

Research Article

Integrated Mediapipe with a CNN Model for Arabic Sign Language Recognition

Ahmad M. J. AL Moustafa ^{1,2} **Mohd Shafry Mohd Rahim** ^{1,3} **Belgacem Bouallegue** ^{2,4}
Mahmoud M. Khattab ² **Amr Mohmed Soliman** ⁵ **Gamal Tharwat** ⁶
and Abdelmoty M. Ahmed ^{2,6}

¹Faculty of Computing, Universiti Teknologi Malaysia, 81310 Johor Bahru, Johor, Malaysia

²College of Computer Science, King Khalid University, Abha, Saudi Arabia

³Media and Game Innovation Centre of Excellence, Institute of Human Centered Engineering, Universiti Teknologi Malaysia, 81310 Johor Bahru, Johor, Malaysia

⁴Electronics and Micro-Electronics Laboratory (E. µ. E. L), Faculty of Sciences of Monastir, University of Monastir, Monastir, Tunisia

⁵Special Education Department, College of Education, King Khalid University, Abha, Saudi Arabia

⁶Department of Systems and Computers Engineering, Faculty of Engineering, Al-Azhar University, Cairo, Egypt

Correspondence should be addressed to Abdelmoty M. Ahmed; abd2005moty@yahoo.com

Received 19 March 2023; Revised 6 July 2023; Accepted 25 July 2023; Published 19 August 2023

Academic Editor: Raid Al-Nima

Copyright © 2023 Ahmad M. J. AL Moustafa et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Deaf and dumb people struggle with communicating on a day-to-day basis. Current advancements in artificial intelligence (AI) have allowed this communication barrier to be removed. A letter recognition system for Arabic sign language (ArSL) has been developed as a result of this effort. The deep convolutional neural network (CNN) structure is used by the ArSL recognition system in order to process depth data and to improve the ability for hearing-impaired to communicate with others. In the proposed model, letters of the hand-sign alphabet and the Arabic alphabet would be recognized and identified automatically based on user input. The proposed model should be able to identify ArSL with a rate of accuracy of 97.1%. In order to test our approach, we carried out a comparative study and discovered that it is able to differentiate between static indications with a higher level of accuracy than prior studies had achieved using the same dataset.

1. Introduction

The majority of deaf and hearing-impaired people continue to prefer using sign language as their primary mode of communication, as well as a natural mode of communication between them and nondeaf people. It is also the only formal means of communication for dumb and hearing-impaired people to express themselves through body language and facial expressions. It is a well-organized code based on hand gestures [1]. The primary difficulty is that the majority of people are unable to comprehend sign language, and the great majority of deaf persons are dumb in both their native language and their ability to read and write it.

This creates an imbalance between the deaf community and the hearing world, limiting access to services and preventing deaf people from achieving their full potential in education and other areas.

In recent years, tremendous advancement has been achieved in the area of human-computer interaction (HCI), which has led to the emergence of a growth mechanism for human communication. As a result of this development trend, technology has contributed greatly to social applications that assist people with special needs, such as automatic sign language translation. The availability of sign language explanations determines deaf and dumb people's access to various aspects of life and education. In the field of

servicing special needs with modern technological means, Arabic research is severely lacking. In addition, the Arabic computer applications that help people with special needs in education, training, and communication in general have a flaw. There is also a lack of automation in ArSL. In general, there is a scarcity of Arabic research in this field [2].

The development of fully automated sign language interpretation systems is in its infancy. All translation services are currently human-based, which makes them extremely costly due to the personal experience required. As a result, this highlights the need to increase automatic sign language recognition in order to assist and serve the deaf and dumb in understanding their mother tongue. These systems act as an interpreter among persons who are deaf, dumb, and normal people, allowing them to communicate more effectively [3]. To convey information, sign language recognition systems primarily use hand gestures. Hand gesture recognition is a rapidly expanding field with numerous applications in fields such as entertainment, computer games, object grasping, and command descriptions.

The majority of gestures are made with the hand, but they can also be made with the face and the body. A hand gesture is created not only by the shape of the hand but also by the gesture of the hand and where it is positioned in relation to other parts of the body. Communication using hand gestures is the most significant component of sign language since they are employed in every element of human communication. For example, they can be used to accompany speech or to communicate on their own in a setting with a lot of background noise because signers communicate the majority of their information with their hands [4]. So, in order to incorporate all members of the community, regardless of their abilities, we intend to construct a robust automatic sign language translation and recognition system to address these issues. In addition, the proposed automatic translation of ArSL to Arabic text is intended to increase the quality of Arabic text translation, which is determined by pattern recognition theories.

In accordance with the World Health Organization (WHO), and its 2021 report, there are more than 466 million people worldwide who suffer from hearing loss (roughly 5% of the world's population). Those with crippling hearing loss, 32 million of whom are children, and 1.1 billion young adults "between 12 and 35 years old" are at risk of hearing loss due to music noise. According to the report, that number is expected to double over the next 30 years. It is mentioned that the number of deaf people may reach 90 million by 2050, which is a large number, according to the statistics of the "International Federation of the Deaf." 80% of the deaf and dumb live in developing countries, so governments give priority to helping these people to integrate and participate in societies [5, 6].

Even ArSL has regional and linguistic variations that make it difficult to generalize about how it is used. This study focuses on the most widely understood variety of ArSL among Arabs. Figure 1 shows the expectations of an increase in the number of people with disabling hearing loss (DHL) during the next thirty years [5, 6].

In Arab society, in order to understand the deaf and dumb, it is necessary to learn the language and design of deep models for ArSL, teaching this language through these models to more categories of the deaf community and dumb as a family, friend, and neighbor to be integrated into society [1]. The aim of the design of the deep models was to complement the proposed system design that aimed to substitute human interpreters to some extent and also to develop a system to translate video frames or images based on sign language into text based on the Arabic language to enhance their communications.

The motivations of this research are to help deaf and dumb people to understand Arabic text, help interested people to learn the sign language and interact with deaf and dumb people, and translate sign language to Arabic text without a human translator. In response to this problem, a new sign language recognition approach is proposed to improve the quality of Arabic sign translation to Arabic text. In addition, the proposed system is based on the theory of pattern recognition to construct a translation system using computer vision techniques, digital image processing, and machine learning [1].

Due to its central social importance and inherent ambiguity, developing a framework for deaf people to understand and interpret sign language is a difficult task. The following are some of the problems and difficulties associated with ArSL.

In this situation, sign language is important for communicating with hearing-impaired persons. Sign language recognition systems are getting attention. Proposing a new sign language recognition model and considering alternate methods are crucial. Any sign language recognition system may use a sensor-based or image-based approach [7]. Sensor-based systems interpret gestures by connecting many sensors to a glove. Image-based systems identify signs via image processing. This work employed image-based analysis.

Sign language has the most structured movement expressions of any language [8], yet its unique expressive mode makes it most challenging to learn and use. There are two major methods to demonstrate words in sign language, which are complimentary to each other. The first method employs body movements (such as the hands and arms of humans being used for transmitting meanings) [9] and facial emotions (such as eyebrows lifting and mouth shaping) [10–14], whereas the second approach utilizes fingerspelling, which takes into account orientation, hand pose, and trajectory [15–18].

Developing tools to translate sign language into text or voice to help nondeaf and dumb individuals communicate is crucial. Many sign language recognition systems are based on Indian, Japanese, British, Brazilian, Chinese, Australian, and American sign languages [19–30]. However, sign language research has also focused on Asian [31], English, Turkish [32], and German sign languages [33], as well as Arabic, albeit without achieving the desired results in terms of system performance. Unfortunately, no sign language is universal. The alphabet, where each letter has its own symbol, varies by nation and language. Researchers may not have a reliable ArSL database because the Arabic language is

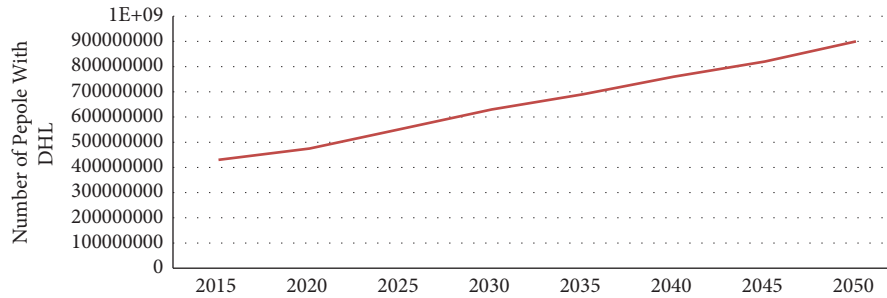


FIGURE 1: Projections of the number of people with DHL.

complicated. Researchers had to create datasets manually, which is cumbersome. Table 1 provides further information on several international sign languages.

Sign language recognition systems automate communication between hearing-impaired people and ordinary people [48], and ArSL recognition detects and classifies. The detection phase preprocesses images, finds interesting areas, and classifies them. Recognition uses each segmented hand sign's characteristics. These traits may distinguish indications. Each study has cost, image preprocessing, and sign classification restrictions [49, 50]. To understand ArSL, a deep convolutional neural network is fed hand motion images from different angles and lighting.

A substantial collection of sign language videos or images tagged with sign language labels is required to construct an ArSL recognition system. Classification-stage machine learning models may be trained and tested on this dataset. Figure 2 depicts two possible methods of data collection for ArSL recognition: (1) a vision-based approach and (2) a sensor-based approach [2].

Using computer vision methods, a vision-based approach to ArSL recognition analyzes and interprets sign language motions from image or video data. Gestural detection, feature extraction, and classification are only a few of the common intermediate phases in this process. To detect gestures, images or videos must be analyzed to identify and follow the hands or other important body components. Methods including skin color segmentation, hand form classification, and optical flow analysis may do this. After the hands and other important portions of the body have been recognized, feature extraction methods are employed to glean information from the video data, such as the hands' shape, motion, and orientation. This may be accomplished through the utilization of techniques such as CNNs, the histogram of oriented gradients (HOG), and local binary patterns (LBPs). Finally, the obtained data are used by machine-learning methods including support vector machines (SVMs), decision trees (DTs), and artificial neural networks (ANNs) to classify the sign language gesture [4].

Using a vision-based technique for ArSL detection is complicated by the fact that sign language motions may seem quite different depending on the signer and the situation. Methods including data augmentation, multimodal fusion, and transfer learning, which may boost the

recognition system's resilience and generalizability, have been investigated by researchers as potential solutions to this problem [2].

Using sensors to detect and record information about sign language motions is at the heart of the sensor-based method to ArSL recognition. There are a number of benefits and drawbacks to using this method instead of or in addition to vision-based methods. In low-light or obstructed environments, for example, sensor data may be more trustworthy than that obtained using vision-based methods. Sensors can record data that the human eye cannot, such as pressure, force, or electromyography (EMG) signals [52].

There are two primary types of systems: the first kind of system [52] is based on gloves, while the second type of system [7] is based on Microsoft Kinect. Gloves [52] may be used in systems that read hand gestures because they include electronic and mechanical components. In spite of the fact that this method may generate satisfactory outcomes [53], wearing a glove that is attached to a variety of sensors in order to collect information may be unpleasant for signers who have hearing or speech impairments. Kinect sensors are put to use in order to identify the indicators that fall into the second group. These sensor devices are first developed by Microsoft for its Xbox game as an alternative to using a controller to play video games [54]. Recently, the application of this technology has expanded to encompass additional recognition technologies, such as the capability to identify sign language. The current trend is predicted to continue for the immediate future.

This research utilizes 28 Arabic sign images to illustrate how preprocessing images may assist in describing motions in feature extraction and how our suggested architecture outperforms previous techniques. Contributions are as follows:

- (i) We need about 7,057 Arabic alphabet sign language images to train our model. In partnership with Resala Association in Egypt, we developed a database of 10,000 red-green-blue (RGB) images and combined them with a 5600-image collection from our previous work [50, 51].
- (ii) Our ArSL comprehension technique was evaluated.
- (iii) Training and comparison of our architecture and others for static Arabic sign letter interpretation were performed.

TABLE 1: Databases of the primary standard for sign languages in several countries.

Sign languages	Description	Reference
American sign language (ASL)	RVL-SLLL database includes ASL words, phrases, gestures, and sentences. 14 singers did it. This library has 2,576 movies of 39 motion primitives. 62 hand forms, and phrases RWTH-BOSTON-400 develops isolated ASL recognition. 843 sentences. Four signers performed	[34] [35]
Australian sign language (Auslan)	Auslan-Sign-bank has 7415 terms and over 1,000 video clips. 100 signers performed The ArSL database's eighty-word lexicon is utilized to generate forty sentences with unconstrained grammar and sentence length and repeated 19 times. Continuous sentences in ArSL are fragmented and completely defined. Many signatories recorded these phrases utilizing a glove-based technique. Labelled and segmented sentences are captured. The vision-based continuous ArSL system employed the same words	[36] [37, 38]
Arabic sign language (ArSL)	Images and videos are in the Signs International ArSL database. The authors updated and expanded the database to test their posture detection and real-time ArSL algorithms. Over 500 static gestures (manual signs) are in the database, including "fingerspelling and hand movements" (nonmanual signs). Facial expressions, body language, and lip reading are important aspects A huge fully labelled ArSL dataset is accessible to academics for free. The ArSL2018 database comprises 54,049 images of thirty-two ArSL signs and alphabets, contributed by 40 volunteers of varying ages. Preprocessing may eliminate noise, center the image, and reduce dimensions and differences in images AASL RGB dataset. AASL contains 7,856 raw and fully tagged RGB images of Arabic sign language alphabets to help construct real-world classification algorithms. AASL is gathered from over 200 individuals in various lighting, backdrop, image orientation, size, and resolution conditions. For a high-quality dataset, field experts supervised, verified, and filtered the images	[39] [40] [41]
British sign language (BSL)	BSL Corpus contains 249 video talks. The conversational dataset contains 6,330 signer gesture annotations	[42]
Brazilian sign language or Língua Brasileira (Libras)	LIBRAS-HC-RGBDS has sixty Libras one-hand setups. Kinect collected the data. DB has 610 five-signer videos	[43]
German sign language or Deutsche Gebärdensprache (DGS)	SIGNUM has 450 DGS motions. This DB has 780 sentences. Twenty-five signers performed	[44, 45]
Indian sign language (ISL)	ISL is accomplished by nine signers, each of whom performed the gesture twenty times in a variety of environments, for a grand total of eighty possible movements	[46]
Turkish sign language (TSL)	BUHMAP has 132 dynamic gesture videos. Eleven signers performed on this DB	[47]

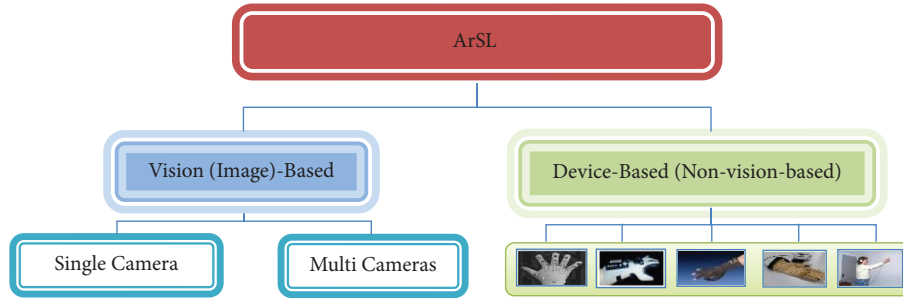


FIGURE 2: Arabic sign language recognition approaches [51].

(iv) The produced datasets were used to create a new algorithm for Arabic alphabet sign identification. The recognition step uses 70% of the dataset for training and 30% for testing. It determined performance evaluation.

(v) The model was tested using the ArSL alphabet alone.

The remaining parts of this work are divided as follows: In Section 2, we give a literature overview of the related studies. Recognition of the hand signs is shown using the method described in Section 3. Section 4 presents the results of the experiments. The authors conclude and provide some final analysis in Section 5.

2. Related Work

This section reviews ArSL recognition methods. ArSL recognition system replaces sign language interpreters. This area pioneered sign language recognition preprocessing, feature extraction, and classification.

In [55], the authors evaluated classification algorithms used in recognition, such as traditional machine learning and deep learning, and discussed previous work on differentiating between static alphabetic and dynamic sign languages for Arabic and non-Arabic sign languages. In [56], the authors developed a fully automated method for recognizing 28 Arabic signs for letters and numbers. There are a total of 7869 images used. Several cycles of training and testing on various combinations of training and testing data are used to perfect the 7-layer model. In conclusion, the authors provided evidence that the proposed model is superior to KNN- and SVM-based approaches. In [50], the idea of using computers to translate between Arabic sign language and spoken language is presented. For the ATASAT system to function, it needs two datasets of Arabic alphabet movements. Arabic sign movements are extracted from images or movies using manual detection based on hand coverage, then statistical classifiers are used, and the results are compared to further refine the classification.

In [51], the authors offered a machine-learning-based alphabet recognition system for ArSL. Each group of 10 students examines 2,800 pictures and 28 different alphabets. There are 2,800 different pictures for all 26 letters (28×100). Features are extracted from a hand shape description, and classification is accomplished via k-nearest neighbors (KNN) and multilayer perceptron (MLP) algorithms.

The precision is 97.548%. In [57], the authors proposed a deep transfer learning-based recognition system for ArSL. Specifically, they used data augmentation to lessen the likelihood of model overfitting and boost the system's overall performance. To accomplish the job of target identification, several network architectures are explored. The dataset for ArSL used in this study is made accessible to the public (ArSL2018).

According to the research of Hisham and Hamouda [58], ArSL may be identified in real time using Microsoft Kinect. The system's recognition skills have been vastly improved using a decision tree, a Bayesian network, and Ada-boosting. In the study [59], the authors provided a novel ArSL recognition approach that makes direct use of microscopic images and is based on an unsupervised deep-learning method, a deep belief network (DBN). It has been used to recognize and sort the Arabic alphabet. Saleh and Issa [60] improved the accuracy of detecting 32 hand gestures in ArSL by using transfer learning and fine-tuning deep CNNs. Aldahri et al. [61] used a CNN model fed with a grayscale image to create a system that can automatically recognize 28 letters in ArSL.

In order to address the challenges that are specific to sign language, the authors of [62] applied ontology to this field of study. They employed simple, permanent signs composed of Arabic letters. During training and testing the deep CNN architecture, both a preexisting and newly collected dataset of ArSL are used. ArSL is difficult to recognize automatically; however, a new framework suggested by Duwairi and Halloush. [63] may help. Popular deep-learning models (such as AlexNet, VGGNet, and Inception Net) are used in this framework to do image processing through transfer learning. As the VGGNet architecture outperformed previously trained models, they proposed using it to automatically recognize Arabic alphabets in sign language. This is done to put their plan into action.

There might be many positive outcomes for society as a whole as a result of the development of this system, some of which are summarized as follows:

- (i) There are not many people who are fluent in ArSL, so creating a system like this would greatly help the hearing-impaired community communicate with individuals who are not native speakers.
- (ii) The community of deaf people be expanded so that those who are deaf are able to speak with hearing

people and take part in the activities of hearing people and simplify the process of communicating with deaf and hard-of-hearing people for those who are able to hear and increase the quality and availability of educational options for the deaf community.

3. Proposed Methodology

This section explains the procedure that we have developed for recognizing the static alphabet used in the ArSL language. The primary objective of this research is to produce deep CNN that is competent in recognizing the ArSL alphabets to a high degree of precision. Figure 3 presents a conceptual overview of the system for your reference. The proposed system has a basic workflow that is made up of two stages: the initial stage is image preprocessing, and the following stage is a proposed model for classification and text generation of ArSL. In the next part of this article, we discuss the various phases.

The proposed approach for recognizing the ArSL alphabet is depicted in Figure 3. In the proposed approach, the ArSL of the alphabet is used to recognize gestures. The first step, which can be covered in the following section, is to gather the dataset for the system and do any necessary image preprocessing [64, 65]. In the process of preparing the dataset, the high-fidelity hand and finger tracking framework known as Mediapipe is utilized. It uses multiple machine-learning methods to deduce the 21 3D landmarks of a hand from a single frame in real time.

3.1. Collect Dataset and Image Preprocessing Phase. When the images of ArSL have been extracted from the camera, they are sent on to the preprocessing phase, which is comprised of a number of primary processes [66, 67]. The images used as data are those of a depth that feature a hand. We used the Mediapipe architecture as a method for detecting the hand in order to accomplish our goal of determining the precise location of the hand [68]. The latter makes use of a variety of methods for detecting and following its prey. One of its methods is called the Mediapipe hand. It is made up of two separate models that collaborate with one another. The first of these is the palm detection model (PDM), which analyzes the whole image to provide a palm or fist bounding box in the correct orientation. This model can recognize both self- and other-occluded hands. The hand landmark model (HLM) is the second model, and it uses the palm detector's clipped image region to generate a high-quality 3D hand key. During the localization step, we scale the images to 240 by 240 pixels and then cut the image by changing it into black images that include just the hand landmarks and a line linking them (Figure 4). Last but not least, we cut our dataset up into three parts: training, validation, and testing for the purpose of training our model. Figure 4 presents the image that is received as input in Figure 4(a), as well as the results of the image preprocessing that included edge detection in Figure 4(b).

3.2. Mediapipe. Machine learning is used in the Mediapipe to recognize the user's hand's key points. One of the models in the Mediapipe machine learning pipeline is a palm identification model that can analyze the whole picture. It takes in the whole image and spits back a bounding box that you can manually rotate, a model for identifying landmarks on the human hand that may provide extremely accurate 3D hand keypoints by using the cropped image area produced by the palm detector [68]. In addition, the hand landmarks from the previous frame are used by the machine-learning pipeline to crop the hand, and palm detection is only employed to localize the hand when the landmark model fails. Mediapipe's hand landmark detector can identify 21 different points on the user's hand, as shown in Figure 5.

The work of Arshedy Alvin et al. [69], in which hand movements are identified using the Mediapipe and k-nearest neighbors technique, should be consulted in order to get an understanding of how Mediapipe can be used for the detection of hand gestures. Figure 6 shows the detection of hand landmarks and the drawing of points.

The machine-learning solutions provided by Mediapipe are open source, compatible with several platforms, and fully configurable. In this research, we have used Mediapipe for the purpose of recording the hand keypoints and TensorFlow for the purpose of both training and detecting the machine-learning algorithm. The Mediapipe application may run on either the central processing unit (CPU) or the graphical processing unit (GPU). The operation of Mediapipe does not need any extra computing power.

3.3. Classification and Text Generation Phase by the Proposed Model Architecture. In this research, we use a model of CNN. This model is made up of numerous layers. The proposed design of CNN is shown in Figure 7, and it comprises an input layer that can handle images that are 64 pixels wide by 64 pixels high by 3 pixels depth. Within the image, it is a representation of the maximum file size that can be utilized by the system. The dimensions of the images which can be uploaded to this system are displayed within the image itself.

Each convolution filter in the feature extraction portion has dimensions of three by three, and the feature extraction section as a whole comprises three convolutional layers (Conv2d_1, Conv2d_2, and Conv2d_3). Hence, there are sixteen filters for Conv2d_1, thirty-two filters for Conv2d_2, and sixty-four filters for Conv2d_3. Take note that after each convolution operation comes an activation layer, which applies the recalculated linear unit function (ReLU). After the activation of each neuron, we will employ a max-pooling layer with a size of 3×3 in order to minimize the size of the pictures without altering the essential qualities that distinguish them. Because of this, the amount of processing power that is needed by the model will be greatly reduced. Next, in order to deactivate 25% of the neurons, we will first utilize a flatten layer and then go on to a dropout layer. By periodically turning off neurons across the network, this particular sort of layer serves to prevent the network from accumulating too much knowledge. In the end, we will use

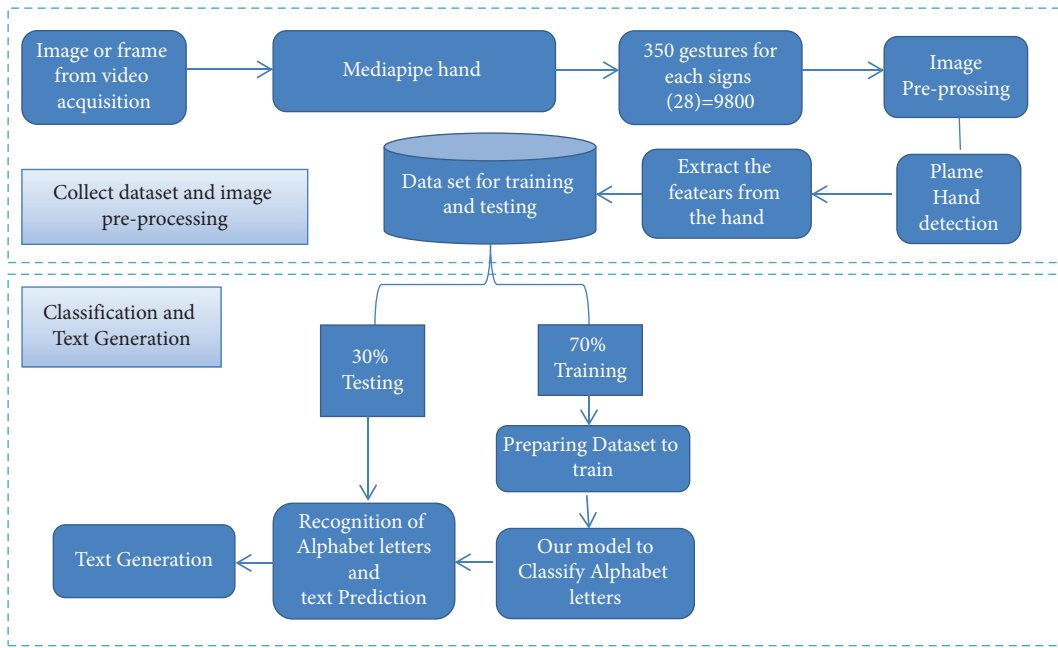


FIGURE 3: Overview of the proposed system.

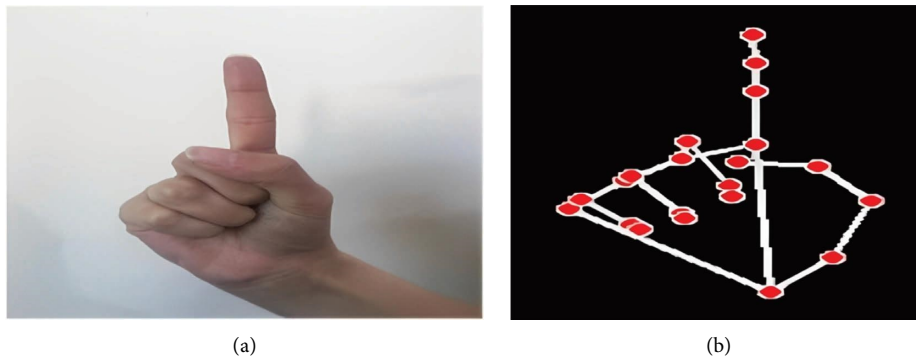


FIGURE 4: An example of the image fed into the system (a) and a sample of the image extracted after processing (b) for the letter Beh “ب.”

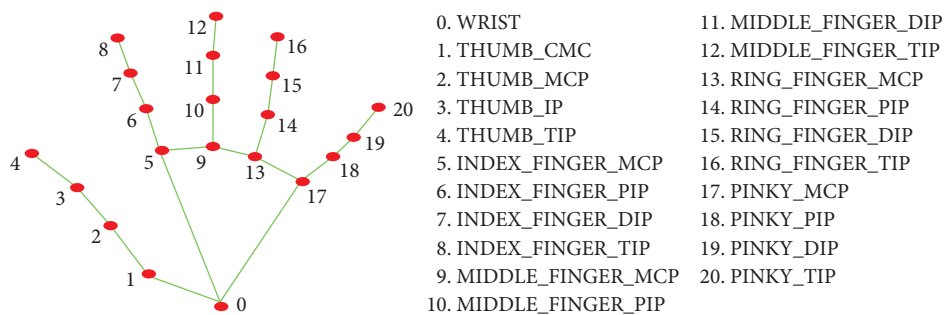


FIGURE 5: Hand landmarks representation used in Mediapipe.

an activation layer constructed using the ReLU function, which will be followed by a dense classification layer constructed with the softmax activation function. The information shown in Table 2 provides a summary of the parameters found in each layer of the proposed network, as well as the overall number of parameters.

4. Experimental Results

4.1. Datasets. To evaluate our system, we make a combination of the ArSL dataset [41] and 7,057 Arabic alphabet sign language images. In partnership with Resala Association in Egypt, we combined them with a 5600-image collection from

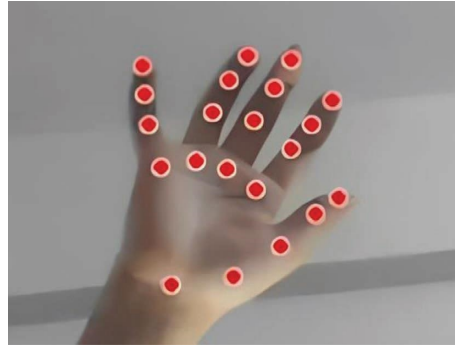


FIGURE 6: Hand landmarks detection.

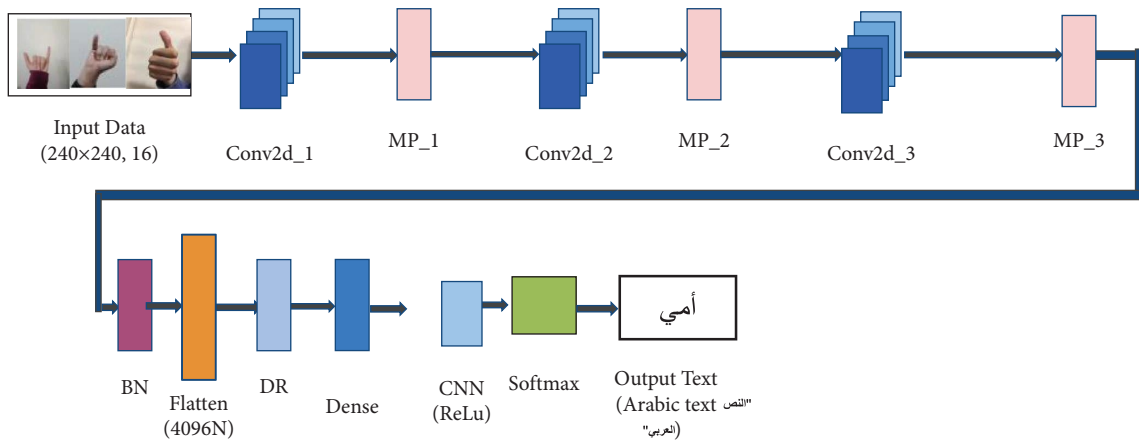


FIGURE 7: Classification and text generation proposed model architecture.

TABLE 2: Summary of the proposed model.

Network layers	Output shape	Parameters
conv2d_1 (Conv2D)	(None, 240, 240, 16)	448
Max_pooling2d (max-pooling2)	(None, 80, 80, 16)	0
conv2d_2 (Conv2D)	(None, 80, 80, 32)	4,640
max_pooling2d_2 (max-pooling2D)	(None, 26, 26, 32)	0
conv2d_3 (Conv2D)	(None, 26, 26, 64)	18,496
max_pooling2d_3 (max-pooling2D)	(None, 8, 8, 64)	0
batch_normalization (BatchNormalization)	(None, 8, 8, 64)	256
flatten (Flatten)	(None, 4,096)	0
dropout (Dropout)	(None, 4,096)	0
dense (Dense)	(None, 256)	1,048,832
Dense_1 (Dense)	(None, 28)	7,196
Total parameters		1,079,868

our previous work [50, 51]. Our dataset is compiled by using professional cameras using data from a variety of signers with varying levels of luminance. These data are then used to train our model.

The dataset for ArSL is primarily made up of two folders: the first folder has 7,057 images, each of which is 240 pixels by 240 pixels in size, and the second folder contains 7,057 text files that correspondingly describe the content of each image in the first folder. The second dataset is constructed in such a manner that includes 28 classes, each of which have around 252 images depicting an Arabic letter. In order to

integrate the two datasets, it is required to make certain adjustments to the structure of the dataset that is in ArSL. So, in order to get rid of the requirement for text files and to unify the structure of both datasets, we group the images of the same letter together and place them in the same folder. Because of this, later on, we are able to merge the two datasets with more ease. It is important to note that the dataset for ArSL had an issue with the images of the letter "Thal and Jeem" not existing in the proper class. This problem needs to be manually corrected by shifting the images back to where they belong in the dataset. Figure 8

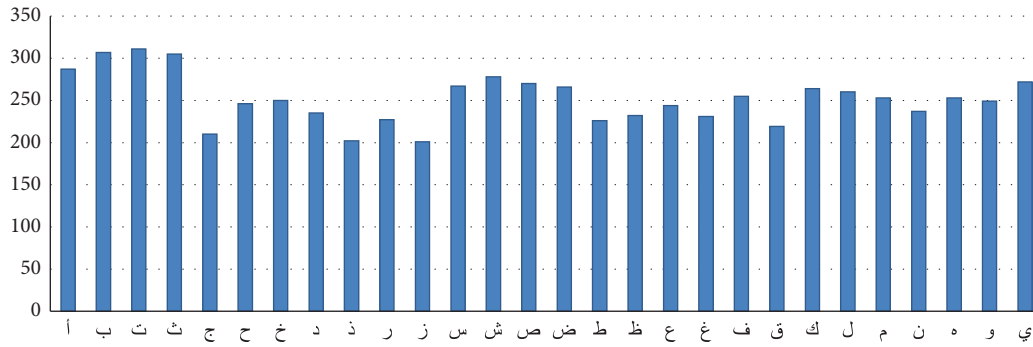


FIGURE 8: The total number of samples used in our model.

TABLE 3: The similarity between some ArSL characters according to Mediapipe’s hand landmark from the authors’ point of view.

























Alphabets			Similarity
			The index finger indicates a single point under the shape of “ب” The index and middle fingers indicate two points above the shape of “ت” The index, middle, and ring fingers indicate three points above “ث” The similarities can occur in Mediapipe’s hand landmark for index, middle, and ring fingers
			Thumb position indicates a point inside the letter “ج” Thumb equal to the level of the other fingers as there is no point to the letter “ح” Thumb height above other fingers indicating a point above the letter “خ” The similarities can occur in Mediapipe’s hand landmark for the thumb finger
			Matches the shape of a letter “د” with a sharp angle between the thumb and the index finger on the left Matches the shape of the alphabet. There is a sharp angle between the thumb and the index finger on the left, and the presence of the middle on the top of the index finger indicates the presence of a point on the letter “ذ” The similarities can occur in Mediapipe’s hand landmark for the thumb, index, and middle fingers
			Corresponds to the alphabet, where the thumb with the index finger curvature and closed fingers represents the letter “ر” Matching the alphabet, with the thumbs curving, the index finger on top, and the rest of the fingers closed, the shape of the letter “ز” 4 The similarities can occur in Mediapipe’s hand landmark for the thumb, index fingers, and wrist
			Matches the alphabet “س, ك, ش” with the numerator and the convergence or spacing of the fingers The similarities can occur in Mediapipe’s hand landmark for all fingers and wrist
			Approximate at the format with the letter “ص” with a note placed thumb Likeness as close when the thumb raised indication of a point above letter “ض” The similarities can occur in Mediapipe’s hand landmark for the thumb
			Close match with the letter “ط” and raise the index finger to the top Close match with thumbs above the rest of the fingers to indicate a point above the letter “ظ” and the index finger up The similarities can occur in Mediapipe’s hand landmark for the thumb, index fingers, and wrist
			Match with alphabet letter “ع” Align your hand with the form of the letter and raise your thumb over your palm to signify the existence of a point above the letter. “غ” Mediapipe’s thumb, index, middle finger, and wrist landmarks may seem familiar

TABLE 3: Continued.

Alphabets			Similarity
			Corresponds with the shape of the alphabet “ف” and the index finger position over the thumb to indicate a point above the letter Match with the shape of alphabet “ق” with the index and middle fingers above the thumb to indicate 2 points above the letter Match with the alphabet letter “ه”
Feh	Qaf	Kaf	The similarities can occur in Mediapipe’s hand landmark for the thumb, index fingers, and wrist
			Match with the alphabet letter “ل” Match with the alphabet letter “ن” with the absence of a point above the letter The similarities can occur in Mediapipe’s hand landmark for the thumb and index fingers
Lam	Noon		

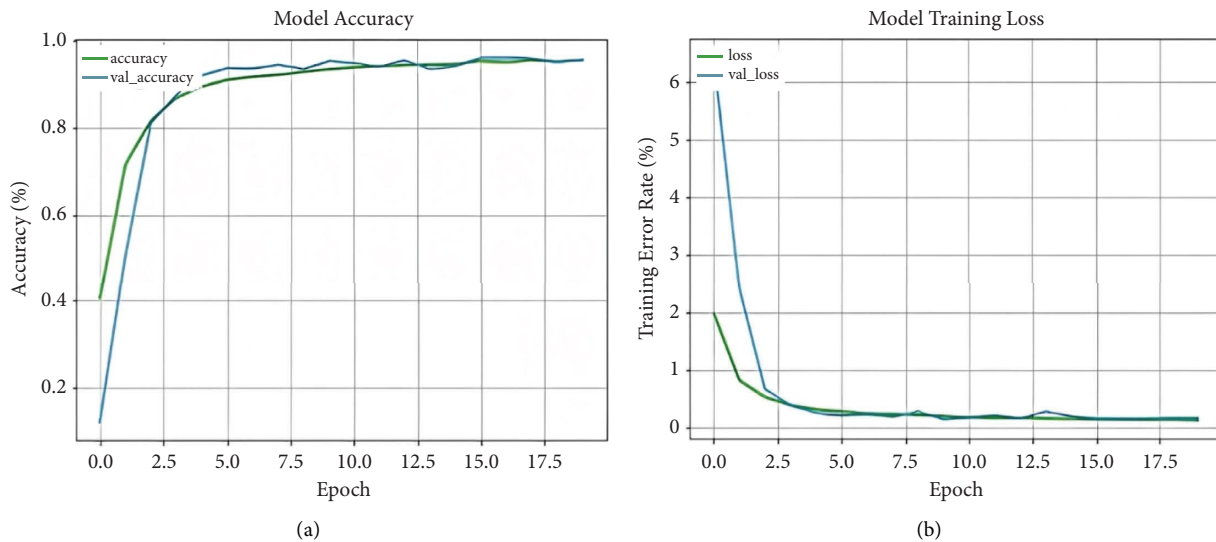


FIGURE 9: Accuracy and validation accuracy (training and testing) progress (a) and (b) training loss vs. validation loss progress.

provides a visual representation of the breakdown of the total number of samples into several classes, each of which corresponds to a letter of the Arabic alphabet.

According to the study of the databases of signs of the Arabic alphabet in sign language, we notice that the same characters have strong similarities in gestures (Beh, Teh, Theh, . . .) which probably influence our model in the feature extraction process, complicating their classification process. Table 3 shows the similarity between some ArSL characters according to Mediapipe’s hand landmark from the authors’ point of view. It also describes the similarity between its representation in sign language and its writing style in Arabic.

4.2. Results and Discussion. The dataset that is utilized for training consisted of more than 7,000 images, which are organized into 28 different classes of ArSL motions using a consistent file structure. This information is presented in the section that came before this one. The dataset is then divided into a training set, a validation set, and a testing set, with the training set receiving 60% of the data, the validation set receiving 20% of the data, and the remaining 20% being reserved for the testing set. The following are the actions that

would then make up our proposed method: We began by performing the preprocessing stage, from which we obtained black and white images consisting of only the hand landmarks connected by a line. After that, we fed the data into our model, where iterations (epochs) are run through our model and accuracy values are provided at the conclusion of each iteration. After that, we ran our model. After all of the training epochs have been completed, the final accuracy is calculated and shown.

Figure 9 displays the accuracy and validation accuracy obtained through training and testing. Figure 9(a) displays the training loss and validation loss obtained through training. Figure 9(b) displays accuracy and validation accuracy obtained through testing.

The CNN’s architecture is trained for 32 batches over 20 epochs. By the 10th epoch, the model achieved a validation accuracy of 94.66%, and at the 20th epoch, it achieved a record-high 97.1% accuracy. Figure 9(a) displays the development of the model throughout training, revealing that the model maintains its lead during each training iteration, while the loss decreases, as shown in Figure 9(b). Based on the outcome, we do not see any evidence of overfitting.

TABLE 4: Results for our approach.

	Precision	Recall	F1 score	Support
Alef	0.98	1.00	0.99	50
Beh	0.95	0.95	0.95	38
Teh	0.98	1.00	0.99	40
Theh	1.00	1.00	1.00	59
Jeem	0.92	1.00	0.96	46
Hah	0.98	0.96	0.97	45
Khah	1.00	0.98	0.99	41
Dal	0.94	0.87	0.91	39
Thal	0.98	0.98	0.98	43
Reh	0.88	0.97	0.93	39
Zain	0.97	0.93	0.95	40
Seen	1.00	1.00	1.00	37
Sheen	0.97	1.00	0.99	38
Sad	1.00	1.00	1.00	28
Dad	0.97	1.00	0.99	37
Tah	0.98	1.00	0.99	43
Zah	1.00	0.97	0.99	38
Ain	0.97	0.95	0.96	38
Ghain	0.98	0.98	0.98	42
Feh	0.92	0.96	0.94	25
Qaf	1.00	0.93	0.96	44
Kaf	0.97	1.00	0.99	33
Lam	0.96	1.00	0.98	25
Meem	0.94	0.97	0.96	34
Noon	0.97	0.90	0.93	39
Heh	0.97	1.00	0.99	37
Waw	0.97	0.93	0.95	40
Yeh	1.00	0.97	0.99	35
Accuracy			0.97	1093
Macro avg	0.97	0.97	0.97	1093
Weighted avg	0.97	0.97	0.97	1093

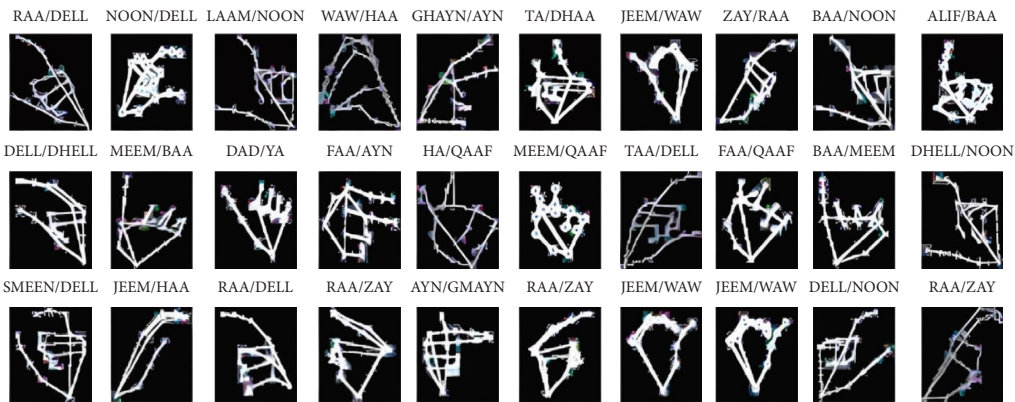


FIGURE 10: An extract of the incorrectly classified images.

Table 4 shows the classification report and indicates that the error rate is quite low and that the vast majority of classes have been assigned to the correct groups. Letters such as “Alef,” “Beh,” “Seen,” and “Noon” have a high accuracy of 95–100%, while letters such as “Jeem,” “Dal,” “Reh,” and “Feh” have a low accuracy of 88%–94%. One major cause of this variation is that the position of the fingers is different for each letter. This demonstrates the significance of gesture representation in feature extraction for obtaining reliable outcomes.

According to the findings, we are able to determine that there are around 32 indicators that are not correctly detected overall. The fact that the same letters show strong similarities of gestures (RAA/DELL, LAAM/NOON, ALIF/BAA...) can be attributed to the high rate of misclassification. This has probably prompted our model to extract comparable features, which has made the classification process more difficult. A portion of the images that are incorrectly labelled is displayed in Figure 10.

TABLE 5: Evaluation of the proposed model in relation to existing models.

Reference (Years)	Dataset size and type	Classifier	Accuracy in %
[70] (2020)	A hundred Arabic signs derived from standard ArSL There are two sets of data for the static alphabet: one set has 700 samples for 28 letters typed with bare hands, while the other set contains 700 samples for 28 characters typed using colored gloves	Euclidian distance	95%
[50] (2016)		C4.5(J48), MLP, K-NN(IBK), and naïve-Bayesian classifiers	80.67, 88.66, 90.7%, and 84.4% for dataset1 and 89.5, 94.11, 97.5, and 96.63 for dataset2
[71] (2018)	Thirty people really take images using their phones. These 30 ArSL alphabets are gestured by volunteers. A total of 900 images are used throughout 30 letters	SVM	63.5%
[72] (2017)	A total of 25 words from the unified ArSL dictionary are included in the dataset	3D CNN-softmax layer	98% accuracy for observed data and 85% average accuracy for new data
[73] (2021)	There are a total of 54,049 images in ArSL2018, representing the 32 ArSL alphabets and signs contributed by 40 signers	CNN	97.29%
[74] (2018)	450 full-colored ArSL videos recorded at 30 frames per second	Euclidean distance classifier	97%
[75] (2011)	Six motions producing 6,000 images of gestures	Max-pooling CNN (MPCNN)	96%
[56] (2019)	28 Arabic letters and digits from 0 to 10 are represented in all of the 7,869 images	CNN	90.02%
Our model	7,057 images for recognizing 28 Arabic letters	CNN with Mediapipe	97.1%

4.3. *Comparative Study.* We compare our findings to those found in the current literature to demonstrate the efficacy of the proposed model. The outcomes of our model and many others in the literature are shown in Table 5.

In contrast to the findings of the aforementioned models, our proposed architecture obtained the best result, and this architecture has been shown to be efficient in sign language recognition, since a validation accuracy of 97.1% is achieved.

5. Conclusion

We use deep learning to recognize ArSL from deaf people's hands in real time to transform signs into alphabets and increase communication between deaf and hearing people. First, we collect data and create our own dataset. Next, we apply a preprocessing step that includes clipping the hand from the image and retrieving its shape. Finally, we propose four distinct CNN models that each have their own unique architecture: three from scientific papers and one from ourselves. We apply them to our datasets to compare them and choose the best model. Our architecture performs best in the comparative research, scoring 97.1%. In the future, we hope to apply new techniques to convert gestures of the hands into written form, employ natural language processing (NLP) to process and display the writings, and extend the dataset's variety with regards to noise, orientation, and so on.

Data Availability

The datasets used and/or analyzed during the current study are available from the corresponding author Abdelmoty M. Ahmed, Email: abd2005moty@yahoo.com on reasonable request.

Conflicts of Interest

The authors declare that there are no conflicts of interest.

Acknowledgments

The authors extend their appreciation to the Deanship of Scientific Research at King Khalid University for funding this work through Large Groups Research.

References

- [1] H. A. AbdElghfar, A. M. Ahmed, A. A. Alani et al., "QSLRS-CNN: qur'anic sign language recognition system based on convolutional neural networks," *The Imaging Science Journal*, vol. 2023, pp. 1–13, 2023.
- [2] A. M. Soliman, M. M. Khattab, and A. M. Ahmed, "Arabic sign language recognition system: using an image-based hand gesture detection method to help deaf and dumb children to engage in education," *Journal of Qena Faculty of Arts*, vol. 32, no. 58, pp. 1–28, 2023.
- [3] A. M. Soliman, M. M. Khattab, and A. M. Ahmed, "Recognize the alphabet of fingerspelling using statistical classifiers to facilitate communication between hearing-impaired persons and others," *Journal of Qena Faculty of Arts*, vol. 31, no. 57, pp. 1–30, 2022.
- [4] H. A. AbdElghfar, A. M. Ahmed, A. A. Alani et al., "A model for qur'anic sign language recognition based on deep learning algorithms," *Journal of Sensors*, vol. 2023, Article ID 9926245, 13 pages, 2023.
- [5] Who, *Addressing the Rising Prevalence of Hearing Loss*, World Health Organization, Geneva, Switzerland, 2018.
- [6] S. Chadha, K. Kamenov, and A. Cieza, "The world report on hearing," *Bulletin of the World Health Organization*, vol. 99, no. 4, pp. 242–242A, 2021.
- [7] K. Assaleh and M. Al-Rousan, "Recognition of Arabic sign language alphabet using polynomial classifiers," *EURASIP Journal on Applied Signal Processing*, vol. 13, Article ID 507614, 10 pages, 2005.
- [8] S. C. Agrawal, A. S. Jalal, and R. K. Tripathi, "A survey on manual and non-manual sign language recognition for isolated and continuous sign," *International Journal of Applied Pattern Recognition*, vol. 3, no. 2, pp. 99–134, 2016.
- [9] B. Demircioglu, G. Bülbül, and H. Köse, "Turkish sign language recognition with leap motion," in *Proceedings of the 2016 24th Signal Processing and Communication Application Conference (SIU)*, IEEE, Zonguldak, Turkey, May 2016.
- [10] E. Gubin Mounq, C. Chuan Wooi, M. Mohd Sufian, C. Kim On, and J. Ahmad Dargham, "Ensemble-based face expression recognition approach for image sentiment analysis," *International Journal of Electrical and Computer Engineering*, vol. 12, no. 3, p. 2588, 2022.
- [11] S. P. Das, A. K. Talukdar, and K. K. Sarma, "Sign language recognition using facial expression," *Procedia Computer Science*, vol. 58, pp. 210–216, 2015.
- [12] E. A. Elliott and A. M. Jacobs, "Facial expressions, emotions, and sign languages," *Frontiers in Psychology*, vol. 4, p. 115, 2013.
- [13] P. Kumar, P. P. Roy, and D. P. Dogra, "Independent Bayesian classifier combination based sign language recognition using facial expression," *Information Sciences*, vol. 428, pp. 30–48, 2018.
- [14] C. Viegas, "Including facial expressions in contextual embeddings for sign language generation," 2022, <https://arxiv.org/abs/2202.05383#:~:text=Including%20Facial%20Expressions%20in%20Contextual%20Embeddings%20for%20Sign%20Language%20Generation,-Carla%20Viegas%2C%20Mert&text=State%2Dof%2Dthe%2Dart,semantic%20functions%20of%20facial%20expressions>.
- [15] D. Waters, R. Campbell, C. M. Capek et al., "Fingerspelling, signed language, text and picture processing in deaf native signers: the role of the mid-fusiform gyrus," *NeuroImage*, vol. 35, no. 3, pp. 1287–1302, 2007.
- [16] C. Lucas, *Turn-taking, Fingerspelling and Contact in Signed Languages*, Gallaudet University Press, Washington, DC, USA, 2002.
- [17] D. A. Morere and R. Roberts, "Fingerspelling," *Assessing Literacy in Deaf Individuals: Neurocognitive Measurement and Predictors*, APA, Washington, DC, USA, pp. 179–189, 2012.
- [18] S. Aly, "Arabic sign language fingerspelling recognition from depth and intensity images," in *Proceedings of the 2016 12th International Computer Engineering Conference (ICENCO)*, IEEE, Cairo, Egypt, December 2016.
- [19] Y. I. Rokade and P. M. Jadav, "Indian sign language recognition system," *International Journal of Engineering and Technology*, vol. 9, no. 3, pp. 189–196, 2017.
- [20] T. Nanaware, "Fingerspelling-Indian sign language training tool," in *Proceedings of the 2018 IEEE 18th International Conference on Advanced Learning Technologies (ICALT)*, IEEE, Mumbai, India, July 2018.

- [21] H. Hosoe, S. Sako, and B. Kwolek, "Recognition of JSL finger spelling using convolutional neural networks," in *Proceedings of the 2017 Fifteenth IAPR International Conference on Machine Vision Applications (MVA)*, IEEE, Nagoya, Japan, May 2017.
- [22] S. Liwicki and M. Everingham, "Automatic recognition of fingerspelled words in british sign language," in *Proceedings of the 2009 IEEE computer society conference on computer vision and pattern recognition workshops*, IEEE, Miami, FL, USA, June 2009.
- [23] K. Prajwal, "Weakly-supervised fingerspelling recognition in British sign language videos," 2022, <https://arxiv.org/pdf/2211.08954.pdf>.
- [24] B. Nicodemus, L. Swabey, L. Leeson, J. Napier, G. Petitta, and M. M. Taylor, "A cross-linguistic analysis of fingerspelling production by sign language interpreters," *Sign Language Studies*, vol. 17, no. 2, pp. 143–171, 2017.
- [25] D. F. Lima, "Using convolutional neural networks for fingerspelling sign recognition in brazilian sign language," in *Proceedings of the 25th Brazillian Symposium on Multimedia and the Web*, Rio de Janeiro Brazil, October 2019.
- [26] X. Jiang, "Fingerspelling identification for Chinese sign language via AlexNet-based transfer learning and Adam optimizer," *Scientific Programming*, vol. 2020, Article ID 3291426, 13 pages, 2020.
- [27] G. Major, J. Napier, L. Ferrara, and T. Johnston, "Exploring lexical gaps in Australian Sign Language for the purposes of health communication," *Communication and Medicine*, vol. 9, no. 1, pp. 37–47, 2013.
- [28] A. Schembri and T. Johnston, "Sociolinguistic variation in the use of fingerspelling in Australian Sign Language: a pilot study," *Sign Language Studies*, vol. 7, no. 3, pp. 319–347, 2007.
- [29] B. Shi, "Fingerspelling detection in american sign language," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Nashville, TN, USA, June 2021.
- [30] B. Shi, "American sign language fingerspelling recognition in the wild," in *Proceedings of the 2018 IEEE Spoken Language Technology Workshop (SLT)*, IEEE, Athens, Greece, December 2018.
- [31] S. Fischer and Q. Gong, "Marked hand configurations in Asian sign languages," *Formational units in sign languages*, vol. 3, p. 19, 2011.
- [32] S. S. Taşçi, "Phonological and morphological aspects of lexicalized fingerspelling in Turkish Sign Language (TİD)," *Sosyal Bilimler Enstitüsü*, 2012.
- [33] S. Ebling, "Evaluation of animated Swiss German sign language fingerspelling sequences and signs," in *Proceedings of the Universal Access in Human-Computer Interaction. Designing Novel Interactions: 11th International Conference, UAHCI 2017, Held as Part of HCI International 2017*, Springer, Vancouver, BC, Canada, July 2017.
- [34] A. M. Martínez, "Purdue RVL-SLLL ASL database for automatic recognition of American Sign Language," in *Proceedings. Fourth IEEE International Conference on Multimodal Interfaces*, IEEE, Pittsburgh, PA, USA, October 2002.
- [35] P. Dreuw, "Benchmark databases for video-based automatic sign language recognition," in *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*, LREC, Marrakech, Morocco, May 2008.
- [36] T. Johnston, "Auslan SignBank: the auslan lexical database," *Sign Language and Linguistics*, vol. 4.
- [37] K. Assaleh, T. Shanableh, and M. Zourob, "Low complexity classification system for glove-based Arabic sign language recognition," in *Proceedings of the Neural Information Processing: 19th International Conference*, Springer, Doha, Qatar, November 2012.
- [38] K. Assaleh, "Continuous Arabic sign language recognition in user dependent mode," *Journal of Intelligent Learning Systems and Applications*, vol. 2, 2010.
- [39] S. M. Shohieb, H. K. Elminir, and A. M. Riad, "Signsworld atlas; a benchmark Arabic sign language database," *Journal of King Saud University-Computer and Information Sciences*, vol. 27, no. 1, pp. 68–76, 2015.
- [40] G. Latif, N. Mohammad, J. Alghazo, R. AlKhalaf, and R. AlKhalaf, "ArASL: Arabic alphabets sign language dataset," *Data in Brief*, vol. 23, Article ID 103777, 2019.
- [41] M. Al-Barham, "RGB Arabic alphabets sign language dataset," 2023.
- [42] A. Schembri, "Building the British sign language corpus," *Language documentation and reseervation*, vol. 7, 2013.
- [43] A. J. Porfirio, "Libras sign language hand configuration recognition based on 3D meshes," in *Proceedings of the 2013 IEEE International Conference on Systems, Man, and Cybernetics*, IEEE, Manchester, UK, October 2013.
- [44] U. Agris and K. F. Kraiss, "Signum database: video corpus for signer-independent continuous sign language recognition," in *Proceedings of the LREC2010 4th Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies*, Valletta, Malta, May 2010.
- [45] U. Von Agris and K.-F. Kraiss, "Towards a video corpus for signer-independent continuous sign language recognition," *Gesture in Human-Computer Interaction and Simulation*, vol. 11, p. 2, 2007.
- [46] P. Kishore, "Video audio interface for recognizing gestures of indian sign," *International Journal of Image Processing*, vol. 5, no. 4, p. 479, 2011.
- [47] M. Koestinger, "Annotated facial landmarks in the wild: a large-scale, real-world database for facial landmark localization," in *Proceedings of the 2011 IEEE international conference on computer vision workshops (ICCV workshops)*, IEEE, Barcelona, Spain, November 2011.
- [48] M. Mohandes, J. Liu, and M. Deriche, "A survey of image-based Arabic sign language recognition," in *Proceedings of the 2014 IEEE 11th International Multi-Conference on Systems, Signals & Devices (SSD14)*, IEEE, Barcelona, Spain, February 2014.
- [49] R. Anderson, R. Anderson, F. Wiryana, M. C. Ariesta, and G. P. Kusuma, "Sign language recognition application systems for deaf-mute people: a review based on input-process-output," *Procedia Computer Science*, vol. 116, pp. 441–448, 2017.
- [50] A. M. Ahmed, "Automatic translation of Arabic sign to Arabic text (ATASAT) system," *Journal of Computer Science and Information Technology*, vol. 6, pp. 109–122, 2016.
- [51] G. Tharwat, A. M. Ahmed, and B. Bouallegue, "Arabic sign language recognition system for alphabets using machine learning techniques," *Journal of Electrical and Computer Engineering*, vol. 2021, Article ID 2995851, 17 pages, 2021.
- [52] N. Tubaiz, T. Shanableh, and K. Assaleh, "Glove-based continuous Arabic sign language recognition in user-dependent mode," *IEEE Transactions on Human-Machine Systems*, vol. 45, no. 4, pp. 526–533, 2015.
- [53] A. Tharwat, "Sift-based Arabic sign language recognition system," in *Afro-european Conference for Industrial Advancement*, Springer, Berlin, Germany, 2015.

- [54] J. Han, Ling Shao, Dong Xu, and J. Shotton, "Enhanced computer vision with microsoft kinect sensor: a review," *IEEE Transactions on Cybernetics*, vol. 43, no. 5, pp. 1318–1334, 2013.
- [55] M. H. Ismail, S. A. Dawwd, and F. H. Ali, "A review on Arabic sign language recognition," *Journal of Advances in Computer and Electronics Engineering*, vol. 6, no. 12, pp. 1–12, 2021.
- [56] S. Hayani, "Arab sign language recognition with convolutional neural networks," in *Proceedings of the 2019 International Conference of Computer Science and Renewable Energies (ICCSRE)*, IEEE, Agadir, Morocco, July 2019.
- [57] A. Shahin and S. Almotairi, "Automated Arabic sign language recognition system based on deep transfer learning," *IJCSNS Int. J. Comput. Sci. Netw. Secur.*, vol. 19, no. 10, pp. 144–152, 2019.
- [58] B. Hisham and A. Hamouda, "Supervised learning classifiers for Arabic gestures recognition using Kinect V2," *SN Applied Sciences*, vol. 1, no. 7, pp. 768–821, 2019.
- [59] A. Hasasneh, "Arabic sign language characters recognition based on A deep learning approach and a simple linear classifier," *Jordanian Journal of Computers and Information Technology*, vol. 6, no. 3, 2020.
- [60] Y. Saleh and G. Issa, *Arabic Sign Language Recognition through Deep Neural Networks fine-tuning*, University of Petra, Amman, Jordan, 2020.
- [61] E. Aldhahri, R. Aljuhani, A. Alfaidi et al., "Arabic sign language recognition using convolutional neural network and MobileNet," *Arabian Journal for Science and Engineering*, vol. 48, no. 2, pp. 2147–2154, 2022.
- [62] E. K. Elsayed and D. R. Fathy, "Sign language semantic translation system using ontology and deep learning," *International Journal of Advanced Computer Science and Applications*, vol. 11, no. 1, 2020.
- [63] R. M. Duwairi and Z. A. Halloush, "Automatic recognition of Arabic alphabets sign language using deep learning," *International Journal of Electrical and Computer Engineering*, vol. 12, no. 3, p. 2996, 2022.
- [64] M. M. Khattab, "Multi-frame super-resolution: a survey," in *Proceedings of the 2018 IEEE International Conference on Computational Intelligence and Computing Research (ICIC)*, IEEE, Madurai, India, December 2018.
- [65] M. M. Khattab, A. M. Zeki, A. A. Alwan, and A. S. Badawy, "Regularization-based multi-frame super-resolution: a systematic review," *Journal of King Saud University-Computer and Information Sciences*, vol. 32, no. 7, pp. 755–762, 2020.
- [66] M. M. Khattab, A. M. Zeki, A. A. Alwan, B. Bouallegue, S. S. Matter, and A. M. Ahmed, "Regularized multiframe Super-Resolution image reconstruction using linear and nonlinear filters," *Journal of Electrical and Computer Engineering*, vol. 2021, Article ID 8309910, pp. 1–16, 2021.
- [67] M. M. Khattab, A. M. Zeki, A. Alwan, B. Bouallegue, S. Matter, and A. Ahmed, "A hybrid regularization-based multi-frame super-resolution using Bayesian framework," *Computer Systems Science and Engineering*, vol. 44, no. 1, pp. 35–54, 2023.
- [68] F. Zhang, "Mediapipe hands: on-device real-time hand tracking," 2020, <https://arxiv.org/abs/2006.10214>.
- [69] A. Alvin, N. H. Shabrina, A. Ryo, and E. Christian, "Hand gesture detection for sign language using neural network with Mediapipe," *Ultima Computing: Jurnal Sistem Komputer*, vol. 13, no. 2, pp. 57–62, 2021.
- [70] A. M. Ahmed, R. Abo Alez, G. Tharwat, M. Taha, B. Belgacem, and A. M. J. Al Moustafa, "Arabic sign language intelligent translator," *The Imaging Science Journal*, vol. 68, no. 1, pp. 11–23, 2020.
- [71] R. Alzohairi, R. Alghonaim, W. Alshehri, and S. Aloqeely, "Image based Arabic sign language recognition system," *International Journal of Advanced Computer Science and Applications*, vol. 9, no. 3, 2018.
- [72] M. ElBadawy, "Arabic sign language recognition with 3d convolutional neural networks," in *Proceedings of the 2017 Eighth international conference on intelligent computing and information systems (ICICIS)*, IEEE, Cairo, Egypt, December 2017.
- [73] A. A. Alani and G. Cosma, "ArSL-CNN: a convolutional neural network for Arabic sign language gesture recognition," *Indonesian journal of electrical engineering and computer science*, vol. 22, 2021.
- [74] N. B. Ibrahim, M. M. Selim, and H. H. Zayed, "An automatic Arabic sign language recognition system (ArSLRS)," *Journal of King Saud University-Computer and Information Sciences*, vol. 30, no. 4, pp. 470–477, 2018.
- [75] J. Nagi, "Max-pooling convolutional neural networks for vision-based hand gesture recognition," in *Proceedings of the 2011 IEEE international conference on signal and image processing applications (ICSIPA)*, IEEE, Kuala Lumpur, Malaysia, November 2011.