Machine Learning Approach for Gesture Based Arabic Sign Language Recognition for Impaired People

Mahyudin Ritonga (✉ mahyuritonga@gmail.com)

Muhammadiyah University of Sumatera Utara: Universitas Muhammadiyah Sumatera Utara

Rasha M.Abd El-Aziz

Jouf University

Varsha Dr.

Ct. institute of Engineering

Maulik Bader Alazzam

Ajloun National Private University

Fawaz Allassery

Taif University

NagaJyothi A.

Vignan's Institute of Information Technology

Yousef Methkal Abd Algani

College of Sakhnin for Teacher Education

Kiran Bala B.

K Ramakrishnan College of Engineering

Balaji S

Panimalar Engineering College

Research Article

Keywords: Convolutional Neural Network (CNN), Deep Learning, Wearable Sensors, Arabic Sign Language Recognition, cyber security

Posted Date: November 15th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-942259/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Exceptional research activities have been endorsed by the Arabic Sign Language for recognizing gestures and hand signs utilizing the deep learning model. Sign languages refer to the gestures, which are utilized by hearing impaired people for communication. These gestures are complex for understanding by normal people. Due to variation of Arabic Sign Language (ArSL) from one territory to another territory or between countries, the recognition of Arabic Sign Language (ArSL) became an arduous research problem. The recognition of Arabic Sign Language has been learned and implemented utilizing multiple traditional and intelligent approaches and there were only less attempts made for enhancing the process with the help of deep learning networks. The proposed system here encapsulates a Convolutional Neural Network (CNN) based machine learning technique, which utilizes wearable sensors for recognition of the Arabic Sign Language (ArSL). The model suits to all local Arabic gestures, which are used by the hearing-impaired people of the local Arabic community. The proposed system has a reasonable and moderate accuracy. Initially a deep Convolutional network is built for feature extraction, which is extracted from the collected data by the wearable sensors. These sensors are used for recognizing accurately the 30 hand sign letters of the Arabic sign language. DG5-V hand gloves embedded with wearable sensors are used for capturing the hand movements in the dataset. The CNN approach is utilized for the classification purpose. The hand gestures of the Arabic sign language are the input and the vocalized speech is the output of the proposed system. The results achieved a recognition rate of 90%. The proposed system was found highly efficient for translating hand gestures of the Arabic Sign Language into speech and writing.

1. Introduction

The sign language indicates the hand gestures or signs that help a person for expressing the expressions, words, and letters of a particular language. Communication could be made easier with the introduction of systems, which could recognize the hand gestures and signs (Shahin and Sultan 2019). The interaction between machines and humans provides huge benefits for improving the sign language recognition techniques. Building of the sign language datasets is highly important for constructing automated systems. Sign language recognition is a challenging process as it is difficult for representing the primary features such as hand configuration, movement etc. (Escobedo et.al 2019). The approaches used in sign language recognition can be of three types. Hybrid based approach, vision-based approach and sensor-based approach. The difficulties in these approaches arise out of the integration of hardware sensors. Traditional machine learning techniques are employed in computer vision systems, which are used for recognition, detection, and feature extraction, segmentation and preprocessing. These systems had limited datasets that leads to generalization and robustness. The deep machine learning algorithms were found as highly efficient. Convolutional Neural Networks (CNN) is one among the best tools, which is implemented for various recognition processes. The proposed system employs a Convolutional Neural Network (CNN) based machine learning technique using wearable sensors for recognition of sign gestures of the Arabic sign language.

2. Literature Review

(Bencherif et.al 2021) built a new Arabic Sign Language (ArSL) video-based sign database. It was an online sign recognition system, where the inverse efficiency score was used for determining optimal successive frames used in recognition decisions. Best results were gathered in the dependent mode for both static and dynamic signs.

(Hassan et.al 2019) presented an efficient comparison between two recognition methods such as Hidden Markov Model (HMM) and the modified k^{th} Nearest Neighbor (m-KNN), which was contemplated on two unique tool kits. Glove based data sets and ArSL datasets were introduced in their study. It was found that the accuracy of classification, which were got for the collected sign sentences were equivalent to the accuracy got by utilizing the sensor gloves. An inferior difference was observed between the two classification techniques.

(Al-Hammadi et.al 2020) proposed a novel system for hand gesture recognition, which is dynamic with the use of various deep learning architectures, which were used for sequence feature recognition and globalization, global and local feature representations and for hand segmentation. The results of their system indicated that their system outperformed the other sign language recognition approaches by indicating its efficiency.

(Mustafa 2021) reviewed the sign language recognition system, which is contemplated on various classifier approaches. His work focused on the deep learning methods mainly and on the recognition systems of the Arabic sign language with the objective of finding out the best classifier model for sign language recognition. In his review the deep-learning classifiers were found the best model for sign language recognition.

(Kamruzzaman 2020) developed a system, which automatically detects the hand sign letters and converts it into speech of the Arabic language. 90% accuracy was obtained with his proposed system. His system was found efficient in recognizing the Arabic sign letters with the application of the deep learning models. However more advance systems are recommended for improving the accuracy of his system.

3. The Proposed System

The proposed system comprises of the data gloves for getting the data set encapsulated with the sign letters of the Arabic Sign Language (ArSL), Convolutional Neural Network for classification of the extracted features in the data set, evaluation stage, which involves the conversion of hand gestures of the Arabic Sign Language (ArSL) into speech, which is the output of the proposed system. The following figure indicates the block diagram of the proposed system.

Wearable Sensors

Wearable sensors are devices, which are embedded into materials, which could be worn in a body in a comfortable manner. The wearable sensors make use of the measurements, which has direct contact with the human hands. These sensors are in the form of gloves, position tracking instruments etc. In the proposed system, DG5-V Hand data gloves are utilized for capturing the hand movements in the data set. This glove comprise of a software module, microcontroller unit and a communication unit.

ArSL Dataset

The Arabic Sign Language data set comprises of 30 sign letters of the Arabic sign language. This Arabic Sign Language dataset is used for training the proposed system. The 30 sign letters of the Arabic Sign Language (ArSL), which is used for the proposed system is shown below.

CNN Classification

The proposed system makes use of a CNN based machine learning technique. The CNN classifier is used here, which is used for extracting the features of the ArSL dataset through the input image processing utilizing the Convolutional layers, which are designed for producing a features' map. Convolutional Neural Networks (CNN) refers to the system that makes use of Perceptron, machine learning algorithms in the implementation of their functions for data analysis. CNN comes under the artificial neural network (ANN) category. The two important processes of CNN classification are extraction of features and classification of features.

Evaluation

The proposed model is evaluated based on 30 sign letters of the Arabic sign language. The images of these sign letters were used for training the Convolutional Neural Network. The dataset is divided in 80:20 ration with 80 for the training phase and 20 for the testing phase. The results of the proposed system provide an accuracy of 90%, which is high when compared to the previous studies.

4. Materials And Methods

In the proposed system, a wearable sensor-based approach is employed, which makes use of a Convolutional Neural Network (CNN) based machine learning technique for the recognition of the Arabic Sign Language (ArSL). The wearable sensor-based recognition method would process the data gathered from the hand data gloves, which is embedded with the wearable sensor. Cyber gloves, data gloves and power gloves are commonly employed for the recognition of the Arabic Sign Language. In the proposed method, the data gloves are employed. The information regarding the finger bending, hand orientation, movement, rotation and position are provided by these data gloves. Multiple features could be extracted from the acquired data, got from the data gloves. These features could be utilized with a suitable classifier for recognizing the Arabic Sign Language (ArSL). The classifier used for this purpose in the proposed system is the Convolutional Neural Network (CNN) classifier.

4.1 Data Preprocessing

The initial step for building an efficient deep learning model is the data preprocessing step. This step is step is utilized for transforming the raw data into an efficient and useful format. The following figure indicates the flow of processes in the data preprocessing step.

4.1.1 Raw images

The raw images of the proposed system are the hand sign images, which are got using wearable sensors for executing the proposed system. The ways used for capturing these images are by altering the distance and size of the object, with a good focus and quality, by altering lighting conditions and from various angles. The focus for creating these raw images is for creating the dataset for training phase and testing phase. The raw images of the Arabic Sign Language (ArSL), which are used in the proposed system are indicated in figure (1).

4.1.2 Classification of images

The images are classified in the proposed system into 30 letters of the Arabic Alphabet, which comes under 30 categories. One subfolder is created and utilized for storing one category images for implementing the system. All the subfolders that encapsulate the classes are together placed in one important folder called the dataset in the system presented here.

4.1.3 Formatting of images

The hand sign images are unequal usually and have various backgrounds. Thus it is essential for erasing the unnecessary data from these images for fetching the hand section. Then resizing of the extracted images is done with pixel size of 128*128 and then the resized images are converted into RGB.

4.1.4 Dividing images for training and testing

The 30 Arabic alphabets have 125 pictures in each of them. The dataset of these Arabic alphabets are sectioned into two datasets such as the dataset for learning or training and the dataset for testing. These datasets are divided in the ratio 80:20; where 80 represents the training dataset and 20 represents the testing dataset. The images of the training set are 100 in number and the number of images in the testing set is 25 in number, which are allotted for every handsign.

4.1.5 Augmentation

The data in real-time is usually unpredictable and inconsistent. This is because of the existence of many transformations such as moving, rotating etc. The image augmentation is utilized for enhancing the performance of the deep network. Artificial images are created with multiple processing methods like rotation, shear, flips, shifts etc. The proposed system's images are rotated randomly up to 360 degrees with the help of the image augmentation technique. Fewer images are randomly sheared in the range of 0.2 degrees and were horizontally flipped. The following table indicates the accuracy and loss without and with augmentation of images.

Table 1
Accuracy and Loss without and with Augmentation

Size of the batch	Image Augmentation	Loss	Accuracy
64	True	0.58	85%
	False	0.83	83%
128	True	0.51	90%
	False	0.54	86%

4.1.6 File creation for recording images and generation of training dataset

There is a need for the creation of a list, which encapsulates all the images that are placed in a unique folder for getting the information regarding file name and labels. This would be useful for generating the training and testing data needed for the proposed system implementation.

4.2 Architecture of ArSL using CNN

The following figure shows the architecture of the Arabic Sign Language (ArSL) recognition with the help of the Convolutional Neural Network (CNN).

The CNN classification is mostly employed in the computer vision field. It assists mainly in the recognition and classification of images. The two eminent components of CNN are extraction of features and classification of features. Every component possesses their own features, which has to be discovered. The components of the CNN architecture, which is employed in the proposed system, are as follows;

4.2.1 Extraction of features

Multiple building blocks are encapsulated within the Convolutional Neural Network (CNN). The convolution layer is the important building block in the CNN network. This layer indicates the mathematical representation of functions for yielding another third function. The convolution has to be done within the input with the help of a kernel or a filter to create a feature map. The implementation of the convolution comprises of sliding of every filter with a suitable input. Matrix multiplication is performed at every place, and the output is added to the feature map. Each of the images is converted to a 3D matrix with a specified depth, height and width. The depth is considered as the dimension as the image comprises of the color channels. Multiple convolutions are done on the input dataset various filters that forms various feature maps. The convolution layer's output is got with the combination of the various feature maps. An activation function is performed in the output for performing a non-linear output. The activation function used here is the Rectified Linear Unit (ReLU) that functions with the function $f(k) = \max(0, k)$. At the threshold condition, the activation function = 0. This activation function is highly reliable and fragile in its functioning. The learning rate can be adjusted to overcome this drawback. A step, which is frequently performed by the convolution filter function, is the stride. The stride size is equal to 1 usually. This indicates the moving of the convolution filter pixel by pixel. When the stride size is increased, then the convolution filter moves over the high interval input and thus a small overlapping occurs in the cells. The input size is always greater than the size of the feature map and hence the shrinking of the feature map should be stopped. Thus, a padding technique is used here. Then the zero valued pixel layers must be added to the inputs for preventing the inputs of the feature map. The padding process assists in retaining the spatial dimension as constant post convolution such that the size of the stride and the kernel is equal to the input. Thus, this improves the entire system performance. Three essential parameters require adjustments within the CNN for modifying the convolution layer's behavior. These three parameters are padding, stride and size of the filter. The output size of the convolution layer is calculated as follows;

$$\text{Output}_{\text{size}} = (2 * \text{padding}_{\text{amount}} - \text{filter}_{\text{size}} + \text{input}_{\text{size}}) / \text{Stride}_{\text{size}} \dots \dots \dots (1)$$

Here, $\text{Output}_{\text{size}}$ indicates the output size of the convolution layer; $\text{input}_{\text{size}}$ indicates size of the input image, $\text{filter}_{\text{size}}$ indicates the size of the filter, $\text{padding}_{\text{amount}}$ indicates the amount of padding and $\text{stride}_{\text{size}}$ indicates size of the stride.

4.2.2 Pooling layer

In between the convolution layer, the pooling layer is included. But, the important purpose of the pooling layer is to decrease constantly the dimensions and computations having less parameter. It helps in regulating the training time and regulating the over-fitting of the layers. Max pooling is a commonly employed pooling technique and thus it is included in the proposed system as well. It makes use of the largest value in the windows and thus minimizes the feature map size and retains the needed information. There is a requirement for specifying the sizes of the windows earlier for determining the output size of the pooling layer. The output size of the pooling layer is calculated as follows;

$$\text{Output}_{\text{size}} = 1 + (\text{input}_{\text{size}} - \text{filter}_{\text{size}}) / \text{stride}_{\text{size}} \dots \dots \dots (2)$$

In all the situations, invariance in the translation is offered by the pooling layer, which represents the identity of a specific object with respect to its visibility within the frame.

4.2.3 Classification of features

The next eminent CNN component is the classification. The classification part comprises of fewer layers that are connected with each other. These are called as fully connected (FC) layers. The neurons within these layers have their own connection to the previous layers' activations. The mapping between the output and input is assisted with the help of the FC layer. The function of this layer is similar to that of a neural network. The fully connected layer accepts only the data having one dimension. For converting the 3D into one dimensional data, Python's flattening function is utilized in the proposed system.

5. Results And Discussion

The testing of the proposed system is done with the usage of two convolution layers. Then each convolution layer is followed by 2×2 maximum pooling layers. In the first layer, different structures are exhibited by the convolution layers. 32 kernels are there in this first layer. 64 kernels are there in the second layer. But, the kernel size in these two layers is equal to 3×3. Each pair of pooling and convolution layers are verified with two various drop out values, which are regularized. These values are 25% for the convolution layer and 50% for the pooling layer. This helps in erasing one input out of four inputs in the convolution layer and two inputs of the pooling layer. Softmax and ReLu are used by FC layer's activation function for determining whether there is firing of the neuron or not. The training was done for 100 epochs with the help of an RMS prop optimizer, which is having a cost function and is contemplated on the categorical cross entropy as there is a convergence before the 100 epochs such that the system has the storage of the weights, which is to be used in the following phase. In the testing phase optimistic test accuracy was presented by the proposed system having less loss rates. These loss rates could be further reduced utilizing augmented images with the accuracy as kept constant. Every new image is processed in the testing phase, before they are employed in the proposed system. The vector size that is got from the proposed system is equal to 10. Here, one among the ten values is equal to one and the remaining value indicates the data's predicted class value. The system is then connected with the signature part that involves the conversion of hand signs into speech of the Arabic language. There are two phases of this process. The initial phase is the translation phase that involves the translation of hand signs into letters in Arabic. This is done with Google Translator (API). Better results are achieved in the system with the built up of different combinations of the convolution and pooling layers. For the convolution layers having drop-out values of 50% and 25% having 64 and 32 kernels, 83% accuracy is attained and the validation loss is 0.84. The convolution layers of batch size 128 and 64 are also tested in the proposed system. The accuracy rate could be raised with the help of Image augmentation. For the convolution layer having batch size 128, the validation loss was found to be 0.50. These results are indicated in Table 1. It was observed that additional Convolutional layers were not added in the system. The confusion matrix indicates the proposed system's performance based on right or wrong classification, which is developed here. The confusion matrix in the presence and absence of image augmentation is indicated in Table 2 and 3 respectively.

Table 2
Presence of Image Augmentation

PR	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
AC																										
1	18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
2	0	20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	1	18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	2	1	0	13	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	1	0	0	0	0	0
5	0	0	0	0	18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	1	19	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	18	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0	19	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0
11	0	0	0	0	0	0	0	0	0	0	20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
12	0	0	0	0	0	0	0	0	0	0	0	20	0	0	0	0	0	0	0	0	0	0	0	0	0	0
13	0	0	0	0	0	0	0	0	0	0	0	0	18	0	0	0	0	0	0	0	1	0	0	0	0	0
14	0	0	0	0	0	0	0	0	0	0	0	0	0	20	0	0	0	0	0	0	0	0	0	0	0	0
15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	16	0	0	0	0	4	0	0	0	0	0	0
16	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	12	6	0	0	0	0	0	0	0	0	0
17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	20	0	0	0	0	0	0	0	0	0
18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	10	4	0	0	0	0	0	0	0
19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	19	0	0	0	0	0	0	0
20	0	2	0	0	0	1	0	0	0	0	0	0	0	1	5	0	1	0	1	9	0	0	0	0	0	0
21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	18	0	0	0	0	0
22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	20	0	0	0	0
23	0	0	0	0	0	0	0	2	0	0	0	0	1	0	0	0	0	0	0	0	0	0	17	0	0	0
24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	18	0	0
25	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	4	0	0	0	0	0	0	0	10	0
26	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
27	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
28	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
29	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
30	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0

Table 3
Absence of Image Augmentation

PR	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
AC																										
1	18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
4	0	0	0	14	0	0	0	0	0	0	0	0	7	0	0	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	18	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	1	18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	19	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	1	0	1	0	19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
11	0	0	0	0	0	0	0	0	0	0	20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
12	0	0	0	0	0	0	0	0	0	0	0	20	0	0	0	0	0	0	0	0	0	0	0	0	0	0
13	0	0	0	0	0	0	0	0	0	0	0	0	18	0	0	0	0	0	0	0	0	0	0	0	0	0
14	0	0	0	0	0	0	0	0	0	0	0	0	0	20	0	0	0	0	0	0	0	0	0	0	0	0
15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	17	0	0	0	0	0	0	0	0	0	0	0
16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	17	3	0	0	0	0	0	0	0	0	0
17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	20	0	0	0	0	0	0	0	0	0
18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	14	6	0	0	0	0	0	0	0
19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	18	0	0	0	0	0	0	0
20	0	0	0	0	0	0	0	0	0	0	0	0	0	1	4	0	1	0	0	10	0	0	0	0	0	0
21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	17	0	0	0	0	0
22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	20	0	0	0	0
23	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	1	0	0	0	18	0	0	0
24	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	1	0	0	0	0	0	0	18	0	0
25	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	3	0	0	0	0	0	0	0	14	0
26	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1
27	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
28	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
29	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
30	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

6. Conclusion

Thus, the main purpose of this proposed research is to develop a wearable sensor based CNN model for recognition of sign letters of the Arabic Sign Language (ArSL) and to reduce the implications caused by the sign languages. The proposed model could also be employed in the recognition of hand gestures, which are used for computer-human interactions. The proposed model is in its earlier development stage and still it is highly efficient in the proper identification of gestures and hand digits, which are converted into Arabic speech with accuracy higher than 90%. The quality and accuracy of the proposed model could further be improved by using highly advanced recognition devices such as Xbox Connect or Leap Motion. Also the data set size could be increased further in the future research studies. The output of the proposed system is the speech in Arabic language, which is got with the recognition of the Arabic sign language. Further, the system proposed here would be a best hearing remedy for the hearing-impaired people.

Declarations

Acknowledgement

We deeply acknowledge Taif University for supporting this study through Taif University Researchers Supporting Project Number (TURSP – 2020/150), Taif University, Taif, Saudi Arabia.

Conflict of Interest

This paper has not communicated anywhere till this moment, now only it is communicated to your esteemed journal for the publication with the knowledge of all co-authors.

Ethical approval

This article does not contain any studies with human participants or animals performed by any of the authors.

Informed Consent

All authors have seen the manuscript and approved to submit it to the journal.

Author contributions

All authors have seen the manuscript and approved to submit it to the journal.

References

1. Shahrin and Sultan Almotairi. *Automated Arabic Sign Language Recognition System Based on Deep Transfer Learning*. no. 10, Oct. 2019, pp. 144–52.
2. Bencherif, Mohamed A., et al. "Arabic Sign Language Recognition System Using 2D Hands and Body Skeleton Data." *IEEE Access*, vol. 9, 2021, pp. 59612–27. *IEEE Xplore*, <https://doi.org/10.1109/ACCESS.2021.3069714>.
3. Hassan, Mohamed, et al. "Multiple Proposals for Continuous Arabic Sign Language Recognition." *Sensing and Imaging*, vol. 20, no. 1, Jan. 2019, p. 4. *Springer Link*, <https://doi.org/10.1007/s11220-019-0225-3>.
4. Al-Hammadi, Muneer, et al. "Deep Learning-Based Approach for Sign Language Gesture Recognition With Efficient Hand Gesture Representation." *IEEE Access*, vol. 8, 2020, pp. 192527–42. *IEEE Xplore*, <https://doi.org/10.1109/ACCESS.2020.3032140>.
5. Mustafa, Mohammed. "A Study on Arabic Sign Language Recognition for Differently Abled Using Advanced Machine Learning Classifiers." *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, no. 3, Mar. 2021, pp. 4101–15. *Springer Link*, <https://doi.org/10.1007/s12652-020-01790-w>.
6. Kamruzzaman, M. M. "Arabic Sign Language Recognition and Generating Arabic Speech Using Convolutional Neural Network." *Wireless Communications and Mobile Computing*, vol. 2020, May 2020, p. e3685614. *hindawi.com*, <https://doi.org/10.1155/2020/3685614>.
7. "A Deep Convolutional Neural Network Approach for Static Hand Gesture Recognition." *Procedia Computer Science*, vol. 171, Jan. 2020, pp. 2353–61. *sciencedirect.com*, <https://doi.org/10.1016/j.procs.2020.04.255>.
8. Saggio, Giovanni, et al. "Sign Language Recognition Using Wearable Electronics: Implementing k-Nearest Neighbors with Dynamic Time Warping and Convolutional Neural Network Algorithms." *Sensors (Basel, Switzerland)*, vol. 20, no. 14, July 2020, p. 3879. *PubMed Central*, <https://doi.org/10.3390/s20143879>.
9. Can, Celal, et al. "A Deep Convolutional Neural Network Model for Hand Gesture Recognition in 2D Near-Infrared Images." *Biomedical Physics & Engineering Express*, vol. 7, no. 5, Sept. 2021, p. 055005. *org (Crossref)*, <https://doi.org/10.1088/2057-1976/ac0d91>.
10. Kapuscinski, Tomasz, and Marian Wysocki. "Recognition of Signed Expressions in an Experimental System Supporting Deaf Clients in the City Office." *Sensors*, vol. 20, no. 8, Jan. 2020, p. 2190. *mdpi.com*, <https://doi.org/10.3390/s20082190>.
11. Rao, G. Anantha, et al. "Deep Convolutional Neural Networks for Sign Language Recognition." *2018 Conference on Signal Processing And Communication Engineering Systems (SPACES)*, 2018, pp. 194–97. *IEEE Xplore*, <https://doi.org/10.1109/SPACES.2018.8306344>.
12. Shahriar, Shadman, et al. "Real-Time American Sign Language Recognition Using Skin Segmentation and Image Category Classification with Convolutional Neural Network and Deep Learning." *TENCON 2018 - 2018 IEEE Region 10 Conference*, 2018, pp. 1168–71. *IEEE Xplore*, <https://doi.org/10.1109/TENCON.2018.8650524>.
13. Escobedo, Edwin, et al. "Dynamic Sign Language Recognition Based on Convolutional Neural Networks and Texture Maps." *2019 32nd SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*, 2019, pp. 265–72. *IEEE Xplore*, <https://doi.org/10.1109/SIBGRAPI.2019.00043>.
14. Hwang, Kyo-Min, and Sang-Chul Kim. "A Study of CNN-Based Human Behavior Recognition with Channel State Information." *2021 International Conference on Information Networking (ICOIN)*, 2021, pp. 749–51. *IEEE Xplore*, <https://doi.org/10.1109/ICOIN50884.2021.9333879>.
15. Makantasis, Konstantinos, et al. "Deep Learning Based Human Behavior Recognition in Industrial Workflows." *2016 IEEE International Conference on Image Processing (ICIP)*, 2016, pp. 1609–13. *IEEE Xplore*, <https://doi.org/10.1109/ICIP.2016.7532630>.
16. "Human Action Recognition Using 3D Convolutional Neural Networks with 3D Motion Cuboids in Surveillance Videos." *Procedia Computer Science*, vol. 133, Jan. 2018, pp. 471–77. *sciencedirect.com*, <https://doi.org/10.1016/j.procs.2018.07.059>.

Figures

Figure 1

Block Diagram of the Proposed System

Figure 2

ArSL dataset (About Ella Hassanein 2010)

Figure 3
Data Preprocessing-Flow Diagram

Figure 4
Architecture of ArSL Recognition with CNN