# Large Class Arabic Sign Language Recognition

Zakia Saadaoui[1,2][a], Rakia Saidi[1][b] and Fethi Jarray[1,2][c]
[1]*LIMTIC Laboratory, UTM University, Tunis, Tunisia*
[2]*Higher Institute of Computer Science of Medenine, Gabes University, Medenine, Tunisia*

Keywords: Arabic Sign Language, Recognition, CNN, Gesture Recognition.

Abstract: Sign languages are as rich, complex and creative as spoken languages, and consist of hand movements, facial expressions and body language. Today, sign language is the language most commonly used by many deaf people and is also learned by hearing people who wish to communicate with the deaf community. Arabic sign language has been the subject of research activities to recognize signs and hand gestures using a deep learning model. A vision-based system by applying a deep neural network for letters and digits recognition based on Arabic hand signs is proposed in this paper.

## 1 INTRODUCTION

Sign language is a system of communication set up by deaf and hard of hearing people to communicate with each other, but also with the hearing world. It is a natural visual and non-verbal language that functions as a language in its own right, with its own alphabet, lexicon and syntax. According to the World Federation of the Deaf, there are around 70 million deaf people in the world. Deaf people collectively use over 300 different sign languages. These are natural languages in their own right, structurally distinct from spoken language. Letters detection and recognition is the first step in any pipeline of automatic sign language processing. In the present paper, we focus on identifying sign language gestures that correspond to letters in Arabic languages. The contribution is based on Convolution Neural Network CNN algorithm, a deep learning algorithm that automatically recognizes 32 letters and from 0 to 9 digits using a CNN model feed the ARSL dataset 2018. We organize the rest of this paper on six sections: section two introduces the related works achieved in this field. Section three presents the used dataset. Section four exposes the proposed model. Section five discusses the experimental results and section six presents a general conclusion and some future work.

[a] https://orcid.org/0000-0001-8695-2034
[b] https://orcid.org/0000-0003-0798-4834
[c] https://orcid.org/0000-0003-2007-2645

## 2 STATE OF ART

The Arabic sign language (ASL) approaches can be divided into two categories sensor and vision based approaches. Sensor-based methods, such as e-gloves and the Leap Motion Controller, are needed to track hand movements.The glove-based method, seems a bit uncomfortable for practical use, despite an accuracy of more than 90%. Vision-based method, classified into static and dynamic recognition. Static is the detection of static gestures (2D images) while dynamic is a real-time capture of gestures. This involves the use of a camera to capture the movements. In this paper, we adopt the vision-based approach.

Intrinsically, an image representing a sign languages is composed by three elements: Finger spelling,World level gesture vocabulary and Non-manual characteristics. Finger spelling: spelling out words character by character, and word level association which involves hand gestures that convey the meaning of the word. The static image dataset is used for this purpose. World level gesture vocabulary concerns the recognition of The entire gesture of words or alphabets (Dynamic input / Video classification). Non-manual characteristics contains facial expressions, tongue, mouth, body positions.

In the following, we present a summary of the static approaches of hand gesture recognition since letters are mainly expressed by hands. FASIHUDDIN et al. (Fasihuddin et al., 2018) proposed an interactive System for able-bodied learners to learn sign language. They detect and track hands and fingers move-



Saadaoui, Z., Saidi, R. and Jarray, F.
Large Class Arabic Sign Language Recognition.
DOI: 10.5220/0011539800003335
In *Proceedings of the 14th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2022) - Volume 2: KEOD*, pages 165-168
ISBN: 978-989-758-614-9; ISSN: 2184-3228

ments based on a sensor then the processing is done by a the Kplus algorithm for classification and recognition of signs. In (Ibrahim et al., 2018) an automatic visual system has been designed, it translates isolated Arabic word signs into text. This translation system has four main steps: segmentation and tracking of the hand by a skin detector, feature extraction and finally classification is done by Euclidean distance. Another model for the recognition of Arabic sign language alphabets was designed in (Al-Jarrah and Halawani, 2001) the work was done by training a set of ANFIS models, each of them being dedicated to the recognition of a given gesture. Without the need for gloves, an image of the gesture is acquired using a camera connected to a computer. The proposed system is robust to changes in the position, size and/or direction of the gesture in the image. depends on the calculation of 30 vectors between the center of the gesture area and the useful part of the gesture edge. These vectors are then fed into ANFIS to assign them to a specific class (gesture). The proposed system is powerful when faced with changes in the position, size and/or direction of the gesture in the image. This is due to the fact that the extracted features are assumed to be invariant in terms of translation, scale and rotation. The simulation results showed the model was able to achieve a recognition rate of 93.55%.

Shanableh et al (Shanableh et al., 2007) proposed a system based on gesture classification with KNN method. KNN has proved its performance with an accuracy of 97%. AL-Rousan et al. (Maraqa and Abu-Zaiter, 2008) proposed an automatic Arab sign language (ArSL) recognition system based on Hidden Markov models (HMMs). Experimental results on using real ArSL data collected from deaf people demonstrate that the proposed system has high recognition rate for all modes.

For signer-dependent case, the system obtains a word recognition rate of 98.13%, 96.74%, and 93.8%, on the training data in offline mode, on the test data in offline mode, and on the test data in online mode respectively. On the other hand, for signer-independent case the system obtains a word recognition rate of 94.2% and 90.6% for offline and online modes respectively. The system does not rely on the use of data gloves or other means as input devices, and it allows the deaf signers to perform gestures freely.

In (Alzohairi et al., 2018) the authors propose the use of a system based on feed-forward neural networks and recurrent neural networks with its own architectures; partially and fully recurrent networks. they obtained results with an accuracy of 95% for static signs recognition. Alzohairi and al. in (Al-Rousan et al., 2009) introduced an Arabic alphabet recognition system. This system determines the HOG descriptor and transfers a set of features to the SVM. The proposed system achieved an accuracy of 63.5% for the Arabic alphabet signs. Recently, several researchers have developed a deep CNNs that identify ArSL alphabets with a high level of accuracy. The table 1 presents a summary of the methods based on the CNN architecture.

## 3 DATASET

In this paper, we used the ArSL2018(Latif et al., 2019) dataset which is composed of 54 049 gray scale images with a size of 64x64.Variations of images have been introduced with different lighting and backgrounds. The dataset was randomly divided into eighty percent training set and twenty percent test set. The total number of output classes is 32, ranging from 0 to 31, each representing an ArSL sign, as shown in Figure 1.

## 4 PROPOSED MODEL FOR ASL RECOGNITION

We propose a convolutional neural network(CNN) for Arabic sign letters recognition, inspired by the great success of CNN for image analysis. CNN is a system that utilizes perception, algorithms in machine learning (ML) in the execution of its functions for analyzing the data. This system falls in the category of artificial neural network (ANN). CNN is mostly applicable in the field of computer vision. It mainly helps in image classification and recognition. Our model is based on centralized DL techniques, (Boughorbel et al., 2019) and clean corpora (Boughorbel et al., 2018). Our proposed architecture CNN-5 is composed with 5 convolution layers. Then maximum pooling layers follow each convolution layer. The convolution layers have different structure in the first layer and the second layer there are 64 kernels the size of each kernel is similar 5*5 . Each pair of convolution and pooling layers was checked with a regularization value of elimination which was 50% . The activation function of the fully connected layer uses ReLu and Softmax to decide whether the neuron fires or not. The system was trained for hundred epochs by the RMSProp optimizer with a cost function based on categorical cross-entropy, as it converged well before hundred epochs, so the weights were stored with the system. We have several parameters to set for the model:the number of epochs during training and the

Table 1: Summary of the sign recognition methods based on the CNN architecture.

| Method | Goal | Dataset | Accurancy |
|---|---|---|---|
| Salma Hayani et al(Hayani et al., 2019) | Recognition of Arab sign numbers and letters | Set of images is collected by a set of students | 90,02% |
| M. M. Kamruzzaman(Kamruzzaman, 2020) | Detection of hand sign letters and speaks out the result with the Arabic languageb | Raw images of 31 letters of the Arabic Alphabet | 90% |
| Shroog Alshomrani et al (Alshomrani et al., 2021) | CNN-2 consisting of two hidden layers produced the best results | Arsl dataset | 96,4% |
| Ghazanfar Latif et al (Latif et al., 2020) | CNN-4 are used to obtain the best results | Arsl dataset | **97,6%** |

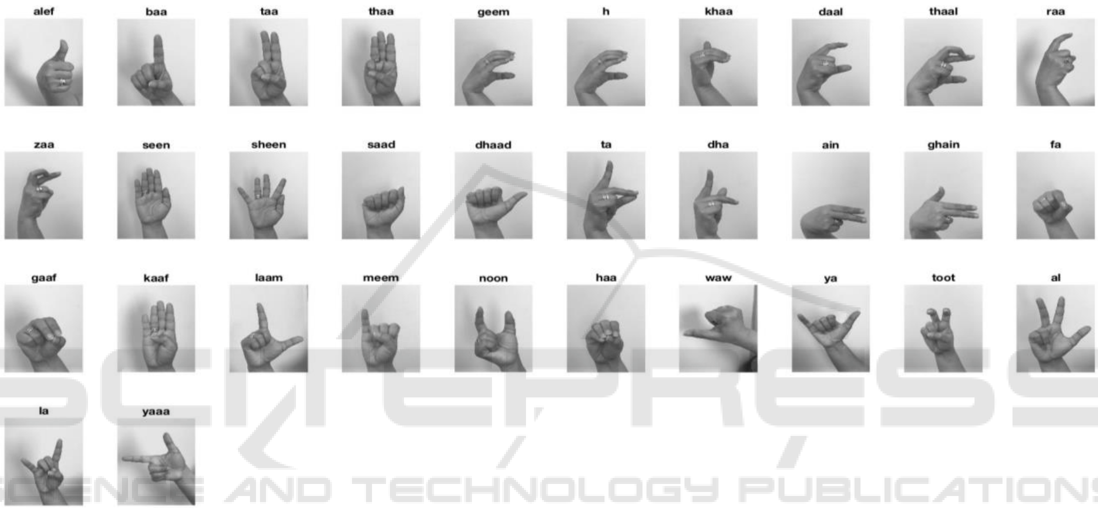


Figure 1: Representation of the Arabic Sign Language for Arabic Alphabet.



batch size. In the testing process, we randomly select 80% of the dataset as a training set and the remaining 20% as a test set.

## 5 RESULTS AND DISCUSSION

We used different experiments, firstly, we evaluated our CNN-5 with different dataset of images , we use ARSL 2018 composed with 32 classes then we combine it with digits [0-9] we obtained 42 classes. In this table 2 we tried both collections of images Alphabets and Alphabets combined with [0-9] digits We applied a grid search to optimize the number of epochs which is the number of times that the entire training dataset is shown to the network during training. The results are presented in the table 3. Generally, the number of classes has a great influence on the efficiency of the system so each time the number increases the precision will decrease and especially for a comparison between systems which use the same dataset but the number of classes is different. However, (Tharwat et al., 2021) reached a higher accuracy for 28 classes. This is due to the use of 28 classes unlike our approach which uses different number of classes despite we cannot compare with this work because they do not use the same dataset as ours and also its dataset not available.

## 6 CONCLUSION

In this paper we implemented a CNN-5 model for ArSL and we validated it through 32 classes of signs with 98.02% and for 42 classes with 97.96% in terms of accuracy. some limitations of static datasets are declared for exemple Fingerspelling for big words and sentences is not a feasible task and Temporal properties are not captured. so as a future work, we aim to interest to dynamic (or video) Datasets.

Table 2: Variations of number of classes.

| Dataset | architecture CNN | Prediction accuracy |
|---|---|---|
| Alphabets 32 classes | CNN-5 | 98.02% |
| Alphabets+[0-9] digits 42 classes | CNN-5 | 97.96% |

Table 3: Variation of the number of epochs with 32 classes.

| Epoch | Predicting accuracy |
|---|---|
| 50 | 88% |
| 100 | 97.23 % |
| 150 | 97.77% |
| 200 | **98.02%** |
| 250 | 96.95% |

## REFERENCES


Al-Jarrah, O. and Halawani, A. (2001). Recognition of gestures in arabic sign language using neuro-fuzzy systems. *Artificial Intelligence*, 133(1-2):117–138.

Al-Rousan, M., Assaleh, K., and Tala'a, A. (2009). Video-based signer-independent arabic sign language recognition using hidden markov models. *Applied Soft Computing*, 9(3):990–999.

Alshomrani, S., Aljoudi, L., and Arif, M. (2021). Arabic and american sign languages alphabet recognition by convolutional neural network. *Advances in Science and Technology. Research Journal*, 15(4).

Alzohairi, R., Alghonaim, R., Alshehri, W., and Aloqeely, S. (2018). Image based arabic sign language recognition system. *International Journal of Advanced Computer Science and Applications*, 9(3).

Boughorbel, S., Jarray, F., Venugopal, N., and Elhadi, H. (2018). Alternating loss correction for preterm-birth prediction from ehr data with noisy labels. *arXiv preprint arXiv:1811.09782*.

Boughorbel, S., Jarray, F., Venugopal, N., Moosa, S., Elhadi, H., and Makhlouf, M. (2019). Federated uncertainty-aware learning for distributed hospital ehr data. *arXiv preprint arXiv:1910.12191*.

Fasihuddin, H., Alsolami, S., Alzahrani, S., Alasiri, R., and Sahloli, A. (2018). Smart tutoring system for arabic sign language using leap motion controller. In *2018 International Conference on Smart Computing and Electronic Enterprise (ICSCEE)*, pages 1–5. IEEE.

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

Ibrahim, N. B., Selim, M. M., and Zayed, H. H. (2018). An automatic arabic sign language recognition system (arslrs). *Journal of King Saud University-Computer and Information Sciences*, 30(4):470–477.

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

Latif, G., Mohammad, N., Alghazo, J., AlKhalaf, R., and AlKhalaf, R. (2019). Arasl: Arabic alphabets sign language dataset. *Data in brief*, 23:103777.

Latif, G., Mohammad, N., AlKhalaf, R., AlKhalaf, R., Alghazo, J., and Khan, M. (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):715–724.

Maraqa, M. and Abu-Zaiter, R. (2008). 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)*, pages 478–481. IEEE.

Shanableh, T., Assaleh, K., and Al-Rousan, M. (2007). Spatio-temporal feature-extraction techniques for isolated gesture recognition in arabic sign language. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 37(3):641–650.

Tharwat, G., Ahmed, A. M., and Bouallegue, B. (2021). Arabic sign language recognition system for alphabets using machine learning techniques. *Journal of Electrical and Computer Engineering*, 2021.