

# A Deep Learning based Approach for Recognition of Arabic Sign Language Letters

Boutaina Hdioud<sup>1</sup>, Mohammed El Haj Tirari<sup>2</sup>

Department of Computer Science, ENSIAS, Mohammed V University, Rabat, Morocco<sup>1</sup>

Department of Statistic, INSEA, Rabat, Morocco<sup>2</sup>

**Abstract**—No one can deny that the deaf-mute community has communication problems in daily life. Advances in artificial intelligence over the past few years have broken through this communication barrier. The principal objective of this work is creating an Arabic Sign Language Recognition system (ArSLR) for recognizing Arabic letters. The ArSLR system is developed using our image pre-processing method to extract the exact position of the hand and we proposed architecture of the Deep Convolutional Neural Network (CNN) using depth data. The goal is to make it easier for people who have hearing problems to interact with normal people. Based on user input, our method will detect and recognize hand-sign letters of the Arabic alphabet automatically. The suggested model is anticipated to deliver encouraging results in the recognition of Arabic sign language with an accuracy score of 97,07%. We conducted a comparison study in order to evaluate proposed system, the obtained results demonstrated that this method is able to recognize static signs with greater accuracy than the accuracy obtained by similar other studies on the same dataset used.

**Keywords**—Deep learning; hand landmark model; convolutional neural network; Arabic sign language recognition

## I. INTRODUCTION

According to the statistical information issued by the World Health Organization in 2021, people with hearing disabilities and deaf persons represent more than 5% of people on the planet [1]. Additionally, it is predicted that in the following 30 years, this number will double. As a result, governments and researchers therefore place a high priority on helping those individuals to participate in their societies. Nowadays, these individuals are the primary users of sign language recognition typically used to communicate both inside and outside of their community.

In this context, we can understand the importance of sign language as a form of non-verbal communication that based on signs and gestures. So, Sign Language Recognition systems (SLR) have been given attention recently. Despite the fact that movement expressions in sign language are the most structured [2], the difficulty of sign language recognition lies in the way that it is expressed. There are two main ways to represent words in sign language; these two ways can be considered complementary to one another: The first way uses particular body movements (e.g. human hands and arms to convey meanings) [3] and facial expressions (e.g. eyebrow raising and mouth shaping) [4]-[6], whereas the second way uses a fingerspelling approach (e.g. orientation, hand posing, and trajectory) [7]. Unfortunately, there is no universal sign language. It's typically different from one country to another

and from one language to another, including the level of the alphabet, where each letter has its own conventional sign. But, with the recent advance in the domain of sign recognition, the researchers have prompted the creation of robust sign language recognition systems for various sign languages, for instance, Brazilian Sign Language [8], British Sign Language [9], Chinese Sign Language [10] and the American Sign Language [11]. However, Asian, English, and Latin sign languages have all been the subject of substantial research, whereas Arabic has received relatively little attention. This can be due to the Arabic language's complexity or the absence of a widely used dataset for the Arabic sign language available to academics. As a result, researchers were forced to build their own datasets, which is a tedious task.

It is crucial to create systems that can translate sign languages into text or speech to help people, who are not deaf or mute, to communicate. Typically, to create any sign language recognition system we can use an image-based approach or a sensor-based technique [12]. Specifically, sensor-based systems work by connecting a glove to a number of sensors to read the gesture, which the system can then recognize; such solutions suffer from a lack of usability. On the other hand, image-based solutions have alleviated this issue and offered a solution in which signs are translated using the cameras. Ideally, those systems can reduce the need for human intervention in the exchange of information between normal people and deaf [13]. Image-based solutions relying on image processing can be carried out in two stages: detection and classification. Each image is first pre-processed during the detection phase, after which the regions of interest are found. The classification process can then be carried out using the results of the preceding procedure. During the recognition stage, each segmented hand sign has a set of features that are extracted in order to perform in the recognition process. Therefore, it is possible to understand the differences between the different signs by using these features as a guide.

Despite extensive research in this field, and as indicated in [14], each study has limitations in terms of the image processing, cost, and sign classification. In general, the system performance depend on the accuracy of the image pre-processing stage, which separates the hand region from the full image, therefore, the main focus of this work is on the recognition of Arabic Sign Language (ArSL) alphabets based on fingerspelling [15]. In order to improve the images of hand gestures accuracy, we proposed to implement a novel ArSL recognition system based on deep convolutional neural network, which incorporates a novel pre-processing stage that

can detect sign gestures by feeding it images of hand gestures performed in varied lighting and orientation, this stage can also alleviate in a small way the problem of strong similarities within the sign language alphabet that present highly similar visual properties, for example, the letter pairs DELL/DHELL, RAA/ZAY, Ayn/Ghayn, ...

Using a dataset of 28 classes of different images of Arabic signs, the objective of this research is to demonstrate how the pre-processing images can be aided in the representation of gestures in feature extraction and how proposed architecture can offer improved accuracy in comparison to other existing methodologies. This work has the following contributions:

- Create a new database of 5000 RGB images and combining them with existing Arabic Sign Language ArSL dataset to input into proposed model and use them as training data.
- Integrate proposed pre-processing stage to enhance the representation gestures.
- Training and comparison of others' works as well as the interpretation of static Arabic sign letters using proposed architecture.
- Analysis of performance of the proposed model in terms of Arabic sign language interpretation.

The rest of the paper is divided into the following sections. An overview of the connected works is given in Section II. The proposed methodology for recognizing hand signs is presented in Section III. The experimental results are described in Section IV. Finally, conclusions and perspectives are presented in Section V.

## II. RELATED WORK

Recently, several research papers focused on ArSLR systems. For recognizing the ArSL with gestures, there are three partitions of the method: recognizing the Arabic alphabet, recognizing individual words, and recognizing entire sentences. In this research we focused on the recognizing of the Arabic alphabet. In this context, we provide a summary of some previously employed methods in this field. Automatic Sign Language Recognition is a field to replace sign language interpreters. There have been many studies performed in this way, and several technologies have been built, such as pre-processing to extract hand gesture, feature extraction to reduce an input data into relevant features, and classification to identify the class of each hand sign. In [16] the authors have summarized the work carried out previously, which focuses on making a difference between dynamic sign language and isolated, static, alphabetic non-Arabic and Arabic sign languages, as well as classification methods used in recognition and that relied on traditional machine learning methods or deep learning methods.

An automatic system for recognizing numbers from 0 to 10 and Arabic letters was implemented in [17]. The authors used a dataset with 7869 images overall. Seven layers composed the suggested model, which was trained repeatedly using various training-testing configurations. Finally, the authors offered a comparison with various strategies based on

KNN and SVM, demonstrating the benefit of the suggested model. An automated method for translating Arabic sign language was suggested by the authors in [18]. This system relies on the building of two datasets for Arabic alphabet gestures. In order to extract Arabic sign gestures from images or videos, this system suggests a hand coverage-based manual detection approach. It then uses a range of statistical classifiers, compares the results, and produces a more accurate classification.

With the use of Microsoft Kinect, Hisham et al. [19] introduced a dynamic Arabic sign language recognition. Two machine learning methods, Bayesian Network and Decision Tree are used for recognition, and the Ada-Boosting methodology is then used to improve system recognition. In [20], the authors used deep transfer learning for developing a robust recognition system for Arabic sign language. To prevent overfitting and improve overall system performance, they used data-augmentation. For the target recognition task, several network architectures have been studied. The Arabic sign language (ArSL2018) public dataset was used in this experiment. Other researchers, such as Saleh et al. [21], enhanced the accuracy of Arabic sign language hand gesture recognition through the use of deep CNN and transfer learning. A novel ArSLR system was proposed in [22] to recognize and classify Arabic alphabetical letters, which uses microscopic images along with an unsupervised deep learning algorithm built on a deep belief network (DBN). In [23], authors developed a system that automatically recognizes 28 letters in Arabic Sign Language using a CNN model with a grayscale image as input. In [24], the authors applied ontology to the sign language domain in order to address some sign language problems. They used simple static signs composed of Arabic alphabets. An architecture of CNN was trained and evaluated using a dataset that was collected and a pre-made dataset for Arabic sign language. A new framework for the automatic recognition of Arabic sign language was proposed by Duwairi et al. [25] and is based on transfer learning applied via popular deep learning models (AlexNet, VGGNet, and Inception Net) for image processing. They proposed using VGGNet architecture, which performed better than previous trained models for automatically recognizing Arabic alphabets in sign language. The authors in [26] suggested a system that can be used by the impaired people. The proposed system converts hand gestures in Arabic sign language into vocalized speech. They used Deep Convolutional Network to extract features from the data gathered by the sensor devices and they employed DG5-V hand gloves to capture the hand movements in the dataset. Finally, they applied CNN method for classification. Furthermore, recent advances in sign language have produced a number of models that are suitable for a variety of tasks; however neither of them actually possesses the necessary generalization capability.

## III. PROPOSED METHODOLOGY

Using two-dimensional images, we seek to recognize the hand gestures of the ArSL and convert them into alphabet in LSA language. The main objective of this research is to build a deep CNN capable of accurately recognizing the ArSL alphabets. An overall view of the representation of the system is presented in Fig. 1.

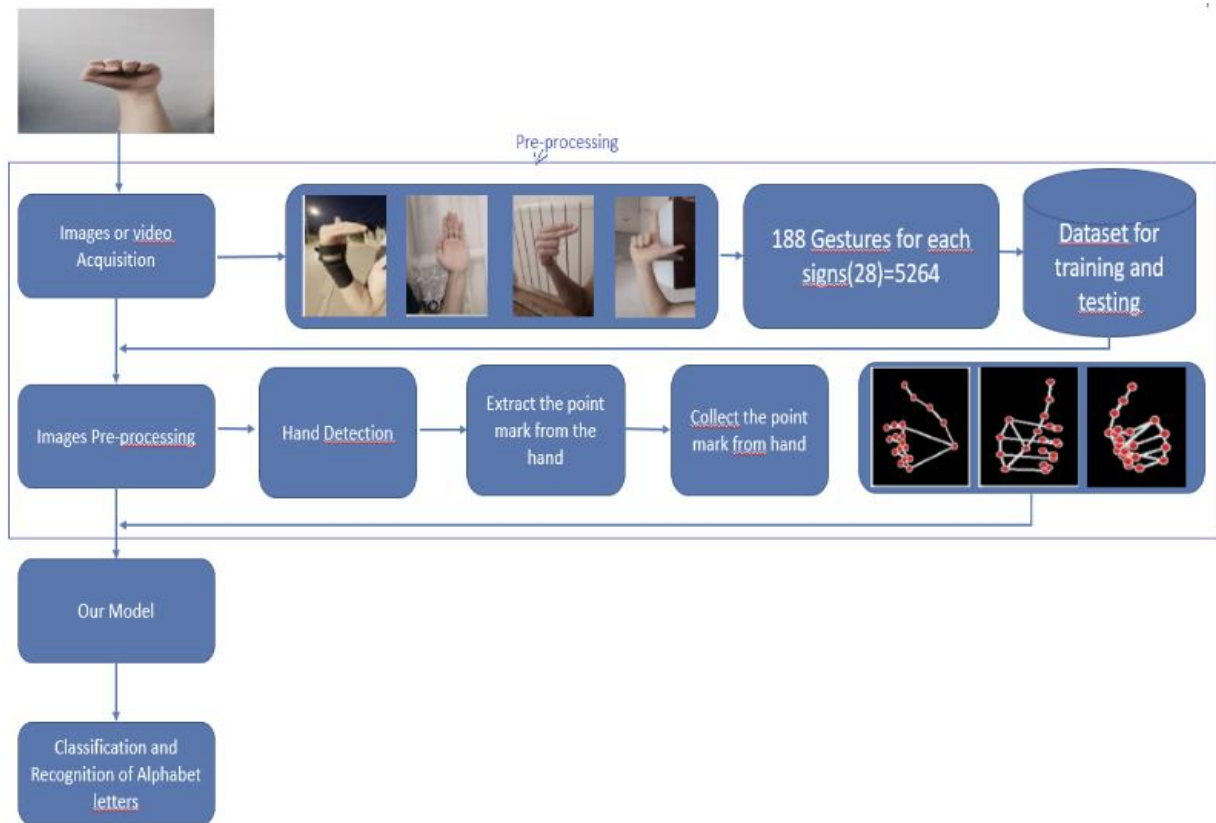


Fig. 1. Overview of the proposed method.

The general architecture of the proposed system comprises two stages: image pre-processing and a proposed model for recognizing Arabic sign language. In the following section we will cover different stages.

#### A. Proposed Pre-processing

After the Arabic sign language images are captured by the camera, they are subjected to a number of main processes during the pre-processing stage. The data are images depth that contain a hand. In order to extract the exact position of the hand we used the mediapipe framework as a solution to detect the hand [27]. The latter has a number of techniques for detection and tracking. Mediapipe Hand is one of its techniques. It consists of two different models working together, the first is namely Palm Detection Model which uses the whole image, produces an oriented hand bounding box made up of rigid objects (e.g. palms and fists) this model has the ability to detect occluded and self-occluded hand, and the second is namely Hand Landmark Model which produces high-fidelity 3D hand keypoints using the cropped image area defined by the palm detector. After the localization phase we cut the image by transforming it into black images that contain just the hand landmarks with a line connecting them after we resize the images to  $240 \times 240$  (Fig. 2.). Finally, we divided our dataset into three parts- training, validation and test for training our model. Fig. 2 shows the input image Fig. 2(a) and the outcomes of the image pre-processing with edge detection in Fig. 2(b).

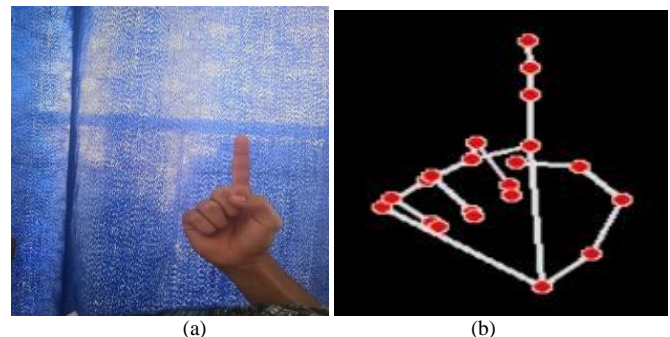


Fig. 2. Input image (a) Before preprocessing and Sample of characters extracted (b) After preprocessing.

#### B. Proposed CNN Architecture

The convolutional neural network CNN model we used in this study is composed of several layers. Fig. 3. shows the proposed architecture of the CNN, which consists of an input layer that accepts images with a size of  $240 \times 240 \times 3$  pixels, which corresponds to the size of the images that the system accepts.

The feature extraction section consists of three convolutional layers (Conv2d, Conv2d\_1, Conv2d\_2) in which each convolution filter has dimensions of 3 by 3. There are thus 16 filters for Conv2d, 32 filters for Conv2d\_1, 64 filters for Conv2d\_2.

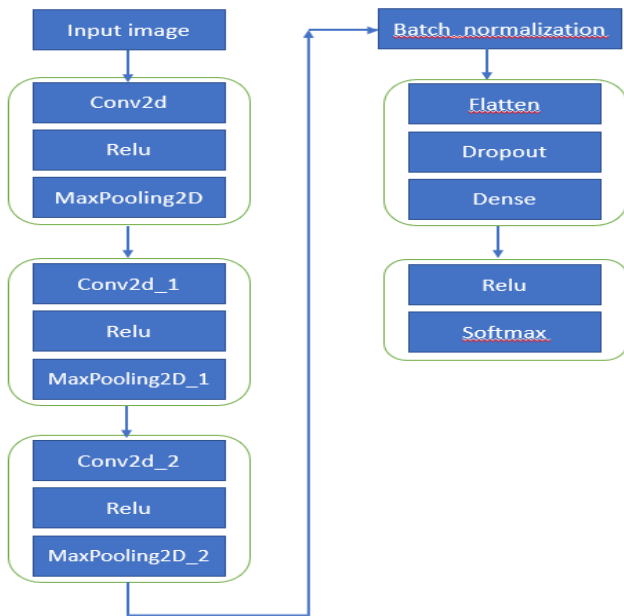


Fig. 3. Proposed CNN network.

Note here that each convolution operation is followed by an activation layer which uses the corrected linear unit function (ReLU). After each activation of the neurons, we will use a Maxpooling layer with a size of 3x3 in an attempt to reduce the size of the images without modifying their important characteristics. This will help to significantly reduce the computational power required by the model. Next, we will use a flatten layer followed by a dropout layer to disable 25% of the neurons. This type of layer helps to reduce overlearning by randomly disabling neurons in the network. Finally, we will use an activation layer using the RELU function followed by a dense classification layer using the softmax activation function. An overview of the parameters in each layer and the overall parameters in the proposed network are shown in Table I.

TABLE I. SUMMARY OF THE PROPOSED CNN NETWORK

Network Layer	Output Shape	Parameters
conv2d(Conv2D)	(None, 240, 240, 16)	448
Max_pooling2d(MaxPooling2D)	(None, 80, 80, 16)	0
conv2d_1(Conv2D)	(None, 80, 80, 32)	4640
max_pooling2d_1(MaxPooling2D)	(None,26,26,32)	0
conv2d_2(Conv2D)	(None, 26, 26, 64)	18496
max_pooling2d_2(MaxPooling2D)	(None, 8, 8, 64)	0
batch_normalization(BatchNormalization)	(None, 8, 8, 64)	256
flatten(Flatten)	(None, 4096)	0
dropout(Dropout)	(None, 4096)	0
dense(Dense)	(None, 256)	1048832
Dense_1(Dense)	(None, 28)	7196
Total parameters	1,079,868	

#### IV. EXPERIMENT AND RESULTS

##### A. Dataset

To evaluate proposed system, we create a combination of the Arabic Sign Language ArSL dataset [28] and our dataset created by professional camera from different signers and different luminosity intensities.

The Arabic Sign Language ArSL dataset is mainly composed of two folders, the first one containing 1160 images each having a size of 416x416 pixels, and the second one containing 1160 text files describing respectively the content of each image in the first folder. The second dataset was created in such a way as to have 28 classes, each containing about 135 images describing an Arabic letter. In order to combine the two datasets, a modification to the format of the Arabic Sign Language ArSL dataset was necessary. We therefore grouped the images of the same letter under the same folder in order to get rid of the need for text files and to unify the format of both datasets. This made it easier for us to combine the two datasets later on. Note here that the Arabic Sign Language ArSL dataset had a problem with the non-existence of the images of the letter "Noun" in their appropriate class which required a manual correction of this by moving them to their original class.

##### B. Results

As previously mentioned, the dataset used for training contained more than 5000 images, distributed in a unified format on 28 different classes of Arabic Sign Language gestures. Three distinct sets of data were created from the dataset, with 60% of the data used for training, 20% for testing, and 20% for validation. The proposed technique would then involve the following steps: We started by pre-processing stage which we outcome black images that contain just the hand landmarks with a line connecting. After that, we inputted the data into proposed model, where the proposed model is iterated through several epochs and at the end of each iteration, accuracy values are provided. Finally, after running the last training epoch, the final accuracy is shown, and the training model is finished. The training and validation accuracy can be seen in Fig. 4(a) and training and validation loss in Fig. 4(b).

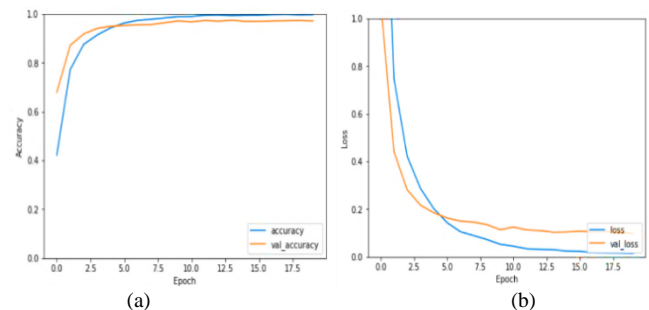


Fig. 4. (a) Training accuracy vs. validation accuracy, (b). Training loss vs. validation loss.

The architecture of CNN was trained for 20 epochs with a batch size of 32. At the 10th epoch, the model was able to validate with an accuracy of 97%, while the 20th epoch saw the highest accuracy of 97.07%. In Fig. 4(a), it can also be

noted how the model training progressed, which demonstrates also how the model stays ahead in every iteration of training and how the loss decreased (see Fig. 4(b)). According to the obtained result we observe no sign of the overfitting.

The confusion matrix shows how well the system performs in terms of correctly and incorrectly classifying developed data. The CM obtained by the CNN model using our methodology is shown in the following Fig. 5. The results obtained show that almost all gestures are correctly classified by our model.

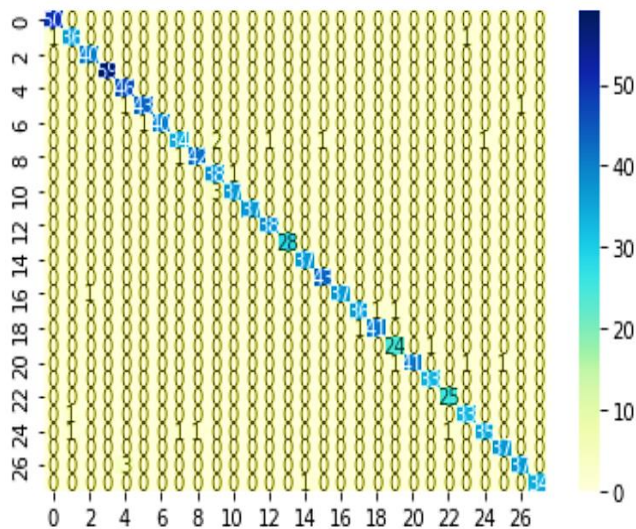


Fig. 5. Confusion matrix.

The classification report, as evident in Fig. 6, confirms what the confusion matrix states. Thus, it can be seen that the rate is actually low, and the majority of the class is properly categorized. The accuracy for the letters ("THA", "KHAA", "SEEN", "SAD", "ZAY") is high and ranges between 95 and 100%, whereas the accuracy for other gestures, such as the letters ("FAA", "JEEM", "RAA", "DEEL", "MEEM"), ranges between 88% and 94%. The principal reason for this difference is that each letter requires a different finger position. This confirms that accurate results depend greatly on how gestures are represented when features are extracted.

According to the obtained result, we can observe that the proposed method allows us to reduce the total number of incorrectly recognized signs to about 32. The misclassification rate is related as previously cited, to the fact that some letters have strong similarities in gestures (RAA/DELL, DELL/DHELL, RAA/ZAY, etc.), which has probably forced proposed model to extract similar features, rendering their classification process more difficult. The following figure (Fig. 7) displays an extract of misclassified images.

C. Comparative Study

To show the performance of the proposed system, we have compared the results obtained from recent literature [18],[29] and [30] with our method. We tested all methods on the same dataset used in this work. Table II reports the results of the three methods under consideration.

	precision	recall	f1-score	support
ALIF	0.98	1.00	0.99	50
BAA	0.95	0.95	0.95	38
TA	0.98	1.00	0.99	40
THA	1.00	1.00	1.00	59
JEEM	0.92	1.00	0.96	46
HAA	0.98	0.96	0.97	45
KHAA	1.00	0.98	0.99	41
DELL	0.94	0.87	0.91	39
DHELL	0.98	0.98	0.98	43
RAA	0.88	0.97	0.93	39
ZAY	0.97	0.93	0.95	40
SEEN	1.00	1.00	1.00	37
SHEEN	0.97	1.00	0.99	38
SAD	1.00	1.00	1.00	28
DAD	0.97	1.00	0.99	37
TAA	0.98	1.00	0.99	43
DHAA	1.00	0.97	0.99	38
AYN	0.97	0.95	0.96	38
GHAYN	0.98	0.98	0.98	42
FAA	0.92	0.96	0.94	25
QAAF	1.00	0.93	0.96	44
KAAF	0.97	1.00	0.99	33
LAAM	0.96	1.00	0.98	25
MEEM	0.94	0.97	0.96	34
NOON	0.97	0.90	0.93	39
HA	0.97	1.00	0.99	37
WAW	0.97	0.93	0.95	40
YA	1.00	0.97	0.99	35
accuracy			0.97	1093
macro avg	0.97	0.97	0.97	1093
weighted avg	0.97	0.97	0.97	1093

Fig. 6. Results for our approach.

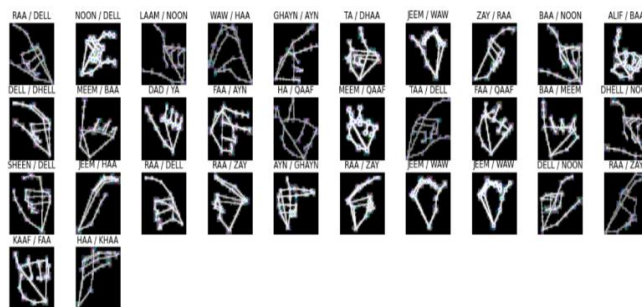


Fig. 7. An extract of the incorrectly classified images.

TABLE II. COMPARISON OF THE PROPOSED SYSTEM WITH OTHER METHODS

Methods	Accuracy
HAYANI et al. [18]	79.27%
KAMRUZZAMAN et al. [29]	83.91%
Ghazanfar Latif et al. [30]	90.61%
proposed model	97.07%

As compared with the results of the models listed above. The best result was reached with our suggested architecture, which has proven effective in the recognition of sign language due to a validation accuracy achieved of 97.70%.

## V. CONCLUSION

In this paper, we used deep learning techniques to perform Arabic sign language recognition based on the hands of deaf people in real time in order to translate the signals into alphabets and improve communication between deaf and non-deaf people.

We started by collecting data and creating our own dataset, then applying a preprocessing that consists in cropping the hand from the image, and extracting its shape, then we proposed four CNN models with different architectures of which three architectures are proposed in three different scientific papers and the last one is proposed by ourselves, and applied them afterwards on our datasets with the aim of making a comparative study in order to choose the best model. The comparative study allowed us to conclude that our architecture is the best performing, where we obtained a score of 97.07%.

In the future work, we would like to extend the dataset for example, increase the variety of images in terms of noise, orientation, etc. and implement new techniques to convert hand movements into written text, and to use NLP (Natural Language Processing) techniques in order to process the obtained texts and present them with a comprehensible and adequate way.

## REFERENCES

- [1] W. H. Organization, "Deafness and hearing loss", 2021. <https://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss>.
- [2] S. Agrawal, A. Jalal, R. Tripathi, "A survey on manual and non-manual sign language recognition for isolated and continuous sign", *International Journal of Applied Pattern Recognition*, vol. 3, no. 2, pp. 99-134, 2016, doi: 10.1504/IJAPR.2016.079048.
- [3] B. Demircioglu, G. Bulbul, and H. Kose, "Turkish Sign Language recognition with Leap Motion", 24th Signal Processing and Communication Application Conference (SIU), no. 2015, pp. 589-592, 2016.
- [4] E. G. Mounq, C. C. Wooi, M. M. Sufian, C. K. On, and J. A. Dargham, "Ensemble-based face expression recognition approach for image sentiment analysis", *Int. J. Electr. Comput. Eng.*, vol. 12, no. 3, pp. 2588-2600, Jun. 2022.
- [5] B. Hdioud and M. E. H. Tirari, "Facial expression recognition of masked faces using deep learning", *IAES Int. J. Artif. Intell.*, vol. 12, no. 2, pp. 921-930, Jun. 2023.
- [6] S. P. Das, A. K. Talukdar, and K. K. Sarma, "Sign Language Recognition Using Facial Expression", *Procedia Comput. Sci.*, vol. 58, pp. 210-216, Jan. 2015.
- [7] D. Waters, R. Campbell, C. M. Capek, B. Woll, A. S. David, P. K. McGuire, M. J. Brammer, and M. MacSweeney, "Fingerspelling, signed language, text and picture processing in deaf native signers: The role of the mid-fusiform gyrus", *Neuroimage*, vol. 35, no. 3, pp. 1287-1302, 2007.
- [8] E. P. da Silva, K. M. O. Kumada, and P. D. P. Costa, "Analysis of Facial Expressions in Brazilian Sign Language (Libras)", *Eur. Sci. Journal*, ESJ, vol. 17, no. 22, pp. 1-1, Jul. 2021.
- [9] KumarKrishan, "DEAF-BSL: Deep IEarning Framework for British Sign Language recognition", *Trans. Asian Low-Resource Lang. Inf. Process.*, vol. 21, no. 5, pp. 1-14, Aug. 2022.
- [10] X. Jiang, S. C. Satapathy, L. Yang, S. H. Wang, and Y. D. Zhang, "A Survey on Artificial Intelligence in Chinese Sign Language Recognition", *Arab. J. Sci. Eng.*, vol. 45, no. 12, pp. 9859-9894, Dec. 2020.
- [11] K. Bantupalli and Y. Xie, "American Sign Language Recognition using Deep Learning and Computer Vision", *Proc. - 2018 IEEE Int. Conf. Big Data, Big Data 2018*, pp. 4896-4899, Jan. 2019.
- [12] K. Assaleh and M. Al-Rousan, "Recognition of Arabic sign language alphabet using polynomial classifiers", *EURASIP J. Adv. Signal Process.*, vol. 2005, no. 13, pp. 2136-2145, Jan. 2005.
- [13] M. Mohandes, J. Liu, and M. Deriche, "A survey of image-based Arabic sign language recognition", 2014 IEEE 11th Int. Multi-Conference Syst. Signals Devices, 2014.
- [14] Suharjo, R. Anderson, F. Wiryana, M. C. Ariesta, and G. P. Kusuma, "Sign Language Recognition Application Systems for Deaf-Mute People: A Review Based on Input-Process-Output", *Procedia Comput. Sci.*, vol. 116, pp. 441-448, 2017, doi: 10.1016/J.PROCS.2017.10.028.
- [15] O. Al-Jarrah and F. A. Al-Omari, "IMPROVING GESTURE RECOGNITION IN THE ARABIC SIGN LANGUAGE USING TEXTURE ANALYSIS", vol. 21, no. 1, pp. 11-33, Jan. 2007, <http://dx.doi.org/10.1080/08839510600938524>.
- [16] M. H. Ismail, S. A. Dawwd, and F. H. Ali, "Arabic Sign Language Recognition : A Review", *Int. J. Adv. Comput. Electron. Eng.*, vol. 6, no. June, pp. 1-12, 2021.
- [17] S. Hayani, M. Benaddy, O. El Meslouhi, and M. Kardouchi, "Arab Sign language Recognition with Convolutional Neural Networks", *Proc. 2019 Int. Conf. Comput. Sci. Renew. Energies, ICCSRE 2019*, Jul. 2019, doi: 10.1109/ICCSRE.2019.8807586.
- [18] A. MAhmed, R. Abo Alez, M. Taha, and G. Tharwat, "Automatic Translation of Arabic Sign to Arabic Text (ATASAT) System", pp. 109-122, 2016.
- [19] B. Hisham and A. Hamouda, "Supervised learning classifiers for Arabic gestures recognition using Kinect V2", *SN Appl. Sci.*, vol. 1, no. 7, pp. 1-21, 2019, doi: 10.1007/s42452-019-0771-2.
- [20] A. I. Shahin and S. Almotairi, "Automated Arabic Sign Language Recognition System Based on Deep Transfer Learning", *IJCSNS Int. J. Comput. Sci. Netw. Secur.*, vol. 19, no. 10, p. 144, 2019.
- [21] Y. Saleh and G. F. Issa, "Arabic sign language recognition through deep neural networks fine-tuning", *Int. J. online Biomed. Eng.*, vol. 16, no. 5, pp. 71-83, 2020, doi: 10.3991/IJOE.V16I05.13087.
- [22] A. Hasasneh, "Arabic sign language characters recognition based on a deep learning approach and a simple linear classifier", *Jordanian J. Comput. Inf. Technol.*, vol. 6, no. 3, pp. 281-290, 2020, doi: 10.5455/jicit.71-1587943974.
- [23] A. Althagafi, G. Althobaiti, T. Alsubait, and T. Alqurashi, "ASLR : Arabic Sign Language Recognition Using Convolutional Neural Networks", vol. 20, no. 7, pp. 124-129, 2020.
- [24] E. K. Elsayed and D. R. Fathy, "Sign language semantic translation system using ontology and deep learning", *Int. J. Adv. Comput. Sci. Appl.*, vol. 11, no. 1, pp. 141-147, 2020, doi: 10.14569/ijacsa.2020.0110118.
- [25] R. M. Duwairi and Z. A. Halloush, "Automatic recognition of Arabic alphabets sign language using deep learning", *Int. J. Electr. Comput. Eng.*, vol. 12, no. 3, pp. 2996-3004, 2022.
- [26] El Rwelli, R. Shahin, O. R. and Taloba, A. I. (2021). Gesture based Arabic Sign Language Recognition for Impaired People based on Convolution Neural Network," *Int. J. Adv. Comput. Sci. Appl.*, 12(12), 574-582, doi: 10.14569/IJACSA.2021.0121273.
- [27] F. Zhang et al., "MediaPipe Hands: On-device Real-time Hand Tracking", 2020, [Online]. Available: <http://arxiv.org/abs/2006.10214>.
- [28] S. BELMADOUI, "Arabic Sign Language ArSL dataset | Kaggle", <https://www.kaggle.com/datasets/sabirbelmadoui/arabic-sign-language-augmented-dataset/code>.
- [29] M. M. Kamruzzaman, "Arabic Sign Language Recognition and Generating Arabic Speech Using Convolutional Neural Network," *Wirel. Commun. Mob. Comput.*, vol. 2020, 2020, doi: 10.1155/2020/3685614.
- [30] G. Latif, N. Mohammad, R. AlKhalaf, R. AlKhalaf, J. Alghazo, and M. A. Khan", "An automatic arabic sign language recognition system based on deep CNN: An assistive system for the deaf and hard of hearing", *Int. J. Comput. Digit. Syst.*, vol. 90, no. 4, pp. 715-724, 2020, doi: 10.12785/ijcds/090418.