Facial Emotions over Static Facial Images Using Deep Learning Techniques with Hysterical Interpretation

Dr A Viswanath Reddy, A Aswini Reddy and C A Bindyashree

1 Solutions Architect, Arohak Inc., New Jersey, USA.
2 UI/UX Developer, Arohak Inc, Bangalore/India.
3 Dept. of ISE, Assistant Professor, The Oxford College of Engineering, Bangalore/India.

Abstract

Recognition of facial expression has many potential applications that have attracted the researcher's attention during the last decade. Taking out of features, is an important step in the analysis of expression that contributes to a quick and accurate recognition of expression, i.e., happiness, surprise and disgust, sadness, anger, and fear are expressions of the faces. Facial expressions are most frequently used to interpret human emotions. Two categories contain a range of different emotions: positive emotions and non-positive emotions. Face Detection, Extraction, Classification, and Recognition are major steps used in the proposed system. The proposed segmentation techniques are applied and compared to determine which method is appropriate for splitting the mouth region, and then the mouth region can be extracted using techniques for stretching contrasts and segmenting the image. After the extraction of the mouth area, the facial emotions are graded in the face picture region of the extracted mouth based on white pixel values. The Supervisory Learning Approach is widely used for face identification algorithms and it takes more computation time and effort. It may also give incorrect class labels in the classification process. For this reason, supervised learning and reinforcement learning is being used. In general, it will be like a trial-and-error method that is, in the training process it tries to learn and produce expected results. It was specified accordingly. Reinforcement learning always tries to enhance the results.

Keywords-- Facial Expression, Face detection, Classification, Segmentation, Supervised learning, Reinforcement learning.

I INTRODUCTION

Automatic emotion recognition is a challenging task that has gained significant scientific interest in recent years due to its applications in crowd analytics, social media, marketing, event detection and summarization, public safety, human-computer interaction, digital security surveillance, street analytic, and image retrieval. The challenge in automated facial expression recognition recognizes each different facial expression and classify it into its respective emotion classes. This topic has a wide implementation area, such as entertainment, education, e-commerce, health and security. This paper aspires to use deep Convolutional neural networks to categories the feeling on an individual's face into one of seven classifications. The scale was formulated using the FER-2013 informational collection, which was distributed in 2013.

International Conference on Machine Learning (ICML). This informational index comprises of 35887grayscale, 48x48 estimated face pictures with seven feelings angry, disgusted, fearful, happy, neutral, sad, and surprised. Our paper's focus is to classify the different types of emotions on a human face into one of seven classes utilizing a Convolutional neural organization. Any human emotion can be identified by observing the movements of eyes, mouth, nose, etc. So here we are implementing three stages to know the emotion of the person. A video or image is given as input to the system. Then the first stage is to detect the face in the image and extracts features like size, background, brightness then the face is identified. The second stage is to detect the emotion from the image then detected features are given to the classifier to identify the expression of the person. The detecting of emotion by facial expression has always been a simple task for humans but doing the same task with a computer system is rather difficult. It is now possible to discern emotions from photos because to recent advances in machine learning and computer vision. In this fields. For example, security agencies can track emails/messages/blogs, etc., and detect suspicious activities. The business communities nowadays prefer to use
emotional marketing. In emotional Convolutional Neural Network (CNN): The first part removes the background from the picture, and the second part concentrates on the facial feature vector extraction. However, the problem of emotion recognition for a group of people has been less extensively studied, but it is slowly gaining popularity due to the massive amount of data available on social networking sites containing images of groups of people participating in social events. Usually, emotion recognition is accomplished through the analysis of a set of facial muscle movements that correspond to a displayed emotion (also known as action units). After detecting and separating the subject face, feature extraction can be applied to predict the contribution of each action unit. However, this approach becomes impractical in unregulated environments where factors like head and body pose variations, occlusions, variable lighting conditions, the variance of actors, varied indoor and outdoor settings, and image quality affect their cognition of emotions.

Overview:
In our daily life, we go through different situations and develop a feeling about them. Emotion is a strong feeling about a human’s situation or relation with others. These feelings and express Emotion is expressed as facial expression. The primary emotion levels are of six types namely, Love, Joy, Anger, Sadness, Fear, and Surprise. Human expresses emotion in different ways including facial expression, speech, gestures/actions, and written text. This article mainly focuses on two expressions namely, written text and speech. As technology progresses, the internet is now commonly used on PCs, tablets, and smartphones. This generates a huge amount of data, especially textual data. It has become impossible to manually analyze all the data for a specific purpose. New research directions have emerged from automatic data analysis like automatic emotion analysis. Emotion analysis has attracted researchers’ attention because of its applications in different paper, we propose a novel technique called facial emotion recognition using Convolution neural networks (FERC). The FERC is based on two-part marketing, they try to stimulate customers ‘emotions to buy products or services. Some emotional marketing advertisements are given at the link.

Background:
A Facial expression is the visible manifestation of the affective state, cognitive activity, intention, personality, and psychopathology of a person and plays a communicative role in interpersonal relations. Human facial expressions can be easily classified into 7 basic emotions: happy, sad, surprise, fear, anger, disgust, and neutral. Our facial emotions are expressed through the activation of specific sets of facial muscles. These sometimes subtle, yet complex, signals in an expression often contain an abundant amount of information about our state of mind. Automatic recognition of facial expressions can be an important component of natural human-machine interfaces; it may also be used in behavioral science and clinical practice. It has been studied for a long period and obtaining progress in recent decades. Though much progress has been made, recognizing facial expressions with a high accuracy remains to be difficult due to the complexity and varieties of facial expressions. On a day-to-day basis, humans commonly recognize emotions by characteristic features, displayed as a part of a facial expression. For instance, happiness is undeniably associated with a smile or an upward movement of the corners of the lips. Similarly, other emotions are characterized by other deformations typical to a particular expression. Research into automatic recognition of facial expressions addresses the problems surrounding the representation and categorization of static or dynamic characteristics of these deformations of face pigmentation. In machine learning, a convolutional neural network (CNN, or ConvNet) is a feed-forward artificial neural network whose connection pattern is inspired by the arrangement of the animal visual cortex. Individual cortical neurons react to upgrades in the responsive field, which is a constrained area of space. The visual field is tiled so because receptive fields of distinct neurons partially overlap. The response of a shooting guard and it is also known as shift-invariant or space-invariant artificial neural network (SIANN), which is named based on its shared weights architecture and translation in variance characteristics.

Statement of problem:
The problem of emotion recognition for a group of people has been less extensively studied, but it is slowly gaining popularity due to the massive amount of data available on social networking sites containing images of groups of people participating in social events. A model must be created that determines the emotion of many people at a time, which is still a challenge in these facial emotion detection papers.
**Purpose of the paper:**

Facial detection has become a largely interesting concept in Deep Learning, now emotion detection is the new concept that is grabbing the interest of people and researchers. Our model's main purpose is to recognize the emotion of the person in front of the camera, which does not just give the emotion but also the percentages of accuracy for each emotion. This percentages accuracy information can be used for analytics purposes also. Our model also determines the emotion of many people at a time, which is still a challenge in these facial emotion detection papers. This model can be used in various situations like in examination or virtual interview processes, to know the frank review of a person while using a product, can also be added as an additional feature to the existing systems in entertainment fields, etc. This model can be deployed in many places and papers which could be an additional feature.

**Motivation:**

Neuron to stimuli within Humans can quickly and even subconsciously assess a multitude of indicators such as word choices, voice inflections, and body language to discern the sentiments of others. This analytical ability likely stems from the fact that humans share a universal set of fundamental emotions. Significantly, these emotions are exhibited through facial expressions that are consistently correspondent. This means that regardless of language and cultural barriers, there will always be a set of fundamental facial expressions that people assess and communicate with. After extensive research, it is now generally agreed that humans share seven facial expressions that reflect the experiencing of fundamental emotions. These fundamental emotions are anger, contempt, disgust, fear, happiness, sadness, and surprise. Unless a person actively suppresses their expressions, examining a person’s face can be one method of effectively discerning their genuine mood and reactions. Facial emotion recognition is a task that can also be accomplished by computers. Furthermore, like many other important tasks, computers can provide advantages over humans in analysis and problem-solving.

**Deep learning:**

Deep learning is a machine learning technique that trains PCs to do what falls into place without any issues for people. Deep Learning is also known as Deep Structured Learning. If we consider an example: Profound learning is a critical innovation behind driver less vehicles, empowering them to perceive a stop sign or to recognize a walker from a light post. It is the way to voice control in customer gadgets like telephones, tablets, televisions, and sans hands speakers. Profound learning is getting loads of consideration of late and in light of current circumstances. It’s accomplishing results that were impractical previously. In profound learning, a PC model figures out how to perform arrangement assignments straightforwardly from pictures, text, or sound. Profound learning models can accomplish cutting edge exactness, now and again surpassing human-level execution. Models are prepared by utilizing an enormous arrangement of named information and neural organization designs that contain many layers. In 2016 there was talk going where it is said that “Deep Learning for Building Intelligent Computer Systems" was When you hear the term deep learning, just think of a large deep neural net. Deep refersto the number of layers typically and so this kind of the popular term that’s been adopted in the press. I think of them as deep neural networks generally.

The difference we see while we learn machine learning and deep learning is that in ML, we see that the term machine learning refers to a technology that enables a device to perform a task without any human intervention. In other words, machine learning is that field of data science that consists of the algorithms that perform the learning procedure without human assistance. And if we see in Deep Learning is that while deep learning is referred to the procedure which is used to implement machine learning. Deep learning is a configuration of machine learning which is inspired by the framework of the human brain and is peculiarly efficient in feature detection.
CNN:
The agenda for this field is to enable machines to view the world as humans do, perceive it similarly and even use the knowledge for a multitude of tasks such as Image & Video recognition, Image Analysis & Classification, Media Recreation, Recommendation Systems, Natural Language Processing, etc. The advancements in Computer Vision with Deep Learning have been constructed and perfected with time, primarily over one particular algorithm—a Convolutional Neural Network. The following operations are available by default.

A Convolutional Neural Network (ConvNet/CNN) is a Deep Learning algorithm that can take in an input image, assign importance (learnable weights and biases) to various aspects/objects in the image, and be able to differentiate one from the other. The pre-processing required in a ConvNet is much lower as compared to other classification algorithms. While in primitive methods filters are hand-engineered, with enough training, ConvNets can learn these filters/characteristics.

2. LITERATURE SURVEY

Facial expression is the common signal for all humans to convey the mood. There are many attempts to make an automatic facial expression analysis tool as it has application in many fields such as robotics, medicine, driving assist systems, and lie detector. Since the twentieth century, Ekman et al. [11] defined seven basic emotions, irrespective of culture in which a human grows with the seven expressions (anger, feared, happy, sad, contempt [12], disgust, and surprise). There are many existing systems related to this field like face detection, age prediction, gender prediction, etc. Some of the models are mentioned below.

A facial recognition system is a technology capable of matching a human face from a digital image or a video
frame against a database of faces, typically employed to authenticate users through ID verification services, works by pinpointing and measuring facial features from a given image. While initially a form of computer application, facial recognition systems have seen wider uses in recent times on smartphones and in other forms of technology, such as robotics. Because computerized facial recognition involves the measurement of a human's physiological characteristics, facial recognition systems are categorized as biometrics. Although the accuracy of facial recognition systems as biometric technology is lower than iris recognition and fingerprint recognition, it is widely adopted due to its contactless process. Facial recognition systems have been deployed in advanced human-computer interaction, video surveillance, and automatic indexing of images. They are also used widely by law enforcement agencies. This system detects the face of a person from different objects in the input the face detection performs three steps like face identification, feature extraction, classification. This system only specifies if the face is detected or not. It is implemented by using LBP (Local Binary Patterns) algorithm. This algorithm is less accurate by comparing to the HAAR cascade. So, still, accuracy is a problem and it can't tell difference between two identical twins. Person detection by recognizing the face. This paper detects the name of the person, gender of the person by using the person's facial features.

The face is one of the most widely used biometrics for human identity authentication. Facial recognition has remained interesting and active research is in the past ever al decades due to its ever-growing applications in biometric authentication, content-based data retrieval, video surveillance, access control, and social media. Unlike other biometric systems, facial systems work independently without involving the individual, due to which it does not add unnecessary delay. Its ability of recognizing multiple persons at a time further adds to its speed. There are many face recognition methods based on traditional machine learning that are available in the literature. Improvements are being made with the constant developments in computer vision and machine learning. However, most of the traditional methods lack robustness against varying illumination, facial expression, scale, occlusions and pose. With the advent of big data and graphical computing, deep learning has impressively advanced the traditional computer vision systems over the past decade. In this paper, we present a Convolutional neural network-based face recognition system which detects faces in an input image using Viola Jones face detector and automatically extracts facial features from detected faces using a pre-trained CNN for recognition. A large database official images of subjects is created, which is augmented in order to increase the number of images per subject and to incorporate different illumination and noise conditions for optimal training of the Convolutional neural network. Moreover, an optimal pretrained CNN model along with a Set Of hyper parameters is experimentally selected for deep face recognition. Promising experimental results, with an overall accuracy of 98.76%, are obtained which depict the effectiveness of deep face recognition in automated biometric authentication systems. Biometric identification is the automatic verification of any person based on some features of the person. In this face recognition they have used SVM classification, so by using this classification it is not suitable to perform large datasets. CNN classification is better than SVM classification. By using this CNN classification, it automatically detects the important features without any human action.

Age detection has become famous as it’s interesting to the people as an additional feature in various applications. In this, the age of the person will be detected after the face detection. The age will depend on the facial points like forehead, cheeks, wrinkles, eyes, etc. features. Since deep CNN has been proved to be rich in scene and target representation ability, it has intelligent and automatic performance compared with the traditional manual feature extraction method. In recognition-based this paper, the aging network (agent) is used to extract the face age descriptor and then the age estimation is carried out by using the divide-and-rule strategy. The Age Net uses an approach is based on regression and classification to construct an age-estimated deep CNN. To reduce the complexity of the network and training time, we only use the regression-based age estimation deep CNN.

3. PROPOSED ARCHITECTURE

The proposed model determines the emotion of the face of a human in front of the camera. We chose Deep Learning to implement the model as it is better for classification and image processing problems. The model uses Convolutional Neural Network (CNN) to determine the emotion. We have implemented two models for the emotion detection analysis, which gives the same emotion output but in a unique way of its own. This model does the face recognition and emotion prediction along with the accuracy for each emotion. The accuracies of the
emotions will be displayed in another frame which will be beside the face detection frame. But, this model can detect only one face emotion at a time, due to the overlapping problem. The accuracies or probabilities of the face overlap if more than one face is detected so it just detects single facial emotion. Our secondary model determines the emotion, for any number of faces in one frame and gives the emotion on the face with the bounding box. It also produces curves for the accuracy of the model. This model can detect the emotion of more than one face and it produces the curves of the tested and trained data, i.e.; the expected accuracy and the accuracy received while testing the model. We have constructed two models for the Purpose of obtaining emotion percentage of every emotion for many faces. But if we consider multiple faces we face the overlapping problem. Our model uses Keras, Tensor flow for the data processing. The model can be developed in different ways by machine learning, using just computer vision but we choose Keras as it is powerful in processing when compares to others in Deep Learning Implementation.

![Various emotions](image)

**Figure 3: Various emotions**

*Architecture:*  
Different steps in the architecture:

*Input face image:*  
In this phase, the video input will be taken from the user, i.e.; live capture from the camera, and the video will be made into multiple images. This can be done using Open Cv inbuilt methods.

*Pre-processing:*  
In this step, the image will be further filtered into the required form. The image will be made into 48 X 48-pixel size and will be turned into grayscale. Grayscale conversion is important as the colors add the additional weights while processing which leads to the diversion from the main aim and will require to be handled separately.

*Feature extraction:*  
In this step, the features of the face will be identified. We use Haar Cascade Classifier to achieve this. The face, eyes, mouth are detected using Haar Classifier. Using the weight soft these features further emotion classification can be processed.

*Classification:*  
The emotion classification will be done based on the values achieved from the training data. The features will be further processed using CNN and pooling multiple times until the required result is achieved and then, the final dense and flattened layer will be obtained. The final flattened layer will be used for predicting emotions. The emotion classes should be priorly defined before this processing. Then the image will be again turned into its original form from 48 X 48 and turned into a color scale. This image will consist of the
emotion name along with the boundary box in the output stage.

- **Modules:**
  - **Input and Pre-processing:**
    This is the first step of the model construction. Here the image input will be taken from the user using the OpenCV `read()` method. Then the video will be converted into images of 48X 48pixel size and into grey scale.
  - **Face detection and feature extraction:**
    This step involves the HaarCascadeClassifier.xml file to be included in the paper file. Here the cascade classifier should be applied for the eyes, face, and mouth.
  - **Emotion detection:**
    Here the Convolutional Neural Network is applied by multiple convolutions and pooling using the activation functions like ReLU, SoftMax. The model should be first prepared and then the images will be passed to the model.
  - **Output with estimated emotion:**
    The output will be given using the boundary box around the face of the person detected and emotion above the box & a separate frame having rectangles representing the accuracies of each emotion. The compilation of the paper must be done through the anaconda command, but not through the anaconda spyder platform.

![Figure 4. Sequence diagram of the model](image)

![Figure 5. Collaboration diagram of the system](image)
Figure 6. Module 1. Input and Pre-processing

Figure 7. Face detection and feature extraction

Figure 8. Emotion Detection
4 IMPLEMENTATION

In our model we used this algorithm for face, eye, mouth detections. Haar Cascade is machine learning Object Detection algorithm. It was proposed by Paul Viola & Michael Jones in 2001 in their paper “Rapid Object Detection using a Boosted Cascade of Simple Features”. In this algorithm the cascade classifier is trained with the positive images and negative images. Where positive images are the images containing object whereas negative images are those which doesn’t contain the object in the images. So, after this the algorithm gets trained and determines the object from the image. We have the pre-trained algorithm in Computer Vision package, which was used in our model creation process. In our model we use the Haar Cascade Classifier.xml file and get it into model using Cascade Classifier () function, which is present in OpenCV.

Convolutional Neural Network:

The pre-processed image is inserted into the layer. Convolutional Neural Network is algorithm present in Deep Learning for image processing, by using the weights in the image. This network contains different layers where each layer has input and output. Convolutional layer has the input and output in the form of array vectors. Convolutional Layer scan be 1d, 2d, 3d varying by the input. For the voice and text input the layer will be 1d and for images it is 2d. So, in the model we use 2d convolutional layer. Each layer uses an activation function for processing that layer. Activation functions are like Soft Max, linear, sigmoid. Here we use only Soft Max as sigmoid function returns the output values only between 0&1. SoftMax Activation function: It is a mathematical function, which converts a vector of numbers into a vector of probabilities, where the probabilities of each value are proportional to the relative scale of each value in vector. It normalizes the output and converts the weights of into probabilities which sum to one value.
Figure: 10. Convolutional Neural Network

- **Pooling:**
  Pooling layer is applied after the convolutional layer, to remove the overfitting of model i.e.; removal of unnecessary values of the data. Pooling layer filters the input given. Here pool size should be chosen to pool the values of that layer. For every pool the pooling function is applied.
  
  Pooling is of mainly two types:
  - Max Pooling: In this pooling method the maximum value from the input values is given as output for every pool.
  - Average Pooling: This method finds the average value of every pool and gives it as output.

**Example:** Input data is of size 5x5 then pool size can be 2x2, 3x3 etc. So, for every 2x2 (or) 3x3 of the current pool will be processed according to the type of function. The inbuilt functions for these algorithms are available in ‘Keras’ package.

- **Implementation Steps:**
  - Import all the required packages like TensorFlow, Keras, NumPy, pandas, os and Haar Cascade classifier.
  - Convert the image input into images, resizing images, converting images to grayscale.
  - Creating a model of convolutional layers along with the pooling layers and flattening the final layer.
  - Training the model with the existing data set and producing, saving the output.
  - Face, eyes, mouth detection using Haar Classifer algorithm and passing them to the model.
  - Testing model with the input data and comparing the results and determining the emotion.
  - Producing the curves for the trained and input data.
  - The output with emotion and bounding box on face and separate frame where the accuracy for each emotion will be displayed.
  - The above steps are to be done for obtaining the final output. After the steps the graphs, determined emotion will be the final output. The output curves give the information about the accuracy of the model. And neural networks i.e. convolutional networks are created as in the following image while processing the image.
  - Dataset: For this project we are using FEB-2013 dataset. The FER-2013 data set consists of 28,709 labelled images in the training set, 35,887 labelled images in the development set, and 7,178 images in the test set. In this dataset each image belongs to any of these emotions: happy, sad, angry, afraid, surprise, disgust, and neutral, with happy being the most prevalent emotion, providing a baseline for random guessing of 24.4%. The images in FER-2013 consist of grayscale and 48x48 pixels. The FER-
2013 dataset consists of the results of a Google image search of each emotion and related images of the emotions.

- Our data set is as below:

Figure 11. Project folder

Figure 12. Dataset Folder

Figure 13. Train and Test
Figure 14. Images of the dataset

- **Source Code:**
  
  ```python
  import numpy as np
  import argparse
  import matplotlib.pyplot as plt
  import cv2
  from tensorflow.keras.models import Sequential
  from tensorflow.keras.layers import Dense, Dropout, Flatten
  from tensorflow.keras.layers import Conv2D
  from tensorflow.keras.optimizers import Adam
  from tensorflow.keras.preprocessing.image import ImageDataGenerator
  import os
  os.environ['TF_CPP_MIN_LOG_LEVEL'] = '2'

  # command line argument
  ap = argparse.ArgumentParser()
  ap.add_argument("--mode", help="train/display")
  Mode = ap.parse_args().mode

  # plots accuracy and loss curves
  def plot_model_history(model_history):
      """ Plot Accuracy and Loss curves given the model_history """
      fig, axs = plt.subplots(1, 2, figsize=(15, 5))
      axs[0].plot(range(1, len(model_history['accuracy'])+1), model_history['accuracy'])
      axs[0].plot(range(1, len(model_history['val_accuracy'])+1), model_history['val_accuracy'])
      axs[0].set_title('Model Accuracy')
      axs[0].set_ylabel('Accuracy')
      axs[0].set_xlabel('Epoch')
      axs[0].set_xticks(np.arange(1, len(model_history['accuracy'])+1), len(model_history['accuracy'])/10)
      axs[0].legend(['train', 'val'], loc='best')

  # summarize history for accuracy
  axs[0].plot(range(1, len(model_history.history['accuracy'])+1), model_history.history['accuracy'])
  axs[0].plot(range(1, len(model_history.history['val_accuracy'])+1), model_history.history['val_accuracy'])
  axs[0].set_title('Model Accuracy')
  axs[0].set_ylabel('Accuracy')
  axs[0].set_xlabel('Epoch')
  axs[0].set_xticks(np.arange(1, len(model_history.history['accuracy'])+1), len(model_history.history['accuracy'])/10)
  axs[0].legend(['train', 'val'], loc='best')
  # summarize history for loss
axs[1].plot(range(1,len(model_history.history['loss'])+1),model_history.history['loss'])
axs[1].plot(range(1,len(model_history.history['val_loss'])+1),model_history.history['val_loss'])
axs[1].set_title('Model Loss')
axs[1].set_ylabel('Loss')
axs[1].set_xlabel('Epoch')
axs[1].set_xticks(np.arange(1,len(model_history.history['loss'])+1),len(model_history.history['loss'])*10)
axs[1].legend(['train', 'val'], loc='best')
fig.savefig ('plot.png')
plt.show ()

# define data generators
train_dir = 'data/train'
val_dir = 'data/test'
um_train = 28709
num_val = 7178
batch_size = 64
num_epoch = 50
train_datagen = ImageDataGenerator (rescale=1./255)
val_datagen = ImageDataGenerator (rescale=1./255)
train_generator = train_datagen.flow_from_directory(  
    train_dir,  
    target_size=(48,48),  
    batch_size=batch_size,  
    color_mode="grayscale",  
    class_mode='categorical')
validation_generator = val_datagen.flow_from_directory(  
    val_dir,  
    target_size=(48,48),  
    batch_size=batch_size,  
    color_mode="grayscale",  
    class_mode='categorical')

# Create the model
model = Sequential()  
model.add(Conv2D(32, kernel_size=(3, 3), activation='relu', input_shape=(48,48,1)))  
model.add(Conv2D(64, kernel_size=(3, 3), activation='relu'))  
model.add(MaxPooling2D(pool_size=(2, 2)))  
model.add(Dropout(0.25))  
model.add(Conv2D(128, kernel_size=(3, 3), activation='relu'))  
model.add(MaxPooling2D(pool_size=(2, 2)))  
model.add(Conv2D(128, kernel_size=(3, 3), activation='relu'))  
model.add(MaxPooling2D(pool_size=(2, 2)))  
model.add(Dropout(0.25))  
model.add(Flatten())  
model.add(Dense(1024, activation='relu'))  
model.add(Dropout(0.5))  
model.add(Dense(7, activation='softmax'))

# If you want to train the same model or try other models, go for this
if mode == "train":  
    model.compile(loss='categorical_crossentropy',optimizer=Adam(lr=0.0001, decay=1e-6),metrics=['accuracy'])  
    model_info = model.fit_generator(  
        train_generator,  
        steps_per_epoch=num_train // batch_size,
epochs=num_epoch,
validation_data=validation_generator,
validation_steps=num_val // batch_size)
plot_model_history(model_info)
model.save_weights('model.h5')

# emotions will be displayed on your face from the webcam feed
elif mode == "display":
    model.load_weights('model.h5')

# prevents openCL usage and unnecessary logging messages
cv2.ocl.setUseOpenCL(False)

# dictionary which assigns each label an emotion (alphabetical order)
emotion_dict = {0: "Angry", 1: "Disgusted", 2: "Fearful", 3: "Happy", 4: "Neutral", 5: "Sad", 6: "Surprised"}

# start the webcam feed
    cap = cv2.VideoCapture(0)
    while True:
        # Find haar cascade to draw bounding box around face
        ret, frame = cap.read()
        if not ret:
            break
        facecasc = cv2.CascadeClassifier('haarcascade_frontalface_default.xml')
        gray = cv2.cvtColor(frame, cv2.COLOR_BGR2GRAY)
        faces = facecasc.detectMultiScale(gray, scaleFactor=1.3, minNeighbors=5)
        for (x, y, w, h) in faces:
            cv2.rectangle(frame, (x, y-50), (x+w, y+h+10), (255, 0, 0), 2)
            roi_gray = gray[y:y+h, x:x+w]
            cropped_img = np.expand_dims(np.expand_dims(cv2.resize(roi_gray, (48, 48)), -1), 0)
            prediction = model.predict(cropped_img)
            maxindex = int(np.argmax(prediction))
            cv2.putText(frame, emotion_dict[maxindex], (x+20, y-60), cv2.FONT_HERSHEY_SIMPLEX, 1, (255, 255, 255), 2, cv2.LINE_AA)
        cv2.imshow('Video', cv2.resize(frame, (1600,960), interpolation = cv2.INTER_CUBIC))
        if cv2.waitKey(1) & 0xFF == ord('q'):
            break
    cap.release()  
    cv2.destroyAllWindows()

5. RESULT

The FER-2013 dataset consists of 28,709 labeled images in the training set, 35,887 labeled images in the development set, and 7,178 images in the test set; each image belongs to any of these emotions: happy, sad, angry, afraid, surprise, disgust, and neutral, with happiness being the most prevalent emotion, providing a baseline for random guessing of 24.4%.
Figure 15. Test and Train folder

- Estimation Plots:

  After training and testing the dataset, it will generate two plots are Model Accuracy and Model Loss by these plots we can estimate paper.

  ![Model Accuracy](image)

  **Figure 16. Model Accuracy**

  ![Model Loss](image)

  **Figure 17. Model Loss**

  After this, we can detect the emotion of a given input image.
6. CONCLUSION

In this paper, research is to divided the emotions of natural facial detection over the consistent facial image model using deep learning techniques was developed. This is a compound issue that has previously been addressed. Many times, with other strategic methods. Here the results have been achieved using feature engineering, this paper is engrossed on feature learning, which is one of Deep learning assurance. While the outcome performed were not advanced, they were moderately better than other methods including feature engineering. It means that in the due course, Deep Learning approach will be able to resolve this issue by giving ample stamped illustrations. While feature engineering is not mandatory, image pre-processing uplift the categorization of efficiency. Hence, it lowers the noise on the input data. These days, facial emotion detection software includes the use of feature engineering. A emulsion based on feature learning does not seem near yet because of a high drawback: the deficiency of an immeasurable dataset of emotions. For exemplar, Image Net contest uses a dataset containing 1419712 images. By having a enormous data set, networks with a huge potential to streak attribute quality could be expected. Thus, emotion classing could be accomplished using deep learning methods very effectively.

REFERENCES

[1] Y. Lecan, Y. Bengo, and G. Hintin, “Deep learning,” Nature, vol. 521, no. 7553, pp. 436–444, 2015.
[2] J. Ngiam, A. Khosla, M. Kaml, J. Nam, H. Lee, and A. Y. Ng, “Multimodal Deep Learning,” Proc. 28th Int. Conf. Mach. Learn., pp. 689–696, 2011.
[3] J. Schmidhuber, “Deep Learning in Neural Networks: An Overview,” Neural Networks, vol. 61, pp. 85–117, 2015.
[4] C. Busso, Z. Deng, S. Yildirim, M. Bulut, C. Lee, A. Kazemzadeh, et al., “Analysis of emotion recognition using facial expressions speech and multimodal information”, ACM International Conference on Multimodal Interfaces, 2004.
[5] Campadelli P, Lanzarotti R, Lipori G (2006) Eye localization: a survey. In: Esposito A et al (eds) Fundamentals of verbal and nonverbal communication and the biometric issue. IOS Press BV, Amsterdam, pp 234–245
[6] Ludmila I. Kuncheva, Thomas Christy, Iestyn Pierce, Sa’ad P. Mansoor, “Multi-modal Biometric Emotion Recognition Using Classifier Ensembles”
[7] Maheen Zulfiqar, Fatima Syed, Muhammad Jaleed Khan, Khurram Khurshid Deep Face Recognition for Biometric Authentication.
[8] Y Li, S Gong, J Sherrah, H Liddell - “Support vector machine based multi-view face detection and recognition” Image and Vision Computing, 2004 – Elsevier

[9] P.Devika “A smart information system for counting people” JARDCS ISSN 1943-023X 2018

[10] K. Sai Prasad & Rajender Miryala “Histopathological Image Classification Using Deep Learning Techniques” International Journal on Emerging Technologies 10(2): 467-473 (2019).

[11] G. sowmya “Machine Learning and mining for social media analytics” Springer; ISBN978- 981-13-1580-0; 2019.

[12] Sowmya G, Navya K, Divya Jyothi G.” Machine learning and mining for social media analytics” Advances in Intelligent Systems and computing 978-981(2019).

[13] Subhashini Valluru “Stock market Analysis and Prediction using Machine Learning”.

[14] B.Schaller, G.Rigull, and M.Ling,”Emotion recognition combining acoustic features and information in a neural network-belief network architecture,” Acoust. Speech, Signal Process, vol. 1, pp. 577–580, 2004.

[15] Kishor Kumar Reddy C and Vijaya Babu B, “ISPM: Improved Snow Prediction Model to Nowcast the Presence of Snow/No-Snow”, International Review on Computers and Software, 2015

[16] Kishor Kumar Reddy C, Rupa C H and Vijaya Babu B, “SLGAS: Supervised Learning using Gain Ratio as Attribute Selection Measure to Nowcast Snow/No-Snow”, International Review on Computers and Software, 2015

[17] Kishor Kumar Reddy C, Rupa C H and Vijaya Babu B, “A Pragmatic Methodology to Predict the Presence of Snow/No-Snow using Supervised Learning Methodologies”, International Journal of Applied Engineering Research, 2014.

[18] Link: https://scholarworks.sjsu.edu/cgi/viewcontent.cgi?article=1643&context=etd_projects

[19] Link: https://ieeexplore.ieee.org/document/9155948

[20] Link: https://www.slideshare.net/AshwinRachha/facial-emotion-recognition-a-deep-learning-approach

[21] Link: https://www.slideshare.net/Sandywakchaure/facial-expression-recognition-system-using-deep-convolutional-neural-networks

[22] Link: https://www.pantechsolutions.net/face-emotion-recognition-with-deep-learning

[23] Link: https://visionintelligence.github.io/FER

[24] Link: https://www.scitepress.org/papers/2017/63579/63579.pdf

[25] K. A. S. V. Rathnayake, W. K. I. L. Wanniarachchi, W H K P Nanayakkara. (2017, November). Voice and Gesture