www.ijaceeonline.com ISSN: 2456 - 3935

IJACEE

International Journal of Advances in Computer and Electronics Engineering
Volume 6, Issue 12, December 2021, pp. 1 – 12

A Review on Arabic Sign Language Recognition

Mohammad H. Ismail
Research Scholar, Department of Computer Engineering, University of Mosul, Iraq
Email: mohammadhaqqi@gmail.com

Shefa A. Dawwd
Professor, Department of Computer Engineering, University of Mosul, Iraq
Email: shefa.dawwd@uomosul.edu.iq

Fakhrulddin H. Ali
Assistant Professor, Department of Computer Engineering, University of Mosul, Iraq
Email: fhazaa@uomosul.edu.iq

Abstract: *Automatic Sign Language Recognition is a comprehensive research area in the field of human-computer interaction to replace sign language interpreters. With the development of image processing and artificial intelligence techniques, there are many technologies that have been developed, including pre-processing for sign language recognition, feature extraction, and classification. In this research, previous studies were identified that include distinguishing isolated static alphabetic and dynamic sign language for Arabic and non-Arabic sign language and surveying classification methods adopted in recognition, whether they depend on deep learning methods or traditional machine learning methods.*

Keyword: *ArSL, machine learning; Deep learning; Sign language Recognition*

1. INTRODUCTION

Sign Language (SL) is a natural and effective means for deaf and hard of hearing people to communicate with their peers. To communicate with them, you'll need to learn sign language or another kind of communication. With the fast growth of multimedia communication technology, academics have been more interested in sign language in order to promote social communication among its users. Vision-based systems employ depth cameras to record color or depth images/videos in a more realistic setting, which is one of the primary developments in sign language recognition. Several machine learning algorithms have been created to analyze and categorize video data. Deep Learning has been a dominant feature of recent years' state-of-the-art solutions for Computer Vision, after the success of Machine Learning in general and Convolutional Neural Networks (CNNs) in particular. CNNs are giving the best and most accurate results when solving real-world problems. One of its applications is image classification, which is the process of capturing an image as an input and producing the image's class. A critically important good prediction can be obtained through CNNs role in reducing images to a form that is easy to process without losing features**.** Many researchers have used different methods to identify sign language non Arabic or Arabic, and some of them will be presented.

2. ALPHABET SIGN RECOGNITION

2.1 Arabic Sign Language

In [1], authors developed an appearance-based feature-based strategy for solving the sign language recognition challenge. Skin color detection was done on the hands and head using the YCbCr color space. The texture and structure of sign language visuals are described using Local Binary Patterns (LBP). Principal Component Analysis is used to further decrease the feature vector generated by the LBP operator (PCA). Hidden Markov Models are used to classify the characteristics based on their appearance (HMM). The suggested method's performance is evaluated using the Arabic Sign Language (ArSL) database. For signer dependent recognition, a 99.97% recognition rate was attained using LBP and PCA features. In [2], authors offer an automated translation model for the combination of using facial expressions and manual alphabet motions in the Arabic sign language. The placement of the user's mouth, nose, and eyes has a role in facial expressions. It works using images of the signer's hands without the need for gloves or visible

Cite this paper:

Mohammad H. Ismail, Shefa A. Dawwd, Fakhrulddin H. Ali, "A Review on Arabic Sign Language Recognition", International Journal of Advances in Computer and Electronics Engineering, Vol. 6, No. 12, pp. 1-12, December 2021

markings. The user may interact with the surroundings in a natural manner of two components. The first component of the model is concerned with signs. It is included in three phases: preprocessing, skin detection (which transforms RGB images to YCbCr images), and feature extraction (dependent on the Centroid). Face detection and tracking are the stages of the model's second component, which deals with facial expressions. In the case of facial expressions, the proposed model has a 90% accuracy utilizing minimum distance classifier (MDC) and absolute difference authors, and 99% accuracy in the case of signer's hands.

In [3], authors created a method for translating Arabic Sign Language to text automatically. This system proposes a hand detection approach for detecting and extracting Arabic Sign hand gestures from images or videos. This system works by combining two datasets of image characteristics for Arabic sign language gestures alphabets from two sources: an Arabic Sign Language dictionary and gestures from various signers and gesture detection algorithms to enable the user to communicate with the outside world. The suggested system is divided into five stages: image and video capture, segmentation and hand edge detection for each image and video, construction of hand signs, classification, and text transformation and interpretation. They employed a collection of acceptable features in step hand sign construction and classification, comparing the outcomes of several classification algorithms such as KNN, MLP, C4.5, Voting Feature Intervals VFI, and Sequential Minimal Optimization SMO. Using two datasets, the experimental findings suggest that the MLP classifier yields greater accuracy outcomes (dataset-1 and dataset-2). In [4], authors proposed a deep convolutional neural network-based offline recognition system for Arabic sign numerals and letters. They utilized an actual dataset of 5839 for 28-character images and 2030 numeric images (from 0 to 10). This system is built on the LeNet-5 inspired CNN architecture. It is made up of seven layers that are stacked on top of each other. The first four layers extract deep features from images, followed by the classification layers. The class is given by the final layer softmax, which is made up of 39 neurons. The suggested system has a 90.02% recognition rate. The results of their system's examination reveal that it is effective at identifying both numbers and letters, and that it outperforms other current systems based on the KNN and SVM.

A deep learning-based method for ArSL recognition was suggested in [5]. They used InceptionV3, ResNet18/50/ 101, MobileNet, DenseNet, GoogleNet, GGNet16/19, SqueezeNet, and AlexNet to do transfer learning based on fine-tuning of existing pre-trained networks. Deep features are selected by processing input images with various layers. Finally, the SoftMax function is used to divide target classes into categories and compute a normalized probability score for each. They utilized a public dataset for Arabic sign language (ArSL2018), which contains 54,049 images divided into 32 groups. The proposed method based on residual network ResNet101 had the highest accuracy, with 99.52% accuracy. In [6], authors used a CNN model by taking grayscale images as input to a system that automatically recognizes 28 letters for Arabic Sign Language recognition. They used an image dataset named ArSL2018. Their model included 300 training epochs, two dropout 10% layers, the activation function ReLU, and one extra dense 128 layer before the dense 32 layer. They achieved 92.9% of recognition accuracy on 10810 tested images, around 20% of the whole dataset.

In [7], authors trained and tested Deep CNN architecture on an Arabic sign language dataset. The CNN model is made up of 5 convolutional layers, each of which is followed by a max-pool layer with a 2 x 2 stride filter. In all convolution layers, batch normalization and dropout layers are used. Following the convolutional layers is the fully-connected (FC) layer. It contains 1024 channels and a soft-max layer enabling classification over a large number of classes. Standard RGB images with a size of 64 by 64 pixels are used as input. The training rate of the CNN architecture was measured using (42960 images) nearly 80% of (ArSL2018) sign images of Arabic gestures to assess CNN model performance. Their experimental results show that the training set's classification accuracy was 98.6 %, while the testing set was 94.31 %, according to the collected dataset. In [8], authors suggested a system that recognizes the Arabic alphabet's signs in real-time. A database of more than 50000 images for 32 standard Arabic signs and alphabets was used to train and test the Deep CNN architectures. Since avoiding the additional processing required for three color channels in RGB images, all sign images are converted to grayscale 64x64 dimensions before being sent to the network. Several trials are carried out to determine the highest recognition rates by changing CNN architectural design parameters. Three convolutional layers, three pooling layers, and a fully connected layer made up the proposed Deep CNN architecture. The accuracy of the experimental results is 97.6%.

The accuracy of recognizing 32 hand gestures from the Arabic sign language is improved using transfer learning and fine-tuning deep convolutional neural networks (VGG16, ResNet152) in [9]. The implementation of the presented model was accomplished by reducing the size of the training ArSL dataset while increasing accuracy. The networks were fed by images of various Arabic Sign Language data. The dataset was subjected to random under-sampling to decrease the imbalance caused by irregularity in class sizes, resulting in a reduction in image size from 54,049 to 25,600. The resultant model could achieve 99.4%

validation accuracy for the VGG 16 and 99.6% for the Resnet152. CNN used the proposed system by [10] to recognize Arabic hand sign-based letters and translate them into Arabic speech. The suggested approach classifies the images into 31 different classes. Each one has 125 RGB images with a size of 128 by 128 pixels in the data. The initial part of an Arabic Sign Language Recognition system employing CNN was with two convolution layers. After that, each convolution layer is followed by two maximum pooling layers. The next part consists of a few FC layers for classification. The suggested system has a 90% accuracy rate in recognizing Arabic hand gesture-based letters, indicating that they are reliable. Detecting the Arabic hand gesture-based letters is fed into a "text into speech" engine, producing Arabic language audio.

2.2 Non-Arabic Sign Language

Real-time BdSL detection from images was shown by [11]. Their solution uses a convolutional neural network object detection algorithm to identify and classify signs in an image. They built a dataset called BdSLImset and used a method called Faster Region-based Convolutional Networks. There are ten categories of label sign letters in their BdSLImset dataset. Each gesture was captured approximately 100 times. A 98.2% accuracy rate and a recognition time of 90.03 milliseconds were achieved. In [12], authors offer an RGB and RGB-D static gesture detection system based on a fine-tuned VGG19 model. They re-established a fresh FC layer of 24 classes and eliminated the topmost layer of the pre-trained vgg19 model. They develop an approach for fine-tuning the model by freezing the top 16 layers and training the remainder of the model. They developed two VGG19 models, VGG19-v1 and VGG19-v2, using the technique outlined above. VGG19-v1 was trained only with RGB images, whereas VGG19-v2 was trained exclusively using Depth images. To improve the accuracy of the neural network, they apply a feature concatenate layer of RGB and RGB-D images. The authors conducted a comparison study with various models and found that the suggested model outperforms classic machine learning approaches such as H3DF+SVM, SIFT+PLS, and Gabor+RDF. Furthermore, the model was compared without fine-tuning to other CNN models such as Inception V3, VGG19, CaffeNet, and VGG16. The highest recognition rate among all four models was 88.15%, which is much lower than the suggested model's recognition rate of 94.8% when evaluated on an American Sign Language (ASL) Recognition dataset.

In [13], authors suggested a two-stage fusion network based on CNN architecture for hand gesture recognition. In the first stage of the network, they proposed hand segmentation architecture. When there is a similarity between skin colour and background colour, the hand segmentation model performed well in difficult conditions, according to their data. They designed a two-stream CNN for the network's second level until classification, it can learn to merge feature representations from both the RGB image and its segmentation map. Their system runs at a frame rate of 23 ms per frame. Convolutional Neural Networks were used by [14] to create a robust model that understands 26 alphabets and three special ASL characters. They used a real-time video interface to host their model locally, which delivers real-time predictions and displays the matching English letters on the screen like subtitles. Multiple convolutional layers, pooling layers, ReLU layers, and a fully connected layer make up the model. While reading the data into memory, random shuffling was used to ensure that the training and testing images were not the same every time. On 87,000 ASL images, the CNN model had an accuracy of 96.03%.

The method was created by [15] as a resource for those who are new to Sign Language and incorporates hand sign identification based on skin color modeling. The images were fed into the image categorization model via a CNN. CNN provided the image classification model with images. The approach was used to do real-time computations. A method for recognizing static sign signals and converting them into words was created utilizing a camera. Numbers, fundamental static signs, and the alphabets of the American Sign Language (A–Z) were all included in their research. The system's capacity to construct words by fingerspelling without the need of sensors or other external technologies was one of the study's primary features. The system achieved an overall testing accuracy of 93.67%. The system achieved an accuracy of 90.04% and an average duration of 4.31 seconds for alphabet recognition. The static word recognition had an accuracy of 97.52% and took an average of 2.9 seconds. With an average time of 3.93 seconds, the number recognition accuracy was 93.44%. American Sign Language is used to recognize sign language. The user must be able to record images of hand gestures using a web camera in [16], and the system must predict and show the name of the acquired image. They recognize the hand gesture using the HSV color algorithm and set the background to black. The images are processed using various computer vision methods, including grayscale conversion, dilatation, and masking. And the area of interest is segmented, which in their instance is the hand motion. The binary pixels of the images are the features extracted. CNN is used to train and categorize the images. They have a high level of accuracy in recognizing ten American Sign gesture alphabets. Their model has an accuracy rate of more than 90%.

The system is a re-training VGG system in [17] for real-time ASL fingerspelling recognition with CNNs networks to classify a total of 26 alphabets, as well as two classes for space and delete. The first part of this system represented the gathering of color image data. The second part was a multi-class recognition using CNN. The third part was the writing system, which represented the communication between the computer and the user. The system had a training set accuracy of 98.53 percent, a validation set accuracy of 98.84 percent, and a testing data accuracy of 98.6506 percent that had not been used in training. For the comparison among traditional machine learning models which mentioned above with Deep Learning models. In [18] authors also proposed CNN, and LeNet-5 models. The proposed CNN model's architecture was close to the VGGNet, but it only had six convolutional layers instead of the VGGNet's minimum of 13. Google's Colab online GPU was used to train the model. With an input color image size of 64×64, the various sign language models were trained. The CNN network, which was set similarly to VGGNet, obtained the greatest testing accuracy of 97.62% with the least amount of training time and the highest training accuracy of 99.94%. A dataset of 20,000 sign images of 10 static digits were used in this research to build the BSL digits recognition system in real-time using a webcam. They proposed three traditional machine learning models Logistic regression, KNN, and SVM used to assess the performance of the BSL digit recognition system. The models obtained the testing accuracy of 67.4%, 79.0%, and 70.0% for Logistic regression, KNN, and SVM models respectively.

In [19], authors employed deep learning-based convolutional neural networks to recognize static sign language (CNN). They gathered a total of 35,000 images of 100 static signs from a variety of people. They presented a system based on about 50 CNN models, with regard to changes in parameters such as the number of layers and the number of filters. The suggested system was put to the test using several optimizers, and SGD surpassed Adam and RMSProp optimizers in terms of training and validation accuracy. Furthermore, it was shown that the suggested method had the greatest training accuracy of 99.72% on color images and 99.90% on grayscale images. In [20], authors provided a method for grid-based recognition of Indian Sign Language gestures and poses. This technology offers both real-time detection and accuracy. Object stabilization and skin color, as well as Face detection, are used. Tracking and hand detection are two further applications of segmentation. The research discusses a technique for correctly classifying all 33 Indian Sign Language hand poses. One hand gesture is taken into account. Using approaches such as Skin color extraction, object stabilization, hand extraction, and face removal. A Hidden Markov Model chain is employed for each sign, and a KNN model is used to classify each hand posture. The results show that the system can recognize hand postures and signs in Indian Sign Language with high accuracy and in real time. The proposed system is able to accurately identify a user's hand motions. With 96.4% accuracy, the algorithm identifies all Sign Language hand poses.

To simplify sign language alphabets recognition authors in [21] used MediaPipe's open-source framework and machine learning algorithm to present an approach. They indicated that the MediaPipe could accurately recognize complicated hand gestures and that model can identify 21 hand-knuckle coordinates. The prediction model was simple to use and adaptable to different types of smart devices. Many sign language datasets such as American, Indian, Italian, and Turkish were employed for claimed to be to test the framework's capacity. The suggested model was efficient, accurate, and robust, with an average accuracy of 99%. The usage of this technology was made more comfortable and straightforward by accurate real-time detection utilizing the Support Vector Machine algorithm without wearable sensors.

Using a deep multi-layered CNN by [22] is recommended. Convolution filters with 3x3 kernels, LeakyReLU activation functions, and 2x2 max pooling operations have been used in the suggested technique to recognize and categorize sign languages from hand gesture images. The output layer has made use of the SoftMax activation algorithm. Dynamic (23 categories and 49613 images) and Static (36 categories and 54000 images) hand gestures have been used to test the suggested method on a database of hand gestures. The suggested method's effectiveness in detecting sign language has been shown experimentally. The suggested method achieved a blind test accuracy of 99.89%. There is no implementation of real-time recognition and categorization of hand motions in this study.

In [23], authors introduced a system to recognize hand gestures in real-time. Hand segmentation in the YCbCr color space was used for gesture identification, followed by the suggested CNN model. The initial step in recognizing gestures is to segment the hand area from the depth map. The segmentation procedure assumes the hand is near the camera and starts with depth value thresholding, excluding samples with a distance greater than a pre-defined threshold depending on the application chosen. After capturing the depth images, the hand data were segmented using the YCbCr color space. The YCbCr value of the depth image has been converted. Then each pixel's YCbCr value was compared to the standard values. The segmented hand images were scaled to 256x256 after they were obtained. The suggested CNN model consists of three convolution layers, two max-pooling layers, and two fully connected layers. The output layer has 11 nodes for identifying the dataset's 11 gestures. For 11 gestures from depth images, this proposed technique provided an accuracy of 94.61%. A dataset containing 1320 sample images was used. Table (I) illustrates some algorithms used in static sign recognition for the presented previous work.

TABLE (I): SOME ALGORITHMS USED IN STATIC SIGN RECOGNITION.

Alphabet Sign Recognition					
References	**Model**	**Modality**	**Dataset**	**Accura-cy**	**Signs/ signer**
Ahmed and Aly,2014 [1]	Features extraction using LBP and PCA and HMM classifier	RGB With Skin detection	own ArSL	99.97%	23/3
Fathy et al., 2015 [2]	MDC	RGB With Skin detection	own ArSL	99.00%	30/-
Ahmed et al., 2016 [3]	Contour-based+ KNN	RGB	own ArSL	99.85%	28/-
	Contour-based +MLP			100%	
Hayani et al. 2019 [4]	2D-CNN	RGB	own ArSL	90.02%	39/-
Shahin and Almotairi ,2019 [5]	2D-CNN	Gray	ArSL 2018	99.52%	32/40
Althagafi1 et al., 2020 [6]	2D-CNN	Gray	ArSL 2018	92.90%	32/40
Elsayed and Fathy, 2020 [7]	2D-CNN	Gray	ArSL 2018	94.31%	32/40
Latif et al., 2020 [8]	2D-CNN	Gray	ArSL 2018	97.60%	32/40
Saleh and Issa, 2020 [9]	2D-CNN	Gray	ArSL 2018	99.60%	32/40
Kamruzzaman, 2020 [10]	2D-CNN	RGB	own ArSL	90.00%	31/-
Hoque et al., 2018 [11]	Faster R-2D-CNN	RGB	own BdSL	98.20%	10/10
Khari et al., 2019 [12]	Combined 2D-CNN	RGB+Depth	ASL	94.80%	24/5
Dadashzade et al. ,2019 [13]	2D-CNN	RGB	OUHANDS	88.10 %	23/10
		Binary		93.75 %	
Shobhit Sinha et al., 2019 [14]	2D-CNN	RGB	ASL standard	96.03%	29/-
Tolentino et al., 2019 [15]	2D-CNN	RGB With Skin segmenta-tion	own ASL(Alphabet)	90.04%	26/-
			own ASL(Number)	93.44%	11/-
Hurroo and Walizad, 2020 [16]	2D-CNN	RGB With Skin segmenta-tion	own ASL	90.00%	10/-
Kadhim and Khamees, 2020 [17]	2D-CNN	RGB	ASL standard	98.60%	29/
Wangchuk, et al., 2020 [18]	2D-CNN	RGB	own BSL	97.62%	10/
	Logistic regression			67.40%	
	KNN			78.95%	
	SVM			70.20%	
Wadhawan & Kumar, 2020 [19]	2D-CNN	RGB	own INDIAN SL	99.7%	100/-
		Gray		99.9%	
Patil et al, 2020 [20]	HMM + KNN	RGB With Skin segmenta-tion	own Indian SL	96.40%	33/
Halder and Tayade, 2021 [21]	*Landmarks using MediaPipe* +SVM	RGB	ASL(alphabet)	99.15%	26/-
			Indian(alphabet)	99.29%	26/-
			Italian(alphabet)	98.19%	22/-
Bhadra and Kar, 2021 [22]	2D-CNN	RGB	Indian SL	99.89%	36/
Tasmere, et al., 2021 [23]	2D-CNN	RGB With hand segmenta-tion using YCbCr	Hand Gesture	94.60%	11/4

3. ISOLATED SIGN RECOGNITION

3.1 Arabic Sign Language

In [24], authors compared the Log-Gabor, Hartley, and Fourier frequency domain transformations for feature extraction in Arabic sign language recognition. MLP, SVM, and KNN classifiers may be trained and tested using the feature vectors in the classification step. Using the SVM classifier, the acquired results reveal that the Hartley transform is effective, with 99.0% accuracy.

In [25], authors utilized 3D CNN to identify 25 gestures from an Arabic sign language dictionary and using a features extractor with deep behavior for ArSL Recognition. The recognition system was used input data from depth maps. The system obtained 85% accurate for new data, while for observed data, it obtained 98% accurate on average. In [26], authors demonstrated an automated visual SLRS that translate isolated Arabic language signs to text. Hand segmentation, hand tracking, hand feature extraction, and hand classification are all components of the proposed system. Using a dynamic skin detector that is based on the color of the subject's face, we can segment the hands. With the help of the head, the segmented skin blobs are then utilized to recognize and track hands. The feature vector is constructed using geometric features. Finally, at the classification step, a Euclidean distance classifier is used. The suggested system has a 97% recognition rate on a dataset of 30 isolated words, according to the results of the experiments. In [27], authors described a Real-Time system for ArSL recognition. The system recognizes 30 isolated words from standard ArSL signs using the Dynamic Time Warping algorithm and based on the Kinect sensor that compares signs. The system obtains a signer-independent recognition rate of 95.25% and. a signer-dependent recognition rate of 97.58%.

A novel framework that uses several deep learning architectures, was presented in [28] for independent ArSL recognition. It included hand semantic segmentation, hand shape feature representation, and deep recurrent neural network. Sophisticated pixel-labelled hand images are used to train the DeepLabv3+ semantic segmentation algorithm, which extracts hand areas from each input video frame. Due to reducing hand scale variations, the extracted hand areas are cropped and reduced to a set size. A single layer of Convolutional Self-Organizing Maps is used to extract hand form features. The deep BiLSTM recurrent neural network recognizes the sequence of extracted feature vectors. Three BiLSTM layers, one FC layer, and softmax layers make up the BiLSTM network. The suggested technique is tested on 23 isolated terms from three separate users in Arabic sign language. The suggested system obtains an average accuracy of 89.5%.

In [29], authors suggested a method for hand gesture recognition based on deep convolutional neural networks. They tested the suggested method using the KSU-SSL datasets, which included color videos for 40 classes. 3DCNN was employed in two ways for feature learning by the researchers. the spatiotemporal features of Hand gestures were extracted from an entire video using a single 3DCNN instance trained in the first technique. The features of the Hand gesture were extracted from the beginning, middle, and finish of the video sample using three instances of the 3DCNN structure. Those features were then combined and supplied to the classification algorithm. For feature fusion, MLP, LSTM, and an autoencoder were used. In the signer-dependent mode, the technique achieved recognition rates of 96.69% and 98.12% on the dataset for single and parallel 3DCNN, respectively. The recognition rates for single and parallel 3DCNN in the signer-independent mode were 72.32% and 84.38%, respectively.

CNN Inception model with an attention mechanism for extracting spatial features and Bi-LSTM for temporal feature extraction were suggested by [30]. As a first step, the proposed model's preprocessing phase employed a series of RGB frames with a 30 frame duration and an input frame size of 112x112 to convert an input video. They employ the inception network with two attention modules. Each of these attention modules contains a residual link that aids the network in learning dynamic features with spatial context by using the network. For convolution layers, they employ Exponential Linear Unit activations. And there were two layers of BiLSTM, each with 256 stacked LSTM blocks, in the model utilized. The suggested modelling approach utilized only RGB frames to attain the greatest accuracy on three difficult datasets, namely NVIDIA Gesture, Jester, and ArSL. In the KSU-SSL dataset, there are 40 dynamic sign types. ArSL dataset accuracy was 85.6% while processing speed was 130 frames per second for the suggested model. For Egyptian Sign Language movements, in [31], authors introduced a vision-based system that alternatively translated them into their isolated words. For categorization, they employed Inception v3 CNN and Inception v3 CNN-LSTM architectures. The initial architecture had a 90% accuracy rate. A 72% accuracy rate was achieved using the Inception v3 CNN-LSTM architecture. They concluded that CNNs are excellent at recognizing isolated signs, but CNN-LSTMs are excellent at recognizing continuous words. In [32], authors presented a system that utilized 3D convolutional neural networks and long short-term memory to enhance dynamic sign language recognition accuracy on three dynamic gesture datasets extracted from colour videos, with an average recognition accuracy of 97.4%.

3.2. Non Arabic Sign Language

In [33], authors used a Microsoft Kinect, CNNs, and GPU acceleration to examine a recognition system that uses this technology. Automated feature building is possible using CNNs rather than sophisticated handmade features. The model's architecture comprises two CNNs, one for hand features and one for upper body features. There are three layers of CNNs in each one. After combining the results of the two CNNs, a traditional ANN with a single hidden layer performs classification. The first and second layers use local contrast normalization, and all artificial neurons are rectified linear units. They can accurately identify 20 Italian gestures. An accuracy of 91.7% means that the prediction model may generalize to users and environments not present during training. In [34], authors proposed Motion Fused Frames

MFFs, a data level fusion technique for hand gesture classification that fuses motion information (optical flow frames) into RGB images. Using just optical flow and colour modalities, they assessed the suggested MFFs on numerous current datasets and obtained competitive results. Their findings suggest that combining additional motion information enhances performance in every scenario. The improvement in performance at the first attached optical flow frame is particularly notable. On the Jester and ChaLearn benchmarks, their technique achieves extremely competitive results, as they stated with classification accuracies of 96.28% and 57.4%, respectively.

Continuous dynamic sign language categorization using multi-modal data, including infrared data, contour data, and skeleton data, was presented by [35]. Spatiotemporal motion information learned by their 3D CNN model may be used to identify all dynamic signs. Three modes of recording are used in their model; therefore, each mode has its strategy for handling data. Sign language recognition is more accurate when skeleton data is utilized to monitor upper limb trajectory. They use single-channel infrared data to increase computing efficiency. Infrared data categorization mistakes are compensated by using synchronous contour data. They also employed the late fusion approach to integrating two sub-network classifications' findings into a single result. Contour-hands, infrared-hands, contour-body, and infrared-body, feature maps represent the results. Their deep neural network has 11 layers, and each feature map has 32 stacked frames with a size of 64x64. They mentioned for individuals that the contour module was with an accuracy of 87.6%, while infrared module data was with an accuracy of 88.3%. Contour modalities combined with Infrared images generally perform better than utilizing only a single model, with an accuracy of 89.2%. In [36], authors proposed a vision-based system to translate isolated hand gestures from the Argentinean Sign Language with an accuracy of 95.217%. Two methods for training the model on temporal and spatial features were utilized, and they differed in how inputs were supplied to the RNN to train it on temporal information. In the Prediction Approach, the inception model (CNN) extracted spatial data for each frame, and the RNN was used to extract temporal information. After that, each video was represented by a series of CNN predictions for each of its constituent frames. Then the predictions were sent into the RNN as input. Frames were retrieved from each video matching each action, and background body elements other than hands were deleted to create a grayscale representation of hands that prevented the model from learning colour-specific information. The CNN was used to train the model on the spatial features in the Pool Layer Approach, and the pool layer output was given to the RNN before being turned into a prediction. The pooling layer returns a 2048-dimensional vector that reflects the image's tangled characteristics but no class prediction. The rest of the steps are the same as in the previous method. The only difference between the two algorithms is the RNN input. Because the RNN is given a larger feature vector every frame, the pool layer approach achieves greater accuracy than the prediction approach.

In reference [37], authors isolated sign language recognition systems include two major phases: tracking and representation of the hand. To pre-train CNN hand models, the hand patches are extracted from an annotated dataset during the tracking phase. Using a particle filter, a joint likelihood observation model is created by combining hand motion with pre-trained hand models. The predicted hand position is in line with the most likely joint location of the particle. Segmenting a square hand area around the predicted location serves as the input to the hand representation phase; this uses the predicted hand position as a starting point. By averaging the segmented parts of the hand, a compact hand representation is created. The "Hand Energy Image" (HEI)" is the name given to the hand representation. Isolated ASL recognition improves with the suggested HEI hand representation. They are training the hand model by using the RWTH-BOSTON-50 data set, where they select 15 isolated words. According to the experiment, the CNN approach with HEI has an 89% recognition rate compared to other methods. With the help of the Montalbano dataset in [38], authors suggested an isolated sign recognition solution utilizing a deep neural network. A Siamese of CNNs and an RNN module is included in their network, along with a pair of CNNs. ResNet-50 with an LSTM employing global max pooling, dropout, and batch normalization yielded the best results, with an accuracy of 93.19%, after several trials with two distinct networks for the CNNs, ResNet-50 and VGG-16.

In [39], authors suggested that for Vietnamese sign language VSL recognition, deep learning techniques be used to determine the dependency of each frame in video sequences. Due to getting extra information about hand movement and location, the data augmentation approach is offered. Pre-processing, visual feature extraction, sequence learning, and majority voting are the four basic processes in the Deep VSL recognition system utilized. LSTM models were used to predict the sign language of each image using CNN features derived from the f7 layer of a pre-trained VGG16 model with 4096 feature dimensions. Finally, the final sign for each input video was computed using majority voting. The trials yielded good results, with 88.5% accuracy (conventional SVM) and 95.83% accuracy (deep learning). It shows that deep learning combined with data augmentation techniques may offer additional information about the hand's orientation or movement and hence increase the VSL recognition system's performance. . The Jester 20BN-JESTER dataset was used by [40] to train the 3D-CNN architecture for hand gesture detection using the 3D-CNN model. The architecture requires 18 images with a resolution of 84*84*3. The input data must first pass through four 3D convolutions and three 3D Max pooling operations, with each output after convolution being activated using the 'Relu' function. Then, the 'Fully Connected' and 'Dropout' layers will be used to address the overfitting issue. There are 27 categories of hand gestures. They obtained a 90% accuracy model with three days of training.

In [41], authors proposed a MultiD-CNN technique for recognizing human gestures in RGB-D videos. Convolutional Residual Networks (ResNets) for training exceptionally deep models and Convolutional Long Short-Term Memory Networks (ConvLSTM) for dealing with time-series connections are used in their architecture to learn high-level gesture representations. They built architecture to learn spatiotemporal features from RGB and depth sequences concurrently using 3D ResNets, which are then coupled to a ConvLSTM to capture the temporal correlations between them, and they demonstrated that they efficiently integrated appearance and motion information. The classification accuracy result is 94.54%. Second, they offer a technique for encoding temporal information into a motion representation, which is then used to extract deep features using a two-stream architecture based on 2D-ResNets, which eliminates distractions from background and other variations. The classification accuracy result is 89.83%. Finally, predictions from both sub-networks are combined in a class score fusion layer (CSF) to provide a final gesture label, and the classification accuracy result is 95.87% for NATOPS dataset. In [42], authors suggested analyzing sixty gestures from American Sign Language based on data provided by the "LeapMotion sensor" using several deep learning and machine learning models, including one named "DeepConvLSTM." DeepConvLSTM aids in the integration of recurrent and convolutional layers with Long Short Term Memory cells in DeepConvLSTM. Also included is a kinematic model of the hand, thumb, fingers, and forearms. The neural network generalization is also augmented using the basic data augmentation approach. Convolutional Neural Networks and DeepConvLSTM have the greatest accuracy (89.3% and 91.1%, respectively) when compared to other models. Convolutional layers are mixed with deep learning models to produce a good solution for sign language recognition using depth sensors data, as opposed to multi-layer perceptron or recurrent neural network. Conv3D-based sign recognition and wake-up modules are included by [43] in a new hierarchical design that is both resource and time-aware. The temporal alignment of the RGB and depth modalities was exploited to obtain a high-performance classification. For the smart home, a small Croatian sign language database with 25 various language signs were built with the participation of the deaf community. The data queue holding a series of depth images is subscribed by the wake-up module. A wake-up module conducts gesture detection based on N consecutive frames throughout each operation cycle. Convolutional blocks are followed by two FC layers in this network. Convolution layers were mixed with batch normalization, a rectified linear unit, and max-pooling layers in after each convolution layer. Hand sign recognition utilizing a cascading architecture of SSD, CNN and LSTM from RGB videos has been suggested by [44]. They used videos from five online sign dictionaries to train an SSD model for hand recognition. Video examples of 10 contributors for 100 words in 10 different settings were utilized in their proposed collection of 10,000 Persian word video samples. There are three sorts of spatial characteristics that may be employed for hand sign recognition using a combinational model of CNN and LSTM. Different pre-train models, spatial features, and temporal models for sequence learning were analyzed thoroughly. For feature extraction from static RGB frames, they used the ResNet50 model. An LSTM has been trained to extract temporal information from hand features such as ESHR (interaction between hands while signing: Slope, Orientation) and hand position. They investigated sequence learning in depth, using a variety of pre-train models, spatial characteristics, and temporal-based models. They used the ResNet50 model to extract features from still RGB frames, which had a prediction time of 2.58 seconds and an accuracy of 86.32%.

In [45], authors developed a two-stream network to detect and identify isolated dynamic hand gestures with changing form, size, and colour of the hand. It is a 3D-CNN network in the two-stream architecture of the optical flow motion template (OFMT) that gathers spatial and temporal information from gesture movies. An OFMT image is used as the input for a 2D-CNN model. Templates of compact motion are fed into 2D-CNN. Optically flowing images and motion-energy (MEI) and motion-history images (MHI) are made utilizing a binarized image acquired by subtracting frames from a previously captured image. Where motion has happened in an image sequence is represented by MEI, while how MHI represents an item move. 3D-CNN was used to extract motion patterns for gesture categorization, while 2D-CNN was used to collect Spatio-temporal information directly from RGB gesture films. Consequently, the predictions of 3D-CNN and 2D-CNN networks are combined at the decision level using a simple probability-based ensemble technique to maximize the final output. They obtained an accuracy of 99.20% when using the fusion model, 97.30% when using only 3D-CNN, and 92.60% when using only 2D-CNN. In [46], authors introduced a new large-scale multi-modal Turkish

Sign Language dataset. They have a total of 38,336 isolated sign video samples in their collection, which includes 226 signs performed by 43 various signers. The samples include a wide range of backgrounds captured in both indoor and outdoor environments. Furthermore, throughout the recordings, the signers' spatial placements and postures change. Each sample comprised color picture RGB, depth, and skeleton modalities and was captured with Microsoft Kinect v2. They created to test and to train sets to allow users to analyze the models independently. They employed CNNs to extract features and unidirectional and bidirectional LSTM models to describe temporal information. They used the dataset to train different deep learning-based models and results evaluation. To increase the performance of their models, they included feature pooling modules and temporal attention. Their models performed with up to 95.95% accuracy.

In [47], authors used an RGB-D camera and a three-dimensional convolution neural network 3D-CNN a novel technique to identify fingers and recognize hand gestures in real-time. Videos of hand motions are included in the dataset, which uses fingertip detection in depth videos. Seven different hand gestures may be seen in the videos. There are 5250 videos of each motion done 15 times by 50 individuals. Border-tracing algorithms are used to recover the outlines of the hands from the skeleton-joint information images from a Microsoft Kinect Sensor version 2 and characterize them. Based on a hand-contour coordinate's model, the K-cosine technique is utilized to identify the fingertip position, and the result of fingertip detection is turned into gesture initialization. The 3D convolutional neural network finally recognizes a gesture. Their network included seven convolutional layers, three max-pooling layers, two fully connected layers, and seven hand gestures. The system obtained 92.6% accuracy and then increased the accuracy to 97.12% by using the Ensemble model with 15 different 3D-CNNs. In [48], authors demonstrated a Sign Language Gesture Detection and Recognition implementation that combined a recurrent neural network (RNN) with a Mediapipe hand tracking framework. Multi-Hand tracking and a deep learning model that can identify motions by Hand Landmark Features per frame with RNN training were used to generate training data from the input video. The dataset included motions for the most frequent Vietnamese words. In terms of word recognition, this model gives quite accurate results. The total accuracy was 63.5%. Table (II) illustrates some algorithms used in dynamic sign recognition for the presented previous work.

TABLE(II): SOME ALGORITHMS USED IN DYNAMIC SIGN RECOGNITION

Isolated Sign Recognition					
References	**Model**	**Modality**	**Dataset**	**Accuracy**	**Sign/ signer**
Sidig et al., 2017 [24]	KNN, SVM, MLP	Image	own ArSL	99%	23/3
ElBadawy et al. 2017 [25]	3D-CNN	RGB	own ArSL	98.0%	25
Ibrahim et al., 2018 [26]	Euclidean distance Isolated words	Videos	own ArSL	97%	30 / -
Abdel et al., 2018 [27]	DTW	Kinect sensor skeleton	own ArSL	97.58%	30/10
Aly and Aly, 2020 [28]	CNN+CSOM+BiLSTM	RGB	own ArSL	89.5%	23/3
Al-Hammadi et al., 2020 [29]	3D-CNN	Videos	*KSU-ArSL*	98.1%	40 / 40
		Images		84.4%	
Abdul et al., 2021 [30]	2D-CNN+BiLSTM	RGB	*KSU-ArSL*	85.6%	40/40
			NVIDIA gesture		19/8
Elhagry and Elrayes, 2021 [31]	CNN-LSTM	RGB	own EGYPTIAN SL	72.0%	9/-
Elsayed and Fathy, 2021 [32]	3DCNN + ConvLSTM	RGB	own ArSL	97.4%	11/1
Pigou et al., 2014 [33]	CNN	Gray+ Depth	ITALIAN SL	91.7%	20/27
Köpüklü et al., 2018 [34]	MFF + CNN	RGB + Flow	Jester	96.28%	27/-
			ChaLearn	57.4%	249/21
Liang et al., 2018 [35]	3D-CNN with Fusion approach	Infrared Skelton	CHINESE SL	89.2%	20
Masood et al., 2018 [36]	CNN+RNN	Gray with Colored Gloves	ARGENTINEAN SL	95.21%	46
Lim et al., 2019 [37]	HEI+2D-CNN	Gray (monochrome)	ASL RWTH-BOSTON-50	89.0%	15
Tur and Keles, 2019 [38]	Siamese of	RGB	Montalbano (ITALIAN)	93.19%	20/27

	CNNs+RNN	Depth			
Vo et al., 2019 [39]	SVM	RGB With skin detection	VIETNAMESE SL	88.5%	12/27
	CNN+LSTM			95.83%	
Zhang & Wang,2019 [40]	3D-CNN	RGB	20BN-JESTER	90.0%	27/-

Elboushaki et al,2020 [41]	3D-CNN(colored+depth)
	2D-CNN(motion representation)
	Fusion (3D-CNN + 2D-CNN)
		RGB, Depth	NATOPS	94.54%
				89.83%
				95.87%
					24/20

			Iso. Gesture	66.27% 52.07% 72.53%	249/21
Hernandez et al., 2020 [42]	SVM	Hand joints from LeapMotion sensor	own ASL	80.6%	60/25
	RF			80.1%	
	KNN			67.1%	
	MLP			83.7%	
	ConvNet			87.4%	
	ConvNet (+DA)			89.3%	
	LSTM			84.5%	
	DeepConvLSTM			87.3%	
	DeepConvLSTM +DA			91.1%	
Kraljevi´c et al, 2020 [43]	3D-CNN Multi-modal fusion	RGB	own CROATIAN SL	69.4%	25/40
		Depth		63.9%	
		Fusion		72.2%	
Rastgoo et al., 2020b [44]	SSD, CNN, LSTM	2D, RGB	own PERSIAN SL	86.32%	100/10
Sarma et al.,2020 [45]	2D-CNN	RGB	own gesture	92.6%	10/10
	3D-CNN			97.3%	
	Class fusion			99.2%	
Sincan et al., 2020 [46]	CNN + FPM + BiLSTM + Attention	RGB	TURKISH SL	95.95%	226/43
Tran et al.,2020 [47]	SVM	RGB	own gesture	60.5%	7/50
	2D-CNN			64.28%	
	3D-CNN			92.60%	
	5x 3D-CNN Ensemble			96.42%	
	10x 3D-CNN Ensemble			96.82%	
	15x 3D-CNN Ensemble			97.12%	
Bach et al. 2021 [48]	keypoints selected+LSTM	RGB	own VIETNAMESE SL	63.5%	15/-

4. CONCLUSION

Through a review, and analysis of presented previous researches with regard to detecting and recognizing of sign language and by discussing and analyzing Tables I and II, which include a survey and comparison of the presented researches in this paper, the following is clarified: The deep learning-based classifier performed better than all the various classifiers in terms of recognition accuracy of sign language. There is no use of the multi-model fusion to recognize Arabic sign language. There is no evaluation of the different methods of fusion models to recognize Arabic sign language. In general, the average accuracy rate of 23 searches to recognize sign language by static hand gesture is 94%, and the average accuracy rate for 25 searches to recognize sign language by dynamic hand gesture is 86%, therefore it is necessary to develop a single model or multiple models to increase the performance and accuracy of the static and dynamic Arabic Sign Language recognition.

REFERENCES

[1] Ahmed, A. A., and S. Aly. (2014), Appearance-based arabic sign language recognition using hidden markov models, In *2014 international conference on engineering and technology (ICET)*, pp. 1-6. IEEE.

[2] Fathy, G. D., E. Emary, and H. N. ElMahdy, (2015), Supporting Arabic Sign Language recognition with facial expressions, In *Proc. 7th Int. Conf. Inf. Technol.(ICIT)*, pp. 164-170.

[3] Ahmed, A. M., R. A. Alez, M. Taha, and G. Tharwat, (2016), Automatic translation of Arabic sign to Arabic text

(ATASAT) system, *Journal of Computer Science and Information Technology* 6, pp: 109-122.

[4] Hayani, S. , M. Benaddy, O. El Meslouhi, & M. Kardouchi, (2019), Arab sign language recognition with convolutional neural networks, In *2019 International Conference of Computer Science and Renewable Energies (ICCSRE)*, pp. 1- 4, IEEE.

[5] Shahin, A. I., and S. Almotairi, (2019), Automated Arabic Sign Language Recognition System Based on Deep Transfer Learning, *IJCSNS Int. J. Comput. Sci. Netw. Secur.* 19, no. 10, pp: 144-152.

[6] Althagafi A., G. Althobaiti, T. Alsubait, and T. Alqurashi, (2020), ASLR: Arabic Sign Language Recognition Using Convolutional Neural Networks, *IJCSNS International Journal of Computer Science and Network Security,* vol.20, no.7, pp. 124-12

[7] Elsayed, E. K., and D. R. Fathy, (2020), Sign language semantic translation system using ontology and deep learning, *International Journal of Advanced Computer Science and Applications,* vol. 11, no.1, pp:141-147.

[8] Latif, G., N. Mohammad, R. AlKhalaf, R. AlKhalaf, J. Alghazo, and M. Khan, (2020), An Automatic Arabic Sign Language Recognition System Based on Deep CNN: An Assistive System for the Deaf and Hard of Hearing, *International Journal of Computing and Digital Systems* 9.4, pp. 715-724, doi: 10.12785/ijcds/090418.

[9] Saleh, Y., And G. Issa, (2020), Arabic Sign Language Recognition Through Deep Neural Networks Fine-Tuning, pp. 71-83. doi: 10.3991/ijoe.v16i05.13087.

[10] Kamruzzaman, M.M., (2020), Arabic Sign Language Recognition and Generating Arabic Speech Using Convolutional Neural Network, Wireless *Communications and Mobile Computing*, doi: 10.1155/2020/3685614.

[11] Hoque, O. B., M. I. Jubair, M. S. Islam, A. Akash, and A. S. Paulson, (2018) Real time bangladeshi sign language detection using faster r-cnn, In 2018 international conference on innovation in engineering and technology (ICIET), pp: 1-6.

[12] Khari, M., Garg, A. K., Crespo, R. G., & Verdú, E. (2019), Gesture Recognition of RGB and RGB-D Static Images Using Convolutional Neural Networks, *Int. J. Interact. Multim. Artif. Intell.*, *5*(7), pp:22-27.

[13] Dadashzadeh, A., A. T. Targhi, M. T., and M. Mirmehdi, (2019), HGR-Net: a fusion network for hand gesture segmentation and recognition, *IET Computer Vision* 13, no. 8, pp: 700-707.

[14] Sinha, S., S. Singh, S. Rawat, and A. Chopra, (2019), Real time prediction of american sign language using convolutional neural networks, In *International Conference on Advances in Computing and Data Sciences*, pp. 22-31, Springer, Singapore. doi: 10.1007/978-981-13-9939-83.

[15] Tolentino, L. K. S., R. O. S. Juan, A. C. Thio-ac, M. A. B. Pamahoy, J. R. R. Forteza, and X. J. O. Garcia, (2019), Static sign language recognition using deep learning, *International Journal of Machine Learning and Computing* 9, no. 6, pp: 821-827.l

[16] Hurroo, M. and M. E. Walizad, (2020), Sign Language Recognition System using Convolutional Neural Network and Computer Vision, International Journal of Engineering Research & Technology (IJERT), ISSN: 2278-0181, Vol. 9 Issue 12.

[17] Kadhim, R., and M. Khamees, (2020), A Real-Time American Sign Language Recognition System Using Convolutional Neural Network for Real Datasets, *Tem Journal* 9.3, pp. 937, doi: 10.18421/TEM93-14.

[18] Wangchuk, K., P. Riyamongkol, and R. Waranusast, (2020), Real-Time Bhutanese Sign Language Digits Recognition System Using Convolutional Neural Network, *Ict Express,* pp.234-238, doi: 10.1016/j.icte.2020.08.002.

[19] [19]. Wadhawan, A., and P. Kumar, (2020), Deep learning-based sign language recognition system for static signs, *Neural Computing and Applications* 32, no. 12, pp: 7957-7968.

[20] [20]. Patil, M., P. Pathole, H. Patil, A. Raut, and S. S. Jadhav, (2020), Indian Sign Language Recognition, International Journal of Scientific Research & Engineering Trends.

[21] Halder, A., and A. Tayade, (2021), Real-time vernacular sign language recognition using mediapipe and machine learning, *Journal homepage: www. ijrpr. com ISSN* 2582: 7421.

[22] Bhadra, R., and S. Kar, (2021), Sign Language Detection from Hand Gesture Images using Deep Multi-layered Convolution Neural Network, In *2021 IEEE Second International Conference on Control, Measurement and Instrumentation (CMI)*, pp. 196-200. IEEE.

[23] Tasmere, D., B. Ahmed, and S. R. Das, (2021), Real Time Hand Gesture Recognition in Depth Image Using CNN, International Journal of Computer Applications, 975-8887, Vol.174, no.16, pp.28-32, doi: 10.5120/20347.

[24] Sidig, A. A. I., H. Luqman, and S. A. Mahmoud, (2017), Transform-based Arabic sign language recognition, *Procedia Comput. Sci.*, vol. 117, pp. 2-9.

[25] ElBadawy M., A.S. Elons , H.A. Shedeed, and M.F.Tolba , (2017), Arabic Sign Language Recognition With 3d Convolutional Neural Networks, In *2017 Eighth International Conference on Intelligent Computing and Information Systems (ICICIS), IEEE*, pp. 66-71.

[26] Ibrahim, N. B., M. M. Selim, and H. H. Zayed, (2018), An Automatic Arabic Sign Language Recognition System (Arslrs), *Journal of King Saud University-Computer and Information Sciences* 30, no. 4, pp: 470-477. doi: 10. 1016/j.jksuci.2017.09.007.

[27] Abdel S., A. Abdel-Rabouh, F. A. Elmisery, A. M. Brisha, and A. H. Khalil, (2018), Arabic Sign Language Recognition Using Kinect Sensor, *Research Journal of Applied Sciences, Engineering and Technology* vol.15, no. 2, pp: 57-67.

[28] Aly S., and W. Aly, (2020), DeepArSLR: A Novel Signer-Independent Deep Learning Framework For Isolated Arabic Sign Language Gestures Recognition, *IEEE Access,* 8:83, pp.199-212.

[29] Al-Hammadi, M., G. Muhammad, W. Abdul, M. Alsulaiman, M. A. Bencherif, and M. A. Mekhtiche, (2020), Hand Gesture Recognition For Sign Language Using 3DCNN, *IEEE Access* 8, pp: 79491-79509.

[30] Abdul, W., M. Alsulaiman, S. U. Amin, M. Faisal, G. Muhammad, F. R. Albogamy, M. A. Bencherif, and H. Ghaleb, (2021), Intelligent real-time Arabic sign language classification using attention-based inception and BiLSTM, *Computers & Electrical Engineering* 95: 107395.

[31] Elhagry A., and R. Gla, 2021, Egyptian Sign Language Recognition Using CNN and LSTM, arXiv *preprint arXiv: 2107 .13647*.

[32] Elsayed E. K., and D. R. Fathy, (2021), Semantic Deep Learning to Translate Dynamic Sign Language, *International Journal of Intelligent Engineering and Systems, Vol.14, No.1, pp: 316-325,* DOI: 10.22266/ijies 2021.0228.30

[33] Pigou, L., S. Dieleman, P. J. Kindermans, and B. Schrauwen, (2014), Sign language recognition using convolutional neural networks, In *European Conference on Computer Vision*, pp. 572-578, Springer, Cham.

[34] Kopuklu, O., N. Kose, and G. Rigoll, (2018), Motion fused frames: Data level fusion strategy for hand gesture recognition, In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 2103-2111.

[35] Liang, Z. j., S. B. Liao, and B. Z. Hu, (2018), convolutional neural networks for dynamic sign language recognition, *The Computer Journal* 61, no. 11, pp:1724-1736.

[36] Masood, S., A. Srivastava, H. C. Thuwal, and M. Ahmad, (2018), Real-Time Sign Language Gesture (Word) Recognition from Video Sequences Using CNN And RNN, *In Intelligent Engineering Informatics*, Springer, Singapore, pp. 623-632.

[37] Lim, K. M., A. W. C. Tan, C. P. Lee, and S. C. Tan, (2019), Isolated sign language recognition using convolutional neural network hand modelling and hand energy image, *Multimedia Tools and Applications* 78, no. 14, pp: 19917-19944.

[38] Tur, A. O., and H. Y. Keles, (2019), Isolated sign recognition with a siamese neural network of RGB and depth streams, In *IEEE EUROCON 2019-18th International Conference on Smart Technologies*, pp. 1-6, IEEE.

[39] Vo, A. H., V. H. Pham, and B. T. Nguyen, (2019), Deep Learning for Vietnamese Sign Language Recognition in Video Sequence, *International Journal of Machine Learning and Computing* 9.4, pp.440-445, doi: 10.18178/ijmlc.2019.9.4.823.

[40] Zhang, W., Wenjin, and J. Wang, (2019), Dynamic hand gesture recognition based on 3D convolutional neural network models, in *2019 IEEE 16th International Conference on Networking, Sensing and Control (ICNSC)*, IEEE, pp. 224-229.

[41] Elboushaki, A., R. Hannane, K. Afdel, and L. Koutti, (2020), MultiD-CNN: A multi-dimensional feature learning approach based on deep convolutional networks for gesture recognition in RGB-D image sequences, Expert *Systems with Applications* 139, pp: 112829.

[42] Hernandez, V., T. Suzuki, and G. Venture, (2020), Convolutional and recurrent neural network for human activity recognition: Application on American sign language, PloS *one* 15, no.2: e0228869.

[43] Kraljević, L., M. Russo, M. Pauković, and M. Šarić, (2020), A Dynamic Gesture Recognition Interface for Smart Home Control based on Croatian Sign Language, *Applied Sciences* 10, no. 7, 2300.

[44] [46]. Rastgoo, R., K. Kiani, and S. Escalera, (2020), Video-Based Isolated Hand Sign Language Recognition Using A Deep Cascaded Model, Multimedia *Tools and Applications* 79, pp: 22965-22987.

[45] Sarma, D., V. Kavyasree, and M. K. Bhuyan, (2020), Two-Stream Fusion Model for Dynamic Hand Gesture Recognition Using 3d-Cnn And 2d-Cnn Optical Flow Guided Motion Template, arXiv *preprint arXiv:2007, 08847*.

[46] Sincan, O. M., and H. Y. Keles, (2020), Autsl: A large scale multi-modal turkish sign language dataset and baseline methods, IEEE *Access* 8, pp: 181340-181355.

[47] Tran, D.S., N. H. Ho, H.J. Yang, E.T. Baek, S.H. Kim, and G. Lee. , (2020), Real-Time Hand Gesture Spotting And Recognition Using RGB-D Camera And 3D Convolutional Neural Network, Applied *Sciences* 10, no. 2, p. 722.

[48] Bach, D. K., D. T. Phung, H. T. Thu Pham, A.N. Bui, and S. T. Ngo, (2021) Vietnamese sign language detection using Mediapipe, In 2021 10th International Conference on Software and Computer Applications, pp: 162-165.

Authors Biography

Mohammad Haqqi Ismail received the BSc and MSc degree in Computer Engineering in 2009 and 2017 from University of Mosul, IRAQ. He is work as assistant lecturer in Technical Computer Engineering, Al-Hadba University College, Mosul, IRAQ. Currently, He is PhD student at research stage in Computer Engineering Department University of Mosul, IRAQ. He researches interests include image processing, deep learning and parallel processing.

Prof. Dr. Shefa A. Dawwd is a professor of computer engineering at the Computer Engineering Department-University of Mosul. He received the B.Sc in Communication Engineering, the M.Sc and the Ph.D in Computer Engineering. He has authored about 40 international journal, conference papers and one-chapter book. His research focus is on the processing acceleration of 1D, 2D and 3D signals, real time applications, deep learning, Convolutional Neural Networks, and hetrogeneous computing. He is a regular reviewer of IEE, Elsevier and other Scopus journals.

Dr. Fakhrulddin H. Ali is assistant professor at the computer engineering Department-University of Mosul. He received B.Sc in Electronic and Communication Engineering-Department of Electrical Engineering-University of Mosul. He received P.G. Diploma and M.Sc from the same Department at 1977-1979. He graduated from university of Bradford-U.K. with a PhD degree at 1989. He has more than 30 scientific papers in journals and conferences. He supervised more than 25 postgraduate M.Sc and PhD. Thesises and dissertations. His field of interest is 3D computer graphics and real time systems.

Cite this paper:

Mohammad H. Ismail, Shefa A. Dawwd, Fakhrulddin H. Ali, "A Review on Arabic Sign Language Recognition", International Journal of Advances in Computer and Electronics Engineering, Vol. 6, No. 12, pp. 1-12, December 2021