HAND TALK

Rahma Ali Bauomi *Shahd Mohamed Elshimy* *Menna Allah Thabet*

Key words: Communication, Deaf, VGG16, Image classification

Group 16313

Grade 12

2020-2021

Red Sea STEM

Hand Talk

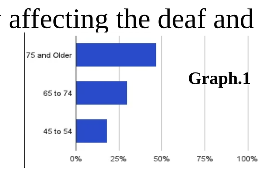
Abstract:

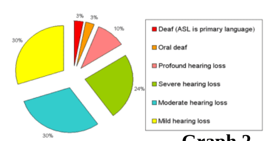
A communication gap between the world and deaf community needs to be bridged. Deaf people can often experience many difficulties because of communication barriers among them and other people in the community. These communication barriers affect their family relationships, work, and school. Surprisingly, this is not the only case! It was discovered that a deaf's mental health care is also exposed to danger. As a result of poor communication between the health care professionals and the deaf, many patients report that they are not properly informed about their disease. They can even face problems before reaching the Clinic rooms. The latest issued statistics by the Egyptian Organization for Disability Challenges in 2000 indicated that their number in Egypt is 3 million. Currently, World Health Organization indicates that there are about 7.5 million deaf people in Egypt. After studying the current state of deaf people, an efficient and innovative method has been introduced for an ease communication based on Artificial intelligence. By applying the techniques of Deep learning, the application aids in understanding deaf people. Therefore, the chosen solution is dedicated for translation of their sign language into text. In addition, the design requirements were provided such as image classification, minimum loss, and high accuracy. As a result of prototype testing, the predicted text had an accuracy of 99.56%, a loss of 0.157, and image classification with 99.87% confidence level. Successfully, this can lead to no need for interpreters, the ability of the deaf to interact more socially, and increase their job opportunities inside and outside the country.

Introduction:

Our goals are to create an application that helps communication between the deaf and hearing people, to help increase the chances of the deaf to work and deal with the external environment for them. The importance of communication between deaf and hearing is a problem that grows day by day, psychologically affecting the deaf and the country by the inability to benefit from the skills of the deaf. (Graph. 1) shows the number of deaf people and their age variation. Failure to benefit from 45% of the deaf whose ages range from 25:40 negatively affects the number of workers present in the country. It can be clearly said that a large percentage of the Egyptian people have hearing problems in general, but only 3% of them - about 90,000 - are completely deaf (graph.2). The idea for the project was rooted in many of the previous solutions. For example, Alexa for deaf people, the principal usage for it is the translate the main 5 motions of the deaf and answer about the question - indicate 5 main questions- that he asked for it. It works by using a complex algorithm and advanced instruments. The disadvantage in Alexa is that there are limitation questions as to the weather in the country – location for regions...etc. This project makes life easier for deaf people only, and its cost 3980L.E. The last solution of hearing aids that helps the deaf to hear and is the best solutions, even if they were priced between 30,000: 45,000$, most importantly by the year 1995 their effect on the electrical signal frequencies was discovered in the brain, which then leads to side effects for the deaf, and also this will help the deaf to hear and it will not help them to communicate with us. By using artificial intelligence (AI) which is the simulation of intelligent behaviors in machinery Especially, deep learning, it has the ability to think or learn from previous experiences that require a mental operation. AI aims to reach systems that have intelligence and act in the way humans behave in terms of learning and understanding and make a decision. The project was based on taking pictures of the deaf's hand movements and translating them into text. Our project is based on the classification of images (the image that the application will capture).The data set is basic images of hand consisting of 87000 images. Our project will lead to an increase in the number of available expressions from 5000 expressions only to infinite expressions. The prototype shown in (fig.1), met the first design requirement as it worked accurately and gave us 99.6% for detect the sign by multiple epochs. It met the second design requirement UI design test which shows the performance of the application and its effect on the phone, and other design requirements like loss test, number of images classified.

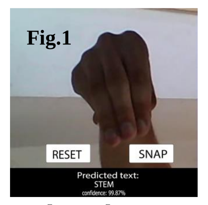
Graph.1





Graph.2

Fig.1



Materials:

Materials for the prototype are:-

1. Laptop .
2. Smart Phone.
3. Python.
4. Dataset.

Methods:

First: - the TF model

The model depended on many blocks:

The first block (1): is shown in fig(2), keras library from the tensor flow platform was used. it also contained two models, in addition, the main functional for classification and detection was imported which is vgg16. Also, The Adam optimizer was imported as well as a train-test split was done.

Fig.2

The second block (2): it included the output classes of the test and the import of the dataset.

The third block (3): it included the loading data and a ratio spilt between training image and testing image.

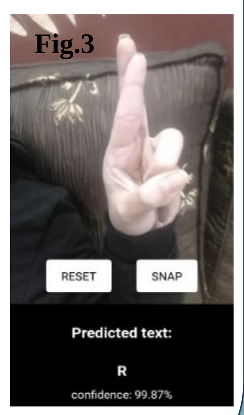
The fourth block (4): it determined the batch number, which was 128, the number of epochs which are 5 epochs, and plotting the graph data.

The fifth block (5): setting the model, and recorded train results of accuracy, learning rate, and loss.

The sixth block (6): the project is saved on E disk as "sign_lang.h5"

Second: -the TF model was converted to Tflite by the official tflite converter, then the tflite model was put in the react native as it has camera view that take photo and put it in the model as shown in figure (3).

Fig.3



Test Plan:

To make sure that the project will succeed, some design requirements were set up to be tested for the prototype:

Image Classification: it was tested by checking the ability of the project in classifying the images which was unrecognizable for it.

Minimum Loss: it will be test by how well the algorithm will model the data by using loss function it will figure out the least loss percentage in the predicted text.

Accuracy: the accuracy of output will be compared to the accuracy of the expert human interpreter, in order to reach a higher accuracy.

UI design: it can be tested after downloading the App on the phone and calculating how it consume energy and space and to reach least data cellular usage.

Results:

For the sake for providing an effective project, we conducted many tests for the design requirements. Then, a trail could not succeed because the dataset was not suitable and affected the training of the project, as well, the predicted text. It has only 64.06% confidence level as shown in fig (4). After changes in the dataset, the results of the successful trail were recorded:

Fig.4

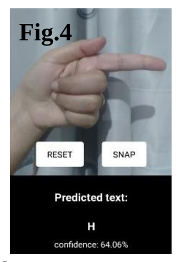
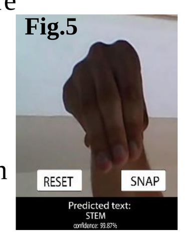


Image Classification Test:

A total of 87000 images were included in our dataset were divided into train and test. A split was performed as the train folder had 78300 images and the test folder had 8700 images. The photo of the hand of the deaf showed the predicted text after test as shown in fig (5), with a confidence level of 99.87%.

This confirmed that the project could classify the images which they were unrecognizable for it before, thus it demonstrates AI basics and learning from inputs and outputs.

Fig.5



Minimum Loss test:

Five epochs were done to train the VGG16 model on the 78300 images. Thus, leaded to reach the least value of loss and the highest value of $\hat{y}$ in the predicted text as shown in table (1). And graph (3).

Graph.3

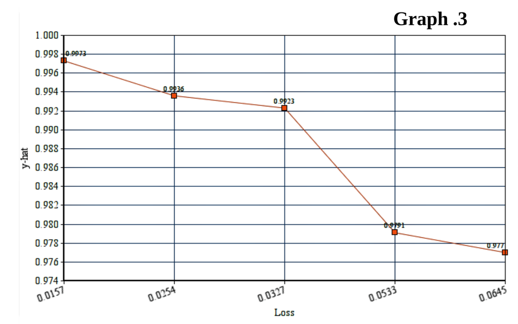


Table.1	$\hat{y}$	Value Loss
Epoch 1	0.9770	0.0645
Epoch 2	0.9791	0.0533
Epoch 3	0.9923	0.0327
Epoch 4	0.9936	0.0254
Epoch 5	0.9973	0.0157

Accuracy test:

In the five mentioned epochs, the accuracy percentage was recorded until we reached the highest accuracy percentage by 99.56% as shown in table (2), and graph (4). Thus, it was higher than the expert human's accuracy percentage which is 88.8%. Accordingly, the application can replace the human expert for more facility and accuracy.

Graph.4

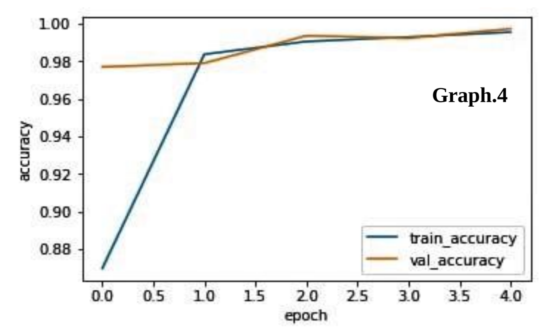


Table.2	Accuracy
Epoch 1	0.8698
Epoch 2	0.9837
Epoch 3	0.9906
Epoch 4	0.9929
Epoch 5	0.9956

UI design Test

The UI design proved its prosperity, as it does not consume a very large space which is 121 MG, as shown in fig (6), and its energy consumption was less than 0.1%. In addition, the application does not need an internet connection, thus, no data cellular usage, as shown in fig (7). It means that Blue Lines has a high performance.

Fig.6

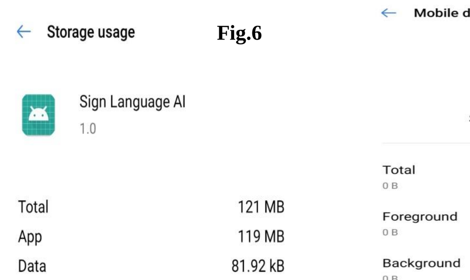
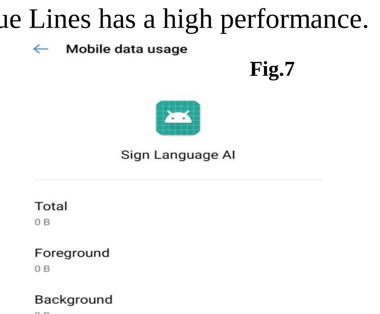


Fig.7



Analysis:

The project aims to break down communication barriers between deaf people and people who are unable to communicate in sign language because this issue significantly affects many fields such as public health. The main goal is to make deaf people communicate easily at any place, specially at hospitals. This is done by translating the sign language into text so people can understand well.

After applying the project in a clinic, the trail could not succeed. This was a result of the unification of the background and image quality of the data set. This unification affected the training of the project, similarly, the predicted text. It meant that the project could not translate the sign language which explains the symptoms, thus, leaded to a wrong Diagnose by the doctor.

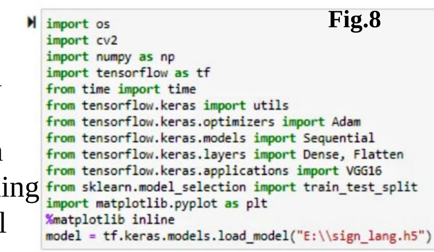
As a result, the dataset was expanded, images with different background were added, and another training for the project took a place. Consequently, a positive trail successfully was done: The symptoms were expressed with no fear of misunderstanding, similarly, the doctor could give the right diagnose.

As the project's main goal is translating expressions included in images, it was dealt with deep learning specifically as the images needed a high recognition level, and this is accurately done by convolutional neural network (CNN).

In **CS 3.2.6**, It was benefited from how neural networks work to mimic human brain. Neural networks are a class of machine learning algorithms used to model complex patterns in datasets using multiple hidden layers and non-linear activation functions.

A neural network takes an input passes it through multiple layers of hidden neurons, and outputs a prediction representing the combined input of all the neurons. After each cycle of training, an error metric is calculated based on the difference between prediction and target. The Python library "tesnsorflow.keras" was imported to provide optimized pre-trained model to be deployed in sign language application, as shown in fig (8).

Fig.8



The imported data set was 87000 images, each one expresses a character in English sign language "ASL" as shown in fig (9)

A 9:1 split ratio of data set in two folders- train and test- was done with 78300 images for training and 8700 images for testing. By this, the model can learn and memorize the information for future prediction. Because Ai models deal with small numbers, the images were resized and a scale of RGB was changed from 0:255 to 0:1. The Ai model was VGG16, as shown in fig (10), it was used because it is a very Deep Convolutional Network for Large-Scale Image Recognition. The model achieved 92.7% top-5 test accuracy in ImageNet, and it was trained for weeks on a dataset of over 14 million images belonging to 1000 classes. So, VGG16 is was a best fit for reaching our purpose.

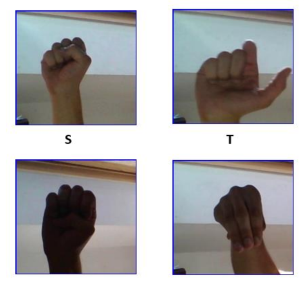
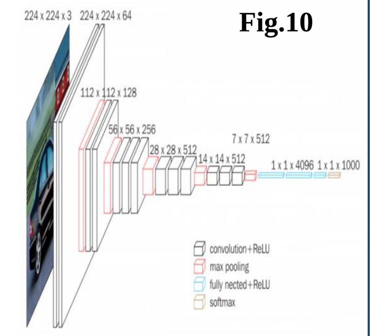
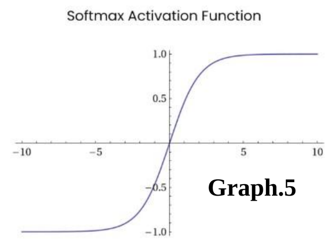


Fig.9

Fig.10



After inputting the images to VGG16, the neurons were activated with SoftMax activation function. The SoftMax activation function as shown in graph (5), is typically placed in output layers of the network, the SoftMax neurons allow the prediction of the text as a mimicking for human brain. In Statistics 2.2.5, as we studied about probabilities, we related this concept with the function of SoftMax. It calculates the probabilities of the event over 'n' different events. In general way of saying, this function will calculate the probabilities of each target class (the character) over all possible target classes (predicted characters), and the calculated probabilities will be helpful for determining the target class for the given inputs. Then, the loss function which is called "Categorical Cross-Entropy" used to calculate the rate of change of the error between **y**(text) and $\hat{y}$ (predicted text) as shown in the equation:



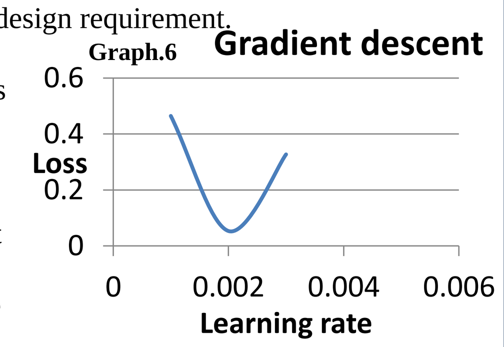
Graph.5

$$\text{Loss}(\hat{y}, y) = -(y \log(\hat{y})) + (1-y)\log(1-\hat{y})$$

In this point, **MATH 3.1.2** was helpful in knowing how the local minima of the loss occurs. As it was studied, the local minima can be calculated when the derivative of a function is zero. Thus, when the loss function is used, the derivative is equal to zero only when the text (y) and the predicted text ($\hat{y}$) are the same, which means the global minima is reached!

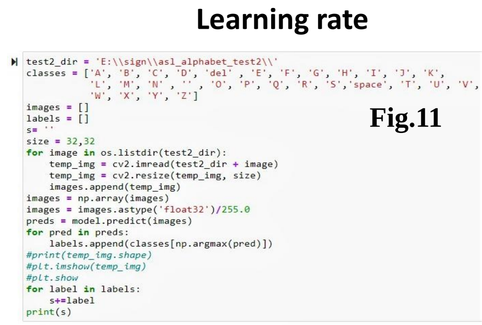
Accordingly, the learning rate was short- only 0.0001- this was for minimizing the gradient decent steps that was took by "Adam" optimizer. Adam is an adaptive learning rate optimization algorithm that's been designed specifically for training deep neural networks. It was imported as shown in fig (8). Gradient descent is used to minimize some function by iteratively moving in the direction of steepest descent to optimize the parameters of our model, after training, the loss of y could reach the minimum value. It was only 0.0157 as shown in graph (6), thus, achieving the **"minimum loss"** design requirement.

Graph.6



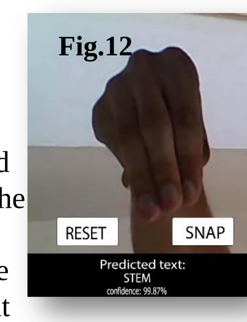
The epochs for training VGG16 were five epochs, and the accuracy of the model was recorded in block 5 as shown in graph (4). By reaching the highest accuracy of 99.56%, which is higher than the 88.8% accuracy of an expert human interpreter, it achieved the **"Accuracy"** design requirement. After training the model, the images in the test folder were tested to check the ability of the project to classify the unrecognizable images for it. The letters of word "STEM" was tested, and it showed the accurate predicted text as shown in fig (11).

Fig.11



It means that the training succeeded to make the model learns from the images of train folder, consequently, the project demonstrates that is AI based. The user interface was created by converting the AI model into tflite model using official tflite converter. The tflite model was put into a react native app which has a camera view. This camera inputted The taken images of hands into the tflite model. Accordingly, the predictions of the characters appear in the prediction text component. To complete our goal of providing facilitated usage of the application, we tested the user interface to make sure that it has no effects on the performance of the phone. As it is specified, the application had less than 0.1% energy consumption, only 121 MG of space and no data cellular usage, and this achieved **"UI design"** as a design requirement. In the end, the application was also tested to recognize sign language of a deaf's hand. The word "STEM" was successfully predicted with a confidence level of 99.87%, as is shown in fig (12). Accordingly, this achieved the **"image classification"** design requirement.

Fig.12



Conclusions:

All the previous illustrations represent the crucial importance of our project and its operative role in enhancing Egypt's communication and health for deaf people. Based on these achievements, it was concluded that:

1- The model is working effectively with training on a data set of images that had a different background to prevent the negative effectiveness on the hand's expression. Thus, it will reach 99.87% of confidence level of the predicted text.

2- The accuracy of the application in translating sign language which is 99.56% is higher than the expert human interpreter which has an accuracy of 88.85%. consequently, the project can effectively replace him.

3- The application has no effect on the mobile phone as it does not need an internet connection, thus, no data cellular usage and it does not consume a very large space which is 121 MGB and its energy consumption was less than 0.1%

4- Sign language ai application will lead to increasing the opportunities for deaf community to interact more socially, thus, increasing the job opportunities.

Recommendations:

The project solved the main problem, Nevertheless the future researchers of the project may work on several aspects:

1- The project can be developed to convert text to sign language by making inverse system and allow the any one to communicate with sign language as text to ASL Generator Tool. It is the primary sign language used by the deaf and people with hearing impairment. This WeCapable Tool easily converts English text into sign language symbols.

2-The sign language can change to speech like Google sign language AI which turns hand gestures into speech or allow change voice into sign language by using voice recognition.

3-Using tensor flow lite instead of tensor flow as the applications developed on **TensorFlow Lite** will have better performance and less binary file size than **TensorFlow** mobile, lightweight version which is specially designed for the mobile platform and embedded devices, much faster and smaller in size.

4- In addition to ASL sign language, "sign language ai" application can translate more than one language. So, the dataset will have all distinguished sign languages, as well, the UI design will have a button to press for the required language.

Literature cited:

Global health: EGYPT. (2020, August 31). Retrieved October 6, 2020, from https://www.usaid.gov/egypt/global-health

Image classification : Tensorflow core. (2016, March). Retrieved November 5, 2020, from https://www.tensorflow.org/tutorials/images/classification

Cook, C., Warisawa, T., Howard, J., Keeble, T., Iglesias, J., Schampaert, E., . . . Davies, J. (2019, July 15). Algorithmic versus expert human interpretation of instantaneous wave-free ratio coronary pressure-wire pull back data. Retrieved November 18, 2020, from https://www.sciencedirect.com/science/article/pii/S1936879819312087

Understanding categorical Cross-Entropy LOSS, Binary Cross-Entropy Loss, SOFTMAX Loss, logistic Loss, Focal loss and all those confusing names. (n.d.). Retrieved December 2, 2020, from https://gombru.github.io/2018/05/23/cross_entropy_loss/

Gonick, L. (1993). **The Cartoon Guide to Statistics**. HarperPerennial.

Hughes, Elisha. Notes on Basic Practices of Statistics– Chapter 5. Regression. Retrieved from https://www.math.utah.edu/~hughes/Chapter_05.pdf on May 25, 2018.

Master the AP Calculus AB & BC, 2nd Edition (Peterson's Ap Calculus) 2nd Edition. by W Michael Kelley (Author), Mark Wilding (Author)

Acknowledgement

First, all the praises and thankful are given to Allah. Giving a special thanks to the people who helped us in our project.

Eng: Mohab Mohamed

Eng. Mahmoud Ashraf.

Teachers: Mr.Naser Gamal (Capstone Leader) - Mr.Ayman Fahmy(Capstone Teacher)

Moreover, our General Capstone leader Madame Heba Shabaan

For Further information:

Rahma Ali Bauomi

Rahma.1618113@stemredsea.moe.edu.eg

Shahd mohammed Elshimy

shahad.1618120@stemredsea.moe.edu.eg

Menna Allah Thabet

Menna.1618131@stemredsea.moe.edu.eg