ARABIC SIGN LANGUAGE CHARACTERS RECOGNITION BASED ON A DEEP LEARNING APPROACH AND A SIMPLE LINEAR CLASSIFIER

Ahmad Hasasneh

ABSTRACT

One of the best ways of communication between deaf people and hearing people is based on sign language or so-called hand gestures. In the Arab society, only deaf people and specialists could deal with Arabic sign language, which makes the deaf community narrow and thus communicating with normal people difficult. In addition to that, studying the problem of Arabic sign language recognition (ArSLR) has been paid attention recently, which emphasizes the necessity of investigating other approaches for such a problem. This paper proposes a novel ArSLR scheme based on an unsupervised deep learning algorithm, a deep belief network (DBN) coupled with a direct use of tiny images, which has been used to recognize and classify Arabic alphabetical letters. The use of deep learning contributed to extracting the most important features that are sparsely represented and played an important role in simplifying the overall recognition task. In total, around 6,000 samples of the 28 Arabic alphabetic signs have been used after resizing and normalization for feature extraction. The classification process was investigated using a softmax regression and achieved an overall accuracy of 83.32%, showing high reliability of the DBN-based Arabic alphabetical character recognition model. This model also achieved a sensitivity and a specificity of 70.5% and 96.2%, respectively.

1. INTRODUCTION

The most natural ways that human beings used to communicate with each other are by using voice, gestures and human-machine interfaces. The last method is still very primitive and forces us to adapt to the machine requirements. Also, the use of voice signals to communicate with hearing-impaired people is impossible or not desirable at all. However, deaf signs or gesture signs of sign language can be desirable and used to communicate with deaf people during their daily life. It's well-known that sign language is the language of deaf people that depends on the body movements or any visual-manual way, particularly the human hands and arms, to convey meanings, where the deaf sign language is typically different from one language to another and from one country to another.

In the Arab society, the community of Arab deaf people is small and very limited; this is due to the fact that only specialists deal with them. Statistics show that over 3% of the Palestinian population are hearing-impaired [1]. Also, according to the Palestinian Central Bureau of Statistics, 19% of disabled Palestinian people are deaf and mute [2]. Indeed, helping those people is very important and thus developing technical systems capable of translating sign languages into text or spoken language is highly needed. Developing such systems will definitely participate in facilitating the communication between the hearing-impaired and hearing people. In addition to that, it has been shown that Arabic sign language (ArSL) is the most difficult recognition task among other foreign sign languages due to its unique structure and complex grammar [3], where the researchers in the Middle East and Arab countries started to pay attention to this problem in the early 1990s. Therefore, developing recognition systems for ArSL is still an open question and a challenging task, which involves three phases of recognition: the first phase is alphabets recognition; the second one is the recognition of an isolated word with one sign and finally, the recognition of a word that contains continuous signs. However, we have seen that most of the existing ArSLR models have been satisfactory for Arabic alphabet recognition with excellent accuracy. Figure 1 shows the 30 letters of Arabic alphabet.

Generally, ArSLR can be performed through two main phases: detection and classification. In the detection phase, each captured image is pre-processed, improved and then the Regions of Interest

---

A. Hasasneh is with Faculty of Eng. and IT, Palestine Ahliya University, Bethlehem, Palestine. Email: ahasasneh@paluniv.edu.ps

(ROI) are identified using a segmentation algorithm. The output of the segmentation process can thus be used to perform the classification process. Indeed, the accuracy and speed of detection play an important role in obtaining accurate and fast recognition process. In the recognition phase, a set of features are extracted from each segmented hand sign and then used to perform the recognition process. These features can, therefore, be used as a reference to understand the differences among the different signs.

Figure 1. **Left:** Original Arabic sign language alphabet [4]. **Right:** The corresponding 32x32 small images.

As mentioned earlier, ArSLR systems have only been paid attention recently [5]–[7]. Some of these attempts are vision-based ArSLR and used the K-nearest neighbor rule [5] or used the hidden Markov model (HMM) [6] and achieved promising results. Some others are sensor-based ArSLR and used the CyberGlove coupled with principal component analysis for feature extraction, followed by a support vector machine (SVM) [7]. Therefore, investigating and developing a new ArSLR model are important and using alternative approaches has to be considered. This paper thus proposes an alternative simpler Arabic sign recognition system based on deep feature extraction methods followed by a simple linear classifier method. Deep learning models have recently shown significant successes in different applications, such as robotics [8], neuroscience [9], traffic sign recognition [10], object detection and recognition [11], audio recognition [12], Bib number detection and recognition [13], Arabic handwritten recognition [10], [14] and image compression and information retrieval [15]-[16].

The rest of the paper is organized as follows. Section 2 presents an overview of the related works. Section 3 presents the proposed model for the recognition of 28 Arabic machine-print characters. Section 4 details the experimental results. Finally, the discussion and conclusions are presented in section 5.

2. RELATED WORKS

Generally, sign language recognition systems for American, British, Indian, Chinese, Turkish and many international sign languages have received much attention compared to the Arabic sign language. A review of the recent development in sign language recognition for foreign languages can be found in [17]–[20]. Also, most of the proposed approaches to the problem of ArSLR have given rise to sensor-based techniques [7], [21]-[22]. Sensor-based model usually employs sensors attached to the hand glove and a look-up table software is usually provided with the glove to be used for hand gesture recognition. However, image-based ArSLR techniques have recently emerged and have been investigated [23]–[29]. The task of image-based ArSLR typically requires firstly producing an appropriate code for the initial data and secondly using this code to classify and learn the alphabet in a fast and accurate manner. Image-based model usually uses video cameras to capture the movements of the hand. However, image-based techniques exhibit a number of challenges, including lighting conditions, image background, face and hand segmentation and different types of noise.

Different classification approaches (generative & discriminative methods) have recently been developed and used to address the problem of ArSLR and most of them focused on recognizing the signs of Arabic alphabet [23], [26]-[27], [30]–[33]. In particular, the authors in [23] developed a neuro-fuzzy system, which includes five main stages, including image acquisition, filtering,

285

Jordanian Journal of Computers and Information Technology (JJCIT), Vol. 06, No. 03, September 2020.

segmentation, hand outline detection followed by feature extraction. The conducted experiment considered the use of the bare hand and achieved a hit rate of 93.6%. The author in [32] has also introduced an automatic recognition of the Arabic sign language letters. For feature extraction, Hu's moments are used and for the classification process, the moment invariants are fed to an SVM. A correct classification rate of 87% was achieved.

The authors in [26] proposed an automatic ArSLR system, that translates isolated Arabic word signs into text, which involves four main stages: hand segmentation, tracking, feature extraction and classification. This proposed model achieved a correct recognition rate of 97% in signer-independent mode. Another recent ArSLR model based on optical flow-based features and HMM was proposed in [30]. In this model, the signs were transformed using four transformation techniques, including: Fourier Transform (MFT), Local Binary Pattern, Histogram of Oriented Gradients (HOG) and combination of HOG and Histogram of Optical Flow. The best classification result was achieved using HMM with MFT features with an accuracy of 99.11%.

Furthermore, the authors in [31] proposed an isolated sign language recognition system that extracts geometric features from a camera for the hand gesture and builds a geometric model for the hand gesture. Using the extracted geometric features, the recognition process of a specific gesture was then performed using the rule-based classifier. The proposed model was tested on seven Arabic words and achieved an overall classification rate of 95.3%. Another work in [32] described two dynamic sign language recognition systems based on two different methods; real-time (online) and offline ones. The comparison was made between the two methods and it was found that the recognition rate for the online hand gesture recognition is lower than the recognition rate for the offline system, which underlines that further enhancements are still needed to improve the real-time recognition performance.

Moreover, several recent attempts have been proposed to use recurrent neural networks (RNNs) [34], convolutional neural networks (CNNs) [27], [31] to achieve ArSLR. More specifically, the authors in [34] proposed to use RNNs together with colored gloves in their experiments. This proposed model achieved an accuracy rate of 89.7%, while a fully recurrent network improved the accuracy to 95.1%; however, it complicated the learning process. The authors in [31] proposed to use 3D CNNs to recognize 25 gestures of Arabic sign dictionary. In this model, the features were extracted with deep behaviour and the proposed model achieved a correct classification rate of 85% for the testing data. Finally, a similar ArSLR approach was proposed in [27] based on CNNs and direct use of softmax regression. When 50% of the dataset was used for training, this similar approach achieved 83.27% of correct classification rate, while when the training dataset was increased to 80%, the correct classification rate increased to 90.02%.

Although most of the current approaches have achieved excellent classification results, some of them are either based on sophisticated classifiers, such as [32], or based on complex learning methods, such as [26]-[27], [31], [33]. In contrast, this paper proposes a deep learning approach that uses a fast learning technique, a Restricted Boltzmann Machine (RBM) and a simple linear classification method. The simplification of the overall classification process of Arabic sign letters must also achieve accurate results. This hypothesis is based on improving the linear separation between signs of Arabic alphabet in the learning phase. This can be achieved by using a powerful machine learning method capable of extracting appropriate features that can be used later to generate an appropriate code for the classification phase. The proposed model is illustrated in the following section and has been recently proposed as a perspective study in our previous publication [35], where it is mainly based on Deep Belief Networks (DBNs) and softmax regression.

3. MATERIALS AND PROPOSED MODEL

3.1 ArSL Dataset and Data Preprocessing

The ArSL dataset used in this paper to test the proposed model was collected by Suez Canal University and used by [4]. In brief, this dataset consists of 210 gray-scale images representing the gestures of 30 Arabic letters; i.e., 7 images for each letter gesture. As stated in [4], the dataset has been captured with different rotations, under different illumination conditions and based on various volunteers who have different hand sizes. In our experiments, we used the first 28 signs of the Arabic

286

" Arabic Sign Language Characters Recognition Based on a Deep Learning Approach and a Simple Linear Classifier," A. Hasasneh.

letters, as shown in Figure 1. 50% of these images were repeated to create a sufficient dataset of a total of 6000 images, which were randomly distributed into groups to make them appropriate for epoch training method. The other sub-set was used for testing the proposed model.

One can see that ArSL images contain a lot of hand features, borders, corners and edges which will contribute and foster the learning network to extract localized and sparse features based on a deep learning approach. It has been shown that the typical input layer of DBN training should be approximately around 1000 pixels, which requires a significant reduction of the original images [8]. Therefore, all images were cropped and significantly reduced to 32×32 = 1024 pixels with a fixed scale, as shown in Figure 1 (right). Despite the big reduction, one can see that the reduced images remain fully recognizable. It has also been shown that DBNs are still capable of extracting interesting features from tiny images [8], [36]. To ensure the network to learn sparse and localized features with higher-order statistics, the tiny images were locally normalized with zero-mean and unit variance. It has recently been shown that the local normalization method achieved better results in terms of feature extraction and classification [8]. Consequently, the normalized tiny images have been used as the input vector to train the network and build the model.

3.2 Deep Learning Model

We developed a novel deep learning approach specifically for the classification of ArSL. The general workflow of the proposed model includes three main stages (see Figure 2), which can be summarized as follows: 1) image pre-processing, 2) unsupervised feature space construction and finally 3) Arabic sign language recognition. The first two steps of the proposed model have recently been carried out and published [35].
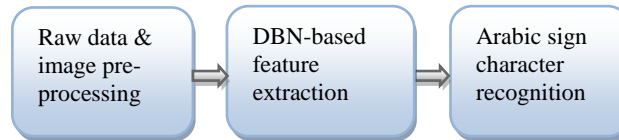

Figure 2. General workflow of the DBN-based model for Arabic sign character recognition using the normalized tiny images.

DBNs are probabilistic generative models composed of multiple RBM layers of latent stochastic variables [8], as illustrated in Figure 3 (right). The RBM is a powerful machine learning algorithm, which can be used to train deep networks in a greedy layer-wise way. The RBM is a bipartite undirected graphical model and consists of two layers (visible (v_i) and hidden (h_j) layers). The two layers are fully connected through a set of weights (w_{ij}) and biases $\{b_i, c_j\}$. Moreover, there is no connection between units of the same layer.

Typically, the input layer of RBM corresponds to the input normalized data. Thus, given the visible vector, the activation probability of the hidden layer can be computed as follows:

$$P\left(h_j = 1 \mid \mathrm{v};\theta\right) = \sigma\left(c_j + \sum w_{ij}v_i\right)$$ (1)

where $\theta = \{w_{ij}, c_j, b_i\}$ and represents the model parameters, σ(x) is the sigmoid function, w_{ij} is the weight matrix between the visible layer and the hidden layer and c_j is the bias of the hidden layer.

Similarly, once the hidden units are computed, the zero-mean Gaussian activation of the visible layer can be recomputed as follows:

$$P\left(v_i = 1 \mid \mathrm{h};\theta\right) \leftarrow \mathcal{N}\left(b_i + \sum_j w_{ij}h_i, \sigma^2\right)$$ (2)

where, b_i is the bias of the visible layer and $\mathcal{N}(\cdot)$ denotes a Gaussian distribution with zero-mean μ and variance σ^2.

As proposed in [8], the contrastive divergence (CD) algorithm can be used to build the model. In other words, utilizing an unsupervised learning strategy, the model parameters are learned by training the

first RBM layer based on a CD learning technique and using the normalized tiny images as stated before. As illustrated in Figure 3 (left), CD learning starts by setting the states of the visible units to a training vector. Then, the binary states of the hidden layer are computed in parallel using Equation 1. Once binary states have been sampled for the hidden units, a "reconstruction" is produced by setting each **v** to 1 with a probability given by Equation 2. Therefore, the model parameters can be updated using the following equations:

$$w_{inc\_ij} = \mu * w_{inc_{ij}} + \eta * \left(((v_0 * h_0) - (v_1 * h_1))/\gamma\right) - \lambda * w_{ij}$$

$$w_{ij} = w_{ij} + w_{inc_{ij}} \tag{3}$$

$$b_i = b_i + \eta * (v_0 - v_1) \tag{4}$$
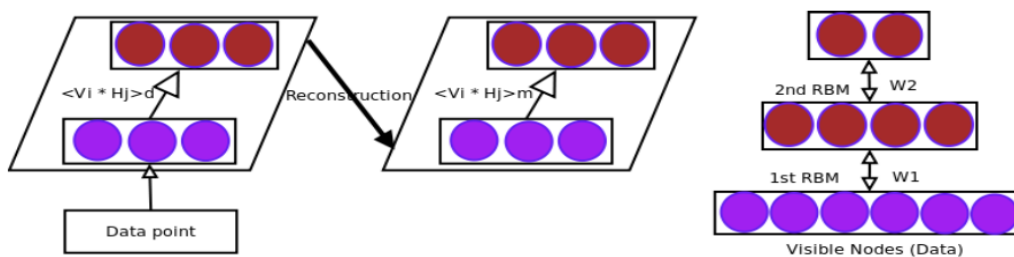
$$c_j = c_j + \eta * (h_0 - h_1) \tag{5}$$

Figure 3. **Left**: Layer-wise training for an RBM with visible and hidden layers using contrastive divergence learning algorithm. **Right**: A deep belief network with two RBM layers.

The convergence of the network is achieved once the difference between the data statistics and the statistics of its representation generated by Gibbs sampling approaches zero, where the training dataset is fed to the network over the epochs. After the convergence of the network, a simple linear classifier, like softmax, is finally used to perform the classification process in the feature space and new samples from the validation dataset are used. It has been shown in many recent studies [8]-[9], [38] that using a deep learning approach plays important roles in: (1) Reducing the dimensionality of the data by using the most significant features to represent an object thus speeding up further tasks; the classification process for instance. (2) Improving the linear separability of the 28 signs of Arabic letters thus simplifying the overall classification process among them.

Therefore, the use of softmax regression was based on the assumption that the data becomes linearly separated in the feature space operated by DBNs. To underline this hypothesis, a non-linear classification algorithm, like SVM, will be used in the classification phase.

4. EXPERIMENTAL RESULTS

4.1 Feature Extraction

Preliminary experiments have shown that the best structure for DBN training in terms of the final classification rate is the complete one (1024-1024). The training protocol is similar to the one proposed in [35] (300 epochs, a mini-batch size of 200, a learning rate of 0.02, an initial momentum of 0.5, a final momentum of 0.9, a weight decay of 0.0002, a sparsity target of 0.02 and a sparsity cost of 0.02).

After the network of the first RBM layer is converged, a set of sparse localized features were learned and extracted, as shown in Figure 4, by training the first RBM layer. One can see that these features represent most of the gestures of Arabic language letters. Some of the extracted features are very localized and represent small parts of the initial hands, like finger edges and hand borders and shapes. Another observation that can be mentioned here is that the thumb and/or index fingers can be seen in most of the extracted features which can be used later as reference features to code the testing images and perform the classification process. We have also seen that training a second RBM layer leads to reduce the classification rate, which might have suppressed some important features. Therefore, we

assume that there is no need to train a second RBM layer, as the extracted features from the first RBM layer represent the shapes of most of the 28 signs of Arabic letters [35]. It can also be seen that some of the extracted features are overrepresented; for example, the signs of the following letters "NOON", "GHAYN" and "LAM". Some other features are underrepresented, such as the signs of the following letters "HA", "JIEM", "QAF", "FA" and "SAD". This might be due to the fact that the overrepresented signs have very sharp shapes which forced the learning network to extract them multiple times over other less sharp signs. The over-representation of high frequencies in the obtained feature will definitely play an important role in improving the linear separation of the initial data in the feature space thus enhancing the classification rate.
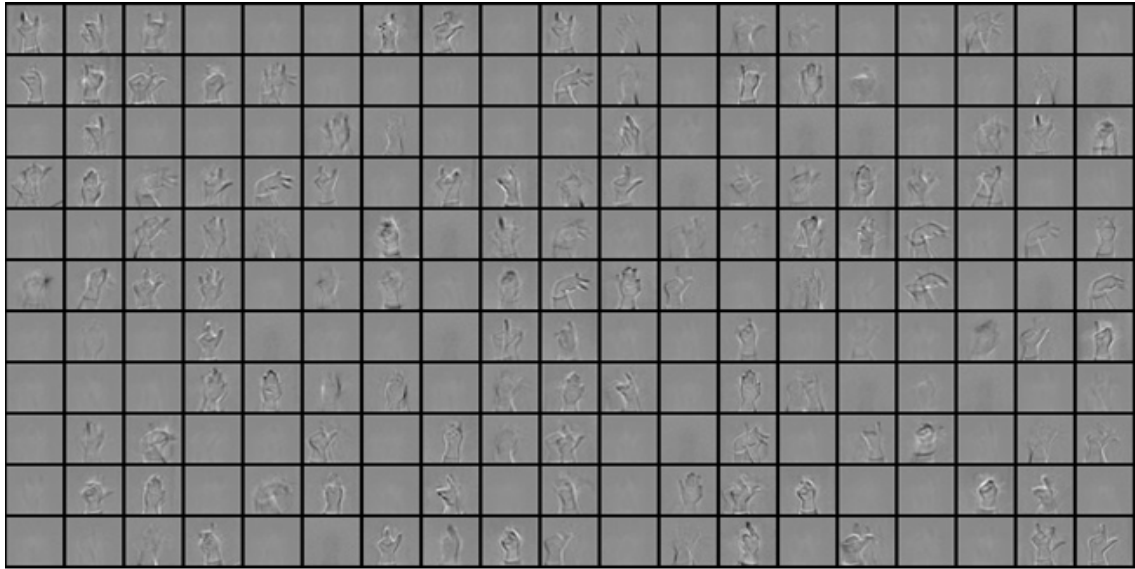

Figure 4. A sample of the 32×32 extracted features obtained by training the 1st RBM layer using the normalized tiny images. A detailed description of the training protocol and its training parameters used in this experiment can be found in [35].
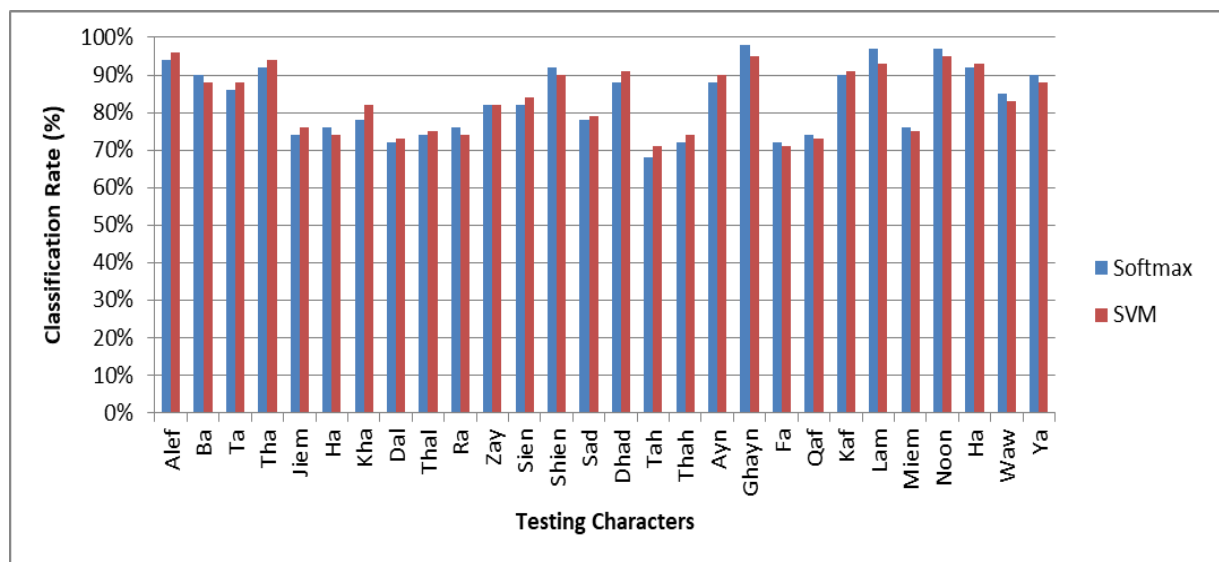

Figure 5. Average classification results for each letter sign of the Arabic language using DBNs coupled with tiny images and followed by softmax regression/ SVM.

4.2 Arabic Gesture Recognition

As mentioned before, after the learning phase and building an appropriate model using the extracted features, the next phase is the classification process for the testing dataset, which was created from the original dataset. The best classification results were achieved using a network of a single RBM layer,

289

Jordanian Journal of Computers and Information Technology (JJCIT), Vol. 06, No. 03, September 2020.

as stated before. Therefore, the real-valued output of the first RBM units is used as an input to a softmax regression to perform the classification process. Figure 5 shows the classification results obtained using DBNs coupled with tiny images and followed by a softmax regression method. The use of a simple classifier was based on the assumption that after an appropriate image coding process is performed using the extracted features shown in Figure 4, the coded images become linearly separated.

For each image from the testing dataset, the softmax network uses the coded sign units to compute the probability of being one of the 28 Arabic letters. Based on the maximum probability value, the system identifies the corresponding character. As illustrated in Figure 5, the correct classification rate for each character of the Arabic language was ranging from 70% to 98% using the proposed model. The overall average of correct classification results for the 28 characters was 83.32% and 83.5% using a softmax regression and an SVM, respectively. These results are consistent and quite comparable when a sophisticated classification algorithm, like SVM, was used instead of softmax. This underlines and demonstrates that the use of DBN has significantly contributed to improving the linear separation between the gestures of Arabic language characters and thus a simple linear classifier was sufficient in the classification stage. These results are also quite comparable to the best recently published image-based approach [27], when the dataset is equally divided for training and testing the model.

5. DISCUSSION AND CONCLUSIONS

In this paper, a simple alternative approach based on deep learning and a linear classifier was proposed to classify the signs of Arabic letters and achieve accurate results. The overall obtained results of this proposed model are comparable to some of the existing approaches; for instance, see [23], [27] based on more complicated learning techniques and sophisticated classifiers and outperformed the results obtained in [29] based on various visual descriptors followed by an SVM. However, these results are still relatively lower than those obtained based on the use of hand-engineered signatures, like SIFT descriptors, followed by the use of a sophisticated classifier, like SVM [38]. Based on the obtained classification results shown in Figure 5, several observations can be discussed. It can be seen that the correct classification rate for some gestures; for example, the letters "NOON", "GHAYN" and "LAM", is high and reached 98%, while the accuracy for other gestures; for example, the letters "HA", "JIEM", "QAF", "FA" and "SAD", was ranging from 70% to 75%. This confirms that the representation of gestures in feature extraction plays an important role in achieving accurate results. It can also be seen that the correct classification rate for the following pair of gestures {("DAL", "THAL"), ("TAH", "THAH") and ("RA", "ZAY")} was ranging from 70% to 80%. The misclassification rate is attributed to the fact that each pair of these letters has strong similarities of gestures, which has probably forced the learning network to extract similar features thus complicating their classification process.

The final result is also illustrated in Table 1, where the recognition evaluation system parameters have been calculated for both models; DBNs followed by softmax regression and DBNs followed by an SVM. The results of these parameters are somehow similar and demonstrate that the linear separation of the initial data has been gained by DBNs and the use of the classification algorithm later will not significantly affect the result. In other words, the use of a simple classifier, like softmax regression, was therefore sufficient to obtain comparable classification results and the use of a sophisticated classifier, such as SVM, slightly improved the classification rate from %83.2 to %83.5. Also, the mean specificity of the proposed model was less when we used softmax regression compared to SVM, which indicates that the true negative rate of the classification results is lower. In addition to that, the mean sensitivity of the proposed model was higher when we used the softmax regression instead of SVM, showing that the true positive rate of the classification results is higher. Finally, the use of

Table 1. Recognition system evaluation parameters for DBNs followed by softmax/ SVM.

Performance parameters/ Classification method	Accuracy	Error	Sensitivity	Specificity	Precision	F1 score
DBNs followed by a softmax	0.832	0.177	0.705	0.962	0.955	0.811
DBNs followed by a SVM	0.835	0.175	0.675	0.995	0.993	0.804

290

" Arabic Sign Language Characters Recognition Based on a Deep Learning Approach and a Simple Linear Classifier," A. Hasasneh.

DBNs followed by softmax regression achieved 81.1% F1 score, while f the use of DBNs followed by an SVM achieved 80.4% F1 score, which underlines that the proposed model is more simple and accurate. The major advantages of the proposed model can be summarized as follows. First, the obtained results demonstrated that a small image-based model followed by an appropriate feature space projection is capable of achieving comparable results to the recent methods [27], based on a complex learning technique (the use of convolutional neural network) and [38], based on a more sophisticated algorithm (the use of SIFT descriptors followed by the use of SVM classifier). Second, based on the obtained results, this paper presents a simple alternative to the existing approaches of ArSLR, by improving the linear separability of the initial data in the feature space thus simplifying the overall classification process. Third, after projecting the gestures onto an appropriate feature space that increases the linear separation between them the use of a nonlinear classifier, like SVM, did not change or improve the results.

Different ways will be investigated in the future to improve the results, including: (1) Assuming sharp signs to force the learning network to extract them multiple times. Thus, studying the effect of normalization and whitening on feature extraction remains an open question. (2) If we assume that extracting sparse features plays an important role in improving the classification results, then studying the sparsity factor with different parameters needs also to be considered. (3) Increasing the size of the dataset to ensure its scalability and to include new gestures that represent the Arabic digits for instance. (4) Increasing the number of images that were underrepresented in feature extraction and have similarities in gestures, thus studying their effect on the learning and classification phases.

REFERENCES

[1]   M. Mohandes, M. Deriche and J. Liu, "Image-based and Sensor-based Approaches to Arabic Sign Language Recognition," IEEE Trans. Human-Machine Syst., vol. 44, no. 4, pp. 551–557, 2014.

[2]   "Palestinian Central Bureau of Statistics," [Online], Available: http://www.pcbs.gov.ps/site/lang__ar/507/default.aspx#ه.

[3]   M. A. Abdel-Fattah, "Arabic Sign Language: A Perspective," Journal of Deaf Studies and Deaf Education, vol. 10, no. 2, pp. 212–221, 2005.

[4]   A. Tiwari, A. B. Narayan and O. Dikshit, "Deep Larning Networks for Selection of Persistent Scatterer Pixels in Multi-temporal SAR Interferometric Processing," arXiv Prepr. arXiv …, 2019.

[5]   R. Naoum, H. Owaied and S. Joudeh, "Development of A New Arabic Sign Language Recognition Using k-nearest Neighbor Algorithm," J. Emerg. Trends Comput. Inf. Sci., vol. 3, no. 8, pp. 1173–1178, 2012.

[6]   N. Saliza et al., "Sign Language to Voice Recognition: Hand Detection Techniques for Vision-based Approach, " Current Developments in Technology-assisted Education, vol. 422, pp. 967-972, 2006.

[7]   M. A. Mohandes, "Recognition of Two-Handed Arabic Signs Using the CyberGlove," Arab. J. Sci. Eng., vol. 38, no. 3, pp. 669–677, 2013.

[8]   A. Hasasneh, E. Frenoux and P. Tarroux, "Semantic Place Recognition Based on Deep Belief Networks and Tiny Images," Proceedings of the 9th International Conference on Informatics in Control, Automation and Robotics (ICINCO-2012), vol. 2, pp. 236–241, 2012.

[9]   A. Hasasneh, N. Kampel, P. Sripad and N. J. Shah, "Deep Learning Approach for Automatic Classification of Ocular and Cardiac Artifacts in MEG Data," vol. 2018, pp. 15–17, 2018.

[10]   A. Hasasneh, Y.-A. Daraghmi and N. M. Hasasneh, "Towards Accurate Real-Time Traffic Sign Recognition Based on Unsupervised Deep Learning of Spatial Sparse Features: A Perspective," Int. J. Comput. Inf. Technol., vol. 06, no. 01, pp. 32–36, 2017.

[11]   L. Liu et al., "Deep Learning for Generic Object Detection: A Survey," International Journal of Computer Vision, vol. 128, no. 2, pp. 261–318, Feb. 2020.

[12]   J. Cornejo and H. Pedrini, "Bimodal Emotion Recognition Based on Audio and Facial Parts Using Deep Convolutional Neural Networks," Proceedings of the 18th IEEE International Conference on Machine Learning and Applications (ICMLA), pp. 111–117, Dec. 2019.

[13]   Y. C. Wong, L. J. Choi, R. S. S. Singh, H. Zhang and A. R. Syafeeza, "Deep Learning-based Racing Bib Number Detection and Recognition," Jordanian Journal of Computers and Information Technology

(JJCIT), vol. 5, no. 3, pp. 181–194, Dec. 2019.

[14] M. Elleuch, N. Tagougui and M. Kherallah, "Optimization of DBN Using Regularization Methods Applied for Recognizing Arabic Handwritten Scripts," Procedia Computer Science, vol. 108, pp. 2292–2297, Jan. 2017.

[15] C. Wan, "Research on Computer Information Retrieval Based on Deep Learning," IOP Conference Series: Materials Science and Engineering, vol. 677, no. 3, DOI: 10.1088/1757-899X/677/3/032101 2019.

[16] Z. Zhao, S. Wang, C. Jia, X. Zhang, S. Ma and J. Yang, "Light Field Image Compression Based on Deep Learning," Proceedings of IEEE International Conference on Multimedia and Expo (ICME), vol. 2018-July, San Diego, CA, USA, Oct. 2018.

[17] A. M. Riad, H. K. Elmonier, S. M. Shohieb and A. S. Asem, "SignsWorld: Deeping into the Silence World and Hearing Its Signs (State-of-the-Art), " Int. J. Comput. Sci. Inf. Technol., vol. 4, no. 1, pp. 189–208, Feb. 2012.

[18] R. Anderson, F. Wiryana et al., "Sign Language Recognition Application Systems for Deaf-Mute People: A Review Based on Input-Process-Output," Procedia Computer Science, vol. 116, pp. 441–448, 2017.

[19] M. Mohandes, M. Deriche and J. Liu, "Image-based and Sensor-based Approaches to Arabic Sign Language Recognition," IEEE Trans. Human-Machine Syst., vol. 44, no. 4, pp. 551–557, 2014.

[20] P. K. Vijay, N. Suhas, C. S. Chandrashekhar and D. K. Dhananjay, "Recent Developments in Sign Language Recognition: A Review," Int. J. Adv. Comput. Eng. Commun. Technol., pp. 21–26, 2012.

[21] N. Tubaiz, T. Shanableh and K. Assaleh, "Glove-based Continuous Arabic Sign Language Recognition in User-Dependent Mode," IEEE Trans. Human-Machine Syst., vol. 45, no. 4, pp. 526–533, Aug. 2015.

[22] P. Kumar, H. Gauba and P. Roy, "A Multimodal Framework for Sensor Based Sign Language Recognition," Neurocomputing, vol. 259, pp. 21–38, 2017.

[23] H. M. Zawbaa, A. Ella Hassanien, K. Nakamatsu, N. El-Bendary and M. S. Daoud, "ArSLAT: Arabic Sign Language Alphabet Translator, " Int. J. Comput. Inf. Syst. Ind. Manag. Appl., vol. 3, pp. 498–506, 2011.

[24] M. Deriche, S. Aliyu and M. Mohandes, "An Intelligent Arabic Sign Language Recognition System Using a Pair of LMCs with GMM Based Classification," IEEE Sens. J., vol. 19, no. 18, pp. 1–12, 2019.

[25] A. A. I. Sidig, H. Luqman and S. A. Mahmoud, "Transform-based Arabic Sign Language Recognition," Procedia Computer Science, vol. 117, pp. 2–9, Jan. 2017.

[26] N. B. Ibrahim, M. M. Selim and H. H. Zayed, "An Automatic Arabic Sign Language Recognition System (ArSLRS)," J. King Saud Univ. - Comput. Inf. Sci., vol. 30, no. 4, pp. 470–477, Oct. 2018.

[27] S. Hayani, M. Benaddy, O. El Meslouhi and M. Kardouchi, "Arab Sign Language Recognition with Convolutional Neural Networks," Proceedings of International Conference of Computer Science and Renewable Energies (ICCSRE 2019), DOI: 10.1109/ICCSRE.2019.8807586, Agadir, Morocco, Jul. 2019.

[28] S. Shohieb and A. M. Riad, "Dynamic Hand Gesture Recognition: Real Time *vs.* Offline Recognition Using Kinect," Advances in Computer Science and Engineering, vol. 11, pp. 39–47, 2013.

[29] R. Alzohairi, R. Alghonaim, W. Alshehri, S. Aloqeely, M. Alzaidan and O. Bchir, "Image-based Arabic Sign Language Recognition System," Int. J. Adv. Comput. Sci. Appl., vol. 9, no. 3, pp. 185–194, 2018.

[30] A. Addin I. Sidig, H. Luqman and S. A. Mahmoud, "Arabic Sign Language Recognition Using Optical Flow-based Features and HMM," Proc. of International Conference of Reliable Information and Communication Technology (IRICT), pp. 297–305, 2018.

[31] M. Elbadawy, A. S. Elons, H. A. Shedeed and M. F. Tolba, "Arabic Sign Language Recognition with 3D Convolutional Neural Networks," Proc. of the 8th IEEE International Conference on Intelligent Computing and Information Systems (ICICIS 2017), vol. 2018-January, pp. 66–71, Cairo, Egypt, 2017.

[32] M. Mohandes, "Arabic Sign Language Recognition," Proc. of Int. Conf. Imaging Sci. Syst. Technol., vol. 1, pp. 753–759, Las Vegas, Nevada, USA, 2001.

[33] M. Maraqa, F. Al-Zboun, M. Dhyabat and R. A. Zitar, "Recognition of Arabic Sign Language (ArSL) Using Recurrent Neural Networks," J. Intell. Learn. Syst. Appl., vol. 04, no. 01, pp. 41–52, 2012.

[34] M. Maraqa and R. Abu-Zaiter, "Recognition of Arabic Sign Language (ArSL) Using Recurrent Neural Networks," Proc. of the 1st International Conference on the Applications of Digital Information and Web Technologies (ICADIWT 2008), pp. 478–481, Ostrava, Czech Republic, 2008.

[35] A. Hasasneh and S. Taqatqa, "Towards Arabic Alphabet and Number Sign Language Recognition," Global Journal of Computer Science and Technology, vol. 17, no. 2, 2017.

[36] G. E. Hinton and R. R. Salakhutdinov, "Reducing the Dimensionality of Data with Neural Networks," Science, vol. 313, no. 5786, pp. 504–507, Jul. 2006.

[37] A. Hasasneh, Robot Semantic Place Recognition Based on Deep Belief Networks and a Direct Use of Tiny Images, Ph.D. Diss., Paris SUD Univ., vol. 2, 2012.

[38] A. Tharwat et al., "SIFT-based Arabic Sign Language Recognition System," Proc. of Afro-European Conference for Industrial Advancement, Springer, Cham, vol. 334, pp. 359–370, 2015.

ملخص البحث:

إن أفضـــل طُـــرق التواصـــل بـــين فاقـــدي السّـــمع والأشـــخاص العـــاديين هـــي لغـــة الإشـــارة أو مـــا يُعـــرف بإيمـــاءات الأيـــدي. وفـــي المجتمعـــات العربيـــة، فـــإنّ الأشـــخاص فاقـــدي السّـــمع والمختصـــين هـــم فقـــط الـــذين باســتطاعتهم التعامـــل مـــع لغـــة الإشـــارة بالعربيـــة؛ الأمـــر الـــذي يجعـــل مجتمـــع فاقـــدي السّـــمع ضـــيقاً ويصـــعب تواصـــلهم مـــع الأشـــخاص العـــاديين. بالإضـــافة الـــى ذلـــك، فـــإنّ دراســـة مشـــكلة تمييـــز لغـــة الإشـــارة بالعربيـــة اكتســـبت اهتمامـــاً فـــي الآونـــة الأخيـــرة؛ الأمـــر الـــذي يؤكـــدّ الحاجـــة الـــى استقصـــاء طـــرقٍ أخـــرى لحـــلّ مثـــل هـــذه المشكلة.

تقتـــرح هـــذه الورقـــة طريقـــة جديـــدة لتمييـــز أحـــرف اللغـــة العربيـــة ورموزهـــا باســتخدام لغـــة الإشـــارة، وذلـــك بنـــاءً علـــى خوارزميـــة تعلُّـــم عميـــق غيـــر مراقبـــة، وشـــبكة ثقـــة عميقـــة (DBN) متصـــلةٍ باســتخدامٍ مباشـــر لصـــور بالغـــة الصِّـــغر: مـــن أجـــل اســتخدامها فـــي تمييـــز أحـــرف اللغـــة العربيـــة وتصـــنيفها. وقـــد أســهم اســتخدام الـــتعلُّم العميـــق فـــي اســتخلاص أهـــم السِّـــمات التـــي تمثـــل بصـــورة متفرقـــة، وكـــان لـــه دور فـــي تبســيط مهمـــة التمييـــز علـــى ســبيل الإجمال.

اســتُخدم مـــا يقـــرب مـــن 6000 عينـــة مـــن الإشـــارات المتعلقـــة بـــأحرف اللغـــة العربيـــة الـــ (28)، بعـــد إجـــراء عمليـــاتٍ تتعلـــق بتكييـــف الحجـــم والتطبيـــع: مـــن أجـــل اســتخلاص السِّـــمات. وقـــد جـــرى استقصـــاء عمليـــة التصـــنيف باســتخدام انحـــدار (ســوفت مـــاكس)، وتـــم الحصـــول علـــى دقـــة إجماليـــة بلغـــت 83.32%: الأمـــر الـــذي يؤكـــد موثوقيـــة النمـــوذج المســتخدم فـــي تمييـــز أحـــرف اللغـــة العربيـــة. أمـــا حساســية النمـــوذج ودقتـــه النوعيـــة، بلغتـــا 70.5% و 96.2%، على الترتيب.