International Journal of Engineering Research and Technology. ISSN 0974-3154, Volume 13, Number 11 (2020), pp. 3328-3334
 https://dx.doi.org/10.37624/IJERT/13.11.2020.3328-3334

A Saudi Sign Language Recognition System based on Convolutional Neural Networks

Alaa H Al-Obodi, Ameerh M Al-Hanine, Khalda N Al-Harbi, Maryam S Al-Dawas, and Amal A. Al-Shargabi

Department of Information Technology, College of Computer, Qassim University, Buraydah, Saudi Arabia
ORCID: 0000-0002-7312-9003 (Amal A. Al-Shargabi)

Abstract

Sign language is the main communication method for deaf people. It is a collection of signs that deaf people use to deal with each other. Deaf people find it difficult to communicate with normal people, as most of them do not understand the signs of the sign language. Sign language recognition systems translate the signs into natural languages and thus shorten the gap between deaf and normal people. Many studies have been done on different sign languages. There is a considerable number of studies on the standard Arabic sign language. In Saudi Arabia, deaf people use the Saudi language, which is different from standard Arabic. This study proposes a smart recognition system for Saudi sign language based on convolutional neural networks. The system is based on the Saudi Sign language dictionary, which was published recently in 2018. In this study, we constructed a dataset of 40 Saudi signs with about 700 images for each sign. We then developed a deep convolutional neural network and trained it on the constructed dataset. To have better recognition, we took images of the signs with different hand sizes, skin colors, lighting, backgrounds, and with/without accessories. The results showed that the recognition model achieved an accuracy of 97.69% for training data and 99.47% for testing data. The model was implemented in two versions: mobile and desktop.

Keywords: Sign language, recognition, convolutional neural networks.

I. INTRODUCTION

Sign language is a visual language used for hand movements and facial expressions by people with speech and hearing disabilities to communicate, and each gesture refers to a specific meaning [1]. As in oral language, each country or even region has its own sign language. There are two main types of sign languages: Alphabet notational hand gesture and Ideographic notational hand gesture. Alphabet's notational hand gesture is translated word letter by letter. The ideographic notational hand gesture expresses each meaningful word by a specific hand gesture. The second type is general and commonly used today [2].

Sign language recognition systems are classified into two categories: Device-based systems and Vision-based systems. Device-based systems use wearable tools for gesture tracking such as Microsoft Kinect, Leap Motion Sensors, and gloves [3]. Vision-based systems are techniques processing and analysis using artificial intelligence of images that are captured by the camera to recognize gestures; this is easier for the deaf because it does not need any device to sensor [4]. These systems are important for the development of sign language to make information accessible to the deaf public, and these systems have varying degrees of success.

Deaf people face problems in social life and depend on other understanding environments. Most people might not understand sign language clearly, even need human experts in sign language translation; this way is very expensive, uncomfortable, and may lead to the isolation of deaf people.

In this paper, we proposed a sign language recognition system designed for Saudi deaf people. It is a vision-based system in which a camera is used to shot the deaf sign and translates it into text. The system was based on a convolutional neural networks model. The model is trained on a dataset that contains 40 Saudi signs. We constructed the dataset so that every class includes around 700 images of different backgrounds and conditions.

The remaining of the paper is organized as follows: Section II presents the related works. Section III presents the proposed model, Section IV presents the system prototype, and Section V concludes the papers.

II. RELATED WORK

The related works are presented in terms of the datasets and the recognition methods used in previous studies.

II.I Datasets

Table 1 shows a summary of the datasets used in the state of the art studies. As shown in the Table, many studies have been done based on the standard Arabic, and most of them were to recognize alphabets, numbers, and a limited number of words.

II.II Recognition Methods

CNN is a powerful artificial intelligence tool in pattern classification. In the study of ElBadawy et al. [3], the authors used a 3D CNN to extract the Spatio-temporal features and then classify 25 classes that exist in the dataset.

In the study of Adithya et al.[4], the authors designed a system based on artificial neural networks for the automatic recognition of fingerspelling for Indian sign language by using digital image processing techniques and artificial neural networks.

The study of Hayani et al.[5] developed a system using convolutional neural networks (CNN), which are multi-layer neural networks that make use of deep learning to analyze images. The proposed model used CNN inspired fbyLeNet-5. They used seven adjacent layers, four layers to extract deep features, and three layers to classify them.

International Journal of Engineering Research and Technology. ISSN 0974-3154, Volume 13, Number 11 (2020), pp. 3328-3334
 https://dx.doi.org/10.37624/IJERT/13.11.2020.3328-3334

Table 1. A summary of the dataset used in previous studies

Description	Size	Language	Ref
Words (25)	200 Images	Arabic	[3]
Alphabet and Numbers from 0 to 10	7869 Images	Arabic	[5]
Words (50)	450 Videos	Arabic	[6]
Alphabets and Numbers from 0 - 10	900 Images	American	[7]
Alphabets (26 letters)	1100 images	Arabic	[8]
Alphabets	900 Images	Arabic	[9]
Six Alphabets	200 Images	Arabic	[10]
Alphabets (26 letters)	1100 images	Arabic	[11]
Alphabets (37 letters)	1147 Images	Bengali	[12]

The work of Ibrahim et al. [6] used Euclidean distance as a feature extraction method for the Arabic sign language. Hand segmentation was applied to the different frames of a video-based dataset. This method tracks the trajectory of an object in the image plane as it moves around a scene, detecting motion with an active camera.

Also, Maraqa and Abu-Zaiter [8] developed a system using a recurrent neural network (RNN) for static images. The algorithm is applied to all frames, and then the segmented skin blobs are used in identifying and tracking the hands with the help of the head.

In the study of Albelwi and Alginahi [10], the authors assessed the signs' classifications by the K-Nearest Neighbor (KNN) algorithm, which is one of the most commonly used methods in sign language recognition systems.

In the work of Hossen et al. [12], the CNN network consists of the following: convolution layer, max-pooling layer, ReLU layer, dropout layer, fully connected layer, softmax layer. The dataset was small, but the authors expanded the dataset by including more samples by scaling and rotating images.

In the study of Sawant and Kumbhar [13], the authors used Euclidean distance to calculate the difference between testing and training images, and then gestures can be recognized.

The study of Rao et al. [14] used the CNN architecture for classifying selfie sign language gestures. The CNN architecture was designed with four convolutional layers, and different filtering window sizes were considered. This improved the speed and accuracy of recognition.

The study of Hartanto and Kartikasari performed feature extraction for each hand gesture image in real-time, which was efficient in terms of computation time and good performance [15].

III. The Saudi Sign Recognition Model

III.I Dataset Construction

The database is based on Saudi Dictionary that was prepared and published by the Saudi Association for Hearing Disability in 2018. The dictionary contains thousands of signs from 28 fields such as medical, social, religious, and others. Fig. 1 shows a picture of the dictionary.

Fig. 1. The standard Saudi sign dictionary

In this study, we scope our dataset to the signs that contain only one gesture, and thus static images are making up the dataset. The signs which include multiple gestures were avoided. The images of the alphabet, numbers, and one word which is 'أنا' which means 'I' in the English language. Sample images of the alphabets' and numbers' signs are shown in Fig. 2 and 3.

It is essential to mention that the images were taken by a mobile camera so that the system can recognize late any images taken by primary cameras. Also, the images were taken by different people and with different backgrounds, lighting, and indoors and outdoors. This makes the proposed CNN model more general as it is trained on images at different conditions. In total, we took around 700 images for each class, resulting in 27,301 images.

International Journal of Engineering Research and Technology. ISSN 0974-3154, Volume 13, Number 11 (2020), pp. 3328-3334
 https://dx.doi.org/10.37624/IJERT/13.11.2020.3328-3334

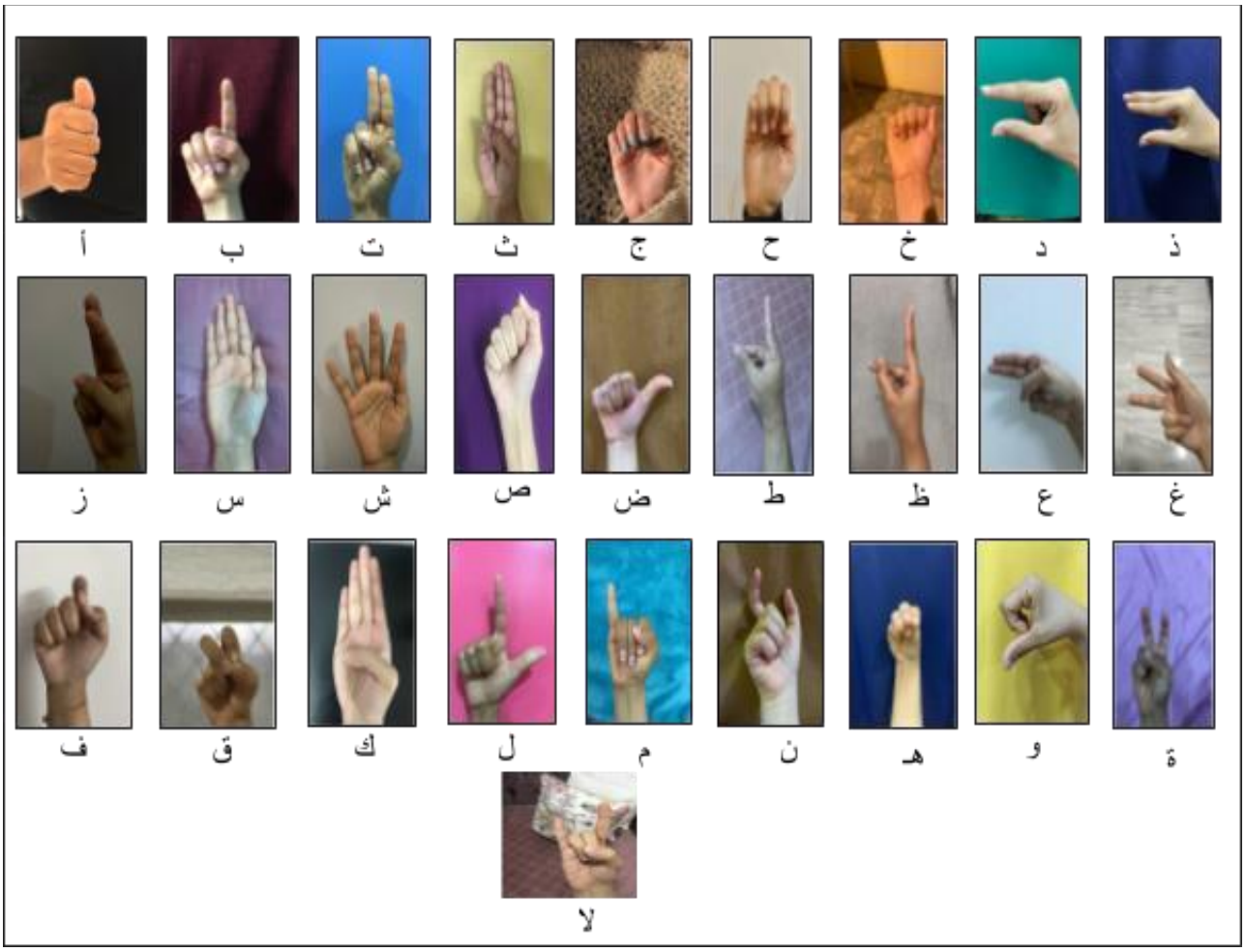

Fig. 2. Sample images of Alphabet Signs

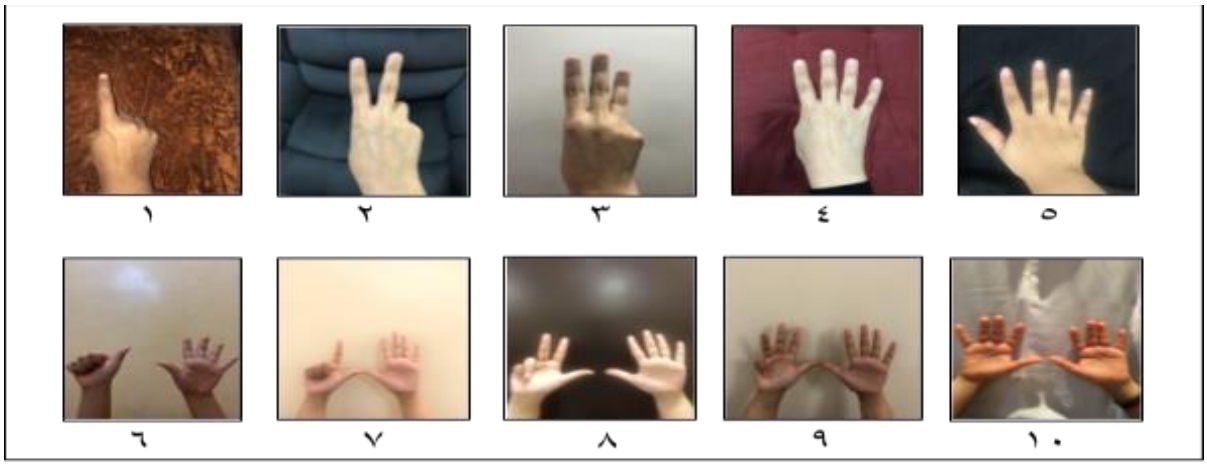

Fig. 3. Sample images of number Signs

International Journal of Engineering Research and Technology. ISSN 0974-3154, Volume 13, Number 11 (2020), pp. 3328-3334
 https://dx.doi.org/10.37624/IJERT/13.11.2020.3328-3334

III.II The Proposed Recognition Model

Neural networks (NN) are machine learning techniques that are modeled based on brain structure. They comprise units called neurons that learn how to convert input, e.g., the image of sign of alphabet 'a' into its corresponding output, e.g., the label of sign of alphabet 'a'.

Convolutional Neural Networks (CNN) are very popular NN algorithms that are often applied to image classification problems, CNNs have proven to be successful in areas like image classification and recognition [7]. CNN does take the image and pass it through a series of layers. The output can be a class label or a probability of classes that best describes the image. In this study, the CNN results in the actual classes, i.e., labels of the sign images. The constructed dataset images were preprocessed to have a unified size.

The CNN architecture used to build the sign language recognition model is shown in Fig. 4. As shown in the figure, the model contains multiple convolutional layers, max pooling layers, dropout layers, and ended with flatten and dense layers.

The constructed dataset of the sign images was used to train the CNN model presented in the previous section.

A total of 21,840 images were used for training, and 5461 images were used for testing.

The accuracies of training and testing were 97.69% and 99.47%, respectively. Fig. 5 shows the confusion matrix for the 40 classes, including numbers, letters, and the word 'أنا' based on validation images. The number of validation images was between 120 and 150 images.

For evaluation, the proposed model was compared with the state of the art models. See Table 2.

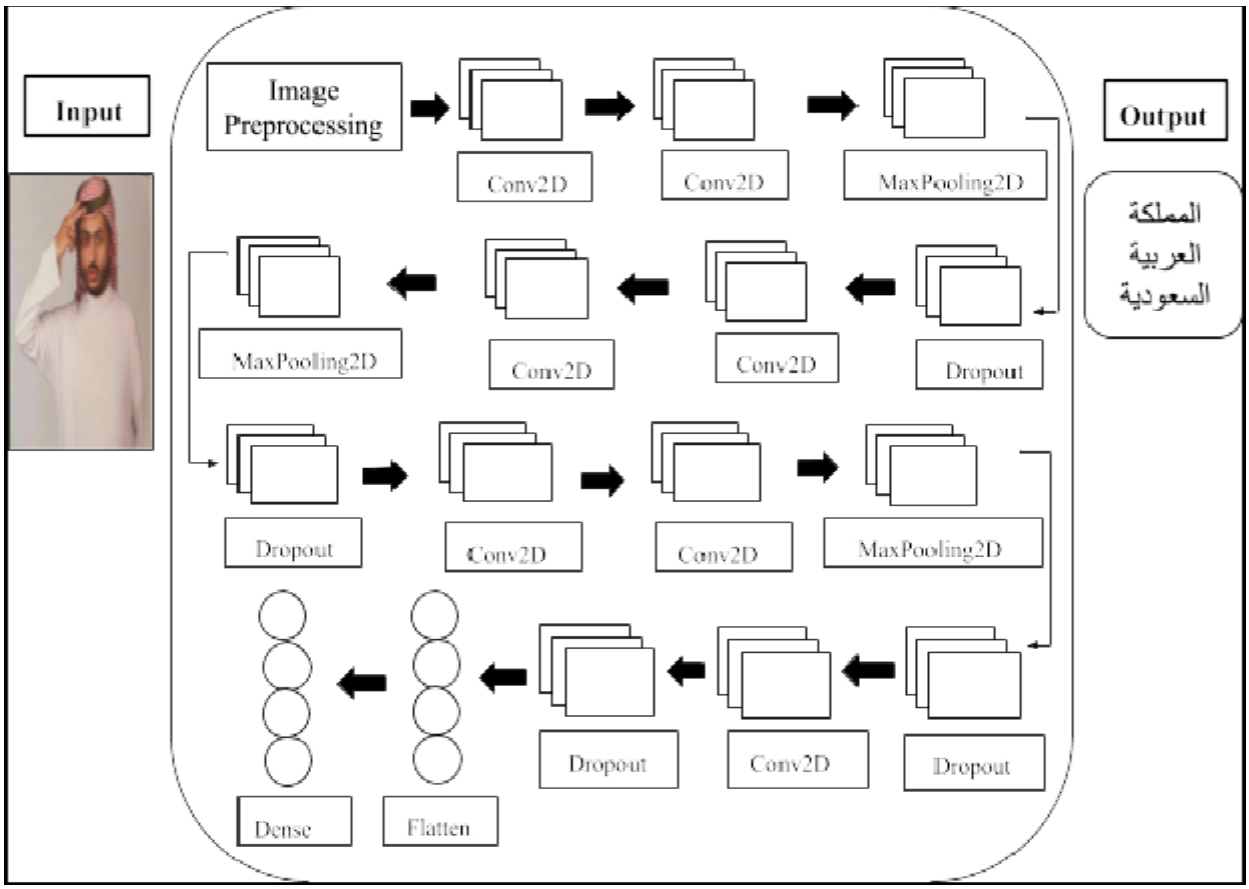

Fig. 4. The CNN model

International Journal of Engineering Research and Technology. ISSN 0974-3154, Volume 13, Number 11 (2020), pp. 3328-3334
 https://dx.doi.org/10.37624/IJERT/13.11.2020.3328-3334

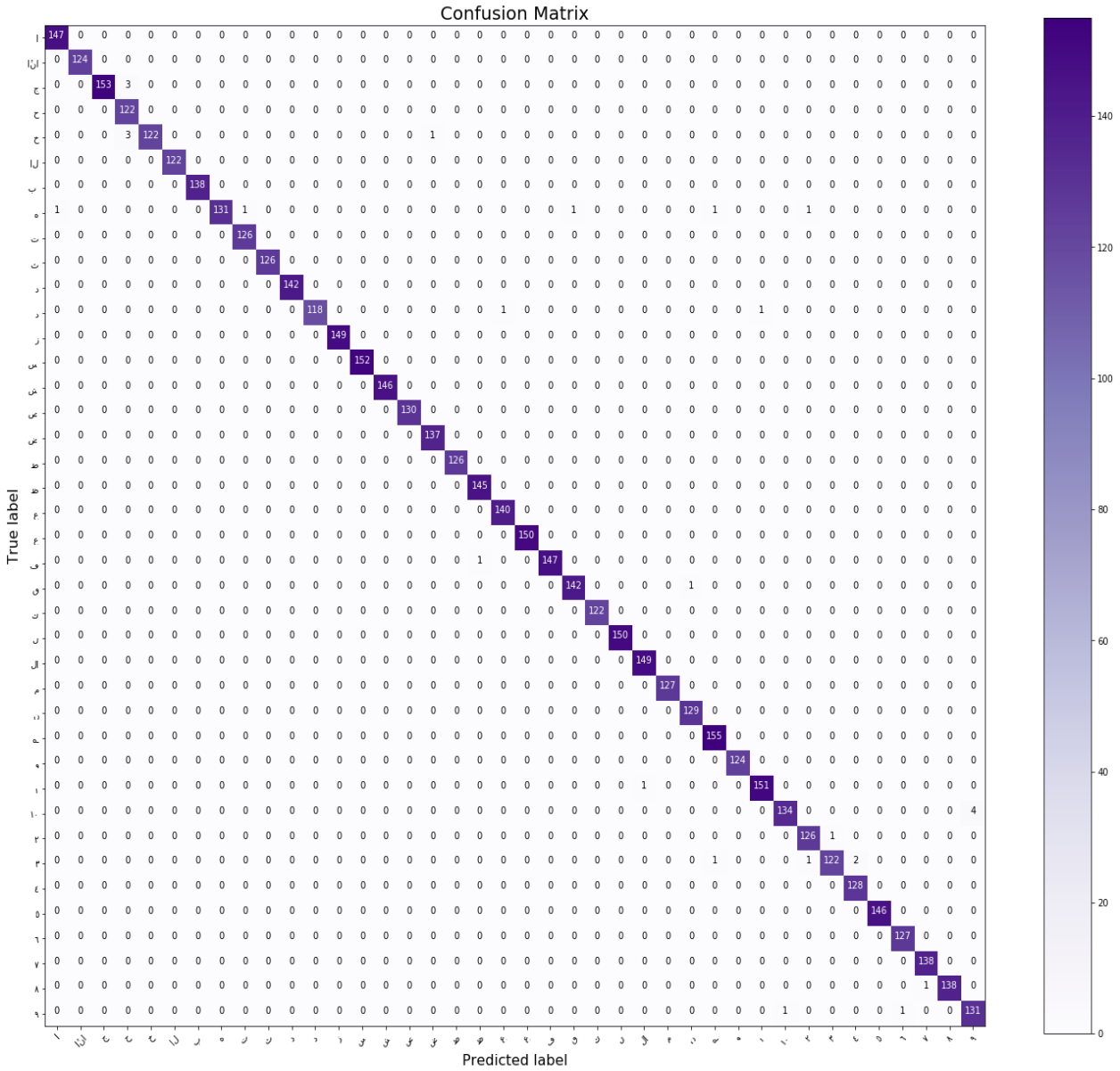

Fig. 5. Confusion Matrix of Testing Images

Table 2. Comparison between the proposed system and the previous studies

Reference	Dataset Description	Size	Training Accuracy	Validation Accuracy
[3]	Word (25)	200 Images	98%	85%
[5]	Alphabets and Number (0-9)	7869 Images	--------	--------
[7]	Alphabets and Numbers (0-9)	900 Images	98.05%	100%
[12]	Alphabets	1147 Images	96.33%	84.68%
[14]	Word (200)	30,000 Video	--------	--------
[16]	Alphabets	900 Images	98%	70-80%
Our System	**Alphabets, Number (1-10) and One Word (40 classes)**	**27,301 Images**	**97.69%**	**99.47%**

International Journal of Engineering Research and Technology. ISSN 0974-3154, Volume 13, Number 11 (2020), pp. 3328-3334
 https://dx.doi.org/10.37624/IJERT/13.11.2020.3328-3334

IV. SYSTEM PROTOTYPE

We have implemented two versions: mobile and desktop versions.

IV.I Mobile App

In the mobile application, we used Flutter with Dart language to program the interface. Flutter is a framework developed by Google that is used to build Android and iOS mobile apps.

The backend was implemented using TensorFlow Lite. Specifically, it was used to integrate the sign recognition model with the user interface. The mobile app interface is shown in Fig. 6, in which the deaf people can open the camera phone and start to take sign images to be translated into texts.

Fig. 7 shows sample signs taken by the app and translated into their corresponding meanings.

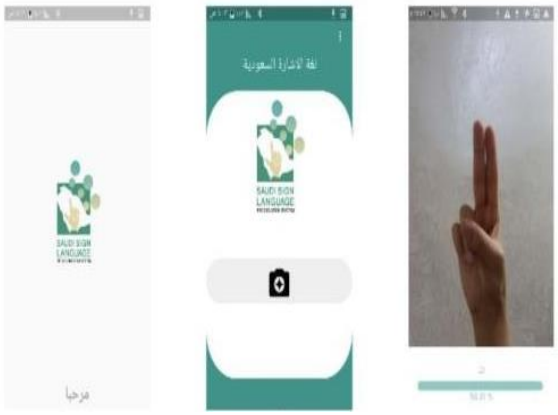

Fig 6. System interface- mobile version

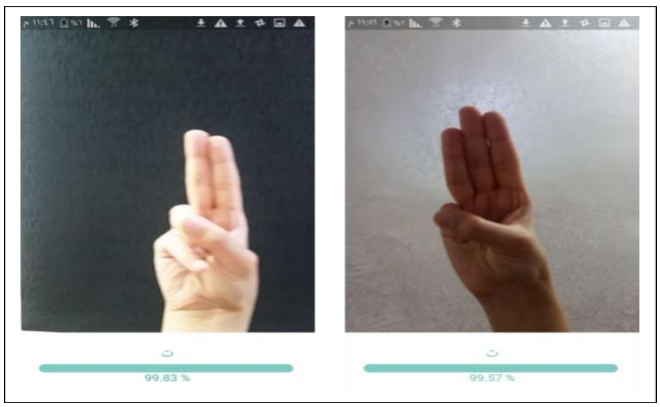

Fig. 7. A sample of a Saudi sign recognized by the developed mobile version

IV.II Desktop Application

For the desktop version, the interface was built using Tkinter, the standard Python's GUI framework. Tkinter is popular due to its simplicity. It is open-source and available under the Python License. Tkinter works on Windows, macOS, and Linux.

The desktop application interface is shown in Fig. 8. As shown in the figure, deaf people can open the camera using the button shown on the screen. Samples of the Saudi signs recognized by the desktop application are shown in Fig. 9 and 10. The signs were taken indoors (Fig. 9) and outdoors (Fig. 10).

Fig. 8. System interface- desktop application version

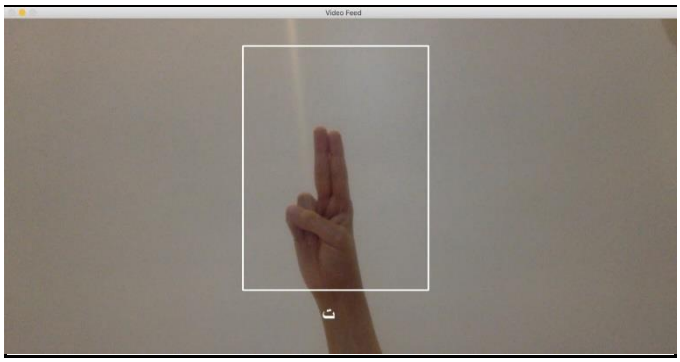

Fig. 9. A sample of a Saudi sign recognized by the desktop application version. The sign was taken indoors

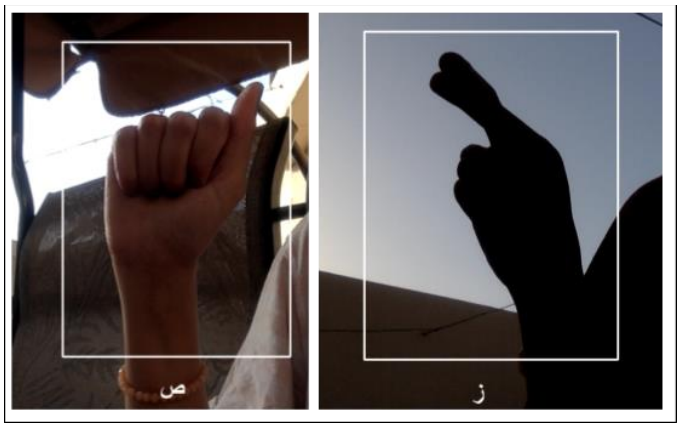

Fig. 10. Samples of Saudi signs are recognized by the desktop application version. The signs were taken outdoors

V. Conclusion

This paper proposed a model and a system for Saudi letters and numbers. To the best of our knowledge, this is the first system meant for Saudi deaf people. This study also constructed a

dataset based on the standard Saudi dictionary. The model was built using convolutional neural networks. The proposed model achieved 97.69% in the training accuracy and 99.47% in the testing accuracy. In the future, we plan to complete the dataset building to include all the signs available in the Saudi sign dictionary. Also, we plan to cover the dynamic signs that are represented as videos rather than images.

REFERENCES

[1] S. Shivashankaraand S. Srinath, "American sign language recognition system: An optimal approach." International Journal of Image, Graphics & Signal Processing, vol. 10, no. 8, 2018.

[2] 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 scenes based on a hierarchical clustering classification method," in2016IEEESecondInternationalConferenceonMulti mediaBigD ata (BigMM). IEEE, 2016, pp. 64–67

[3] 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.

[4] V. Adithya, P. R. Vinod, and U. Gopalakrishnan, "Artificial neural network based method for indian sign language recognition," in 2013 IEEE Conference on Information Communication Technologies, April 2013, pp. 1080–1085

[5] S.Hayani,M.Benaddy,O.ElMeslouhi, and M.Kardouchi, "Arab sign language recognition with convolutional neural networks," in 2019 International Conference Of ComputerScience and Renewable Energies (ICCSR), July 2019, pp. 1–4.

[6] N. B. Ibrahim, M. M. Selim, and H. H. Zayed, "An automatic arabic sign language recognition system (arslrs)," Journal of King Saud University - Computer and Information Sciences, vol. 30, no. 4, pp. 470
– 477, 2018.

[7] M. Taskiran, M. Killioglu, and N. Kahraman, "A real-time system for recognition of american sign language by using deep learning," in the 2018 41st International Conference on Telecommunications and Signal Processing (TSP). IEEE, 2018, pp. 1–5.

[8] 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). IEEE, 2008, pp. 478–481.

[9] R. Al Zuhairi, R. Alghunaim, W. Alshehri, S. Aloqeely, M. Alzaidan, and O. Bchir, "Image based arabic sign language recognition system," International Journal of Advanced Computer Science and Applications (IJACSA), vol. 9, no. 3, 2018.

[10] NR. Albelwi and Y. M. Alginahi, "Real-time arabic sign language (arsl) recognition," in International Conference on Communications and Information Technology, 2012, pp. 497–501.

[11] N.El-Bendary, H. M. Zawbaa, M. S. Daoud, A. E. Hassanien, and K. Nakamatsu, "Arslat: Arabicsignlanguagealphabetstranslator," in2010 International Conference on Computer Information Systems and Industrial Management Applications (CISIM). IEEE, 2010, pp. 590– 595.

[12] M.Hossen, A.Govindaiah, S.Sultana, and A.Bhuiyan, "Bengalisignlanguagerecognition using deep convolutional neural network," in 2018 Joint 7th International Conference on Informatics, Electronics & Vision (ICIEV) and 2018 2nd International Conference on Imaging, Vision & Pattern Recognition (icIVPR). IEEE, 2018, pp. 369– 373.

[13] S. N. Sawant and M. Kumbhar, "Real time sign language recognition using pca," in 2014 IEEE International Conference on Advanced Communications, Control and Computing Technologies. IEEE, 2014, pp. 1412–1415.

[14] G.A.Rao,K.Syamala,P.Kishore,andA.Sastry, "Deep convolutional neural networks for sign language recognition," in the 2018 Conference on Signal Processing And Communication Engineering Systems (SPACES). IEEE, 2018, pp. 194–197.

[15] R. Hartanto and A. Kartikasari, “Android based real-time static indonesian sign language recognition system prototype,” in 2016 8th International Conference on Information Tech-nology and Electrical Engineering (ICITEE). IEEE, 2016, pp. 1–6.

[16] I. Makarov, N. Veldyaykin, M. Chertkov, and A. Pokoev, “Russian sign language dactyl recognition,” in 2019 42nd International Conference on Telecommunications and Signal Processing (TSP). IEEE, 2019, pp. 726–729.