Glove-Based Classification of Hand Gestures for Arabic Sign Language Using Faster-CNN

Ahmed M. D. E. Hassanein, Sarah H. A. Mohamed, and Kamran Pedram

***Abstract* — Recently, American Sign Language has been widely researched to help disabled people to communicate with others. However; the Arabic Sign Language "ASL" has received much less attention. This paper has proposed a smart glove which has been designed using flex sensors to collect a dataset about hand gestures applying ASL. The dataset is composed of resistance and voltage measurements for the bending of the fingers to represent alpha-numeric characters. The measurements are manipulated using normalization and zero referencing methods to create the dataset. A Convolutional Neural Network 'CNN' composed of twenty-one layers is proposed. The dataset is used to train the CNN, and the Accuracy and Loss parameters are used to characterize its success. The dataset is classified with an average success rate of 95% based on the classification accuracy. Loss has decreased from 3 to less than 0.5. The proposed CNN layers have classified ASL characters with a reasonable degree of accuracy.**

***Key words* — Arabic Sign Language, Batch Normalization Layer, Convolution Neural Network, reLU Layer.**

I. Introduction

A lot of recent research has been directed towards improving the living conditions of disabled people. Among the topics researched, the problem of lack of communication between disabled people and the world around them is scrutinized. There are several ways that disabled people can express themselves such as hand gestures. Several sets of hand gestures were designed to facilitate the communication such as American Sign Language and Arabic Sign Language. By doing an intensive literature review, in this section, the efforts made to design systems capable of understanding American Sign Language-based hand gestures are described. Hand gestures include using muscles to cause movement of the fingers and the hand to represent certain signs [1]. A research paper proposed a system that is built on detecting electric signals that are generated from the movement of the muscles and their inertia [1]. The authors use the learning reinforcement approach to build a hand gesture recognition system [1]. The system successfully classify human gestures and outperformed others described in recent research [1]. The system is then used to guide the movement of two kinds of robots with different degrees of freedom [1].

Cabrera *et al*. propose an electronic glove to make a human computer interaction system [2]. The system uses a sensor to show degree of bending of each finger and an accelerometer to measure the position of the hand in three axes [2]. All information is used to generate signals to be able to classify them into information that need to be communicated. Signals are generated using American Sign Language and then the measurements are processed to predict the sign [2].

Abhishek *et al*. used a system of touch sensors to recognize the American Sign Language [3]. The system is designed to facilitate communication between deaf people and others. The system achieves classification accuracy of 92% [3].

Zanghieri *et al*. use special surfaces to detect electrical signals coming out of the human muscles [4]. The aim is to use the surfaces to recognize hand gestures and classify them. Several algorithms have been used such as 1d-CNN which is based on 2d-CNN [4]. Training the users more than one time has proven to help achieve better successful recognition results. The RBF kernel Supported Vector Machine algorithm has given the highest accuracy [4]. Training for five days has shown the best classification results with accuracy 75.9% [4]. Also, two body position training for each gesture has contributed in increasing the classification accuracy to be 81.2% [4].

Bello *et al*. present CaptAinGlove system to recognize hand gestures [5]. The system is textile based and it's a real time one that uses Convolutional Neural Network and Hierarchical Multimodal Fusion to classify readings into nine classes of signs [5]. The authors achieve an accuracy of 80% for the training phase and an accuracy of 67% for the testing phase [5].

Al-Saedi *et al*. survey the research work done in recent years towards accurate recognition of human gestures. The authors divide the most common type of gestures to be made through the face or through the hands [6]. They conclude that the algorithms used in gesture classification are mainly Hidden Markov Model, Condensation algorithm, Fuzzy Clustering algorithm, Artificial Neural Networks, Finite-State Machine and Histogram based feature [6].

In this paper, the methodology used to obtain our results is illustrated in section II. In section III, the measurements obtained using our hardware and the creation of the needed dataset is explained. The theoretical background for our proposed neural network is explained in section IV. The results achieved for the proposed classification algorithm is discussed in section V. Finally, conclusion is derived in section VI.

Submitted on August 03, 2023.
Published on September 29, 2023.
A. M. D. E. Hassanein, Systems and Information Department, Engineering and Renewable Energy Research Institute, National Research Centre (NRC), Egypt.
(e-mail: ahmed.diaa.hassanein@gmail.com)
S. H. A. Mohamed, Electrical Communication and Electronics Systems Dept, School of Engineering, October University for Modern Sciences and Arts (MSA), Egypt.
(e-mail: sara.hesham7@msa.edu.eg)
K. Pedram, Faculty of Engineering and Science, University of Greenwich (Medway Campus), United Kingdom.
(e-mail: Kamran.Pedram@greenwich.ac.uk)

II. Methodology

The hardware implementation which is used to collect our dataset is illustrated in subsection A. The layers of the proposed CNN network which is used to classify our dataset are described in subsection B.

A. Hardware Implementation

A smart glove was designed using flex sensors to take our measurements. The glove produces voltage values for the bending of each of the five fingers of a hand.

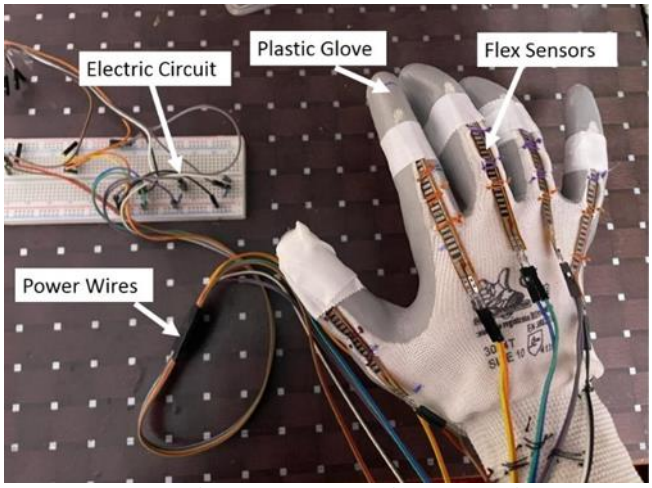

Fig. 1. An image of the implementation of a glove with flex sensors.

Fig.1 shows a plastic glove with flex sensors attached to each of the five fingers. The wires to provide electrical power for the functioning of the sensors and the electric circuit of a potential divider are shown. The flex sensors transfer the bending of the fingers to a change in the resistance. According to the datasheet of the sensor, the more the bending angle increases the higher the resistance of the flex sensor [7]. The resistance of the sensors generated by the bending of the fingers is to be transferred to voltage using the potential divider electric circuit shown in Fig. 1.

B. Neural Network

A CNN network that is composed of twenty-one layers is used. The first layer is used to input the signals to the network. Five layers are used to perform 3×3 convolutions 8 times, 16, 32, 64, and 128 times, respectively. Each of the five convolutional layers is followed by a batch normalization layer and a reLU layer. Two layers of 2×2 Max pooling layers with stride 2×2 are used. Finally, one Fully Connected layer, one Softmax layer and one Classification layer are applied. The batch normalization creates a homogenous range of values in the resulting matrices which leads to a more stable convergence to final results [8]. The reLU layers show generally a higher tendency to converge to results than other functions such as the Sigmoid function [9]. The Max pooling layers decrease the size of the resulting matrices from them which help to decrease complexity of calculations [10]. All three namely Batch Normaliztion, ReLU and Max pooling layers help to make the proposed Neural Network faster. The theoretical background for some of the layers mentioned above is explained in section IV.

TABLE I: The Parameters which are Used in the Carried Out Calculations

Optimizer	SGDM
Learning Rate Initial Value:	0.2
Minimum Batch Size:	20
Maximum Number of Epochs:	100
Learning Rate Drop Period:	10
Learning Rate Drop Factor:	0.2

The parameters used in our calculations are shown in Table I. In our results, the Stochastic Gradient Descent with momentum (SGDM) optimizer is used. It has the benefit of using the all training dataset to update the model's parameters so it gives high accuracy [11]. However; when the number of entries in the dataset is huge the optimization becomes computationally expensive which can affect negatively the speed of obtaining results [11]. The SGDM is very sensitive to the input data due to its accuracy which can lead to instability in obtaining the results [11]. In this paper, the dataset used is relatively not large so the SGDM optimizer leads to the best results as shown later. The initial value of the learning rate is 0.2. The minimum batch size is set to be 20. The maximum number of epochs used is 100. The drop period of the learning rate is set to be 10. The drop factor of the learning rate is set to be 0.2. Values of the parameters are selected based on trial and error until the best accuracies and losses are achieved. The calculations are carried out in Matlab2018a and the graphs are created in Excel sheets.

III. Measurements and Dataset

In this section, the measurements obtained using the hardware implementation and the dataset using these measurements are discussed.

A. Measurements

Few volunteers are asked to wear the glove shown in Fig. 1 and move their fingers to represent nine different hand gestures. Five hand gestures are presented according to the Arabic Sign Language convention [12] which are Dhad, Thaa, Miem, Yaa, Sien. Four hand gestures are adapted from the Arabic Sign Language Convention to the functioning of our glove which are Seven, Laa, Eight and Noon.

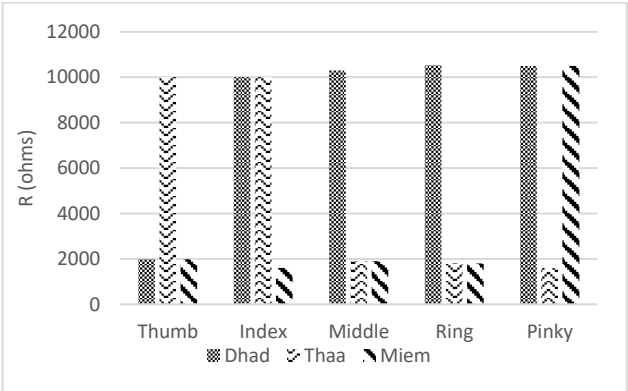

Fig. 2. The resistance values for the three Arabic alphabet: Dhad, Thaa and Miem.

As shown in fig. 2, the resistance measurements for the Arabic letters: Dhad, Thaa and Miem are shown. The Dhad letter is represented by low resistance value for the Thumb

finger but high resistance values for the Index, Middle, Ring and Pinky fingers. The Thaa letter is represented by high resistance values for the Thumb and Index fingers but low resistance values for the Middle, Ring and Pinky fingers. The Miem letter is represented by low resistance value for the Thumb, Index, Middle and Ring fingers but high resistance value for the Pinky finger.

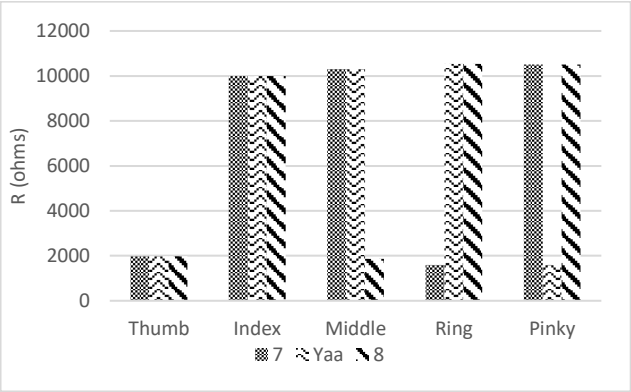

Fig. 3. The resistance values for the three Alpha-numeric Arabic letters: Seven, Yaa and Eight.

As noted in Fig. 3, measurements for the resistance values of the alpha numerical letters: Seven, Ya, and Eight are shown. The Seven number is represented by low resistance values for the Thumb and Ring fingers but high resistance values for the Index, Middle, and Pinky fingers. The Yaa letter is represented by low resistance values for the Thumb and Pinky fingers but high resistance values for the Middle, Index, and Ring fingers. The Eight number is represented by a low resistance value for the Thumb and Middle fingers but high resistance values for the Index, Ring, and Pinky fingers.

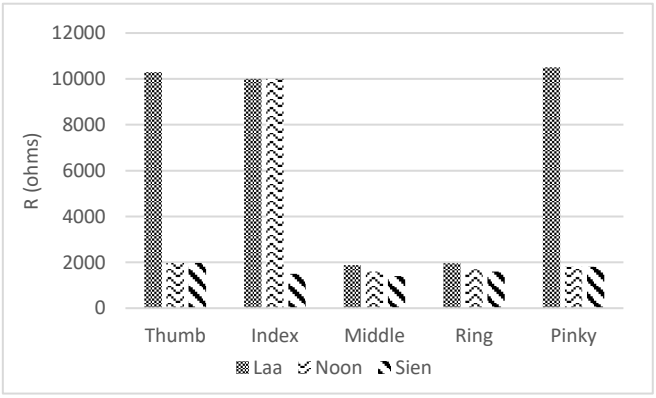

Fig. 4. The resistance values for the three Arabic alphabet: Laa, Noon and Sien.

As shown in Fig. 4, the resistance measurements for the Arabic letters: Laa, Noon and Sien are shown. The Laa letter is represented by high resistance values for the Thumb, Index and Pinky fingers but low resistance values for the Middle and Ring fingers. The Noon letter is represented by low resistance values for the Thumb, Middle, Ring and Pinky fingers but high resistance values for the Index fingers. The Sien letter is represented by low resistance values for the Thumb, Index, Middle and Ring and Pinky fingers but high resistance value for none.

B. Dataset Creation

There are nine alpha numeric Arabic letters used in this paper and the total number of collected measurements from the volunteers are thirty six. The representation of each letter is obtained once using resistance values in kilo ohms and another using voltage values in volts.

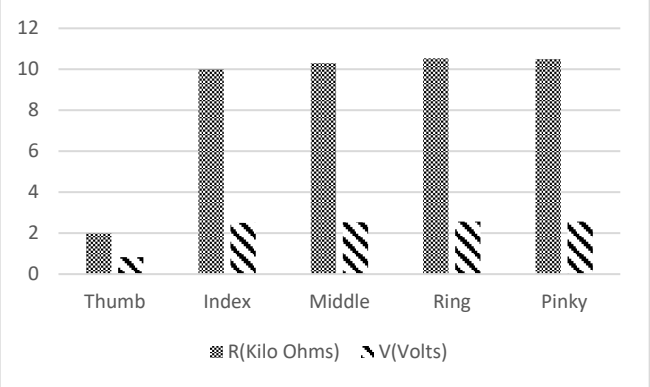

Fig. 5: Two electrical representations for the measurements taken from the flex sensor which are resistance and voltage.

As shown in Fig. 5, the Dhad letter as an example is represented in resistance values and voltage values. It can be seen that both representation have the same pattern. The Thumb finger has the lowest value of resistance and voltage. While, the Index, Middle, Ring, and Pinky fingers have the highest values of resistances and voltages relative to the Thumb finger.

Data augmentation techniques are used to increase the number of inputs in the dataset. Each voltage and resistance representation is calculated in three formats which are Actual, Normalized, and Zero Referenced.

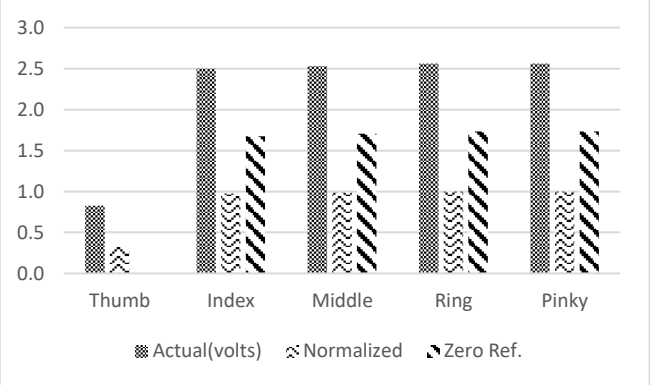

Fig. 6. The different formats of the voltage values which are Actual measurements in volts, Normalized and Zero referenced are plotted.

As shown in Fig. 6, the three formats for the Dhad letter as an example are shown. The Actual format represents the real readings obtained from the glove. The Normalized format represents the Actual readings from the five fingers divided by the maximum value. The resulting Normalized values have a maximum of one. The Zero Referenced (Zero Ref.) format represents the Actual readings of the five fingers subtracted from them the minimum value. The resulting Zero Ref. values have a minimum of zero as shown in Fig. 6.

IV. Theoretical Derivations

In this section, the equations used in our calculations are explained. For the convolution layers, Glorot initializer is

used to increase the stability of the training process and decrease the consumed processing time of the neural network [13]. Convolution equation between two signals A and B is [13]:

$$C(j,k)=\sum_{p}\sum_{q}A(p,q)B(j-p+1,k-q+1) \quad (1)$$

p and q are variables for the signals $A(p,q)$ and $B(j-p+1,k-q+1)$, respectively.

For the Batch Normalization layers, the following equation is used to obtain normalized values x'_i [8]:

$$x'_i=\frac{x_i-\mu_B}{\sqrt{\vartheta_B^2+\in}} \quad (2)$$

where x_i is the value to be normalized, μ_B is the mean of all values, ϑ_B^2 is the variance of all values and $\in$ is a constant used to stabilize the calculations when the variance is very small. The following equation is used to stabilize the calculations [8]:

$$y_i=\gamma x'_i+\beta \quad (3)$$

y_i is a variable that is calculated to avoid very small means and/or variance, γ is scaling factor and β is an offset factor. γ and β are learnable parameters that are to be decided during training.

For the Maxpooling layers, the filter has size of 2 by 2 and a stride of size 2 in the horizontal direction by 2 in the vertical direction. The max pooling function takes the maximum value of each selected filter as it scans the whole input matrix [10].

For the reLU layers, the negative values are replaced by zeros in all the signals. The following function is used [9]:

$$F(x)=\begin{cases}x, x\geq 0\\ 0, x<0\end{cases} \quad (4)$$

which replaces all negative values by zero [9].

For the Softmax layer, the Softmax function applies to multiclass problems [14] as is the case here. The function is [14]:

$$y(z)_i=\frac{\exp(z_i)}{\sum_j \exp(z_j)} \quad (5)$$

where $\sum y=1$ and $0<y<1$.

The Softmax function replaces a range of values with a probability distribution proportional to an exponential function [14]. It is used to get the probability of belonging to each class [14]. The class with the highest probability will be the chosen one.

V. Results

The hand gestures of the Arabic Sign Language are fed to the CNN described in section II. The results are shown in this section. The Accuracy and Loss parameters are used to describe the performance of the network.

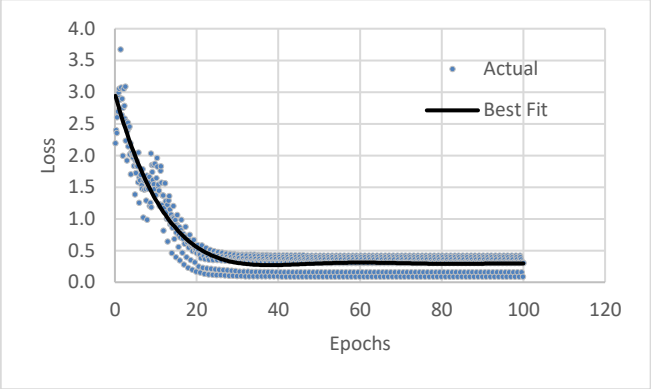

Fig. 7. The Loss in the CNN network and the best fit for it are plotted.

In Fig. 7, the Loss in the CNN network is plotted against the number of epochs. The gray points represent the results of the Loss calculations using the proposed CNN layers. The black line represents a best fit to the Loss points in gray color. It is shown that the Loss drops from 3.5 to 0.5 during the first 20 epochs. It can be seen that after 20 epochs the network Loss is below 0.5. The trend of the points as shown from the best fit is to decrease in value which is very good. The Loss best fit almost saturates at a value below 0.5.

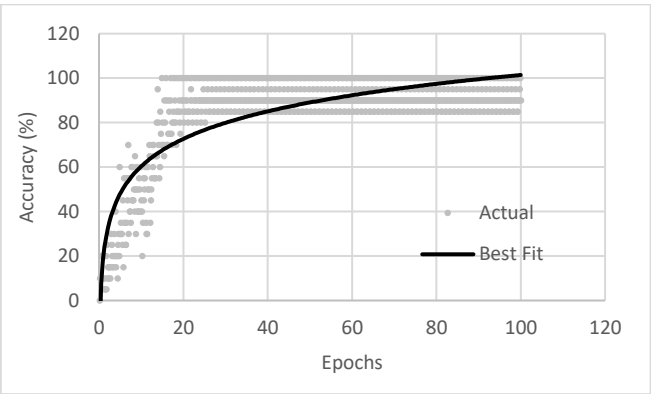

Fig. 8. The Accuracy in the CNN network and the best fit for it are plotted.

In Fig. 8, the percentage of Accuracy achieved in the CNN network is plotted against the number of epochs. The gray points represent the results of the Accuracy calculations using the proposed CNN layers. The black line represents a best fit to the Accuracy points in gray color. The Accuracy increases exponentially as the number of epochs increase. It is shown that the Accuracy increases from 0 to 80% during the first 30 epochs. It can be seen that after 20 epochs the network Accuracy continues to increase to reach almost 100%. The trend of the points as shown from the best fit is to increase in value which is very good. The Accuracy best fit almost saturates at a value of 95%.

VI. Conclusion

A smart glove that is composed of five flex sensors attached to the fingers of a plastic glove. It is electrically connected with a potential divider circuit to create a dataset for the hand gestures of the ASL. Few volunteers move their fingers to represent nine randomly selected alpha-numeric Arabic letters. Measurements are not only presented in voltage and resistance signals but also in different formats which are Actual, Normalized and Zero Ref. The dataset is fed to CNN network that is composed of twenty-one layers. They contain Batch Normalization layers, reLU layers and Max pooling layers to make the CNN network faster. The Accuracy and Loss parameters are used to describe the success of the network in classifying the selected Arabic letters. The Accuracy achieved reaches an average of 95% success in classifying the nine letters. While, the Loss in the network while training drops down to reach 0.5. Both best fits for Accuracy and Loss saturates at their respective values. The Accuracy values show constant increase while the Loss values show constant decrease.

Conflict of Interest

Authors declare that they do not have any conflict of interest.

References

[1] Cruz PJ, Vásconez JP, Romero R, Chico A, Benalcázar ME, Álvarez R, Barona López LI, Valdivieso Caraguay ÁL. A Deep Q-Network based hand gesture recognition system for control of robotic platforms. *Scientific Reports*. 2023;13(1):2045-2322. doi: 10.1038/s41598-023-34540-x

[2] Cabrera, Maria & Bogado, Juan & Fermín, Leonardo & Acuña, Raul & Ralev, Dimitar. Glove-Based Gesture Recognition System. Book: Adaptive Mobile Robotics (pp.747-753). 2012. Doi: 10.1142/9789814415958_0095.

[3] Abhishek KS, Qubeley LCF, Ho D. Glove-based hand gesture recognition sign language translator using capacitive touch sensor. *2016 IEEE International Conference on Electron Devices and Solid-State Circuits* (*EDSSC*), Hong Kong, China, 2016, pp. 334-337, doi: 10.1109/EDSSC.2016.7785276.

[4] Zanghieri M. sEMG-based Hand gesture recognition with deep learning. arXiv:2306.10954v1 [10eess.SP]. 2023. https://doi.org/10.48550/arXiv.2306.10954.

[5] Bello H, Suh S, Geigler D, Ray L, Zhou B, Lukowicz P. CaptAinGlove: Capacitive and inertial fusion-based glove for real-time on edge hand gesture recognition for drone control; arXiv:2306.04319v1 [10cs.LG]. 2023. https://doi.org/10.48550/arXiv.2306.04319.

[6] Al-Saedi AKH, Al-Asadi AHH. Survey of hand gesture recognition systems. *Journal Physics: Conf. Ser*. 2019:1294 042003.

[7] FLEX SENSOR. https://www.sensorprod.com/pdf/flex-sensor.pdf, visited on first of April 2023.

[8] Ogundokun RO, Maskeliunas R, Misra S, Damaševičius R. Improved CNN based on batch normalization and adam optimizer. In: Gervasi O, Murgante B, Misra S, Rocha AMAC, Garau C. (eds). Computational Science and Its Applications – ICCSA 2022 Workshops. ICCSA 2022. Lecture Notes in Computer Science, 2022;13381. Springer, Cham. https://doi.org/10.1007/978-3-031-10548-7_43.

[9] Nair V and Hinton G. Rectified linear units improve restricted Boltzmann machines vinod nair. *Proceedings of ICML*. 2010;27:807-814.

[10] Zreik, Majd & Leiner, Tim & De Vos, Bob & van Hamersvelt, Robbert & Viergever, Max & Isgum, Ivana. Automatic segmentation of the left ventricle in cardiac CT angiography using convolutional neural networks. 2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI 2016). p:40-43. DOI: 10.1109/ISBI.2016.7493206.

[11] Liu Y, Gao Y, Yin W. An improved analysis of stochastic gradient descent with momentum. *34th Conference on Neural Information Processing Systems* (NeurIPS 2020), Vancouver, Canada, 2020.

[12] Latif G, Mohammad N, Alghazo J, AlKhalaf R, AlKhalaf R. ArASL: Arabic alphabets sign language dataset. *Open Access*. 2021. DOI:https://doi.org/10.1016/j.dib.2019.103777.

[13] Glorot X and Bengio Y. Understanding the difficulty of training deep feedforward neural networks. *Journal of Machine Learning Research - Proceedings Track*. 2010;9:249-256.

[14] Gomes D and Saif S. Robust Underwater fish detection using an enhanced convolutional neural network. *International Journal of Image Graphics and Signal Processing*. 2021;13:44-54. 10.5815/ijigsp.2021.03.04.