# Facial Expression Recognition Using Machine Learning

Manthri Bala Kiran

Data Science Master Program, Machine Learning

Högskolan Dalarna

Borlänge, Sweden

Email: h19baman@du.se

**Abstract: The main aim of this project is to find the Facial Expression of humans using Machine Learning Algorithms. CNN classification techniques were used to predict the facial expression of humans. Classifying an image based on its depiction can be a complicated task for machines.**

*Key Words: Machine Learning; Deep Learning; CNN; Hybrid-CNN; Image Classification; Data Mining;*

## I. INTRODUCTION

Facial Expression is the most common thing used by humans to convey various types of feelings or meanings in various contexts. The range of the meanings is basic possibly innate of socio-emotional context such as "Surprise" or "Happy" and can tend to be the most complex meaning such as "Carelessly". Humans will be using Facial Expression in different contexts based on the environment they are been in [1].

Image classification can be done using machine learning algorithms like SVM, DT, and KNN classification. But using those machine learning algorithms can have feature outline problems and core problems [2]. In this paper, to overcome the issues for image classification problems an advanced machine learning algorithms or Neural Networks (Deep Learning Algorithms) was developed. Neural Network can be defined as a series of algorithms which helps to recognize the relationships in the data through a process which replicates the human brain operation. Neural Networks can learn from the wrong outputs and change their weights in the network for better prediction [3].

This paper aims to use different pixel image data and find the facial expression of humans. To identify different classes of the images (as different emotions in the images) image classification technique with CNN is been used. Image classification is a process where the computer will be using an image and identifying the class of the image or the probability of the image for its class. A class is a label to classify the type of image. Convolution Neural Networks are considered as the backbone for image classification where it takes an image and assigns a class and label to that image which makes that image unique [4] [5].

A Convolution Neural Network is a Deep Learning algorithm which takes the image as input, and it assigns importance such as learnable weights and biases to the various aspects of the objects present in that image and which makes that image different from one to other. CNN reduces the images into a form without losing features which are vital for good prediction [6].

## II. LITERATURE REVIEW

Considering Facial Images based on the shape feature using the optimization algorithms. A classification-based similarity finding is been proposed for the classification of the images based on the shape of the image like round or oval [7].

Image classification is considered as the primary domain in the neural network which plays a major important role in medical image analysis. The image classification will be accepting the given input image and produces output classification for identifying whether the disease is present or not. CNN is being considered as the state-of-the-art method for image classification. They are used to predict the development of the disease in the brain [8].

The method of combing Convolution Neural Network (CNN) is eXtreme Gradient Boosting (XGBoost) will be giving a better performance. The two methods are considered as the best classifiers. CNN_XGBoost provides more precise output by integrating CNN as a trainable feature extractor to automatically obtain features from input and XGBoost as a recognizer in the top level of the network to produce results [9].

Image Classification is the core problem in the computer vision field with a large variety of practical applications. Examples like: object recognition for robotic manipulation, pedestrian detention. There is a lot of attention associated with Machine Learning, specifically neural networks such as the Convolution Neural Network [10].

## III. DATA SET

In this paper, the dataset of Facial Expression Recognition has the pixel data of humans with different Facial Expressions, and the dataset is been collected from Kaggle [11]. The attributes of the dataset are listed as follows:

Emotions, Usage, Pixel are the attributes present in the dataset. We have also a test dataset where we have only the pixel data to predict the emotions. The description of the training dataset attributes is:

- Emotion: This column has the emotion of the images in the dataset. The emotions are being classified into seven categories they are: 0=Angry, 1=Disgust, 2=Fear, 3=Happy, 4=Sad, 5=Surprise, 6=Neutral.
- Usage: This column has three categories: Training Set which consists of 28,709 examples, Public Test which consists of 3,589 examples, and Private Test which consists of 3,589 examples.
- Pixels: This column has the pixel data of the image. Which is a series of numbers explaining the grayscale images of the faces. We can consider them as 48*48.

And the Test data file has only the 48*48 grayscale pixel data where we can use the dataset to predict the emotions for the new set of images using our model.

IV.     METHODOLOGY

a. Data Handling:

Firstly, the data is loaded into a data frame using Python. Then the data handling is taken into two steps. In the first step, the training data frame is divided into a training set, a validation set, and a test set. And performed the training for the CNN model using the training and validation sets from the training set. And performed the trained model to predict the emotions for the test set from the training model and predicted the emotions for the test data set where the emotions were missing in the dataset.

In the second part, the downsampling of the dataset is performed to prepare a balanced data set. The balanced data set is divided into training, validation, and test sets. And training is performed for the CNN model using the training, validation sets from the balanced dataset, and performed the prediction for the test set, and the test data.

b. Exploratory Data Analysis:

To understand the insights of the data, the count of each emotion from the dataset is performed. And also tried looking at the images from the dataset concerning the emotion. Also tried to see the count of emotion from training, validation, and test sets from the training dataset.

C. Training and Testing datasets:

The pixel feature data from the training dataset is loaded into variable 'X' and reshaped into 48*48. The emotion data from the training set is loaded into the variable 'y'. The X and y datasets are divided into training, validation, and test sets based on the usage feature counts of training, validation, and test sets. The y dataset is converted into categorical data based on the emotions. And we use the trained model to predict the emotions for the test dataset where the emotions of the images are not given.

d. Techniques Performed:

Pre-defined classes of layers, models, and optimizers from the Keras package were used such as: Convolution2D, Activations, Dense, Sequential, SGD, Adam.

Architecture: The inputs for the model are connected to the CNN layer with 32 cells in it and the input is being reshaped into (48,48,1). Relu activation function is used in all Convolution layers. A max-pooling layer is added after every two convolution layers and after the final max-pooling layer the data is flattened to enter into the dense layer and a dropout layer with 0.4 is added before the final dense layer.

| Layer (type) | Output Shape | Param # |
|---|---|---|
| conv2d_17 (Conv2D) | (None, 48, 48, 32) | 320 |
| conv2d_18 (Conv2D) | (None, 48, 48, 32) | 9248 |
| max_pooling2d_9 (MaxPooling2 | (None, 24, 24, 32) | 0 |
| conv2d_19 (Conv2D) | (None, 24, 24, 64) | 18496 |
| conv2d_20 (Conv2D) | (None, 24, 24, 64) | 36928 |
| max_pooling2d_10 (MaxPooling | (None, 12, 12, 64) | 0 |
| conv2d_21 (Conv2D) | (None, 12, 12, 96) | 55392 |
| conv2d_22 (Conv2D) | (None, 10, 10, 96) | 83040 |
| max_pooling2d_11 (MaxPooling | (None, 5, 5, 96) | 0 |
| conv2d_23 (Conv2D) | (None, 5, 5, 128) | 110720 |
| conv2d_24 (Conv2D) | (None, 3, 3, 128) | 147584 |
| max_pooling2d_12 (MaxPooling | (None, 1, 1, 128) | 0 |
| flatten_3 (Flatten) | (None, 128) | 0 |
| dense_5 (Dense) | (None, 64) | 8256 |
| dropout_3 (Dropout) | (None, 64) | 0 |
| dense_6 (Dense) | (None, 7) | 455 |

Fig 1: Summary of the Model

The same parameters were used for both balanced and unbalanced datasets. The batch size was set to 128 and 14 epochs were used in both cases. After training the model, it is used for predicting the emotions (y values) for the test set from the training dataset. And confusion matrix is drawn to analyze the results and understand is there any bias in the results.

V.     DATA EXPLORATION:

Different insights of the data are drawn to understand the data. Let's see the emotion count from the original dataset which has the training, validation, and test sets.
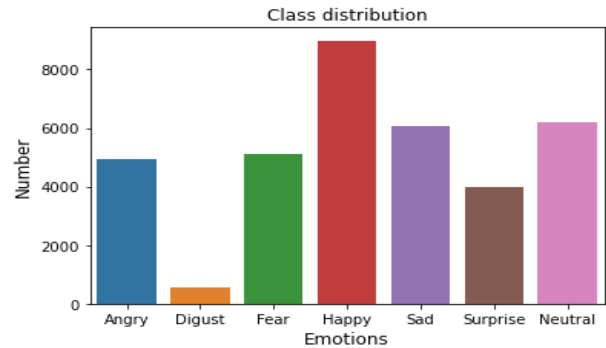


Fig 2: Emotions from the training dataset

To understand the images from the dataset, let's see some sample images from each emotion from the training dataset.
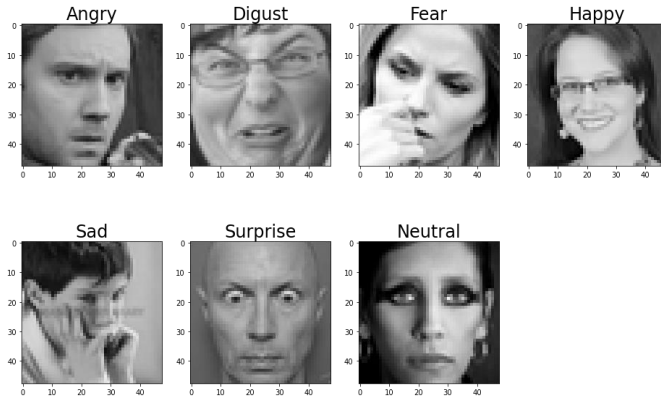
Fig 3: Sample image for each emotion.

Now, let's see how the emotions are being distributed in the training, validation, and test sets.
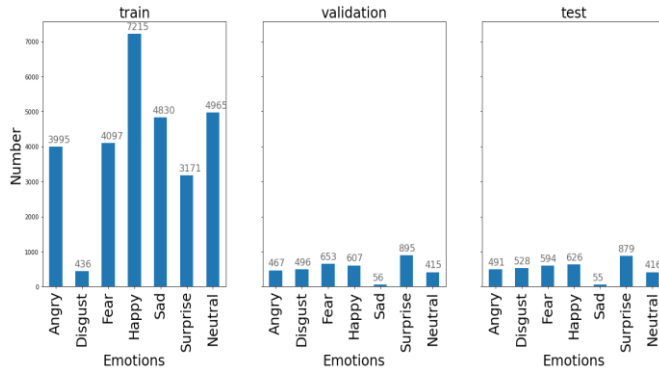

Fig 4: Emotion count of the divided training set

VI.     RESULTS

The model is trained with the training and validation sets from the training dataset. This is the unbalanced dataset. The model is being trained with 90% training accuracy and 89% validation accuracy.
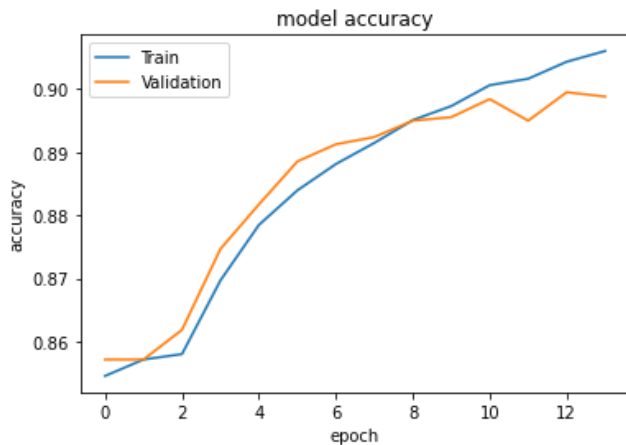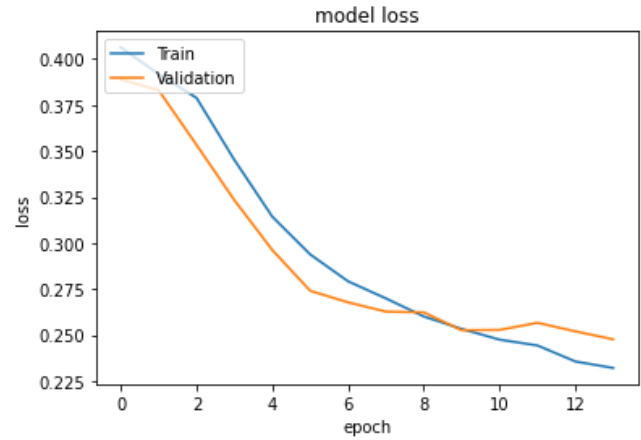

Fig 5: Training and Validation Accuracy (Unbalanced Dataset)


Fig 6: Training and Validation Loss (Unbalanced Dataset)

Prediction for the test set was performed using the model which is trained and achieved the accuracy of 60%. Let's see the confusion matrix for the predicted test set.


Fig 7: Confusion Matrix for the Test set (Unbalanced dataset)

From Fig 7, we can see bias in our results. This may have happened for two reasons. The first might be in the selection of the sample of photos. The pixels might not clearly distinguish the emotions. And the second is due to the overfitting of the data. We can see that emotion 1=Disgust is mainly classified as 0=Angry. And same with another emotion. This may have happened due to the pixel images for all the emotions that might be the same and we can also say that there are not enough images for the emotion 1=Disgust.

In the second part, the dataset is being downsampled for having a balanced dataset. The model with the same parameters are been used for training and validation. Through this, the model has achieved a training accuracy of 86% and the validation accuracy of 86%.
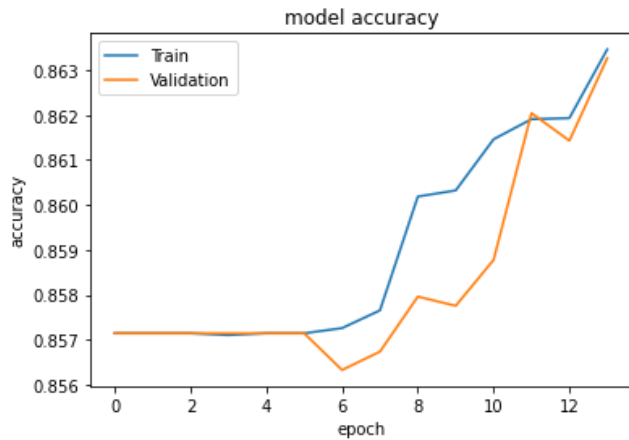
Fig 8: Training and Validation Accuracy (Balanced Dataset)
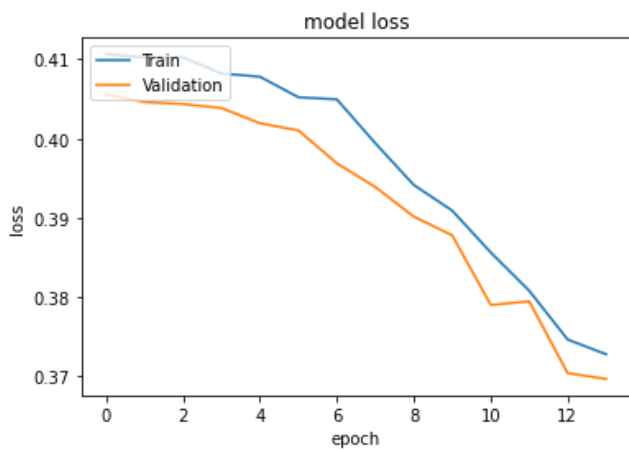

Fig 9: Training and Validation Loss (Balanced Dataset)

From Fig 8 and Fig 9, we can see the training and validation accuracy and loss for the balanced dataset. Using the same model, prediction for the test set was performed and achieved an accuracy of 34%. Let's see the confusion matrix for the test prediction.
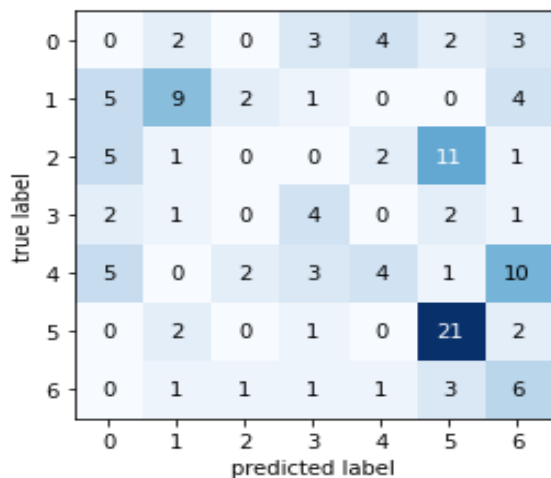

Fig 10: Confusion Matrix for the test set (Balanced dataset)

From Fig 10, we can see that there is still some bias with other emotions from the dataset. Let's see the emotions 0 & 1 which are Angry and Disgust. The model performed well in predicting the Disgust and some of the images were predicted Angry as well. This says that the bias in the prediction comes from the bias of images from the dataset.

Now let's predict the test dataset using the models which are been build. The test dataset has only the pixel data in it. The emotions of the images were not given in the dataset. Now let's see the count of emotions predicted from the test data set using an unbalanced model.
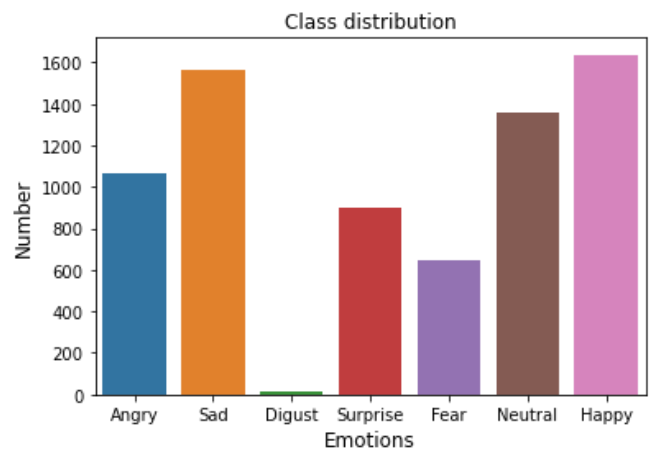

Fig 11: Prediction of Emotions for the Test Set (Unbalanced Model)

Now let's see the sample images from the test dataset which were predicted using the unbalanced dataset model.
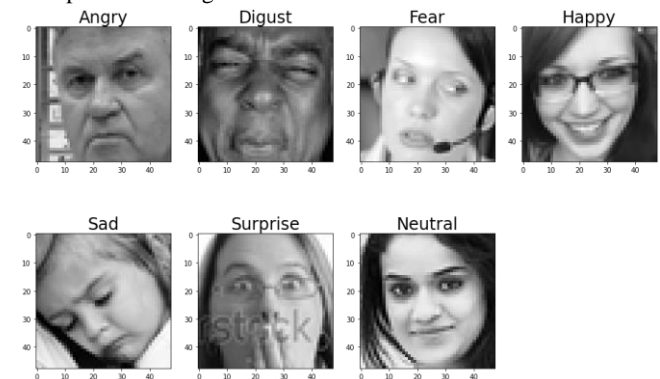

Fig 12: Images from test dataset predicted from Unbalanced model

Fig 12 describes the prediction of the images which were done through the unbalanced dataset. We can see some overfitting or bias with the sad image. It was a complex one we can consider that image as Neutral as we can't clearly say that the girl in the image is sad.

Now let's see the count of emotions that are predicted for the test set using the balanced model.
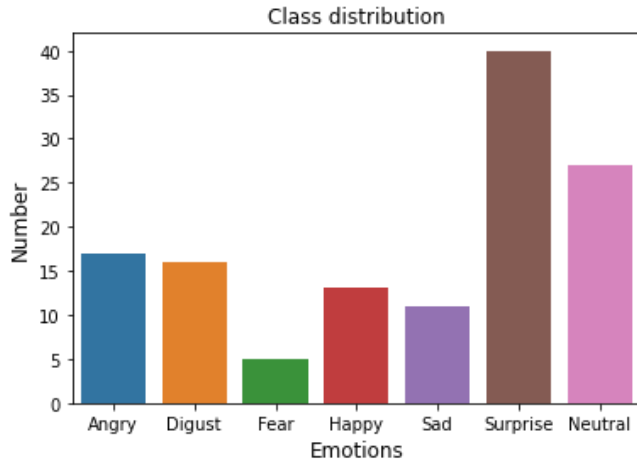
Fig 12: Prediction of Emotions for Test Set (Balanced Model).

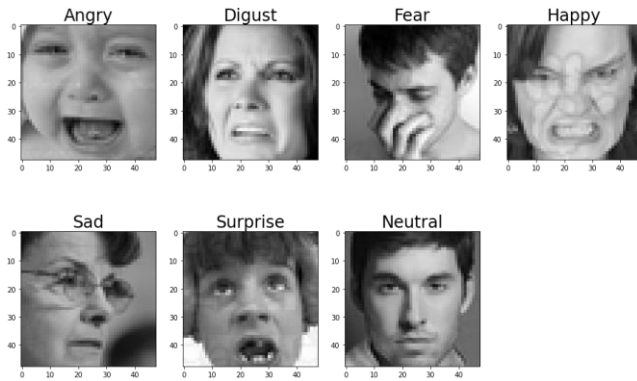Now let us see the sample images from the predicted test dataset.



Fig 13: Images from the test dataset predicted from Balanced Model

Fig 13: describes the prediction of the images which were done through the balanced model. We can see that there might be less training for the emotion of the model or there is some bias in the images which are been selected. The emotion Happy and Angry is wrongly classified. This might be due to the images which are been used to train the model.

Now let's see how the model is predicting the emotion of an image that is taken from an online search engine.
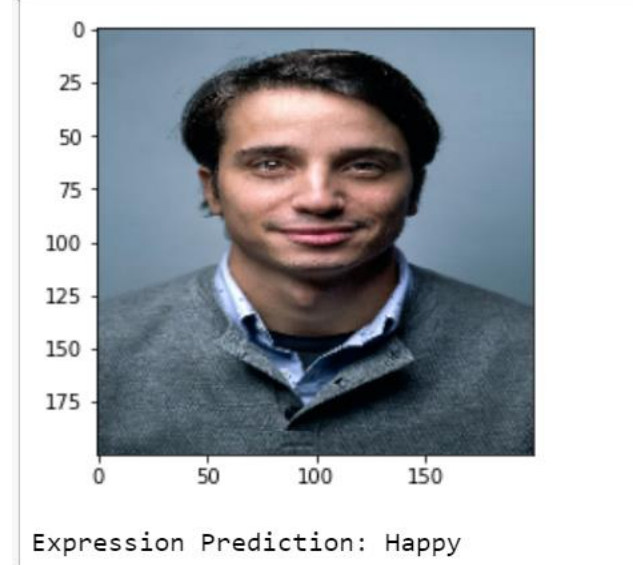


Fig 14: Predicting Emotion of the Image from the Internet

## VII.  DISCUSSION

From the results obtained we can see that the machines can perform image classification like humans. Machines need proper training to classify the images. The images which are selected for the training have a major role in the prediction of the emotions. It would be a different prediction accuracy with different images. The sample and bias of the images in the dataset play a crucial role in the image classification.

## VIII.  CONCLUSION

From the proposed paper we can say that the image classification technique using CNN perform better compared to other models. With the proper training to the models, we can predict all the basic emotions of humans using images. In this paper, we also see how the model will be performing if the dataset has a balanced set of emotions and an unbalanced set of emotions. And it would be changeling if the dataset is biased to certain emotions or if the images for multiple emotions almost look the same.

## IX.  LIMITATIONS

Concerning the project proposal and suggestions given for the project proposal, tried to perform RNN for the Image classification but was not able to compare their performance for the image classification. And understood that the RNN performs better for time series data and sentence prediction and CNN works better for Image classification.

## X.        REFERENCES

[1]    F. Psychol, "Frontiers," [Online]. Available: https://www.frontiersin.org/articles/10.3389/fpsyg.2013.001 15/full. [Accessed 01 11 2020].

[2]    S. Kumar, Z. Khan and A. Jain, "A Review of Content Based Image Classification using Machine Learning Approach," *International Journal of Advanced Computer Research,* vol. 2, no. 3, pp. 55-60, 2012.

[3]    J. CHEN, "Neural Network," investopedia, [Online]. Available: https://www.investopedia.com/terms/n/neuralnetwork.asp#:~:text=A%20neural%20network%20is%20a,organic%20or%20artificial%20in%20nature.. [Accessed 20 11 2020].

[4]    ThinkAutomation, "ELI5: what is image classification in deep learning?," ThinkAutomation, [Online]. Available: https://www.thinkautomation.com/eli5/eli5-what-is-image-classification-in-deep-learning/. [Accessed 01 11 2020].

[5]    P. Vadapalli, "Using Convolutional Neural Network for Image Classification," upgrad, [Online]. Available: https://www.upgrad.com/blog/using-convolutional-neural-network-for-image-classification/#:~:text=Convolutional%20Neural%20Networks%20(CNNs)%20are,part%20of%20machine%20learning%20experiments.. [Accessed 01 11 2020].

[6]    S. Saha, "A Comprehensive Guide to Convolutional Neural Networks — the ELI5 way," towards data science, [Online]. Available: https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53. [Accessed 01 11 2020].

[7]    Dr.S.Vijayarani1, "An Efficient Algorithm for Facial Image Classification," semantic scholar, [Online]. Available: https://pdfs.semanticscholar.org/be8f/5dc361aa247703b073a8951d80be8b571afa.pdf. [Accessed 01 11 2020].

[8]    K.Balaji, "Image Classification," Science direct, [Online]. Available: https://www.sciencedirect.com/topics/computer-science/image-classification. [Accessed 01 11 2020].

[9]    X. Ren, "A Novel Image Classification Method with CNN-XGBoost Model," link.springer, [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-319-64185-0_28. [Accessed 01 11 2020].

[10]   M. Hussain, "A Study on CNN Transfer Learning for Image Classification," link.springer, [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-319-97982-3_16. [Accessed 01 11 2020].

[11]   K. Compete, "Kaggle," [Online]. Available: https://www.kaggle.com/c/challenges-in-representation-learning-facial-expression-recognition-challenge/data. [Accessed 02 10 2020].