DEPRESSION DETECTION BASED ON FACIAL EXPRESSIONS USING MACHINE LEARNING AND DEEP LEARNING TECHNIQUES

Neha Benny1, Prof. Madhuri. J2
MTech student,
Department of Computer Science and Engineering
Bangalore Institute of Technology,
Bangalore, Karnataka, India
Assistant Professor,
Department of Computer Science and Engineering
Bangalore Institute of Technology,
Bangalore, Karnataka, India

ABSTRACT

Depression has played an imperative role in our society and humans have realized the importance of Machine learning and how these techniques can be used to detecting different emotions. Many other ailments can be detected with Machine learning techniques. This model is done to detect 7 emotions using the purpose of the proposed research so as to use Convolutional Neural Network on a human face and depression is highlighted in real-time. An accuracy of 85.78% has been achieved in this work. It is required to find the limit to which we can identify the person with depressed traits to determine the level of depression. The classification plays a main role in finding out the kind of help a depressed person needs.

Keywords: Machine Learning, CNN, Haar Cascade Classifier, Ck+ dataset.

INTRODUCTION TO DEPRESSION

Depression is a disorder depending on mood that causes feelings that is persistent of unhappiness and not having interest. Also, it affects how and what you think, perceive and behave and can be led to many physical and emotional problems and it is called major depressive disorder or clinical depression. Sometimes you may have trouble doing normal diurnal programs, and many a times you might feel very low and life isn't worth to live. Many methods like SVM and FACS were used to detect depression but accuracy was 65% in FACS and speed of detection was very low and then, an idea of combination of FACS along with SVM was used in which the model classifier consisted of a pre-trained and pre-defined neural network for mainly providing two things, they are extraction and SVM for classification of emotions. Data had to be verified before prediction and this was a drawback and accuracy was about 70% so a better model is required that can surmount these drawbacks. This paper proposes Convolutional neural network (CNN) uses artificial neural network that has perceptrons and uses them, a machine learning unit algorithm is used, for analyzing data. ConvNet is also known as convolutional neural network. CNN[13] is one of the unique and best methods to detect facial emotions. It has inherent properties and that's why it does much better than other classic neural network, also it has convolutional layers in it. Networks with Simple feedforward neural don't clock any other inputs in the order. Even if you mix the images in the same or different way, the neural network performance would be constant, training has been given to not shuffled images. The advantage of local spatial images of coherence are taken by CNN on the contrary. This reduces operations number that is needed to process any image with the help of convolution even if the nearer pixels are scattered, because nearer pixels put together are meaningful and they can bring the pixels together. This also called the local connectivity. A small patch of pixels in each map is filled with the result of the convolution, a window is slid over the whole image. Pooling layer also plays an important part in CNN and it doesn't exist in neural networks of classic. The advantage of having this layer is that it reduces all the images to reduce their size which keeps them intact.

1 LITERATURE SURVEY

1.1 MACHINE LEARNING AND ITSTYPES

Machine learning is a major subfield in artificial intelligence (AI). The main goal of machine learning is to comprehend the data's structure and try to fit that data into different models which can be perceived and used by users.

1.1.1 Supervised learning

Supervised learning learning, also known as supervised machine, is another sub-category of artificial intelligence and artificial intelligence. Labelled datasets are been used to train different algorithms that classifies data and predicts outcomes accurately and that is why it is used.

Balaji Balasubramanian et al. [1] covered the datasets and algorithms that are used for the tasks related to FER. So here, the dataset images need to be considered from any natural environment so that they belong to the aura they are executed in. More gathered training examples for less common emotions like Disgust. CNN outshines and performs many Machine Learning algorithms like SVM. The drawback is that it didn't have much pictures of disgust expression so accuracy was low.

Ma Xiaoxi, Lin Weisi *et al.* [2] have implemented many learning methods: DBM and SVM, are all excellent general methods and it construct a system that's called prediction system that is most suitable to any kind of dare. The best detection performance in occurrence of AUs is obtained by facial classification of emotions system along with the SVM, while comparing different experiment results of prediction systems. The frame of the face samples is selected randomly from all the video frames and the dataset is distributed is and is extremely unbalanced.

Boxuan Zhong Zikun Qin*et et al.* [3] proposed a system that demonstrates the temporal information of physiological signals of the temporal information that can mainly improve the recognizing of emotions in terms of its performances and expression obtained in faces and physiological information. Using proper strategies of fusing, the accuracy of automatic emotion recognition can be improved. These experiments also prove that for the proposed TIPF, there is an improvement in multi-scale temporal approach images that are wide-angled were considered and one multi-fusion approach has been used.

Shivam Gupta*et al.* [4] has presented the emotions of fully automatic recognition system using the machine learning algorithms and computer vision that classify eight different emotions. Many algorithms that are liberated, automatic fully real-time coding of facial expressions from the streaming video that plays continuously stream can be considered as achievable goal with the computer's power that is available, few applications are considered with views from front is to be assumed with the help of webcam, the classification came out fine with respect to the results and was endorsing support vectors machines with the accuracy of at around 94-95%. Again, the drawback was that speed was very low. It couldn't detect faces that were clustered.

Tian Xuehong et al [5] proposed a system to reduce the face image in a bilinear interpolation so as to the same narrow-dimensional diagram, it will differentiate the analysis of discriminant in computing process. It didn't detect real time images although.

PRAJAKTA B. KULKARNI *et al.* [6] gives a new way to find the depression in person in a very simple way. Hence, this survey paper helps the reader to identify which depression detection method has be used in their research or study. According to region of interests, it gives many algorithms. The accuracy of detecting depression emotion was quite low.

Adrian Rhesa Septian Siswanto *et al.* [7] Iris and fingerprint methods were used and recognition had low accuracy compared to biometrics and other methods. Face recognition can be a modern system aiming to support biometrics application in multi-modal so as to gain one of the most natural and "easy-to-collect" biometrics. Attendance System Application uses Eigenface for face detection in here.

Roberto Valenti *al.* [8] It significantly has user-friendliness and improved the of the existing facial tracking system, and it extends with automatic face positioning, visualization and emotion classifiers. The Naive Bayes emotion classifier performs quite well. TAN performance classifier is not up to mark with existing research methods. The classifier either lacks enough training data, or has an implementation problem and no accuracy.

Byoung Chul Kong *et al.* [9] proposed approaches that are divided into 2 main streams: FER Conventional approaches that again has 3 steps, they are, face and facial detection of component, extraction of features and classification of expressions. These algorithms of classifications are used in FER which includes SVM, Adaboost, and random forest; on the other hand, deep-learning approaches in FER largely reduce the dependency on models with face-physics based and pre-processing techniques are included by giving “pin-to-pin” from the input images for the learning purpose. As this is a part of deep learning, CNN will visualize the images of input to help fathom the trained model learn with the help various datasets of FER, and also shows the networks that are trained on detection of emotions, in which both of the datasets and different FER related tasks.

Jiequan Li and M. Oussalah et al. [10] proposed an approach with the use of eigenface masks that was developed and it classified facial expressions across available databases publicly. Along with feature vectors of the image, they learned masks that were used to train the SVM. In this method, competitive productive results across large significantly varying data were obtained.

1.1.2 Unsupervised learning

Unsupervised learning, also known as unsupervised machine learning, uses algorithms of learning to analyze and understand clustered datasets that are unlabeled. The algorithms don't reveal or understand the groupings of data properly or hidden patterns.

K. Sasikumar1, P. A. Ashija et al. [11] LBP technique has been applied to the input image sample further it is divided into smaller regions. LBP is linked and derived together as one single vector which is used for representation face images. Processing is fast by using LBP for feature vector. LBP operator is basically used to classify recovery image, texture and helps in perceiving facial image analysis. The elements of the image consist of random variable type, in stochastic process. The Eigen vectors that are scattered is defined as the basic vector by PCA. Whenever the system should store a lot of information related to face, PCA technique helps the system to collect only the important and necessary information by using the representation of mathematics. To remove co-linearities in data and to identify linear regression, PCA technique can be used effectively. It is seen that the hybrid proposed approach gives better rate of recognition and accuracy is high when compared with PC SVM multi-class.

Neel Ramakant Borkar *et al.* [12] He proposed Face recognition system that is based on 2 things: PCA and LDA. The combinations of these two methods gave an accuracy of 97% and has used raspberry pi 3 module. The Raspberry pi 3 is a low weighted module that is used for recognition system and it is also a cost-effective module. This project on Face Recognition has analyzed many face recognition algorithms of face recognition that are used and being currently utilized. In-depth study was required and gained a lot knowledge by combining different methods to increase the Face recognition system accuracy which is important. Multidimensional faces are used and therefore dimensionality was a problem because the face required memory quotient and time in processing.

1.1.3 Deep Learning

This is a type of machine learning that is based on artificial neural networks that has layers of processing in multiple terms and are used to extract higher-level features from data.

Zeynab Rzayeva *et al.* [13] A CNN model was proposed that is trained on Cohn-Kanade and RAVDESS datasets to identify 7 major emotions of different faces. The unique quality of the model is that it's not pre-trained and performs well for datasets with or without few neural layers. The final model has performed better than the other models and it consisted of 6 convolutional layers with the pooling and dropout layers which made it more workable.

Neha S, Nivya *et al.* [14] This system plays a major role in relations that are interpersonal because they can reveal the affective state, personality, intention, cognitive activity affective state and psychological state of a person. There are 3 modules-face detections in the system that are implemented by Haar Cascade and emotion recognition which is implemented by CNN with the help of Keras that widely focuses on emotion detection that can reveals the depression in an individual. In the last module, a chatbot has been used to recognize depression that further helps to differentiate between depression and sadness.

2 PROPOSED ARCHITECTURE

This is a real-time model proposed for emotion-based recommendation for users. This system works as a personalized assistant to each of the users. Since it works under real-time feature extraction and recommendation system, we can extract facial features through different video inputs from people which makes it more interesting. CNN algorithm has been used in the architecture which is more prevailing than other algorithms like SVM[2][9]. This is because CNN's compatibility of feature is less and it can take the ability to arbitrary output/input lengths which affects the efficiency and total computational time. It can be used as a personal assistant for entertainment purposes. Also, a chatbot pops up when the emotion Depressed is detected.

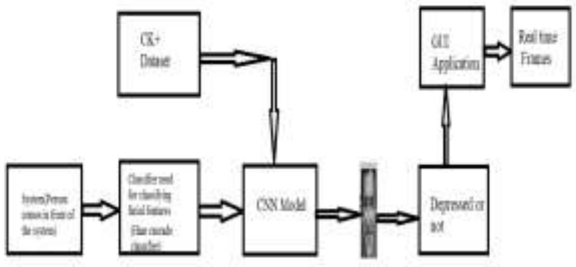

CK+ Dataset
GUI Application
Real time Frames
System(Person comes in front of the system)
Classifier used for classifying facial features (Haar cascade classifier)
CNN Model
Depressed or not

Figure 2.1: Architecture of CNN Model used for Emotion classification

Explanation of the proposed system:

When a person comes in front of the system or any image is brought in front of the system, the webcam starts streaming and from the video, it detects face using Haar-cascade classifier. Once the face has been detected, it compares the expressions of the face to the trained CNN model to detect emotions and if the emotion **Depressed** is detected then, a GUI with chatbot and movies pops-up. The model detects 7 emotions.

2.1 Kaggle

Kaggle is Google LLC' subsidiary, it is an community platform that is online where the machine learning practitioners and scientists take unlabeled data from. Users can find and published datasets, can also build and explore new perfect models in data-science that are web-based environments. They can work with other data scientists engineers and machine learning engineers, and can solve data science challenges and dares in competitions.

Kaggle was started in 2010 by providing machine learning competitions and it is now a public data platform, Artificial Intelligence education and for data science with a cloud-based aura workbench. I have used Cohn-Kanade dataset which consists of 35,887 images. The images are already in grayscale.

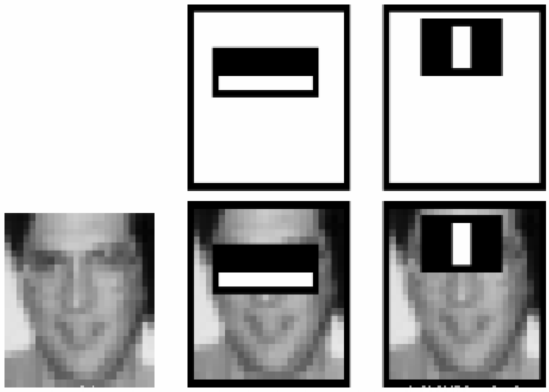
2.2 A Haar classifier, or a Haar cascade classifier, is used in machine learning for detection program that identifies objects or parts of faces in an image and video. This method was proposed by Paul Viola and Michael Jones. Haar Cascade is a machine learning-based approach and has got a lot of images that is positive and negative images and are basically used to train the classifier.

2.2.1 Steps in Haar casacade algorithm for obtaining the classifier:

1. Features are extracted from positive and negative images, for this to be done, the images are divided into different black and white rectangles. A single value is obtained from each feature and it is obtained by subtracting pixels sum that comes under the white rectangle pixels sum under the black rectangle.

2. No matter how large your image is, the calculations will be reduced for a pixel that's given to an operation having just 4 pixels.

Example:



The top of the row shows 2 best features. The feature that is first is selected to focus on the region of the eyes that is often darker than the other regions of the cheeks or nose. The feature that is 2nd is selected depending on the property of the eyes because the eyes region is darker than the nose bridge.

3. Best features are selected with the help of Adaboost training. In this training, each feature has been applied with a set of training images. It finds the best threshold in each feature that will classify the faces to negative and positive. There can be some errors or misclassifications.

4. We can then select the features that has got less rate of error, that means these features are that most classified features accurately on both facial and non-facial images. The process of this is continued until the required accuracy is attained or the error rate is achieved or until the required number of features are found.

2.3 Convolution Neural Network

Convolutional Neural Network (CNN, or ConvNet) is a deep neural network of class, that is most commonly applied and understood and used to analyze and detect imagery of visuals. Space invariant artificial neural networks (SIANN is the short form) are other names are also there to it, depending on the weight shared of the kernels in convolution architecture or filters that are used to slide along with features of input provides equivariant translation responses which are known as feature maps.

Multilayer perceptrons are used to regularize CNNs versions. Perceptrons with multilayers are fully connected networks, that is, each layer is connected to all neurons that is present in each neuron in the next layers. Overfitting data happens because of these networks. CNNs take a unique approach for regularizing and the advantage they take is of data pattern that is hierarchical and patterns are assembled for complexity increase used in their filters of smaller and simpler patterns that are embossed in it.

It has the following layers:

✓ Convolutional layer:

In CNN, the input is of a mathematical object with a shape: (input number) x (height number) x (input width) x (channels of input). Once it has passed through the layer of convolutional, the image becomes habituated to the feature map, and this also called an activation map, with the shape: (inputs number) x (height of the feature map) x(width of feature map)x(feature map channels). A convolutional layer that is within a CNN has the following attributes:

- Width and height define the filters or kernels of CNN (hyperparameters).
- Output channels and input channels numbers are counted (hyper-parameters). Each layer's input channel should be equal to the channel numbers of output (it also called as depth input).
- The convolution operation has additional hyperparameters, like, stride, padding and dilation.

✓ Pooling layers

The dimensions of the data in pooling layers are reduced by combining different outputs of neuron clusters to one single layer into a neuron of single in the next layer. Tiny clusters are there in the local pooling, they tile sizes like 2 x 2 that are used commonly. All the neurons in the feature map are been acted by global pooling.

✓ Fully connected layers

Fully connected layers basically try to connect every neuron to one another in one layer and traditional multi-layer is same as this one, perceptron neural network (MLP) is used in here.

✓ Receptive field

From some number of locations, each neuron receives input in the former layer at the neural networks. Each neuron receives input only from an area that is restricted from the previous layer called the field of neuron's receptive in convolutional layer.

✓ Weights

A specific function is applied to an output value from the values of input received from these receptive field in the previous layer. The input values are applied with functions that is determined by a bias and a vector of weights.

2.4 Evaluation

For evaluation purpose, real time images are given as in, the person comes in front of the system, webcam turns on and starts scanning person's face and detects the emotions. It detects all 7 emotions of a person.

3 IMPLEMENTATION

3.1 Preprocessing

The dataset used was Cohn-Kanade which has got 35,887 images and the images were already in grayscale. Resizing of the images was done to normalize them in the starting part of the code itself. I have used 28709 and for testing. I've used 7178 images. Total number of trainings is 50 because that's when I got maximum accuracy.

3.2 About the libraries used:

NumPy: It is basically used for the conversion of normal image to pixels of 1's and 0's and forming an array.

TensorFlow: TensorFlow is an open-source and free software library used for flow of data and used in programs that are differentiable across a range of tasks.

Keras: A Python interface for artificial neural networks is provided by Keras and it is an open-source. Both Keras and TensorFlow are used for executing the CNN layers.

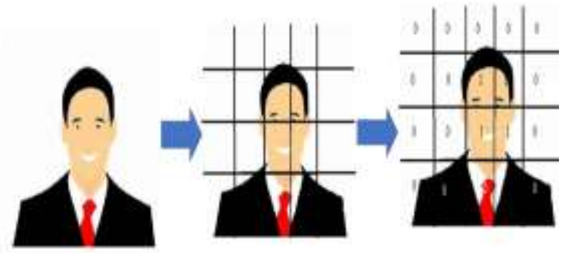
Open CV: OpenCV is a big open-source library used in computer vision, image processing and machine learning algorithms, and now it plays a major role in real-time operations which is very crucial in today's systems. So here, I've used it for turning the webcam on and detecting faces.

3.3 Haar-cascade classifier

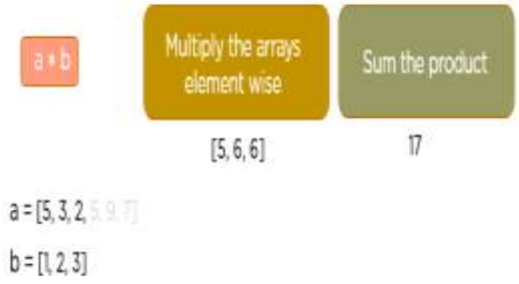
Open CV starts streaming the video and detects the face using bound box. To detect if it's a face or not, it focuses on features and crop them out for comparison. These features are compared with classifications, totally 31 classifications are used. If the features are not recognized, face will not be recognized and the program ends.

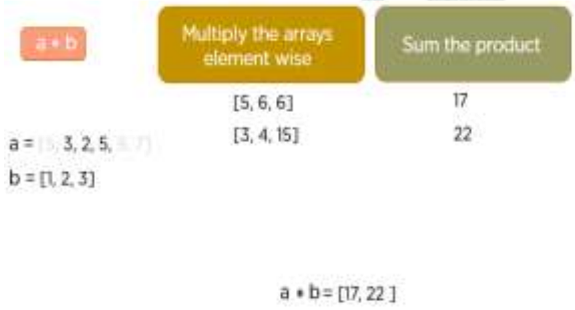
3.4 Number of CNN layers used

Sequential layer: Initially the input images are given in the sequential form, the conversion of normal image to pixels of 1's and 0's and forming an array, is done in this step.

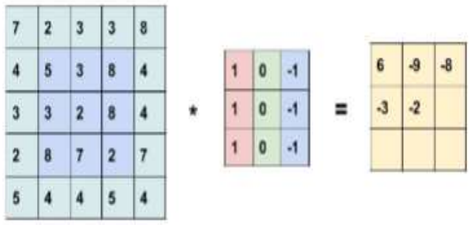


Covn2D layer: Convolutional layer performs convolutional operations, and images are considered to be matrix of pixel values and these matrix arrays are multiplied element wise to form convolute feature maps.



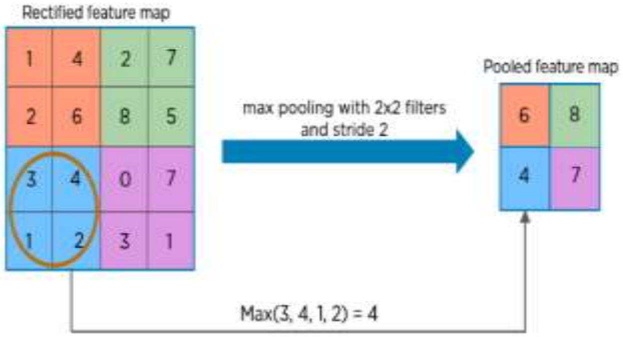


Relu layer: Once the convolute feature map is extracted, we move this to ReLU layer, it performs element wise operation and sets negative pixel to 0(these negative pixels are unwanted parts of the image) and forms another feature map.

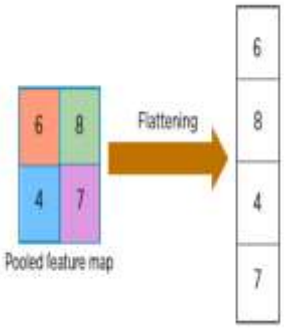


$$\begin{bmatrix} 7 & 2 & 3 & 3 & 8 \\ 4 & 5 & 3 & 8 & 4 \\ 3 & 3 & 2 & 8 & 4 \\ 2 & 8 & 7 & 2 & 7 \\ 5 & 4 & 4 & 5 & 4 \end{bmatrix} * \begin{bmatrix} 1 & 0 & -1 \\ 1 & 0 & -1 \\ 1 & 0 & -1 \end{bmatrix} = \begin{bmatrix} 6 & -9 & -8 \\ -3 & -2 & \\ & & \end{bmatrix}$$

Pooling layer: In this layer, the ReLU feature map is given to pooling layer and it generates pooled feature map, to get this pooled feature map, it extracts features like edges of the image, size of the image, eye size etc.



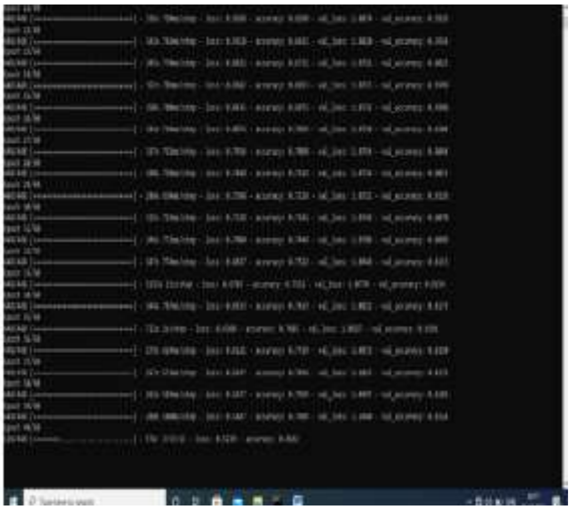
Flattern layer: This will be further sent to flatten layer to convert these features to single long continuous vector.



Dense layer: Then comes the fully connected layer, it consists of Hidden layer. Hidden layer has 3 layers, those are Dense, Dropout and Flatten:

- Dense layer selects images with most prominent dense features
- Dropout layer deletes images those are improper
- Flatten layer is to make the images smooth

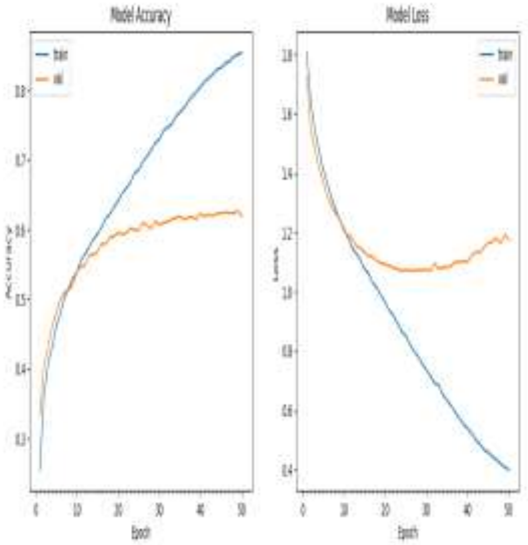
3.5 Output image of training: 50 Epochs was taken as maximum accuracy was obtained at that point.



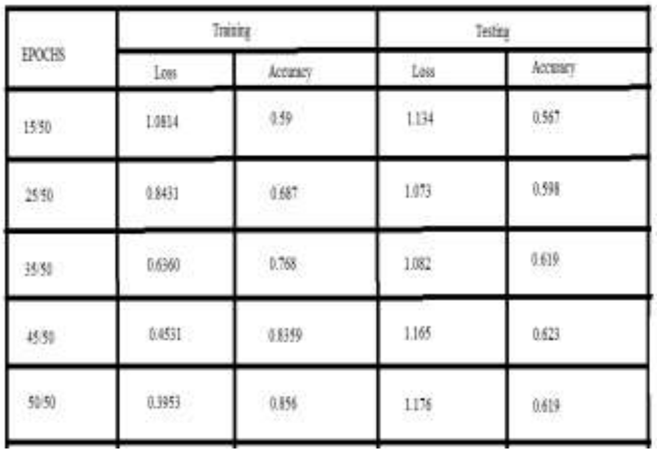
Snapshot 3.1: Training of the Dataset using CNN Model

4 RESULTS AND DISCUSSION

When we see that once the model predicts a certain emotion on an image, low values will be assigned to other emotions. This happens because this is how it has been trained using the dataset. One label is associated with one image so during the training, it was basically taught that one emotion should be detected for one given image. A total of 7 emotions are detected. When compared to approaches like FACS and SVM, there is a high speed at which detection of emotions takes place is high and in SVM, 65% accuracy was achieved and in FACS, 70% accuracy was achieved. The accuracy percentage obtained in this project is 85%. The model can detect both coloured and grayscale images from the dataset and downloaded images from the internet.



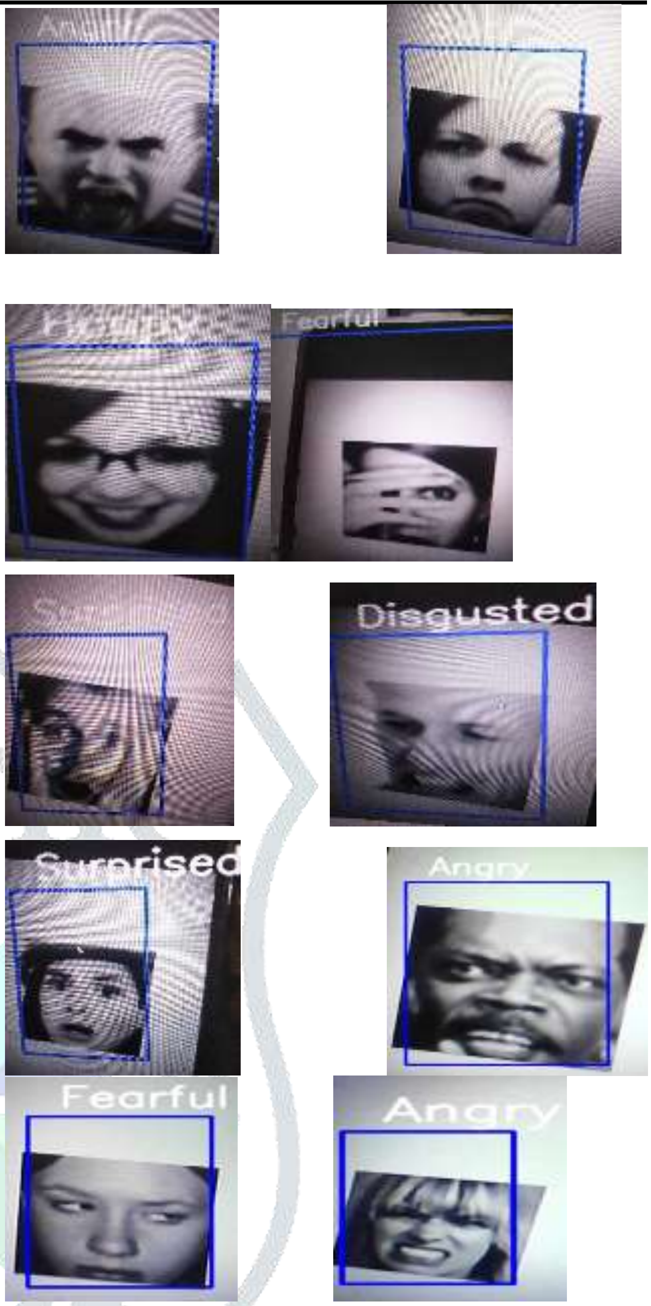
Snapshot 4.1: Graph of Accuracy vs Epoch and Loss vs Epoch



EPOCHS	Training		Testing	
	Loss	Accuracy	Loss	Accuracy
15/50	1.0814	0.59	1.134	0.567
25/50	0.8431	0.687	1.073	0.598
35/50	0.6360	0.768	1.082	0.619
45/50	0.4531	0.8359	1.165	0.623
50/50	0.3953	0.856	1.176	0.619

Table 4.2: Tabular form of Accuracy and Loss of Training and Testing Epochs

The above table shows the loss and accuracy value of training and testing data, maximum accuracy obtained is 85%.



Snapshot 4.3: Emotions detected on test data

The above images show the detected emotions of the dataset used. These images are shown to the webcam via phone and they have been accurately detected.

Snapshot 4.4: Emotions detected on images downloaded from the Internet.

The above images are detected with emotions and they have been downloaded from the Internet.

5 CONCLUSION

In this project, an image processing and classification method have been implemented in which wide range of the face images are used to train the model and it predicts the seven basic human emotions for a given test image. The predictor successfully predicts test data and images that are downloaded from the internet, also from the same dataset used to train the classifiers. The predictor is a little poor at detecting contempt expressions as the model is not trained for this emotion. Also, this is likely due to a combination of lacking training and test images that clearly exhibit poor and contempt pre-training labeling of data. The classifier does predict successfully at predicting emotions for test data that have expressions that do not belong to one of the seven basic expressions, as the model is not been trained for other expressions. Future work should include improving the robustness of the classifiers used in the model by adding more training and testing images from different datasets, investigating and analyzing more accurate detection methods that still maintain efficiency of computation, and considering the classification of more fine and ideal expressions. Also, more images with sad and depressed emotions have to be used in order to get the right prediction of these two emotions.

6 REFERENCES

[1] Balasubramanian, B., Diwan, P., Nadar, R., & Bhatia, A. (2019, April). Analysis of Facial Emotion Recognition. In 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI) (pp. 945-949). IEEE.

[2] Xiaoxi, M., Weisi, L., Dongyan, H., Minghui, D., & Li, H. (2017, August). Facial emotion recognition. In 2017 IEEE 2nd International Conference on Signal and Image Processing (ICSIP) (pp. 77-81). IEEE.

[3] Zhong, B., Qin, Z., Yang, S., Chen, J., Mudrick, N., Taub, M., ... & Lobaton, E. (2017, November). Emotion recognition with facial expressions and physiological signals. In 2017 IEEE Symposium Series on Computational Intelligence (SSCI) (pp. 1-8). IEEE.

[4] Gupta, S. (2018, January). Facial emotion recognition in real-time and static images. In 2018 2nd International Conference on Inventive Systems and Control (ICISC) (pp. 553-560). IEEE.

[5] Tian, X. (2009, December). Face Recognition System and It's Application. In 2009 First International Conference on Information Science and Engineering (pp. 1244-1245). IEEE.

[6] PRAJAKTA B. KULKARNI, International Journal of Industrial Electronics and Electrical Engineering, ISSN: 2393-2835 Volume-5, Issue-6, Jun-2017

[7] Siswanto, A. R. S., Nugroho, A. S., & Galinium, M. (2014, September). Implementation of face recognition algorithm for biometrics based time attendance system. In 2014 International Conference on ICT For Smart Society (ICISS) (pp. 149-154). IEEE.

[8]Azcarate, A., Hageloh, F., Van de Sande, K., & Valenti, R. (2005). Automatic facial emotion recognition. Universiteit van Amsterdam, 1-6.

[9] Ko, B. C. (2018). A brief review of facial emotion recognition based on visual information. sensors, 18(2), 401.

[10] Li, J., & Oussalah, M. (2010, September). Automatic face emotion recognition system. In 2010 IEEE 9th International Conference on Cyberntic Intelligent Systems (pp. 1-6). IEEE.

[11] Sasikumar, K., Ashija, P. A., Jagannath, M., Adalarasu, K., & Nathiya, N. (2006). A Hybrid Approach Based on PCA and LBP for Facial Expression Analysis

[12] Borkar, N. R., & Kuwelkar, S. (2017, July). Real-time implementation of face recognition system. In 2017 International Conference on Computing Methodologies and Communication (ICCMC) (pp. 249-255). IEEE.

[13] Rzayeva, Z., & Alasgarov, E. (2019, October). Facial emotion recognition using convolutional neural networks. In 2019 IEEE 13th international conference on application of information and communication technologies (AICT) (pp. 1-5). IEEE.

[14] Sharma, A. K., Kumar, U., Gupta, S. K., Sharma, U., & LakshmiAgrwal, S. (2018, December). A survey on feature extraction technique for facial expression recognition system. In 2018 4th International Conference on Computing Communication and Automation (ICCCA) (pp. 1-6). IEEE.