International Journal of Electrical and Computer Engineering (IJECE)
Vol. 12, No. 3, June 2022, pp. 2996~3004
ISSN: 2088-8708, DOI: 10.11591/ijece.v12i3.pp2996-3004

Automatic recognition of Arabic alphabets sign language using deep learning

Rehab Mustafa Duwairi, Zain Abdullah Halloush
Department of Computer Information Systems, Faculty of Computer and Information Technology,
Jordan University of Science and Technology, Irbid, Jordan

Article Info

Article history:

Received Jan 26, 2021
Revised Dec 16, 2021
Accepted Jan 19, 2022

Keywords:

AlexNet
Arabic sign language
Assistive technology
GoogleNet/Inception
VGGNet

ABSTRACT

Technological advancements are helping people with special needs overcome many communications' obstacles. Deep learning and computer vision models are innovative leaps nowadays in facilitating unprecedented tasks in human interactions. The Arabic language is always a rich research area. In this paper, different deep learning models were applied to test the accuracy and efficiency obtained in automatic Arabic sign language recognition. In this paper, we provide a novel framework for the automatic detection of Arabic sign language, based on transfer learning applied on popular deep learning models for image processing. Specifically, by training AlexNet, VGGNet and GoogleNet/Inception models, along with testing the efficiency of shallow learning approaches based on support vector machine (SVM) and nearest neighbors algorithms as baselines. As a result, we propose a novel approach for the automatic recognition of Arabic alphabets in sign language based on VGGNet architecture which outperformed the other trained models. The proposed model is set to present promising results in recognizing Arabic sign language with an accuracy score of 97%. The suggested models are tested against a recent fully-labeled dataset of Arabic sign language images. The dataset contains 54,049 images, which is considered the first large and comprehensive real dataset of Arabic sign language to the furthest we know.

Corresponding Author:

Rehab Mustafa Duwairi
Department of Computer Information Systems, Faculty of Computer and Information Technology, Jordan University of Science and Technology
Irbid, Jordan
Email: rehab@just.edu.jo

1. INTRODUCTION

The newly innovated trends these days are guided towards creating novel applications that are of good help to the world around us. The term assistive technology, in many fields, is being brought to the front of researchers' interest. Fingerprints recognition, face detection, and hand gestures recognition are recent applications built on the concepts of machine learning, classification, and image processing.

The need for sign language gestures and special hand movements, make it hard for people with special abilities to communicate freely with people with normal hearing and speaking abilities. This makes the community of people with special abilities secluded from interacting with the rest of the society [1], [2]. The emergence of human and computer interaction (HCI) was the first node for integrating multiple tasks of the human needs and realizing them into easy to use systems and models [3], [4].

The Arabic language is always of huge concern and interest from researchers due to its synaptic nature and traits [5], and with the latest advancements in combining artificial intelligence and computer vision into the needed applications of the human lives [6], everything is becoming more possible and interesting to discover. The Arabic sign language is composed of different hand and fingers movements that represent one of the twenty-eight characters that form the Arabic words, the continuous representation of sequential movements help the deaf person understand what the other person is saying. Due to the lack of knowledge for the normal person to perform these movements, there exists the need to employ the recent advancements in technology to overcome these difficulties [7]. Although sign language is the most structured type of motion expressions [8], the complexity of sign language recognition lies in its way of expression, as there exists two main ways of expressing words in sign language: The first way is by facial expressions and certain body movements (e.g., non-manual components like mouth shaping and eyebrows raising), while the second way is based on fingerspelling technique (e.g., manual components like hand posing, orientation, and trajectory). These two components are complementary to one another. Researchers chose to focus on the area of manual components due to the complex nature of the Arabic language, and the descriptive value obtained from those components for being conventional signs used by many Arab countries [9]. In this paper we will focus on the fingerspelling technique for Arabic words' formation.

The problem of data availability for the Arabic sign language for public use, comparing to other known languages, such as English and Chinese [10]–[12], is an important aspect that challenges the process of Arabic sign language recognition and optimization. Scarcity and lack of quality and quantity terms of available datasets are of constant challenges [13]. This can be referred to as the problem of signer dependency and how each person expresses these signs [2]. Another important challenge in the field of sign language, and especially Arabic, is the co-articulation problem in continuous expression, which can be referred to as the effect of the preceding and latter signs on the targeted sign. Many studies addressed this issue in sign language in general and not in a specific language or culture [14], [15]. The two main approaches used in sign language recognition are summarized in Table 1.

Table 1. Common sign language recognition approaches

	Sign Language Recognition Approaches	
	Vision Based	Sensor Based
Data collection	Images, Videos	Motions and trajectories of fingers and hands
Data collection tools	Different types of cameras	Data gloves, Kinect, EMG, LMC
Data preprocessing	Yes	No
Features extracted	Compact vectors that represent relevant information of hand gestures	3D hand and finger flexes, motions, and orientations
Computational power required	High	Relatively less
Related works	[16]–[18]	[19]–[21]

Sign language is spoken by a wide community of people with hearing and speaking disabilities. Millions of these people worldwide suffer from problematic barriers in the way of expressing themselves due to the lack of automatic systems and tools for recognizing the fingerspelling, hand movements, and facial expressions they use. The importance of humanized computing draws the attention to address the needs of the community of the hearing impaired [22], [23] With more than 70 million people around the world with hearing and speaking disabilities; and with sign language being their only way to communicate with others, the works on automating the process of sign language recognition with the latest technologies have a significant value. Sign languages are as spread as the spoken languages. Every nation has its own way of expressing themselves in words, while also having their own "local" sign language. American sign language [24], Chinese sign language [25], British sign language [26], and Arabic sign language [27], are few examples on standardized forms of sign language. India has a wide range of sign language expressions and dialects that makes it hard to construct a standardized dictionary [28]. These issues raised how widely sign language is used on one hand, while driving the attention to the importance on focusing on the dialectical phenomena of sign language on the other.

While Arabic is one of the top five languages spoken across the globe; the work on Arabic sign language is still in infancy levels and needs more development. This can be referred to the complex nature of Arabic language on one hand, and the lack of a well-constructed database for Arabic sign language on the other hand. For the past few years more work was dedicated to solve the problem of "Diglossia" in Arabic sign language, and the database construction.

The area of computer vision has created a shift in the matter of sign language recognition and helping the involvement of people with disabilities involve in the society with minimum efforts on the side of expresser and receiver, without the need for the inconvenience of an interpreter. The focus of work in this paper is on the vision-based approach in processing Arabic sign language alphabets based on convolutional neural network (CNN). In this paper, a comprehensive study was performed on several deep learning models to evaluate the performance of these models and create a solid comparison criterion between the advancements brought into three consecutive produced models (AlexNet, VGGNet and Inception Net). As a result, a novel approach for Arabic sign alphabets recognition using VGG16 neural network is trained and tested on the latest up-to-date publicly available dataset (ArSL2018) which consists of more than 54,000 images for the alphabets in Arabic sign language. The model was tested and evaluated with an accuracy score of 97%. We intend to provide a robust and evaluated model for sign language recognition in the Arabic language as a benchmark for future optimization.

The rest of this paper is organized in the following: section 2 addresses the related works of sign language recognition using deep learning methodologies, section 3 presents a brief on the different architectures of CNNs. While section 4 discusses the proposed model. Section 5 presents the experiments and results. Finally, section 6 provides the conclusion and suggested future work.

2. RELATED WORKS ON DEEP LEARNING IN SIGN LANGUAGE RECOGNITION

The field of machine learning, specifically deep learning, is the revolutionary state-of-the-art models that enable learning from a set of training data with high futuristic intelligence and preserved back-tracked history, to give accurate predictions and classifications accordingly [29]. Deep learning is now behind many applications of speech recognition, computer vision (CV), and natural language processing (NLP). With its capabilities, deep learning bypasses the traditional techniques of machine learning and provides better results even with unlabeled and unstructured data [30]. It works by processing the input data through multiple layers and performs complex mathematical interactions between the features of the input data. Therefore, the use of deep learning in image processing promises a novel perspective in addressing Arabic sign language.

Elbadawy *et al.* [31], used a 3D CNN to recognize 25 gestures of Arabic sign language dictionary. The input data is fed into the network as videos divided into multiple frames with different rates calculated using scoring algorithm and the frames with the highest priority is chosen to represent the targeted sign. The Softmax layer was used as a features' classifier. Huang *et al.* in [32], also used 3D CNN to extract discriminative spatial-temporal features from raw video stream recorded by Microsoft Kinect. Their work outperformed the Gaussian mixture model with hidden Markov model (GMM-HMM) baseline model which depends on hand-crafted features. Hayani *et al.* [4] proposed a system built on LeNet-5 CNN to recognize Arabic alphabets and digits using a real dataset for a total of 7869 images of size 256x256 pixels. Their system contained four layers to extract features and three layers to classify images. The authors tested their proposed system against well-known machine learning classifiers, such as k-nearest neighbors (KNN) and support vector machine (SVM) and retrieved high recognition rates.

Alzohairi *et al.* in [33] presented an image based Arabic sign language system by investigating visual descriptors from the input data. Those descriptors are later conveyed to one-versus-all support vector machines (SVM) and the experiments showed the significance obtained from using histogram of oriented gradients (HOG) descriptors. In the same context, authors in [34] suggested a semantic oriented approach for detecting and correcting errors based on domain errors. The use of open vocabulary and independent sign recognition lexicon increased the accuracy of the recognition remarkably. Another approach was proposed by [35] in Arabic sign language recognition based on extracting gestures from 2D images using scale-invariant feature transform (SIFT) technique, by first constructing a Gaussian function of varying scales, and then computing the local maximum and the local minimum of key points, while performing latent Dirichlet allocation (LDA) for dimensionality reduction.

Maraqa *et al.* [36] used multilayer feedforward neural network and recurrent neural network (RNN) in a combined model and tested against a dataset consisting of 900 images which were converted into hue-saturation-intensity (HIS) space fro features' extraction. Authors in [37] proposed a glove based Arabic sign language recognition system tested against a manually built lexicon of 80 words to form 40 Arabic sentences. The authors used modified KNN with k=3 and a sliding window approach to reduce fluctuations and reserve long time trends. In the same context, Assaleh *et al.* [38] tested a continuous dataset of the same size to extract spatio-temporal features and hidden Markov model (HMM) that resulted in an average recognition rate of 94%. Kumar *et al.* [39] suggested a multimodal framework for Indian sign language using HMM and bi-directional long short-term memory (Bi-LSTM) sequential classifiers, with a re-sampling technique for linear interpolation of the frames obtained from the Kinect device. In [40] a hybrid model for Arabic sign language detection was introduced using leap motion and digital cameras to detect 20 signs with an accuracy of almost 95%, similar work was also presented in [41]. Koller *et al.* [42] discussed a novel approach in sign language, by introducing multi-stream HMMs constrained by several synchronization constraints with embedded CNN-LSTM in each HMM to recognize weak or unidentified features of hand and lip movements and get over noisy sequence labels. Their work comes as an extension to previous experiment undertaken based on CNN-BLSTM trained end-to-end in several re-alignments [43].

3. CONVOLUTIONAL NEURAL NETWORKS

Convolutional neural networks (or ConvNets) have become the state-of-the-art class used by researchers for visual images and many different applications [44]. The power of the several architectures of CNNs was widely employed to serve image recognition and classification [45], [46]. The structure of the deep multilayer networks, trained in a backpropagation method, helps in obtaining accurate and reliable image classifications. This helps in recognizing visual patters in images with the least pre-processing needed. The advancements in CNNs' structures that are used in image classifications and recognition can be listed as follows:

- Lenet-5: this first multilayer network was designed in 1998 by LeCun *et al.* [47], and is mainly used for hand-writing recognition and documents classification. It is structured with a total of 8 layers including input and output layers for images of 32x32 pixels in grayscale format and is considered the basic form of CNNs.
- AlexNet: in 2012, Krizhevsky *et al.* [48] presented AlexNet. The improvement to the structure of LeNet-5 was introduced by combining maxpooling and rectified linear unit (ReLU) activation functions in a deeper manner with three fully connected layers.
- VGGNet: the power of VGGNets is shown in how they addressed the issue of depth faced by AlexNet [49]. Instead of using large receptive fields; VGG uses a fixed window size of 3×3 and stride of 1 for each layer. This small-sized convolution filters makes the decision function (e.g., rectified linear unit-ReLU) more discriminative, which leads to better performance. In our work, we have employed three deep learning models: AlexNet, VGGNet and Inception network to test the efficiency of these models against the dataset of Arabic sign language images. The use of VGGNet scored remarkable results in Arabic sign language recognition. Figure 1 shows the general structure of VGGNet [50].
- GoogleNet/Inception: another improvement in 2014 was presented by Szegedy *et al.* [51] that addressed the number of parameters. It uses batch normalization, image distortions, and RMSprop to reduce the number of parameters dramatically. Inception Net reduced the number of parameters from 60 million in AlexNet, to 4 million.
- ResNet: residual neural network was introduced in 2015 by He *et al.* [52], to bypass the success of Inception using skip connection approach. The Skip connection approach is the base of RNNs and gated cells method.

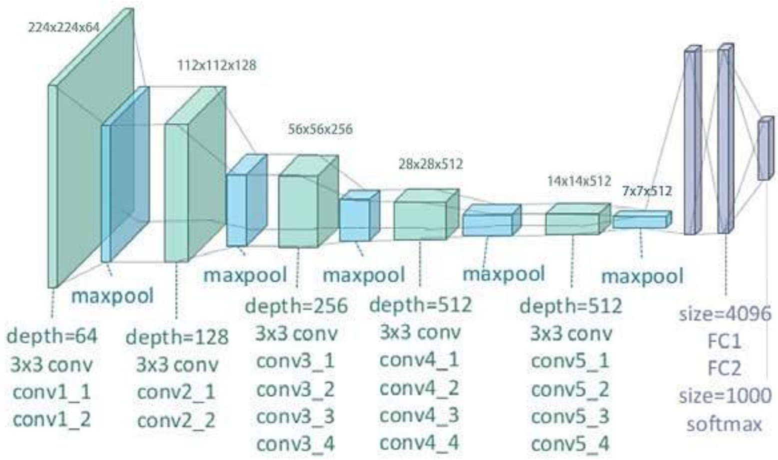

Figure 1. VGG architecture: conv. means convolution, FC means fully connected [50]

4. RESEARCH METHOD

In this section, we discuss the experimental processes to build the proposed model for Arabic sign language recognition. Different models were implemented to compare the results and accuracies obtained. The structure of the model with the highest accuracy is built based on the architecture of the VGGNet and consists of seven layers. The first four layers are used to extract deep features, while the remaining three layers are used for the classification task. The dataset (ArSL2018) is used to experiment the proposed model after being pre-processed using Sobel filter and is fed into the network for training and testing purposes. The VGGNet model scored an accuracy of 97%. Figure 2 shows architectural flow of the employed VGGNet. The layer "Conv_1" is a convolution layer with 32 feature maps. Every neuron performs a convolution with the kernel size of 5x3x3 and adds a bias. The ReLU function was used as an activation function in this case.

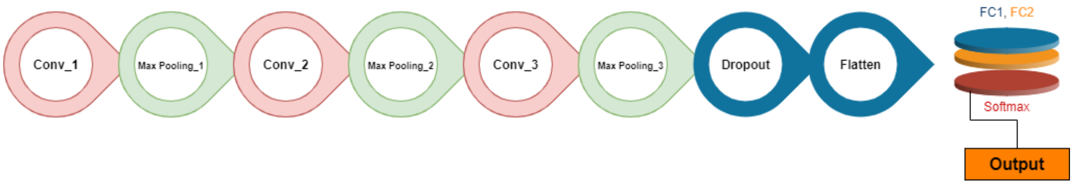

Figure 2. Proposed model architecture

The results obtained from the "Conv_1" layer are then leveraged into the next layer, which is the "Max_Pooling_1", where this layer assembles the results in a conventional 2x2 kernel size (VGGNet), where we next assign a dropout probability of 0.80 for regularization. In "Conv_3", we used a 3x3 kernel size with a 64-feature map. Thereafter, we get a feature map of the size 64 to be flattened and get 576 neurons. While the three remaining layers represent the required classification process, which consists of three fully connected layers of sizes 128 and 84 respectively for the first two, while the last one (the SoftMax layer), is of the size 32 to map the correct class of each input image. The model was trained using Adam optimizer function, using a learning rate of 1e-4. The working environment was built using Tensorflow, Keras, Matplotlib, and Scikit learn.

5. IMPLMENTATION AND DATASET

5.1. Dataset acquisition

The ArSL2018 [53] is the most up to date and comprehensive dataset of Arabic sign language images presented by a research team in Prince Mohammad Bin Fahd University, Al Khobar, Saudi Arabia. The data consists of 54,049 images for the 32 Arabic sign language sign and alphabets collected from 40 participants of different ages, in Grayscale data format and RGB format. The images are 64×64 pixels in (.jpg) format and were captured using a smart phone. Figure 3 shows the images of the used dataset.

Figure 3. Representation of the Arabic sign language in the ArSL2018 dataset

5.2. Data preprocessing and segmentation

For the pre-processing and segmentation phase we used the Sobel operator method which performs a 2D spatial measurement and focuses on the regions with high spatial frequency. Therefore, it is used to find the absolute gradient magnitude in a grayscale image. We have also built a custom ImageDataGenerator for loading, dividing and augmenting the images in the dataset. The generator was used to divide the dataset into batches of the size 64 and feed them sequentially into the neural network for the training process. Thereafter, the dataset is augmented and shuffled to be trained again. The purpose of this process is to expand the variation in the images in the dataset, along with enriching the learning and training process of the neural network. Figure 4 illustrates the distribution of the images in each class. As shown in the graph, the data is normally distributed (Gaussian distribution), and the classes (n=32) contain convergent numbers of images for each letter, which is an important factor for the classification's accuracy. The number of images for each label is relatively the same, which eliminates the bias judgment by the classifier in the recognition process.

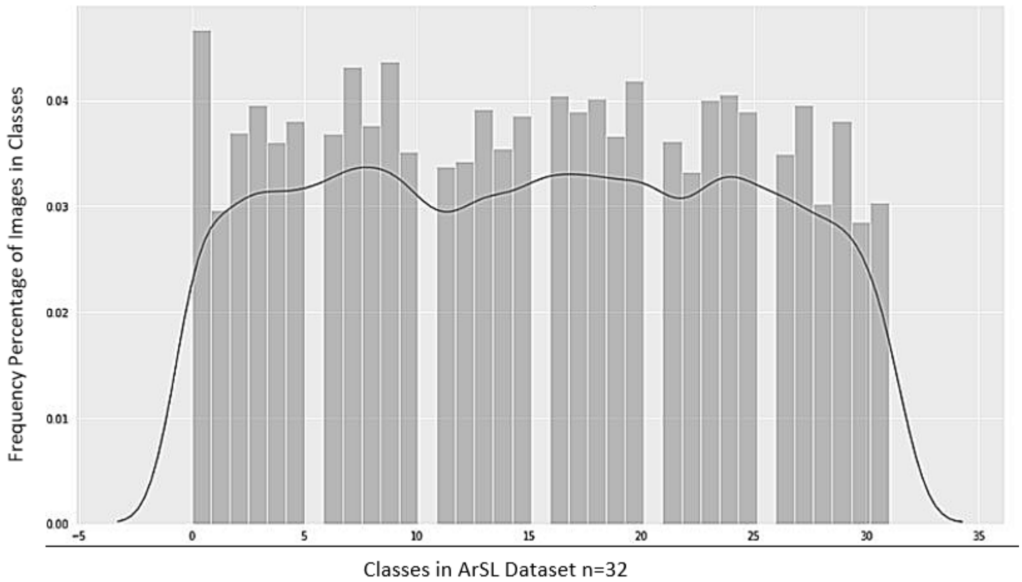

Figure 4. Input images distribution per each letter

5.3. Results and discussion

For the experimental phase and for better improvement of the model validation process. This experiment was applied with k-fold (10) for data splitting in training (training and validation), and testing divisions in an iterative fashion to enhance the performance of the models and tune the hyper parameters and consequently enhance the obtained accuracy. Several experimental phases were run with variation of the train: test splitting criteria, with attest size equals to 0.2, and performed several epochs to justify the accuracy and the error rate of the proposed model, along with testing the model against other structures of known CNNs (e.g. AlexNet and GoogleNet/Inception), while the model built on VGGNet outperformed the previously mentioned two with an accuracy of 97% on test data.

The models were run under unified conditions as follows: the train batch size was 32, and the evaluation batch size was 16. Also, the learning rate was 2e-5, set as a default value. In transfer learning, it was crucial to apply the early stopping method with patience equals to 3. We fine-tuned (unfreeze) the last three layers of the model, the output was then inputted to the average pooling layer with pool size equals to 4, and the output is flattened to be inputted into a dense layer with 64 units. We then applied a dropout of 0.5 probability.

For building a solid argument on the performance of the proposed model for processing the ArSL2018 dataset, different models of CNNs were implemented and evaluated against VGGNet. A model based on the structure of AlexNet and another based on GoogleNet with the application of transfer learning were evaluated and tested against the same dataset. The reason for choosing these three models (i.e. VGGNet, AlexNet, and GoogleNet) is their popularity in image classification and their ease of fine tuning to fit different tasks, along with their chronological development and the advancements brought by each.

5.3.1. AlexNet model

The AlexNet model is pre-trained on the ImageNet dataset and is composed from five convolutional layers, each followed by a Max Pooling layer, followed by three fully connected layers. The accuracy of the model was 93%. AlexNet's model is computationally expensive due to the extensive fully-connected layers, and thus the large number of parameters used. The model of AlexNet only takes images in the input shape of 227×227 in RGB format, thus the first layer, only, will have (227×227×96) output units connected by (11×11×3) input values processed by each filter. This is repeated for all the five layers in a row. The structure of VGGNet solves this problem as it uses a fixed size of kernel size (3×3) in all consecutive layers. This was reflected on the computational consumption in the VGGNet model on one hand, and on the accuracy score on the other. AlexNet has another drawback that was clear in the development phase, which is related to overfitting in the training phase, and a high rate of mislabeling on the test data. The second experiment and comparison were conducted using GoogleNet Model. In this model, the test accuracy obtained was 88%, which is relevantly a small percentage compared to the previous two.

5.3.2. GoogleNet and transfer learning model

Using Tensorflow helps in implementing transfer learning using GoogleNet pre-trained model. GoogleNet was pre-trained on ImageNet dataset, thereafter the last layer of classification was altered with the custom classifier to detect the 32 classes in the dataset and the model was fed with the correctly labeled images of ArSL2018 dataset. The use of transfer learning helps in solving computer vision problems, in a more robust and timesaving manner, due to the pre-training knowledge owned by the model when being trained on large number of images, which is later transferred or fine-tuned to match a similar specific purpose. The pre-trained model (i.e. GoogleNet), was used as a standalone feature extraction to extract relevant features from the images, and then was tested to be integrated for weights initialization, with the allowance of weights optimization for the pre-trained model, along with the weights of our model. It is worth mentioning that weights adjustments process was undertaken in two stages: the first by freezing the pre-defined parameters, while the second was by fine-adjusting the weights and optimize jointly by unfreezing the parameters of base layers. Nevertheless, we justify the degradation in the test accuracy comparing with the previously tested model due to the use of transfer learning and its effect on accuracy rates since the model is not trained on the targeted set of data while keeping the pre-trained weights, even with fine tuning the model for the task. Table 2 provides a summary of the used models and the Accuracy test obtained from each. It is worth mentioning that the proposed system using VGGNet also outperforms the performance of the different classifiers, when using 80% of the dataset for training, when was tested against SVM and KNN classifiers with an approximate value of 45%, as they were used as baseline methods to shed the light on the advancement brought by implementing the state-of-the-art techniques offered by deep learning.

Table 2. CNN models' summary

Test Accuracy Results			
	Model	# of EPOCHS	Accuracy (%)
1	VGGNET	150	97%
2	AlexNet	150	93%
3	GoogleNet	150	88%

6. CONCLUSION AND FUTURE TRENDS

In this paper, a comparative study in offline classification and recognition model for the Arabic sign language alphabets was presented based on the architectures of three common deep learning models (AlexNet, VGGNet and GoogleNet/Inception). The models showed variation in test accuracy with the highest score obtained by VGGNet. The models were tested and trained against the latest publically available dataset of Arabic sign language (ArSL2018) with a size of 54,000 images. The results show the efficiency of the VGGNet model in recognizing the learnt image data with an accuracy of 97%, while the AlexNet model provided an accuracy of 93%, and the accuracy retrieved from using GoogleNet was 88%, with the use of transfer learning approach for image classification. The three models were chosen to compare their performances as they are very common in being used in image classification tasks. For the future, we intent on working on generating real-time sentences and videos using sign language based on CNN models.

REFERENCES

[1] M. Mohandes, M. Deriche, and J. Liu, "Image-based and sensor-based approaches to Arabic sign language recognition," *IEEE Transactions on Human-Machine Systems*, vol. 44, no. 4, pp. 551–557, Aug. 2014, doi: 10.1109/THMS.2014.2318280.

[2] M. Mohandes, J. Liu, and M. Deriche, "A survey of image-based Arabic sign language recognition," in *2014 IEEE 11th International Multi-Conference on Systems, Signals & Devices (SSD14)*, Feb. 2014, pp. 1–4, doi: 10.1109/SSD.2014.6808906.

[3] S. Aly and S. Mohammed, "Arabic sign language recognition using spatio-temporal local binary patterns and support vector machine," *Advanced Machine Learning Technologies and Applications*, pp. 36–45, 2014.

[4] M. Y. Kamil, "Computer aided diagnosis for breast cancer based on the gabor filter technique," *International Journal of*

Electrical and Computer Engineering (IJECE), vol. 10, no. 5, pp. 5235–5243, Oct. 2020, doi: 10.11591/ijece.v10i5.pp5235-5242.

[5] S. Hayani, M. Benaddy, O. El Meslouhi, and M. Kardouchi, "Arab sign language recognition with convolutional neural networks," in *2019 International Conference of Computer Science and Renewable Energies (ICCSRE)*, Jul. 2019, pp. 1–4, doi: 10.1109/ICCSRE.2019.8807586.

[6] L. Zheng, B. Liang, and A. Jiang, "Recent advances of deep learning for sign language recognition," in *2017 International Conference on Digital Image Computing: Techniques and Applications (DICTA)*, Nov. 2017, pp. 1–7, doi: 10.1109/DICTA.2017.8227483.

[7] M. J. Cheok, Z. Omar, and M. H. Jaward, "A review of hand gesture and sign language recognition techniques," *International Journal of Machine Learning and Cybernetics*, vol. 10, no. 1, pp. 131–153, Jan. 2019, doi: 10.1007/s13042-017-0705-5.

[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, 2016, doi: 10.1504/IJAPR.2016.079048.

[9] O. Al-Jarrah and F. A. Al-Omari, "Improving gesture recognition in the Arabic sign language using texture analysis," *Applied Artificial Intelligence*, vol. 21, no. 1, pp. 11–33, Jan. 2007, doi: 10.1080/08839510600938524.

[10] "DGS-corpus," Hamburg University, 2015. https://www.sign-lang.uni-hamburg.de/dgs-korpus/index.php/welcome.html (accessed: Oct. 13, 2021).

[11] "Dictasign project," Hamburg University, 2016. https://www.sign-lang.uni-hamburg.de/dicta-sign/portal/ (accessed: Oct. 13, 2021).

[12] "Bsl corpus project," Economic and Social Research Council, 2016. https://bslcorpusproject.org (accessed: Oct. 13, 2021).

[13] M. Hassan, K. Assaleh, and T. Shanableh, "Multiple proposals for continuous Arabic sign language recognition," *Sensing and Imaging*, vol. 20, no. 1, Dec. 2019, doi: 10.1007/s11220-019-0225-3.

[14] R. Yang, S. Sarkar, and B. Loeding, "Enhanced level building algorithm for the movement epenthesis problem in sign language recognition," in *2007 IEEE Conference on Computer Vision and Pattern Recognition*, Jun. 2007, pp. 1–8, doi: 10.1109/CVPR.2007.383347.

[15] R. Yang, S. Sarkar, and B. Loeding, "Handling movement epenthesis and hand segmentation ambiguities in continuous sign language recognition using nested dynamic programming," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 3, pp. 462–477, Mar. 2010, doi: 10.1109/TPAMI.2009.26.

[16] P. Molchanov, S. Gupta, K. Kim, and J. Kautz, "Hand gesture recognition with 3D convolutional neural networks," in *2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Jun. 2015, pp. 1–7, doi: 10.1109/CVPRW.2015.7301342.

[17] R. Hartanto, A. Susanto, and P. I. Santosa, "Real time static hand gesture recognition system prototype for Indonesian sign language," in *2014 6th International Conference on Information Technology and Electrical Engineering (ICITEE)*, Oct. 2014, pp. 1–6, doi: 10.1109/ICITEED.2014.7007911.

[18] T.-Y. Pan, L.-Y. Lo, C.-W. Yeh, J.-W. Li, H.-T. Liu, and M.-C. Hu, "Real-time sign language recognition in complex background scene based on a hierarchical clustering classification method," in *2016 IEEE Second International Conference on Multimedia Big Data (BigMM)*, Apr. 2016, pp. 64–67, doi: 10.1109/BigMM.2016.44.

[19] M. Mohandes, S. Aliyu, and M. Deriche, "Arabic sign language recognition using the leap motion controller," in *2014 IEEE 23rd International Symposium on Industrial Electronics (ISIE)*, Jun. 2014, pp. 960–965, doi: 10.1109/ISIE.2014.6864742.

[20] C.-H. Chuan, E. Regina, and C. Guardino, "American sign language recognition using leap motion sensor," in *2014 13th International Conference on Machine Learning and Applications*, Dec. 2014, pp. 541–544, doi: 10.1109/ICMLA.2014.110.

[21] K. S. Abhishek, L. C. F. Qubeley, and D. Ho, "Glove-based hand gesture recognition sign language translator using capacitive touch sensor," in *2016 IEEE International Conference on Electron Devices and Solid-State Circuits (EDSSC)*, Aug. 2016, pp. 334–337, doi: 10.1109/EDSSC.2016.7785276.

[22] L. Quesada, G. López, and L. Guerrero, "Automatic recognition of the American sign language fingerspelling alphabet to assist people living with speech or hearing impairments," *Journal of Ambient Intelligence and Humanized Computing*, vol. 8, no. 4, pp. 625–635, Aug. 2017, doi: 10.1007/s12652-017-0475-7.

[23] A.-G. A.-R. Abdel-Samie, F. A. Elmisery, A. M.Brisha, and A. H. Khalil, "Arabic sign language recognition using kinect sensor," *Research Journal of Applied Sciences, Engineering and Technology*, vol. 15, no. 2, pp. 57–67, Feb. 2018, doi: 10.19026/rjaset.15.5292.

[24] C. Zhang, Y. Tian, and M. Huenerfauth, "Multi-modality American sign language recognition," in *2016 IEEE International Conference on Image Processing (ICIP)*, 2016, pp. 2881–2885, doi: 10.1109/ICIP.2016.7532886.

[25] Y. Yu, X. Chen, S. Cao, X. Zhang, and X. Chen, "Exploration of chinese sign language recognition using wearable sensors based on deep belief net," *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 5, pp. 1310–1320, 2020, doi: 10.1109/JBHI.2019.2941535.

[26] J. J. Bird, A. Ekárt, and D. R. Faria, "British sign language recognition via late fusion of computer Vision and leap motion with transfer Learning to American sign language," *Sensors*, vol. 20, no. 18, Sep. 2020, doi: 10.3390/s20185151.

[27] M. Mustafa, "A study on Arabic sign language recognition for differently abled using advanced machine learning classifiers," *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, no. 3, pp. 4101–4115, Mar. 2021, doi: 10.1007/s12652-020-01790-w.

[28] S. C J and L. A, "Signet: a deep learning based Indian sign language recognition system," in *2019 International Conference on Communication and Signal Processing (ICCSP)*, Apr. 2019, pp. 596–600, doi: 10.1109/ICCSP.2019.8698006.

[29] S. Zhang, L. Yao, A. Sun, and Y. Tay, "Deep learning based recommender system: a survey and new perspectives," *ACM Comput. Surv.*, vol. 52, no. 1, 2019, doi: 10.1145/3285029.

[30] Q. Zhang, L. T. Yang, Z. Chen, and P. Li, "A survey on deep learning for big data," *Information Fusion*, vol. 42, pp. 146–157, Jul. 2018, doi: 10.1016/j.inffus.2017.10.006.

[31] M. ElBadawy, A. S. Elons, H. A. Shedeed, and M. F. Tolba, "Arabic sign language recognition with 3D convolutional neural networks," in *2017 Eighth International Conference on Intelligent Computing and Information Systems (ICICIS)*, Dec. 2017, pp. 66–71, doi: 10.1109/INTELCIS.2017.8260028.

[32] J. Huang, W. Zhou, H. Li, and W. Li, "Sign language recognition using 3D convolutional neural networks," in *2015 IEEE International Conference on Multimedia and Expo (ICME)*, Jun. 2015, pp. 1–6, doi: 10.1109/ICME.2015.7177428.

[33] 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, doi: 10.14569/IJACSA.2018.090327.

[34] A. Samir and M. Aboul-Ela, "Error detection and correction approach for Arabic sign language recognition," in *2012 Seventh International Conference on Computer Engineering & Systems (ICCES)*, Nov. 2012, pp. 117–123, doi: 10.1109/ICCES.2012.6408496.

[35] A. Tharwat, T. Gaber, A. E. Hassanien, M. K. Shahin, and B. Refaat, "SIFT-based Arabic sign language recognition system," 2015, pp. 359–370.
[36] M. Maraqa and R. Abu-Zaiter, "Recognition of Arabic sign language (ArSL) using recurrent neural networks," in *2008 First International Conference on the Applications of Digital Information and Web Technologies (ICADIWT)*, 2008, pp. 478–481, doi: 10.1109/ICADIWT.2008.4664396.
[37] 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, Aug. 2015, doi: 10.1109/THMS.2015.2406692.
[38] K. Assaleh, T. Shanableh, M. Fanaswala, F. Amin, and H. Bajaj, "Continuous Arabic sign language recognition in user dependent mode," *Journal of Intelligent Learning Systems and Applications*, vol. 02, no. 01, pp. 19–27, 2010, doi: 10.4236/jilsa.2010.21003.
[39] P. Kumar, H. Gauba, P. Pratim Roy, and D. Prosad Dogra, "A multimodal framework for sensor based sign language recognition," *Neurocomputing*, vol. 259, pp. 21–38, Oct. 2017, doi: 10.1016/j.neucom.2016.08.132.
[40] M. ElBadawy, A. S. Elons, H. Sheded, and M. F. Tolba, "A proposed hybrid sensor a architecture for Arabic sign language recognition," in *Intelligent Systems*, 2015, pp. 721–730.
[41] G. Marin, F. Dominio, and P. Zanuttigh, "Hand gesture recognition with leap motion and kinect devices," in *2014 IEEE International Conference on Image Processing (ICIP)*, Oct. 2014, pp. 1565–1569, doi: 10.1109/ICIP.2014.7025313.
[42] O. Koller, N. C. Camgoz, H. Ney, and R. Bowden, "Weakly supervised learning with multi-stream CNN-LSTM-HMMs to discover sequential parallelism in sign language videos," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 9, pp. 2306–2320, Sep. 2020, doi: 10.1109/TPAMI.2019.2911077.
[43] O. Koller, S. Zargaran, and H. Ney, "Re-sign: re-aligned end-to-end sequence modelling with deep recurrent CNN-HMMs," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jul. 2017, pp. 3416–3424, doi: 10.1109/CVPR.2017.364.
[44] W. Liu, Z. Wang, X. Liu, N. Zeng, Y. Liu, and F. E. Alsaadi, "A survey of deep neural network architectures and their applications," *Neurocomputing*, vol. 234, pp. 11–26, Apr. 2017, doi: 10.1016/j.neucom.2016.12.038.
[45] P. N. Druzhkov and V. D. Kustikova, "A survey of deep learning methods and software tools for image classification and object detection," *Pattern Recognition and Image Analysis*, vol. 26, no. 1, pp. 9–15, Jan. 2016, doi: 10.1134/S1054661816010065.
[46] H. Karim, S. R. Niakan, and R. Safdari, "Comparison of neural network training algorithms for classification of heart diseases," *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 7, no. 4, pp. 185–189, Oct. 2018, doi: 10.11591/ijai.v7.i4.pp185-189.
[47] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," in *Proceedings of the IEEE*, 1998, vol. 86, no. 11, pp. 2278–2324, doi: 10.1109/5.726791.
[48] K. A, S. I, and H. G, "Imagenet classification with deep convolutional neural networks," *Proceedings of the 25th International Conference On Neural Information Processing Systems*, 2012.
[49] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, Sep. 2014.
[50] Y. Zheng, C. Yang, and A. Merkulov, "Breast cancer screening using convolutional neural network and follow-up digital mammography," in *Computational Imaging III*, May 2018, doi: 10.1117/12.2304564.
[51] C. Szegedy *et al.*, "Going deeper with convolutions," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Jun. 2015, vol. 07-12-June, pp. 1–9, doi: 10.1109/CVPR.2015.7298594.
[52] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2016, pp. 770–778, doi: 10.1109/CVPR.2016.90.
[53] G. Latif, N. Mohammad, J. Alghazo, R. AlKhalaf, and R. AlKhalaf, "ArASL: Arabic alphabets sign language dataset," *Data in Brief*, vol. 23, Apr. 2019, doi: 10.1016/j.dib.2019.103777.

BIOGRAPHIES OF AUTHORS

Rehab Mustafa Duwairi received her BSc degree (1989) in computer science from Yarmouk University, Jordan; MSc and PhD degrees in computer science from the University of Wales, Cardiff, UK, in the years 1994 and 1997, respectively. She is currently working as a professor of computer science. Dr. Duwairi research interests include: machine learning, Arabic Language Technologies, social network analysis, and ICT in Education. She is a senior member of ACM. She can be contacted at email: rehab@just.edu.jo.

Zain Abdullah Halloush received her Master Degree of Data Science from Jordan University of Science and Technology, Jordan in 2021. Her Research interests include Deep Learning and Transfer Learning applications in Data Analytics and Social Networks Analysis, Natural Language Processing and Assistive Technology. She currently works as a Research Assistant in Jordan University of Science and Technology. She can be contacted at email: zahalloush18@cit.just.edu.jo.