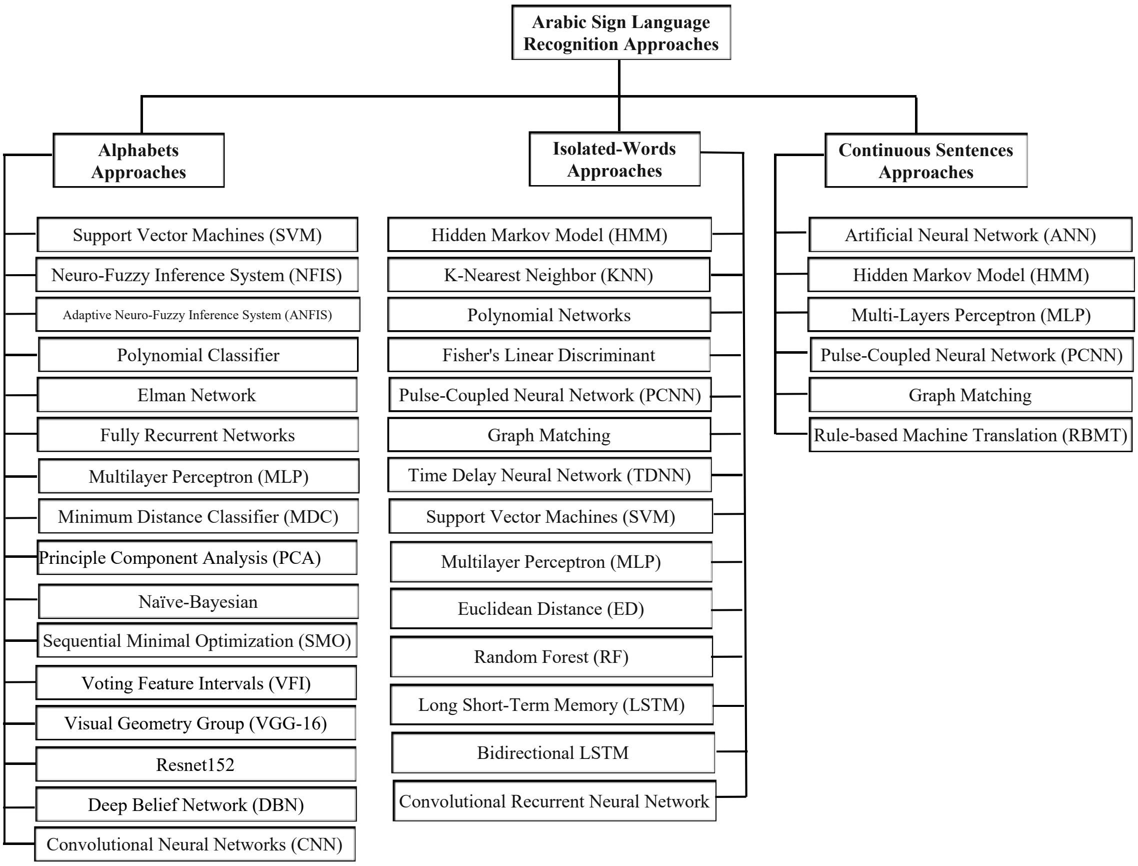
Arabic Sign Language Recognition Approaches
Alphabets Approaches
Isolated-Words Approaches
Continuous Sentences Approaches
Support Vector Machines (SVM)
Neuro-Fuzzy Inference System (NFIS)
Adaptive Neuro-Fuzzy Inference System (ANFIS)
Polynomial Classifier
Elman Network
Fully Recurrent Networks
Multilayer Perceptron (MLP)
Minimum Distance Classifier (MDC)
Principle Component Analysis (PCA)
Naïve-Bayesian
Sequential Minimal Optimization (SMO)
Voting Feature Intervals (VFI)
Visual Geometry Group (VGG-16)
Resnet152
Deep Belief Network (DBN)
Convolutional Neural Networks (CNN)
Hidden Markov Model (HMM)
K-Nearest Neighbor (KNN)
Polynomial Networks
Fisher's Linear Discriminant
Pulse-Coupled Neural Network (PCNN)
Graph Matching
Time Delay Neural Network (TDNN)
Support Vector Machines (SVM)
Multilayer Perceptron (MLP)
Euclidean Distance (ED)
Random Forest (RF)
Long Short-Term Memory (LSTM)
Bidirectional LSTM
Convolutional Recurrent Neural Network
Artificial Neural Network (ANN)
Hidden Markov Model (HMM)
Multi-Layers Perceptron (MLP)
Pulse-Coupled Neural Network (PCNN)
Graph Matching
Rule-based Machine Translation (RBMT)

Fig. 3. The classification of ArSL recognition approaches

In [21], Maraqa and Abu-Zaiter provide a method for recognizing letter signs using Elman and fully recurrent networks. There are 900 different signs presented, with 2 signers demonstrating 30 different gestures. Accuracy of 89.66% is reached by the Elman network, and accuracy of 95.11% is reached by the fully recurrent network. The Arabic alphabet sign translator created by El-Bendary et al. [3] has an accuracy of up to 91.3%. Their method takes a video of signs as input and produces textual representations of the letters as output using features extracted from the video. The characteristics employed are scalable, translational, and rotational invariant. The characteristics employed are scalable, translation-invariant, and rotation-invariant. The feature segmentation process involves a small pause among the signs representing each letter. The video frames and letter numbers are recognized during the pauses. Distances among 3 black pixels in each frame are utilized to create the feature vector. There are three classes into which the alphabet's signs can be classified. Recognizability is achieved by employing a minimum distance classifier (MDC) and a multilayer perceptron (MLP) neural network.

Hemayed and Hassanien [31] present a technique for recognizing signs in the Arabic alphabet and transforming them into spoken language. The system is far more applicable to the actual world, but it lacks the capability to perform recognition in real time. It emphasizes static gestures and simple movable motions. The system receives its input signal from the color images of the gestures. The YCbCr color space is utilized to extract the skin spots. Once the features have been extracted, the Prewitt edge detector is employed to further define the hand's shape. Images are transformed into feature vectors utilizing principal component analysis (PCA). The k-nearest neighbor (KNN) technique is utilized in the classification phase, and it is capable of achieving a success rate of 97%.

In [32], Tolba, M. F., et al., apply PCNN for the recognition of images. The image features that are not affected by a transformation are represented by the signature that PCNN generates. The system achieves rotational, scaling, and color invariance. The suggested model achieves a 90% recognition rate over 28 static postures of the Arabic alphabet. An image-based alphabet signs recognition system is created by Naoum et al. [33]. This system has a hit rate of 50% for with an ungloved hand, 75% with red gloves, 65% with black gloves, and 80% with white gloves. The images of signs are transformed into a histogram in the image processing phase, and the KNN method is then utilized to determine the surface behavior.

A real-time ArSL recognition method from high-resolution video is developed by Nadia Albelwi and Alginahi [34]. The technology performs preprocessing steps such as size normalization and skin detection. The feature extraction phase makes use of Fourier descriptors, and the classification phase employs KNN. This method has a recognition accuracy of 90.55 %.

In [1], Abdelmoty M. Ahmed et al. suggest developing a translation system for ArSL. The development of the Arabic text system, also referred to as ATASAT, relies on the creation of two datasets specifically designed for Arabic alphabet gestures. The researchers introduce a novel manual detection approach that enables the recognition and extraction of Arabic sign gestures from images or videos. This approach primarily focuses on analysing the hand's positioning and coverage within the image or video to identify and extract the relevant Arabic sign gestures. In order to get a more accurate classification, they additionally make use of a variety of statistical classifiers and draw comparisons. Transfer learning is used to a pre-trained network consisting of VGG-16 and Resnet152 by Saleh and Issa [35] in order to increase the accuracy of classifying 32 hand gestures taken from the ArSL dataset (ArSL2018). Random undersampling is employed on the dataset in order to reduce the total quantity of images in it from 54049 to 25600. This is done in order to minimize the imbalance that is caused by the difference in the sizes of the classes. They reach accuracy levels of 99.4% and 99.6% during testing on the VGG16 and Resnet152 techniques, respectively.

A. Hasasneh [36] proposes the ArSL technique. The recognition and classification of Arabic alphabetical letters can be accomplished by unsupervised deep learning by employing a deep belief network (DBN) together with tiny images. Deep learning has simplified recognition and extracted the most important sparsely represented features. 6,000 different samples for every one of the 28 signs that make up the Arabic alphabet are utilized for feature extraction after scaling and normalization. Classification precision is 83.32% using a softmax regression.

In [37], M. Kamruzzaman proposes a vision-based system that employs CNN for the recognition of Arabic hand sign-based letters and the translation of those letters into Arabic speech. Using a model that is based on deep learning, the suggested system is able to automatically recognize hand sign letters and then speak out the result using Arabic. This system has an accuracy rate of 90% when it comes to recognizing the Arabic hand sign-based letters. After the Arabic hand sign-based letters have been recognized, the results are provided to the text that is then fed into the speech engine. Subsequently, the speech engine generates the auditory representation of the Arabic language as its resultant output.

Gamal Tharwat et al. [20] propose an automated ArSL alphabet recognition system established on machine learning. They analyze a total of 2800 images in addition to 28 alphabets, and each class has ten participants. Each letter has 100 corresponding images, for a grand total of 2800. Classification is handled via KNN and MLP methods, while feature extraction makes use of a hand shape-based description in which each hand image is defined by a vector of 15 values representing important point positions. The accuracy of the tests is 97.548%.

For ArSL classification, Alani and Cosma [38] use CNN architecture. They demonstrate how the SMOTE oversampling technique increased the accuracy of the dataset being used. For the ArSL2018 dataset, the largest classification accuracy of the modified ArSLCNN model is 97.29% and 96.59%, without oversampling. Between each convolution, a pooling layer is employed to decrease the computational space's feature mapping dimensions.

Rahaf A. Alawwad et al. [39] utilizes a faster region-based CNN (R-CNN), they develop a unique ArSL recognition system that can locate and detect the ArSL alphabet. To learn the hand's location in an image, quicker R-CNN is modified to perform the process of extraction and mapping image features. Standard phone cameras are used to acquire 15,360 images of hand motions against a variety of backgrounds; these images are then employed to estimate the proposed model. A recognition rate of 93% is obtained for the compiled ArSL images dataset when the proposed model is combined with ResNet and VGG-16 methods.

Noor Azhar et al. [40] create a smartphone application that takes images of sign language and translate them into Arabic letters based on the user's input. They use supervised machine learning using CNN for the detection of shapes and edges when RGB color images are provided. The input is some trained images from internal storage and other is captured from camera (hand gesture image). They used Tensorflow package and Tensorflow Lite APIs which is well-suited with mobile application for testing the model performance. The accuracy is 91.1% with trained image from internal storage and 72.5% with new image captured from camera.

Table 1. Performance comparison of researchers' efforts for alphabet ArSL recognition approaches

| Ref. | Year | Recognition System | Techniques | | Input Source | Recognition Rate | Future Work |
|---|---|---|---|---|---|---|---|
| | | | Classification | Features | | | |
| **[29]** | 2001 | Image-based | SVM | Manual features | Images | 87% | Working on words |
| **[28]** | 2001 | Image-based | Adaptive NFIS (ANFIS) | Manual features | Images of bare hand | 93.55% | Working on words |
| **[30]** | 2001 | Image-based | NFIS | Manual features | Colored images with glove | 95.5% | Working on words |
| **[23]** | 2005 | Image-based | ANFIS and polynomial classifier | Manual features | Colored glove images | 36% and 57% misclassification decrease on training and test data, respectively | Using different feature sets |
| **[21]** | 2008 | Image-based | Elman and fully recurrent networks | Manual features | Colored glove images | 95.11% for fully recurrent network compared with 89.66% for Elman | Working on words |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **[3]** | 2010 | Image-based | MLP and MDC | edge-detection and feature-vector creation | Images of bare hands | 91.3% | Working on words |
| **[31]** | 2010 | Image-based | PCA | Prewitt edge detector | Images | 97% | Convert sign to voice and voice or text to sign |
| **[32]** | 2010 | Image-based | MLP | Discrete Fourier Transform (DFT) | Images | 90% | Apply the model in various problems |
| **[33]** | 2012 | Image-based | KNN | For surface behavior detection, a histogram of the signs is used | Images | Naked hands have a 50% hit rate, red gloves increase it to 75%, black gloves to 65%, and white gloves to 80%. | Working on words |
| **[34]** | 2012 | Image-based | KNN | Fourier transformations and Haar-Like | Real-time video | 90.55% | includes alphabets, numerals, and gestures |
| **[1]** | 2017 | Image-based | Compare between (C4.5, SMO,<br>VFI, MLP and Naïve-Bayesian) | Manual feature | a series of signs' videos or images | 50%,66%,75%,80 % and 90% | Sign language to voice |
| **[35]** | 2020 | Image-based | VGG-16, and Resnet152 | feature vector of images | images of (ArSL2018) with under-sampling | 99.4% for the VGG 16 and 99.6%<br>for the Resnet152 | SL to sound |
| **[36]** | 2020 | Image-based | DBN followed by SoftMax/ SVM | DBN-based feature extraction | image-signs | 83.32% | 1) Testing normalization and whitening<br>2) Studying sparsity factor parameters<br>3) Expanding dataset<br>4) Studying gesture similarities |
| **[37]** | 2020 | Image-based | CNN | CNN feature map | Images | 90% | increase dataset and accuracy |
| **[20]** | 2021 | Image-based | KNN | (KNN) and (MLP) | Images | 97.548% | -- |
| **[38]** | 2021 | Image-based | CNN | Feature Map and Filters in CNN | Images of (ArSL2018) | 96.59% to 97.29% after using SMOTE | Evaluating more datasets |
| **[39]** | 2021 | Image-based | Faster R-CNN | CNN | Images | 93% | increase accuracy and velocity |
| **[40]** | 2022 | Image-based | CNN | Tensorflow and Tensorflow Lite APIs | Images | 91.1% with trained image and 72.5% with new image captured from camera | Increase reliability and accuracy |

### 5.2. *Isolated-Words ArSL Recognition*

Isolated-word recognition, as contrast to alphabet signs recognition, focuses on essential frame extraction from the input video sign [19].

Mohandes and Deriche [41] suggest an image-based system for the recognition of ArSL. An HMM is used to execute the recognition process. The signer's face is identified using a Gaussian skin color model. After a face region has been recognized, the region expanding from the images containing the signs is utilized to follow the hands as they move. After the hand regions across the images have been recognized, a set of characteristics are chosen. The HMM receives these features as input. For a set of 50 signs, the proposed system achieves a recognition accuracy of 98%.

Shanableh and Assaleh [42] provide feature extraction strategies for ArSLrecognition in user-independent mode. They utilize colored gloves since the schemes rely on color segmentation to separate the user's hands. Errors from attempting to predict each successively segmented image from the previous one are compounded into two fixed images. Using accumulated prediction mistakes with a directed and weighted bias maintains motion in its intended direction. By calculating a bounding box around the accumulated prediction errors, we may exclude user-specific motion data. The discrete cosine transform (DCT) coefficients of the bounded images are zone-coded, and these coefficients are used to create the feature vectors. KNN and polynomial networks are used to verify the accuracy of the proposed solution. The classification rate in user-independent mode is reported to be 87%. For feature vector dimensionalities above 30, experiments suggest that KNN performs better than 2nd order polynomial networks due to numerical factors.

Shanableh and Assaleh [43] present a comprehensive range of techniques for extracting spatio-temporal features. These approaches are specifically designed for the purpose of recognizing isolated ArSL movements. The utilization of forward image predictions is employed in order to extract the temporal features of a gesture that is based on video. After thresholding the prediction errors, a single image is created that captures the motion of the entire sequence. In the subsequent steps, spatial domain features are extracted from the gesture representation using techniques like 2D DCT, Zonal coding, or radon transformation. These techniques allow for the analysis and extraction of important spatial characteristics. Afterward, the projected spatial features undergo ideal low pass filtering to enhance the relevant information. To complement the suggested feature extraction strategy, straightforward classification methods such as KNN and Bayesian classifiers are employed. These classification techniques enable the categorization of the extracted features, facilitating the recognition and interpretation of the sign language gestures. The recognition rates achieve in experiments range from 97% to 100%, demonstrating excellent categorization performance. The suggested method is tested using a battery of tests that employ traditional methods of categorizing data with temporal dependencies. Specifically, HMMs. In this case, the features are the differences between successive binarized images, and the strategies for extracting those differences come from the spatial domain. The experimental results indicate that the proposed feature extraction scheme, when combined with basic KNN or Bayesian classification, achieves comparable outcomes compared to the conventional HMM-based approach.

Shanableh and Assaleh [44] propose a feature extraction strategy comprising two tiers to recognize video-based isolated ArSL gestures. In the first tier, they convert the prediction error of the image sequence into a binary representation and produce two images that capture only the accumulated differences along a specific direction. In the second tier, the researchers employ two methods, namely frequency domain transformation and radon transformation, to process the accumulated differences images. These methods facilitate further feature extraction, allowing for the analysis of frequency components and shape/orientation properties of the ArSL gestures captured in the images. The accumulated difference images are subjected to such feature extractions, and the resulting feature vectors are combined. Before doing the second tier of feature extractions, the accumulated differences images are joined together. Fisher's linear discriminants are used to classify the two approaches, and the results are presented in the publication. When compared to previous efforts, the results show that as much as 39% of the incorrect categories have been fixed.

Youssif et al. [45] introduce an automatic ArSL recognition system using HMMs. Twenty isolated words from the standard ArSL have been recognized using a large sample set. The proposed system is signer-independent. Real ArSL videos filmed for deaf persons with a varying clothing and skin tones are used in the experiments. An overall recognition rate of 82.22% is possible with the proposed approach. In [46], Mohandes and Deriche present a signer-independent ArSL recognition system that combines face detection, geometric features, and HMM to achieve accurate and efficient sign language recognition. With the assumption that the signer is wearing gloves in the colors of yellow and orange, the hands areas are divided using a region-growing algorithm. With the expanded data set of 300 signs, the recognition accuracy now exceeds 95%.

Elons et al. [47] present a novel approach for developing a 3D model of hand posture by utilizing two 2D images captured from different viewpoints. These images are weighted and linearly merged to generate single 3D features. The researchers employ PCNN as the initial feature generator technique, followed by a nondeterministic finite automaton (NFA). The aim of this approach is to categorize 50 isolated ArSL words. To determine the most probable meaning of a gesture, a best-match approach is employed. The dataset used in this study comprises 50 isolated words, with an impressive recognition accuracy of 96% achieved by the proposed method.

Feras et al. [48] use a method based on images. Using a time delay neural network (TDNN), they made a system which can certainly recognize individual words in ArSL. In their study, they use gloves of two different colors to do sign. The sign is then changed using a method called image processing. After the image is processed, the features, such as the position of each hand's centroid and the change in horizontal and vertical speed, are taken out. After these features are extracted, they are put into groups using TDNN. During the test, 40 Arabic words are used, and 70% of them are recognized.

ElBadawy et al. [49] use a way to score based on canny edge detector to pick various frames as input to the created system, that employs 3D CNN to retrieve spatial temporal characteristics to detect 25 ArSL signals. The system's input is video, which is separated into frames for down sampling and scoring. The system gets a normalized depth video stream. Spatial-temporal input features are extracted. The 3D deep architecture experimented well. Observed data correctness is 98% and new data accuracy is 85%. Luqman, H. et al. [50] identify 23 ArSL two-handed terms, include variations in clothing, hand size, and camera distance. An accumulated image is transformed by Fourier, Hartley, and Log-Gabor. They compare Fourier, Hartley, and Log-Gabor transforms and KNN, MLP, and SVM models. In uncontrolled settings, they use KNN, SVM, and MLP classifiers. SVM has 98.8% recognition success. The study finds Hartley transform and SVM best.

Ibrahim et al. [51] segment hands using a face color-based dynamic skin detector. Skin-blob tracking is employed to monitor the movement of the hands. They track hands using hand blobs or binary silhouettes. Geometric features construct 97% accurate hand feature vectors. Their technology can only distinguish 30 separate words and continuous sign movements. All these 2D-based modeling algorithms use a single camera and require greater calculation time when identifying motions under different lighting. These approaches suffer when occlusions occur in real-time contexts.

Abdel-Gawad Abdel-Rabouh et al. [52] propose a method for ArSL recognition using a Kinect sensor. The proposed system uses the dynamic time warping matching algorithm for comparing between signs. They classify dynamic time warping data using nearest neighbor. The authors evaluate their system using a set of 30 standard ArSL words that had been isolated from their context. The results show that the system achieves an accuracy rate of 97.58% for signer-dependent online case and 95.25% for signer-independent online case, demonstrating its effectiveness for recognizing ArSL gestures.

Elpeltagy et al. [53] publish a new sign language recognition benchmark dataset and method. The ArSL approach includes not only the segmentation of the hand, but also the succession of hand shapes, a description of body movements, and the classification of signs. Sign classification uses RF classifiers and canonical correlation analysis. The method utilizes 150 signs by Kinectv2 sensor-collected from 21 signers. The total is 7500 samples. Finally, the algorithm's public data set solution is state-of-the-art. The accuracy of recognition is 55.57% based on the use of 150 ArSL signs.

In Basma Hisham et al. [54], they focus on recognizing ArSL using a combination of hardware and machine learning techniques. The hardware components employed in the system include a leap motion controller and latte panda, which enable capturing and processing hand gestures. For the recognition phase, the system utilizes two machine learning algorithms: KNN and SVM. These algorithms play a crucial role in analyzing and classifying the hand gestures captured by the leap motion controller. However, to improve the accuracy of these algorithms, an Ada-boosting technique is applied to create a stronger and more accurate classifier. In addition to the machine learning algorithms, the system incorporates a direct matching technique called dynamic time wrapping (DTW). DTW is used to compare and match the captured hand gestures with known patterns in the dataset. By applying DTW alongside Ada-boosting, the system aims to achieve higher accuracy in recognizing ArSL gestures. The dataset used for evaluation consists of 30 hand gestures, including both twenty single-hand gestures and ten double-hand gestures. The results of the experiments show that the DTW method can recognize single-hand gestures with 88% accuracy and double-hand gestures with 86% accuracy. After using Ada-Boosting, the proposed approach is able to recognize single-hand gestures at a 92.3% accuracy rate and double-hand gestures at a 93% accuracy rate.

In [55], Miada Almasre et al. present a dynamic prototype model (DPM) that is capable of recognizing specific ArSL gestured dynamic phrases and can employ Kinect as a sensor. The recognition and interpretation of ArSL can be enabled and made easier with the use of sensors and natural user interfaces. They compare 11 different predictive models that are constructed using SVM, KNN, and RF, and they discover that SVM with a radial basis kernel produces the best results, with an accuracy rate of 83.01%.

An effective method is created for recognizing ArSL using Kinect data in Mokhtar M. Mohamed [56]. The gesture's form and depth are used as input by the system. The input is processed in order to extract spatial and temporal data by the architecture. The gestures are labeled and separated from the background. The hands are then followed over a set interval of time. The gesture is identified by comparing it to a translation in the databank. In the end, the retrieval procedure completes the transformation from gestural to speech representation. More than 90% accuracy is achieved by the system.

Saleh Aly and Walaa Aly [57] propose an SLR system that is independent of the signer, utilizing deep learning architectures for various tasks. The system incorporates deep recurrent neural networks, hand semantic segmentation, and hand form feature encoding. To efficiently solve the hand segmentation task, the state-of-the-art semantic segmentation model, DeepLabv3+, is employed, which utilizes Resnet-50 as a backbone encoder network along with spatial pyramid pooling. Additionally, a single-layer convolutional self-organizing map (SOM) is employed to learn and represent hand shape features. Evaluating the proposed system on the Arabic benchmark dataset, it achieves an average accuracy of 89.5% when utilizing DeepLabv3+ for hand semantic segmentation and 69.0% without it.

Luqman and El-Alfy [58] suggest a multi-modal ArSL recognition system for recognizing sign gestures. This system can integrate both manual and non-manual motions to recognize sign gestures. Also, they introduce a sign language recognition multimodal video database. This evaluation takes into consideration not just sign-dependent but also sign-independent modalities, involving both manual and non-manual signs. In the first scenario, the researchers utilize color and depth images directly for analysis. However, in the second scenario, they select to extract additional sign-related characteristics by employing optical flow (OF) techniques. By leveraging MobileNet-LSTM with transfer training and fine-tuning, remarkable results are achieved. Specifically, the sign-dependent mode achieves an impressive accuracy of 99.7%, while the sign-independent mode attains a notable accuracy of 72.4%. The using of Kinect sensor to extract face expression key points make this method unsuitable for real-time SLR.

Abdul Wadood et al. [59] propose the utilization of a CNN augmented with an attention mechanism to facilitate the retrieval of spatial data. Additionally, they propose for the adoption of bio-inspired deep learning techniques, namely the bidirectional long short-term memory (BI-LSTM) architecture to extract temporal features. This model is tested with varying lighting, clothing, and camera distances. As a result of having less deep learning layers as well as parameters, the model emerges faster. On the 79-sign ArSL dataset, the 86-sign NVIDIA gesture database, and the 25-sign Jester dataset, the model attains an accuracy of 85.6%, 86.6%, and 95.8%, respectively. The bidirectional model outperforms LSTM models faster and more efficiently.

Abdelbasset Boukdir et al. [60] present an innovative deep learning-based method to classify video sequences depicting Moroccan sign language, known as ArSL. The research employs two distinct classification methods: the two-dimensional convolutional recurrent neural network (2DCRNN) and the three-dimensional CNN (3DCNN). Initially, the 2DCRNN model is utilized for extracting features by leveraging a recurrent network pattern, facilitating the detection of inter-frame relationships. Conversely, the 3DCNN model focuses on learning spatiotemporal features from small patches within the video data. Following the feature extraction by the 2DCRNN and 3DCNN models, a fully connected network is employed to classify the video data into various classes. To measure the effectiveness of the proposed model, a dataset consisting of 224 videos is utilized. These videos feature five individuals performing 56 different signs in ArSL. The dataset is divided into four parts, and the models are repeatedly trained and tested on different subsets of the data using a technique called four-fold cross-validation. The results obtained through the evaluation indicate the strong performance of the proposed model. The 2DCRNN model achieves an accuracy level of 92%, meaning it correctly classifies 92% of the video sequences, while the 3DCNN model achieves an even higher accuracy of 99%.

Mostafa M. Balaha et al. [61] propose a novel method for video classification and recognition by employing CNN and RNN in addition to the standard pre-processing of the raw video. To extract information from video frames, they employ two CNNs, which they then link together. RNN is utilized to recognize the interdependencies among sequential data and provide holistic predictions. With that method, they are able to attain state-of-the-art results, with a validation subset accuracy of 98% and a testing subset accuracy of 92%.

Sarah Al yami et al. [62] develop a hand and face keypoint framework for isolated ArSL recognition. MediaPipe pose estimator extracts sign gesture keypoints from the video stream. Three sign language recognition models - LSTM, temporal convolutional network (TCN), and transformer - are proposed using the retrieved keypoints. Non-manual elements are also examined for SLR systems. Arabic and Argentinian sign language models are tested. The accuracy of the pose-based transformer model is reported to be 99.74% and 68.2%, respectively, while operating in signer-dependent and independent modes, on the KArSL-100 Arabic dataset.

Table 2. Performance comparison of researchers' efforts for isolated-words ArSL recognition approaches

| Ref. | Year | Recognition System | Techniques | | Input Source | Recognition Rate | Future Work |
|---|---|---|---|---|---|---|---|
| | | | Classification | Feature | | | |
| **[41]** | 2005 | Image-based | HMM | Gaussian skin | The dataset consists of 500 signs | 98%. | selecting other relevant features |
| **[42]** | 2007 | Sensor-based | KNN and polynomial networks | Zonal-coded DCT coefficients | Signer wears color gloves | 87% | - |
| **[43]** | 2007 | Video-based | HMM | KNN and Bayesian | Video | 97% to 100% | - |
| **[44]** | 2007 | Video-based | Fisher's linear discriminant | Two tier feature extractions | Sequences of individual gesture | 39% of the misclassifications have been corrected | - |
| **[45]** | 2011 | Video-based | HMM | HMM | Video | 82.2% | Working on continuous sentence |
| **[46]** | 2012 | Image-based | HMM | Gaussian skin | Video | above 95%. | recognize continuous sentence for a larger dataset |
| **[47]** | 2013 | Image-based | Hybrid PCNN and graph matching approach | PCNN | 50 isolated words images | 96% | To distinguish between nearby gestures |
| **[48]** | 2014 | Video-based | TDNN | Manual features | 40 Video | 100% at training phase and 70.0% at testing phase | Working on continuous sentence |
| **[49]** | 2017 | Video-based | CNN and Softmax layer | Spatial-temporal features using CNN | 200 videos | Greater than 90% | Increase dataset |
| **[50]** | 2017 | Video-based | KNN, SVM, and MLP | Fourier, Hartley, and Log-Gabor transforms | Video | 98.8 to 99% | |
| **[51]** | 2017 | Video-based | ED | Geometric features of the spatial domain | 450 videos | 97% | Working on continuous sentences |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **[52]** | 2018 | Sensor-based | Nearest Neighbour | ED | 1200 samples | 97.58% for signer-dependent and 95.25% for signer-independent | Work on sentences |
| **[53]** | 2018 | Video-based | RF | HOG– PCA, CCA, and Cov3DJ+ | 150 videos | 55.57% | |
| **[54]** | 2020 | Sensor-based | KNN and SVM | filter feature selection | 30 hand gestures | 92.3% for single-hand and 93% for double-hand | increasing the accuracy, recognizing full sentences, and reducing computation time |
| **[55]** | 2020 | Sensor-based | SVM, RF, KNN | From sensors | gesture words | 83 % for SVM | produce higher accuracy |
| **[56]** | 2020 | Video-based | CNN | Depth camera | Alphabets and word gestures | 90% | increase accuracy and working on real-time |
| **[57]** | 2020 | Video-based | deep bi-directional LSTM network | Convolutional SOM (CSOM) | 3450 videos | 89.5% using DeepLabv3+ hand segmentation, 69.0% without hand segmentation | Work on sentences |
| **[58]** | 2021 | Video-based | LSTM and Softmax | CNN | 6748 videos | 99.7% for signer-dependent and 72.4% for signer-independent modes | target continuous word |
| **[59]** | 2021 | Video-based | CNN and Bidirectional LSTM | CNN and Bidirectional LSTM | ArSL, Jester, and NVIDIA Gesture datasets | 85.6%, 95.8%, and 86.6% for ArSL, Jester, and NVIDIA Gesture datasets respectively | test it for other domains |
| **[60]** | 2021 | Video-based | 2DCRNN and 3DCNN | 2DCNN with RNN and 3DCNN | 224 videos | 92% for 2DCRNN and 99% for 3DCNN | merging the two patterns into one |
| **[61]** | 2023 | Video-based | CNN and RNN | double CNNs | 8,467 videos | 98% and 92% on validation and testing | enlarge dataset and work on sentences |
| **[62]** | 2023 | Video-based | A fully connected layer with a Softmax | MediaPipe pose estimator | 100 videos | 99.74% and 68.2% in signer-dependent and independent modes | Explore other architectures for ArSL |

### 5.3. *Sentences ArSL Recognition*

When compared to the first two approaches, continuous sign recognition is far more difficult to implement. Problems with this approach include tracking hands, detecting motion, extracting features, and dealing with a large vocabulary. Once the appropriate features vector has been recovered, classification is straightforward thanks to a variety of available methods like KNN and HMM.

M. Saied Abdel-Wahab et al. [63] break continuous gestures down into more static postures, and then using an ANN model for the recognition stage. They represent the sequence of gestures as a graph. The graph matching method is responsible for the recognition of gestures. The final recognition rate for the test set is 90.5%. Assaleh et al. [64] have 40 sentences from a single signer with 19 repetitions. A continuous ArSL recognition system that is user dependent is created by the authors. This system uses HMM and spatial-temporal feature extraction to recognize sentences with a recognition accuracy of 75%.

Tolba and Abul-Ela [65] represent an innovative graph matching method which is intended for the recognition of continuous sentences in ArSL. They come up with a method that uses connected sequence gesture recognition. The accuracy of recognition is greater than 70% for thirty continuous sentences made up of hundred gestures. The evaluation reveals that taking this technique yields some promising results. In [66], Tolba et al. implement PCNN and graph matching to create a continuous ArSL recognition system. Three and four-word sentences are used extensively in the tests. Before employing the graph mapping approach, signs are reduced to their component parts and static postures. The continual gestures that the user inputs are the primary focus of this paper's recognition and categorization efforts. Accuracy of 80% is attained in the Arabic sign dataset used in this study.

In Ala Addin I. Sidig et al. [67], a model for ArSL recognition based on optical flow and HMM is proposed. Fourier transform, local binary pattern, HOG, and optical flow are used to process the signs in this model. The classification accuracy of an HMM trained with modified Fourier transform (MFT) features is 99.11%. Luqman and El-Alfy [68] propose a novel approach to sign language gesture recognition by utilizing motion and spatial features in a cascaded architecture of CNN and LSTM models. The primary goal is to improve the performance of SLR systems by effectively capturing and integrating the temporal dynamics and spatial information of gestures. Experiments on benchmark datasets demonstrate that the suggested method surpasses the state-of-the-art techniques, suggesting it could be used in practical SLR settings. The recognition accuracy of the proposed method outperforms other techniques with over 99% accuracy.

The first ArSL recognition system that can translate ArSL into Arabic sentences is proposed by Hamzah Luqman et al. [69]. The ArSL word is first processed for morphological analysis, then syntactic analysis, as part of the three-stage rule-based machine translation (RBMT) approach. The last step is to switch to Arabic sentences. However, a corpus of statements commonly found in healthcare settings is utilized by the algorithm. There are 600 sentences totaling 3327 sign words (593 unique sign words) and 593 different sign words. The suggested dataset is split into a training dataset (70 percent), a validation dataset (15 percent), and a testing dataset (5 percent). Manual and automatic processes provide the system's output. A manual review by two ArSL specialists, however, reveals that 80 percent of the sentences are translated correctly. The BLEU and TER measures automatically analyze it, yielding consistent scores of 0.39 and 0.45.

In [70], Hamzah Luqman presents a novel multi-modality dataset and benchmark for the purpose of continuous ArSL recognition. The purpose of this dataset is to help researchers in the Arabic-speaking world overcome the difficulties of creating high-quality sign language recognition algorithms. Three pieces of data, namely color, depth, and skeleton joint points, are acquired concurrently for each sentence using a Kinectv2 camera. He has also put out other models for sign language recognition, including encoder-decoder and attention models. Two pre-trained approaches are used to extract spatial information from the phrase frames, which are then used as inputs for the proposed approaches. There is considerable space for improvement, as even the top performing model only managed a word error rate (WER) of 0.50.

Table 3. Performance comparison of researchers' efforts for sentences ArSL recognition approaches

| Ref | Year | Recognition System | Techniques | | Input Source | Recognition Rate | Future work |
|---|---|---|---|---|---|---|---|
| | | | Classification | Feature | | | |
| **[63]** | 2006 | Video-based | ANN | Graph matching | Videos | 90.5% | Using natural languages techniques |
| **[64]** | 2010 | Video-based | HMM | Spatio-temporal feature extraction | Video frames | 75%. | Satisfy higher recognition rates |
| **[65]** | 2012 | Video-based | MLP | PCNN | Video | Until 70% | Increase the recognition accuracy |
| **[66]** | 2013 | Video-based | PCNN and graph matching | PCNN | Video | 70% for 30 consecutive sentences with 100 gestures | Using natural languages techniques |
| **[67]** | 2018 | Video-based | HMMs | MFT, LBP, HOG, and combination of HOF-HOG | Videos | 99.11% | motion and appearance in one feature vector utilizing HOF-HOG |
| **[69]** | 2020 | Image-based | RBMT | morphological and syntactic analysis | Images | more than 80% | -- |
| **[68]** | 2022 | Video-based | Softmax | CNN | Videos | 99% | Continuous sign language recognition |
| **[70]** | 2023 | Video-based and Sensor-based | encoder-decoder model | Pretrained CNN | sentence by Kinect | WER of 0.50 | |

## 6. ArSL Datasets

The availability of large and diverse datasets that capture the variability and complexity of sign language gestures is critical for the development and evaluation of ArSL recognition systems. These datasets typically contain video or sensor data of sign language gestures, along with corresponding annotations that indicate the meaning or label of each gesture. Although adequate video data is accessible online, it will be inappropriate for training ArSL recognition systems since it lacks annotations and the signs have not been segmented. The lack of a large-scale benchmarking dataset is a problem for ArSL recognition systems. It is challenging to locate a comprehensive dataset that meets the requirements of ArSL recognition. This is due in part to the scarcity of qualified ArSL specialists and the time and cost required to collect data on the sign language [71]. In addition, researchers may encounter challenges in obtaining valid ArSL datasets due to the inherent complexity of the Arabic language. Since some studies have created their own data that is generally limited or unavailable to other researchers, it may be difficult to directly compare the recognition accuracies of the various methodologies. Most of these datasets are also camera-based, meaning they lack depth information [71]. Therefore, the majority of researchers are required to manually generate datasets, which is a time-consuming and laborious procedure. Table 4 provides some examples of publicly available ArSL datasets.

Table 4. A publicly available ArSL datasets

| Ref | Dataset Description |
|---|---|
| **[72]/2011** | Six gestures are used to generate 6,000 different sign images. |
| **[73]/2012** | The 80-word vocabulary is used to construct 40 sentences with no restrictions on syntax or sentence length; this process is repeated 19 times. |
| **[47]/2013** | There are 270 postures that make up the 200 gestures, with 189 postures involving two hands and 81 postures comprising only one hand. Every gesture is carried out ten times, each time by a different two person. |
| **[74]/2014** | There are a total of 2800 frames in the dataset generated from a single user's input of 28 alphabets, with 10 samples of each letter. |

| | |
|---|---|
| **[75]/2015** | The database contains about five hundred static gestures, including "finger spelling, hand movements" (non-manual signs). Lip reading, body language, and facial expressions all play significant roles. |
| **[76]/2016** | Two sets of static alphabet data exist: 700 instances for each 28 characters written with naked hands and colored gloves. |
| **[49]/2017** | 200 samples are taken from the unified ArSL lexicon, with each of the 25 signs being performed by two different signers four times. 125 for training and 75 for testing |
| **[77]/2018** | Thirty people are serious mobile photographers. Volunteers gesture these 30 ArSL alphabets. There are a total of 900 images spread across 30 letters. |
| **[78]/2018** | Captured 450 colorful ArSL videos |
| **[79]/2019** | 28 Arabic letters and numerals (0-10) are represented by 7869 images for recognition. |
| **[80]/2019** | The dataset ArSL2018 comprises a collection of 54,049 images, which accurately depict the 32 alphabets and signs of ArSL. These images have been donated by a group of 40 signers. |
| **[81]/2020** | A total of 44 signs (29 single-handed and 15 double-handed) are executed by a group of 5 signers, where 80% are used for training and 20% for testing. |
| **[20]/2021** | There are 9240 images of the Arabic alphabet from 10 places and age groups. These images are organized in four separate datasets. |
| **[71]/2021** | There are eleven chapters totaling 502 signs that make up the words in the ArSL lexicon. Three signers are used for each sign. There are a total of 75300 samples, the result of 50 repetitions of each sign by each signer. |
| **[82]/2021** | There are a total of 220000 images in the dataset, split amongst 44 different classes (32 letters, 11 digits (0-10) and 1). There are 5000 images total, taken by various people, of each of the stationary signs. |
| **[83]/2023** | It contains 7,856 RGB images of ArSL alphabets. Data is collected from over 200 people in a wide range of shooting situations (including but not limited to: lighting, background, image orientation, size, and resolution). |

## 7. Discussion and Analysis

Sign language remains the only simple tool for communication among the hearing-impaired and the world. SLR systems are still in their beginning. Nowadays, the provision of commercial translation services relies predominantly on human translators, resulting in high costs attributed to the requisite expertise and personal involvement. Therefore, there is a need to improve the automatic sign language recognition for helping and serving the deaf and dumb people in interpretation their mother language. These systems work as an interpreter for deaf individuals and normal people to enhance their communication [1].

Multiple techniques have been established for interpreting some signs languages into spoken languages. Yet, it is important to mention that there are a relatively small number of efforts to construct techniques that automate the interpretation for the ArSL recognition [22; 23; 24; 25; 63]. Having software that may identify ArSL and interpret it into Arabic language and vice versa, can help to reduce the gap among Arab deaf community and the rest of the world. It will also improve their education and give them access to sciences using their instinctive language.

Pattern recognition, image processing, natural language processing, and linguistics are just a few of the numerous fields that contribute to the study of SLR. The challenge has several dimensions, including the difficulty of learning to recognize hand movements visually and the wide variety of sign languages. Sign languages, like spoken languages, have a syntax and grammar, but they differ from spoken languages because the sign language structure uses several body motions simultaneously, rather than sequentially, to convey meaning. Sign language's linguistic characteristics result from the use of not just hand gestures but also facial expressions and head movements to convey meaning [5].

There are two primary dimensions to the study of hand gesture and sign recognition: isolated recognition and continuous recognition. In isolated recognition, only one hand gesture from a user is studied for recognition. In continuous recognition, the user is required to produce a series of gestures in rapid succession, with each gesture being individually recognized. Systems that recognize hand gestures and those that recognize sign language present a variant of the continuous recognition issue. The issue may be viewed as a gesture spotting challenge, where the goal is to identify the user's intended gestures and filter out any noise. The co-articulation issue is a subset of the continuous recognition problem when it comes to recognizing signs. Because one sign may have an effect on the next, transitions between signs need to be explicitly modelled and included into the recognition system, making the process more difficult. In addition, language models are employed to enable operations on databases containing extensive vocabularies [20].

The difficulty of representing motion in dynamic hand gestures is another difficult issue. There are two interrelated components to this problem: (i) the methodology for extracting the characteristics inside a single frame, and (ii) the approach for describing and modelling the complete motion trajectory. The dynamical gestures generate sequences of varying lengths, therefore most of the research in the literature employ generative models because of their ability to deal with this kind of data. The next phase in hand gesture recognition is the development of new generative models and the integration of discriminative techniques into current models that can better illustrate the natural dynamics of the gestures while enhancing system performance [2].

Although ArSL recognition research has been underway for a number of years, it remains restricted in many respects. There is a sense that this field is still in its infancy. To the best of our knowledge, no extensive systems have been created so far that cover the current breadth. HCI research in this area will unavoidably have an impact on related fields. This paper presents a concise review of the two predominant techniques employed in the conversion of sign language into written text: the vision-based approaches and the sensor-based approaches. The goal of any strategy for bridging the gap between the hearing and the deaf is to facilitate conversation in everyday life without the use of special equipment or colored gloves. In order to make life easier and lessen these restrictions, researchers have made great efforts to create these methods. Microsoft Kinect is utilized as an interface for ArSL recognition. However, its applicability for ArSL recognition is limited [20].

According to our review of the relevant literature, researchers have had great success in this area by implementing vision-based recognition approaches to alphabet, isolated-words, and sentences for ArSL recognition. It's important to mention that we couldn't discover any work that employed a sensor-based recognition strategy to recognize the alphabet for ArSL recognition. However, sensor-based recognition approaches have seen little utility for recognizing isolated-words and sentences for ArSL recognition. This demonstrates the feasibility of research efforts utilizing both vision-based and sensor-based techniques, both of which have the potential to achieve the best performance in this area.

The lack of a large-scale benchmarking dataset is a problem for ArSL recognition systems. It is challenging to locate a comprehensive dataset that meets the requirements of ArSL recognition. This is due in part to the scarcity of qualified ArSL specialists and the time and cost required to collect data on the sign language [71]. In addition, researchers may encounter challenges in obtaining valid ArSL datasets due to the inherent complexity of the Arabic language. Since some studies have created their own data that is generally limited or unavailable to other researchers, it may be difficult to directly compare the recognition accuracies of the various methodologies. Most of these datasets are also camera-based, meaning they lack depth information [71].

## 8. Conclusion

This paper has given a brief overview from the several perspectives of ArSL recognition available in the open literature. The emphasis has been presenting on the critical features of the language and the models and techniques developed to date. In the last decade, researchers have focused extensively on the problem of recognizing isolated sign languages, proposing methods with high accuracy in the reported datasets of a variety of sign languages from across the world. There are no standard Arabic sign datasets available for researchers to use in testing and comparing their systems. Some datasets have been released to the public, although they have not yet reached the status of "benchmark datasets" among ArSL recognition researchers. The current difficulties in automatic ArSL recognition may be broken down into five categories: continuous signing; huge vocabulary recognition; analysis and integration of non-manual signals; grammatical processes in manual signing; and integrating non-manual signals. Even though a number of researchers have noted these aspects, there is still a restricted amount of study that has been done in these areas. Researchers in ArSL recognition and linguists who specialize in sign language can make significant advances by collaborating closely with one another.

## Acknowledgment

The authors extend their appreciation to the Deanship of Scientific Research at King Khalid University for funding this work through Large Groups [grant number RGP.2/253/44].

## Conflict of Interest

The authors have no conflicts of interest to declare.

## References

[1] Ahmed, A. M., *et al.* (2017). Automatic Translation of Arabic Sign To Arabic Text (ATASAT) System. *Computer Science & Information Technology*, 109.

[2] AbdElghfar, H. A., *et al.* (2023). QSLRS-CNN: Qur'anic sign language recognition system based on convolutional neural networks. *The Imaging Science Journal*, 1-13.

[3] El-Bendary, N., *et al.* (2010). *ArSLAT: Arabic sign language alphabets translator.* Paper presented at the Computer Information Systems and Industrial Management Applications (CISIM), 2010 International Conference on.

[4] Soliman, A. M., *et al.* (2023). Arabic Sign Language Recognition System: Using an Image-Based hand Gesture Detection Method to help Deaf and Dump Children to Engage in Education. مجلة كلية الآداب بقنا, 32(58), 1-28.

[5] Sahoo, A. K., *et al.* (2014). Sign language recognition: State of the art. *ARPN Journal of Engineering and Applied Sciences, 9*(2), 116-134.

[6] Soliman, A. M., *et al.* (2022). Recognize The Alphabet of Fingerspelling Using Statistical Classifiers to Facilitate Communication Between Hearing-Impaired Persons and Others. مجلة كلية الآداب بقنا, 31(57), 1-30.

[7] Han, J., *et al.* (2013). Enhanced computer vision with microsoft kinect sensor: A review. *IEEE transactions on cybernetics, 43*(5), 1318-1334.

[8] AbdElghfar, H. A., *et al.* (2023). A Model for Qur'anic Sign Language Recognition Based on Deep Learning Algorithms. *Journal of Sensors, 2023.*

[9] Ahmed, A., *et al.* (2019). Arabic sign language translator. *Journal of Computer Science, 15*(10), 1522-1537.

[10] Ahmed, A. M., *et al.* (2017). *Towards the design of automatic translation system from Arabic Sign Language to Arabic text.* Paper presented at the 2017 International Conference on Inventive Computing and Informatics (ICICI).

[11] Ahmed, A. M., *et al.* (2015). Propose a New Method for Extracting Hand using in the Arabic Sign Language Recognition (Arslr) System. *International Journal of Engineering Research & Technology (IJERT), 4*(11), 2278-0181.

[12] Ahmed, A. M., *et al.* (2018). *Gestures Arabic Sign Language Conversion to Arabic Alphabets.* Paper presented at the 2018 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC).

[13] Khattab, M. M., *et al.* (2018). *Multi-frame super-resolution: A survey.* Paper presented at the 2018 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC).

[14] Khattab, M. M., *et al.* (2020). Regularization-based multi-frame super-resolution: a systematic review. *Journal of King Saud University-Computer and Information Sciences, 32*(7), 755-762.

[15] Khattab, M. M., *et al.* (2021). Regularized multiframe Super-Resolution image reconstruction using linear and nonlinear filters. *Journal of Electrical and Computer Engineering, 2021*, 1-16.

[16] Khattab, M. M., *et al.* (2023). A Hybrid Regularization-Based Multi-Frame Super-Resolution Using Bayesian Framework. *Computer Systems Science & Engineering, 44*(1).

[17] Organization, W. H. (2018). Addressing the rising prevalence of hearing loss.

[18] Chadha, S., *et al.* (2021). The world report on hearing, 2021. *Bulletin of the World Health Organization, 99*(4), 242.

[19] Ahmed, A. M., *et al.* (2020). Arabic sign language intelligent translator. *The Imaging Science Journal, 68*(1), 11-23.

[20] Tharwat, G., *et al.* (2021). Arabic sign language recognition system for alphabets using machine learning techniques. *Journal of Electrical and Computer Engineering, 2021*, 1-17.

[21] Maraqa, M., & Abu-Zaiter, R. (2008). *Recognition of Arabic Sign Language (ArSL) using recurrent neural networks.* Paper presented at the Applications of Digital Information and Web Technologies, 2008. ICADIWT 2008. First International Conference on the.

[22] Tubaiz, N., *et al.* (2015). Glove-based continuous Arabic sign language recognition in user-dependent mode. *IEEE Transactions on Human-Machine Systems, 45*(4), 526-533.

[23] Assaleh, K., & Al-Rousan, M. (2005). Recognition of Arabic sign language alphabet using polynomial classifiers. *EURASIP Journal on Advances in Signal Processing, 2005*(13), 507614.

[24] Tharwat, A., *et al.* (2015). *Sift-based arabic sign language recognition system.* Paper presented at the Afro-european conference for industrial advancement.

[25] Bauer, B., & Hienz, H. (2000). *Relevant features for video-based continuous sign language recognition.* Paper presented at the Automatic Face and Gesture Recognition, 2000. Proceedings. Fourth IEEE International Conference on.

[26] Al-Rousan, M., *et al.* (2009). Video-based signer-independent Arabic sign language recognition using hidden Markov models. *Applied Soft Computing, 9*(3), 990-999.

[27] Shanableh, T., *et al.* (2007). Spatio-temporal feature-extraction techniques for isolated gesture recognition in Arabic sign language. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 37*(3), 641-650.

[28] Al-Jarrah, O., & Halawani, A. (2001). Recognition of gestures in Arabic sign language using neuro-fuzzy systems. *Artificial Intelligence, 133*(1-2), 117-138.

[29] Mohandes, M. (2001). *Arabic sign language recognition.* Paper presented at the International conference of imaging science, systems, and technology, Las Vegas, Nevada, USA.

[30] Al-Rousan, M., & Hussain, M. (2001). Automatic recognition of Arabic sign language finger spelling. *International Journal of Computers and Their Applications, 8*, 80-88.

[31] Hemayed, E. E., & Hassanien, A. S. (2010). *Edge-based recognizer for Arabic sign language alphabet (ArS2V-Arabic sign to voice).* Paper presented at the Computer Engineering Conference (ICENCO), 2010 International.

[32] Tolba, M. F., *et al.* (2010). Image signature improving by PCNN for Arabic sign language recognition. *Canadian Journal on Artificial Intelligence, Machine Learning & Pattern Recognition, 1*(1), 1-6.

[33] Naoum, R., *et al.* (2012). Development of a new arabic sign language recognition using k-nearest neighbor algorithm.

[34] Albelwi, N. R., & Alginahi, Y. M. (2012). *Real-time arabic sign language (arsl) recognition.* Paper presented at the International Conference on Communications and Information Technology.

[35] Saleh, Y., & Issa, G. (2020). Arabic sign language recognition through deep neural networks fine-tuning.

[36] Hasasneh, A. (2020). Arabic Sign Language Characters Recognition Based on A Deep Learning Approach and a Simple Linear Classifier. *Jordanian Journal of Computers and Information Technology, 6*(3).

[37] Kamruzzaman, M. (2020). Arabic sign language recognition and generating Arabic speech using convolutional neural network. *Wireless Communications and Mobile Computing, 2020.*

[38] Alani, A. A., & Cosma, G. (2021). ArSL-CNN: a convolutional neural network for Arabic sign language gesture recognition. *Indonesian journal of electrical engineering and computer science, 22.*

[39] Alawwad, R. A., *et al.* (2021). Arabic sign language recognition using faster R-CNN. *International Journal of Advanced Computer Science and Applications, 12*(3).

[40] Azhar, N. A. N., *et al.* (2022). Development of Mobile Application for Arabic Sign Language based on Android Studio Software. *JOURNAL OF ALGEBRAIC STATISTICS, 13*(3), 3152-3160.

[41] Mohandes, M., & Deriche, M. (2005). *Image based Arabic sign language recognition.* Paper presented at the Signal Processing and Its Applications, 2005. Proceedings of the Eighth International Symposium on.

[42] Shanableh, T., & Assaleh, K. (2007). *Arabic sign language recognition in user-independent mode.* Paper presented at the Intelligent and Advanced Systems, 2007. ICIAS 2007. International Conference on.

[43] Shanableh, T., & Assaleh, K. (2007). *Video-based feature extraction techniques for isolated Arabic sign language recognition.* Paper presented at the Signal Processing and Its Applications, 2007. ISSPA 2007. 9th International Symposium on.

[44] Shanableh, T., & Assaleh, K. (2007). *Two tier feature extractions for recognition of isolated arabic sign language using fisher's linear discriminants.* Paper presented at the Acoustics, Speech and Signal Processing, 2007. ICASSP 2007. IEEE International Conference on.

[45] Youssif, A. A., *et al.* (2011). Arabic sign language (arsl) recognition system using hmm. *International Journal of Advanced Computer Science and Applications (IJACSA), 2*(11).

[46] Mohandes, M., *et al.* (2012). A signer-independent Arabic Sign Language recognition system using face detection, geometric features, and a Hidden Markov Model. *Computers & Electrical Engineering, 38*(2), 422-433.

[47] Elons, A. S., *et al.* (2013). A proposed PCNN features quality optimization technique for pose-invariant 3D Arabic sign language recognition. *Applied Soft Computing, 13*(4), 1646-1660.

[48] Al Mashagba, F. F., *et al.* (2014). Automatic Isolated-Word Arabic Sign Language Recognition System Based on Time Delay Neural Networks. *Research Journal of Applied Sciences, Engineering and Technology, 7*(11), 2261-2265.

[49] ElBadawy, M., *et al.* (2017). *Arabic sign language recognition with 3d convolutional neural networks.* Paper presented at the 2017 Eighth international conference on intelligent computing and information systems (ICICIS).

[50] Luqman, H., & Mahmoud, S. A. (2017). Transform-based Arabic sign language recognition. *Procedia Computer Science, 117*, 2-9.

[51] Ibrahim, N. B., *et al.* (2017). An Automatic Arabic Sign Language Recognition System (ArSLRS). *Journal of King Saud University-Computer and Information Sciences*.

[52] Abdel-Samie, A.-G. A.-R., *et al.* (2018). Arabic sign language recognition using kinect sensor. *Research Journal of Applied Sciences, Engineering and Technology, 15*(2), 57-67.

[53] Elpeltagy, M., *et al.* (2018). Multi-modality-based Arabic sign language recognition. *IET Computer Vision, 12*(7), 1031-1039.

[54] Hisham, B., & Hamouda, A. (2021). Arabic sign language recognition using Ada-Boosting based on a leap motion controller. *International Journal of Information Technology, 13*, 1221-1234.

[55] Almasre, M. A., & Al-Nuaim, H. (2020). A comparison of Arabic sign language dynamic gesture recognition models. *Heliyon, 6*(3).

[56] Mohamed, M. M. (2020). Automatic system for Arabic sign language recognition and translation to spoken one. *International Journal, 9*(5).

[57] Aly, S., & Aly, W. (2020). DeepArSLR: A novel signer-independent deep learning framework for isolated arabic sign language gestures recognition. *IEEE Access, 8*, 83199-83212.

[58] Luqman, H., & El-Alfy, E.-S. M. (2021). Towards hybrid multimodal manual and non-manual Arabic sign language recognition: MArSL database and pilot study. *Electronics, 10*(14), 1739.

[59] Abdul, W., *et al.* (2021). Intelligent real-time Arabic sign language classification using attention-based inception and BiLSTM. *Computers and Electrical Engineering, 95*, 107395.

[60] Boukdir, A., *et al.* Isolated video-based Arabic sign language recognition using convolutional and recursive neural networks. *Arabian Journal for Science and Engineering*, 1-13.

[61] Balaha, M. M., *et al.* (2023). A vision-based deep learning approach for independent-users Arabic sign language interpretation. *Multimedia Tools and Applications, 82*(5), 6807-6826.

[62] Alyami, S., *et al.* (2023). Isolated Arabic Sign Language Recognition Using A Transformer-based Model and Landmark Keypoints. *ACM Transactions on Asian and Low-Resource Language Information Processing*.

[63] Abdel-Wahab, M. S., *et al.* (2006). *Arabic sign language recognition using neural network and graph matching techniques.* Paper presented at the Proceedings of the 6th WSEAS International Conference on Applied Informatics and Communications.

[64] Assaleh, K., *et al.* (2010). Continuous Arabic sign language recognition in user dependent mode.

[65] Tolba, M., *et al.* (2012). *A proposed graph matching technique for Arabic sign language continuous sentences recognition.* Paper presented at the Informatics and Systems (INFOS), 2012 8th International Conference on.

[66] Tolba, M. F., *et al.* (2013). Arabic sign language continuous sentences recognition using PCNN and graph matching. *Neural Computing and Applications, 23*(3-4), 999-1010.

[67] Sidig, A. a. I., *et al.* (2018). *Arabic sign language recognition using optical flow-based features and HMM.* Paper presented at the Recent Trends in Information and Communication Technology: Proceedings of the 2nd International Conference of Reliable Information and Communication Technology (IRICT 2017).

[68] LUQMAN, H., & ELALFY, E. (2022). Utilizing motion and spatial features for sign language gesture recognition using cascaded CNN and LSTM models. *Turkish Journal of Electrical Engineering and Computer Sciences, 30*(7), 2508-2525.

[69] Luqman, H., & Mahmoud, S. A. (2020). A machine translation system from Arabic sign language to Arabic. *Universal Access in the Information Society, 19*(4), 891-904.

[70] Luqman, H. (2023). *ArabSign: A Multi-modality Dataset and Benchmark for Continuous Arabic Sign Language Recognition.* Paper presented at the 2023 IEEE 17th International Conference on Automatic Face and Gesture Recognition (FG).

[71] Sidig, A. A. I., *et al.* (2021). KArSL: Arabic sign language database. *ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP), 20*(1), 1-19.

[72] Nagi, J., *et al.* (2011). *Max-pooling convolutional neural networks for vision-based hand gesture recognition.* Paper presented at the 2011 IEEE international conference on signal and image processing applications (ICSIPA), Kuala Lumpur, Malaysia.

[73] Assaleh, K., *et al.* (2012). *Low complexity classification system for glove-based arabic sign language recognition.* Paper presented at the Neural Information Processing: 19th International Conference, ICONIP 2012, Doha, Qatar, November 12-15, 2012, Proceedings, Part III 19.

[74] Mohandes, M., *et al.* (2014). *Arabic sign language recognition using the leap motion controller.* Paper presented at the 2014 IEEE 23rd International Symposium on Industrial Electronics (ISIE).

[75] Shohieb, S. M., *et al.* (2015). Signsworld atlas; a benchmark Arabic sign language database. *Journal of King Saud University-Computer and Information Sciences, 27*(1), 68-76.

[76] Ahmed, A. M., *et al.* (2016). Automatic translation of Arabic sign to Arabic text (ATASAT) system. *Journal of Computer Science and Information Technology, 6*, 109-122.

[77] Alzohairi, R., *et al.* (2018). Image based arabic sign language recognition system. *International Journal of Advanced Computer Science and Applications (IJACSA), 9*(3), 185-194.

[78] Ibrahim, N. B., *et al.* (2018). An automatic Arabic sign language recognition system (ArSLRS). *Journal of King Saud University-Computer and Information Sciences, 30*(4), 470-477.

[79] Hayani, S., *et al.* (2019). *Arab sign language recognition with convolutional neural networks.* Paper presented at the 2019 International Conference of Computer Science and Renewable Energies (ICCSRE), Agadir, Morocco.

[80] Latif, G., *et al.* (2019). ArASL: Arabic alphabets sign language dataset. *Data in brief, 23*, 103777.

[81] Alnahhas, A., *et al.* (2020). Enhancing the recognition of Arabic sign language by using deep learning and leap motion controller. *Int. J. Sci. Technol. Res, 9*, 1865-1870.

[82] Ismail, M. H., *et al.* (2021). Static hand gesture recognition of Arabic sign language by using deep CNNs. *Indonesian Journal of Electrical Engineering and Computer Science, 24*(1), 178-188.

[83] Al-Barham, M., *et al.* (2023). RGB Arabic Alphabets Sign Language Dataset. *arXiv preprint arXiv:2301.11932*.

## Authors Profile

**Ahmad M.J. Al Moustafa** received his M.Sc. degrees from UNIVERSITI TEKNOLOGI MALAYSIA, Johor, Malaysia in 2007 and B.Sc. from Sudan University of Science & Technology, Khartoum, Sudan in 2004. He worked for King Khalid University in college of computer science in Saudi Arabia since 2008. He is currently a Ph.D. student in UNIVERSITI TEKNOLOGI MALAYSIA, Johor, Malaysia. His research interests include Computer Vision, Sign Language Recognition, Pattern Recognition, Smart Systems, Machine Learning, Deep Learning and Artificial intelligence.

**Mohd Shafry Mohd Rahim** is a Professor of Image Processing at School of Computing, Faculty of Engineering, University Technology Malaysia, Skudai, Johor, Malaysia. Presently, he has appointed as Deputy Vice Chancellor (Academic & International), University Technology Malaysia. Besides, he has been a member of the Board of Governance (BOG), SPACE College since 2014. He received his Diploma in Computer Science (1997), B.Sc. of Computer Science majoring in Computer Graphics (1999), and MSc. Of Computer Science (2004) from the University Technology Malaysia (UTM), Malaysia and his PhD of Spatial Modelling (2008) from University Putra Malaysia (UPM), Malaysia.

**Mahmoud M. Khattab** is currently serving as a Post-Doctoral Fellow (PDF) and received his Ph.D. in Computer Science (2022) from College of Information and Communication Technology, International Islamic University Malaysia (IIUM), Kuala Lumpur, Malaysia. He earned his M.Sc. (2009) and B.Sc. (2005) degrees in Computer Science from Menofiya University, Egypt. The area of his research interest lies in super-resolution, image processing, pattern recognition, artificial intelligence, and computer vision. He worked in computer science department at King Khalid University (KKU), Abha, Saudi Arabia from 2010 to 2023.

**Akram M. Zeki** is a Professor at Faculty of Information and Communication Technology at International Islamic University Malaysia. He held few administration positions such as Coordinator for Postgraduate Studies at Faculty of Information and Communication Technology, and then Head of Research in the same faculty and then held a position as a Deputy Director of CENTRIS and Acting Director as well. He is a supervisor for more than 30 master and PhD students; he is leading few research grants under the university (International Islamic University Malaysia) or under national grants and international grants. He is an editor of Journal of Science and Technology JST and the International Journal of Islamic applications in Computer Science and Technologies IJASAT. Prof. Akram is a Senior member of IEEE and a Trustee member of International Computing Institute of Quran and Islamic Sciences.

**Safaa S. Matter** received her Ph.D. in Computer Science (2023) from College of Information and Communication Technology, International Islamic University Malaysia (IIUM), Kuala Lumpur, Malaysia. She earned her M.Sc. (2009) and B.Sc. (2005) degrees in Information Technology from Menofiya University, Egypt. The area of her research interest lies in Wireless Computer Networks, Evaluating and improving routing protocols performance using network simulators. She is also working as a lecturer in Computer Science department, at King Khalid University, Saudi Arabia since 2010. She has been Head of Computer Science department since 2013 till 2017.

**Amr Mohmed Soliman** is an Associate Professor of Mental Health at King Khalid University, Abha, Saudi Arabia. He received his Ph.D. (2013), M.Sc. (2009), and B.Sc. (2000) degrees from Ain-Shams University, Cairo, Egypt. His research interests include sign language for the deaf, mental disability, mental health, and Using technologies for people with special needs.

**Abdelmoty M. Ahmed** is currently an Assistant Professor in College of Computer Science, Nahda University, Beni Suef, Egypt. He received his B.Sc., M.Sc., and PhD degrees from Systems and Computers Engineering, Faculty of Engineering, Al- Azhar University, Cairo, Egypt. His research interests include Digital image processing, Artificial intelligent, pattern recognition, Human Computer Interaction, Computer Graphics, machine learning, Deep Learning, E-Learning, Intelligence Systems Engineering, Computer Vision, and IOT systems. He was senior lecturer in computer engineering department at College of Computer Science, King Khalid University, Abha, Saudi Arabia. He is also interested in researching the technical fields that serve the deaf and dumb and also works in the automatic translation of Arabic Sign Language. He is having 20 years of teaching and research experience at various reputed Universities of Egypt and Saudi Arabia. His Ph.D thesis focused on the automatic translation of the Arabic Sign Language. He has published more research articles in reputed SCI and Scopus indexed journals and conferences.