Facial Expression Analysis for Emotion and Behavior of Online Learner and Framework for Content Adaptation

Darshankumar H Hingu
Department of Computer Science and Engineering, SOCSE, SandipUniversity, Nashik, India

Dr. Sajidullah S Khan
HOD, Department of Computer Science and Engineering, SOCSE, SandipUniversity, Nashik, India

Mr. Aditya K Sinha
Associate Director HOD-ACTS CDAC, Pune, India

***Abstract*—In online learning, learners don't have a teacher to watch the facial expression of the learner while learning. We recognize the facial expression of online learners through the Webcam and we identify the emotion and behavior of the learner while learning. We can analyze the learner with respect to their learning style and performance of learning. With the help of object detection techniques, we can detect the facial expression of the learner and classify the learner according to the facial expression while learning. After identifying the facial expression of the learner if learner expression is negative then we can adapt content accordingly to generate interest with the help of content adaptation.**

Keywords—Facial Expression Recognition, Convolutional Neural Network (CNN), Content Adaptation, Leaning Style, Felder-Silverman Learning Style

I. INTRODUCTION

Online learning is the mode of learning that happens over the internet. It's referred as "E-learning". It is just type of distance learning that means learning takes place from the remote location and not in classroom. The main difference between online and offline learning is that in online learning learners can learn from anywhere and at any time while in offline mode learners have to attend classes where teachers can see the expression on the faces of the learner and take the teaching decision accordingly as facial expression of the learner is very useful in learning. It is very important in the online learning system to analyze the facial expression of the learner that gives feedback to the instruction system about the leaner so that system will know the emotions, behavior and accordingly content adaptation will be done. Learning style is a major aspect in online learning so that learning style of the online learner can be identified with the help of the various learning style model and analyze the learner with respect to their learning style and performance of learning. The learning styles are identified by various models such as Felder-Silverman learning style, 4MAT learning style, Kolb learning style etc. and among all models; Felder Silverman model is widely used for the engineering education. In Online learning, facial emotions of the learner can be classifying the learners into different categories of the learner and we can analyze the different learning style with the help of behavior and emotions of the learner while learning. Facial expression analysis is the task of recognition of the facial expression from face images or videos in to different categories such as happy, sad, anger, fear, surprise, contempt and so on. For recognizing facial expression, we can use various object detection techniques such as YOLO, SSD, Fast-RCNN, Faster-RCNN etc. All these detection techniques use convolutional neural network model such as AlexNet, GoogleNet, and VGG16 etc. to recognize, detect and categorize the objects present in the images. And after recognition facial expression we can use various machine learning algorithms such as SVM, Decision tree for classifying the face expression.

II. RELATED WORK

There are various researches and studies have been over the facial expression analysis for emotion and behavior of online learner and framework for content adaptation and its implementation approaches and many more. Some of the studies and implementations from other researchers are described further in this section.

Gupta, S., [3] In this paper, they have used Cohn-Kanade Database (CK) and the Extended Cohn-Kanade (CK+) datasets. Haar-cascade filter from OpenCV is used to crop the face from the image. After the face is cropped and processed for detection of facial landmarks. Then using landmarks, the datasets are trained with the help support vector machine (SVM), was used to classify the emotions and classify datasets such as 0= happy, 1=sadness, 2=fear, 3=anger, 4=surprise, 5=disgust, 6=contempt, 7=neutral. And using SVM they got an accuracy of 94.1%. The emotion detection model is divided into 3 steps: face detection, feature extraction, and emotion classification.

Kumari, J., Rajesh, R. and Pooja, K.M.,[4] In this paper, The popular feature extraction techniques are Local Binary Patterns (LBP), Principal Component Analysis (PCA), Independent Component Analysis (ICA), Linear Discriminant Analysis (LDA), Local Gradient Code (LGC), and Local Directional Pattern (LDP) and Histogram of gradient orientations (HOG). The classification algorithm was used such as Support Vector Machine (SVM) and K-nearest neighbour (K-NN) to classify the feature into the different facial expression classes. The geometry and appearance-based methods were used for the automatic Facial expression recognition systems. JAFFE dataset has been used.

Dewan, M.A.A., Lin, F., Wen, D., Murshed, M. and Uddin,[5] They have used Local Directional Pattern (LDP) and Kernel Principal Component Analysis (KPCA) to detect online learner's engagement through their facial expression. The proposed work is comprised of five different modules: face detection, tracking, face representation, classification, and fusion. They have used

appearance-based techniques for the feature extraction and with geometric based techniques the facial feature point detection was carried out. The experiments were conducted on the Dataset for the Affective States in E-Environments (DAISEE).

In our proposed method, we have Recognized the facial expression of the online learner with the help of Convolutional Neural Network (CNN) which is used for the detection and classification of the facial expression of the learner while learning. For emotion detection initially, Haar-cascade filter from OpenCV is used to crop the faces from the webcam. After the face is cropped and processed for detection of facial landmarks. The CNN is used to extract the features from the input webcam with the help of the number of convolutional layers followed by the number of max-pooling layers and finally, it gets flatten up the input to the dense layers and classified faces into the various expressions and labelled with it. It is faster than the existing system with the real-time detection of the facial expression of the learner.

III. TECHNOLOGIES USED

A. *Convolutional Neural Network (CNN)*

The Convolutional Neural Networks is one of the types of neural networks and it is mainly used in the branch of computer vision. CNN is mainly used for the image classifications, image recognition, object detection, facial expression recognition etc. It takes images/video as an input, process it and classify, recognize, and detect that input image and gives the specified output. It consists of various hidden layers such as a convolutional layer, pooling layer, fully connected layer and normalization layer. It has properties such as sparse connectivity and weight sharing. To train and test CNN models, each input image will pass through a combination of convolution layers with filters, Pooling layers, fully connected layers (FC) and apply SoftMax function to classify an object with probability values ranging between 0 and 1.[12]

1) Convolutional Layer

It is the main part of CNN. It is used to extract features from the given input image by applying various filters on the input image in a sliding window manner from left to right and top to down. The filters such as edge detection, blur image, sharpen image etc. Each filter applying to the 3D image will produce 2D output and putting together the output of multiple filters leads to 3D output. If H and W is the height and width of input image respectively and suppose Fh and Fw is the height and width of the filter respectively then the output image will produce is of the size of height is (H-Fh+1) and width is (W-Fw+1).[12]

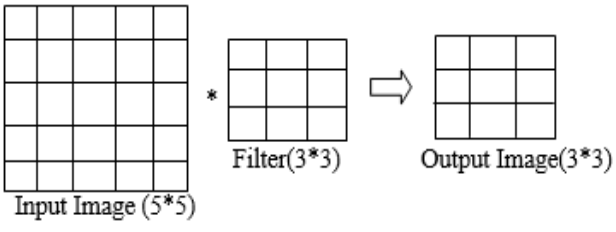

Fig.1. Convolutional Operation [12]

a) Stride

It is the process of number of pixels move over the input matrix at a time in the sliding window process. When stride is 1 then we shift the filter one pixel from left to right and 1 pixel from top to down. When stride is 2 then we shift the filter two pixels from left to right and 2 pixels from top to down. [12]

b) Padding

Sometimes filter does not accurately fit the input image so in that case, padding is used. Zero padding is the concept in which we pad the input image size with an outer boundary of zeros so that it fits can perfectly. Valid padding is the concept in which we don't apply zero padding and it fits the image where the filter fits and drops other parts of the input image. [12]

c) Non-Linearity (ReLU)

ReLU is an activation function and it stands for Rectified Linear Unit. It is mainly used with the convolutional operation. It is used to introduce non-linearity in the convolutional network. Since negative values will be converted into non-negative values. The output is **f(x) = max (0, x).** [12]

2) Pooling Layer

It is used to reduce the size of the input image by half of its original size if the stride is 2 when the image is too large. It is also called as down sampling which is used to reduce the dimensionality of the feature map (input image) but retains the original information. There are different types of pooling operations such as: [12]

a) Max Pooling

In this, for every filter operation with input image (feature map), the largest pixel value will be taken from the feature map.

b) Sum Pooling

In this, for every filter operation with input image (feature map), the sum of all pixel value will be taken from the feature map.

c) Average Pooling

In this, for every filter operation with input image (feature map), the average of all pixel value will be taken from the feature map.

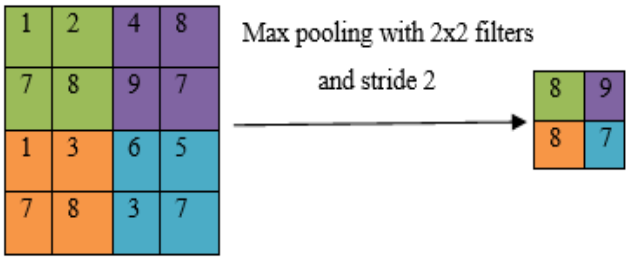

Fig 2. Max Pooling Operation [12]

3) Fully Connected Layer

The optimized output matrix which we will get after the last layer that will be flattened up and converted into a vector and that flattened vector feed into a fully connected neural network and at the end softmax classifier is used to classify the given image input. [12]

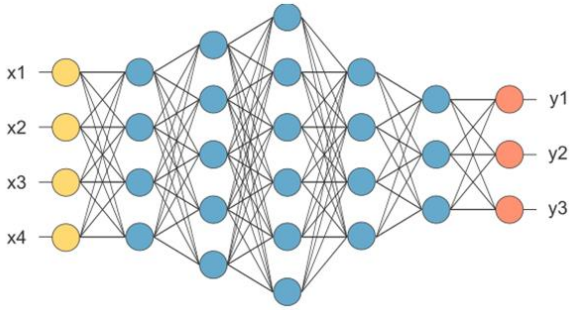

Fig.3. Fully Connected Layer [12]

B. *Classification*

It is a technique comes under supervised learning. Supervised learning means predicting input values based on the label or the target variable of the training examples that you have previously provided. For example, suppose that you have installed bird classifying system. In which you what to predict whether a bird will be a sparrow, peacock. Here your output will be either sparrow or a peacock based on your test data and training example. You are classifying your output as sparrow or peacock. We can use various machine learning algorithms to classify the emotions into different categories such as SVM, Random Forest, K-NN, Naïve Bayes, Logistic Regression etc. In a recent study, SVM gives high accuracy than the others. [3]

C. *Learning Style*

Learning style is how one can learn best. It is based on individual property and preferences. Learning style models are mainly used for content adaptation purposes, as model know how the users learn in an effective way. People can learn best by hearing someone talk, by seeing image or diagram, by text, by the graph, charts etc. The learning styles are identified by various models in E-learning such as Felder-Silverman learning style, 4MAT learning style, Kolb learning style etc. and among all models; felder silverman model is widely used. Personalized learning refers to the guidance in which the rate of learning and the instructional methodology are enhanced for the requirements of every student. [8]

D. *Felder-Silverman Learning Style Model*

The Felder-Silverman model inspects some region of personality that helps to learn. They are active or reflective, sensing or intuitive, visual or verbal, sequential or global. The combinations of these styles construct the learners learning preferences.

Sensing: These learners choose solid thinking, practical, concerned with facts and procedures.

Intuitive: These learners incline toward theoretical reasoning, innovative, concerned with theories and meanings.

Visual: These learners choose visual representations, pictures, graph, and flow charts.

Verbal: These learners choose written and spoken clarifications.

Active: These students like to give things a shot, working with others in gatherings.

Reflective: These students incline toward thoroughly considering things, working alone or with a natural accomplice.

Sequential: These students lean toward direct reasoning, precise, learns in little steady advances.

Global: These learners choose universal thinking, systems thinkers, learns in large leaps. [8]

IV. PROPOSED WORK

The aim of this research and proposed work is recognize the facial expression of online learners through the webcam and identify the emotion and behavior of learner while learning. and based on the expression recognized and the learning style of the leaner the content will be adapted. To recognize the facial expression of the online learner with the help of Convolutional Neural Network (CNN) which is used for the detection and classification of the facial expression of the learner while learning. The learning style of the learner will be identified with the help of questionnaires given by felder-silverman Index of Learning Styles (ILS). Here we have used CNN based approach to detect the facial expression of the online learner while learning. CNN is used to extract the features from the input webcam with the help of the number of convolutional layers followed by the number of max-pooling layers and finally, it gets flatten up the input to the dense layers and classified faces into the various expressions and labelled with it. It is faster than the existing system with the real-time detection of the facial expression of the learner.

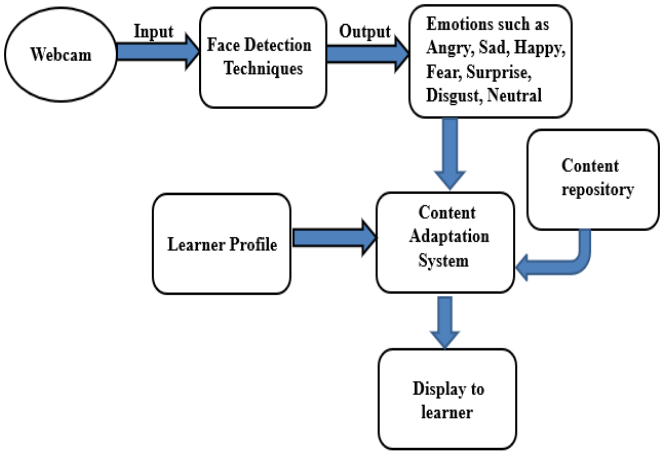

Fig 4. Architecture of Proposed Work

This proposed model will have some major modules, which are:

1. Face Detection Techniques:

This module will be responsible for detecting faces from the webcam input with the help of haar-cascade

classifier and then with the help of deep learning model (CNN) the features are extracted from the detected faces and then classify the detected faces into different emotions of categories such as Neutral, Happy, Sad, Angry, Surprise etc.

2. Learning Profile:

This module will be responsible for storing the learning style of the learner based on the questionnaires of ILS.

3. Content Adaptation:

This module will be responsible for adapting the content of the course stored in the content repository based on the learning preferences and finally, it will be displayed to the learner.

A. Dataset

The FER2013 dataset was created during the ICML challenges in representation learning in 2013. The dataset huge scale-free database gathered with the assistance of Google image API. Also, it comprises of 35887 thousand examples of gray-scale images, the 28,709 thousand examples for training images, 3,589 for validation and 3,589 for testing. The size of images in the dataset comprises of 48 x 48 pixel. The emotion task is to classify each face dependent on the stored emotion shown in facial expression in seven categories. The labels of categories are (0-6) where 0=Angry, 1=Disgust, 2=Fear, 3=Happy, 4=Sad, 5=Surprise, 6=Neutral).

Fig 5. Random images illustrate facial expressions from FER2013 Dataset

B. Proposed Model

The Convolutional Neural Networks (CNN) was the proposed sequential model. It is used to extract features from the given input image by applying various filters on the input image in a sliding window manner from left to right and top to down. The filters such as edge detection, blur image, sharpen image etc. The input of the model is 48 x48 grayscale images and the output shows classification result in one of the classes from (0-6) categories. So here we have trained the model with the 5 convolutional layers each with the filter size of 3x3 but with different numbers of filters, 2 Max-Pooling layers of 2x2 and 2 fully connected dense layers of 128 and 7 respectively. Each convolutional layer has relu activation function and Batch Normalization that will be used to build the stability of a CNN, it standardizes the output of a past activation layer by subtract the batch mean and divided by the batch standard deviation. The dropout is used to off the number of neurons depending the condition to avoid overfitting. The classification has been done with the softmax activation function which is based on the maximum probability of the class hence the final dense layer has 7 neurons. Compile the model with loss function as categorical_crossentropy and to minimize the loss using the optimizer such as Adam gradient descent algorithm.

```
Model: "sequential_1"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
conv2d_1 (Conv2D)            (None, 46, 46, 64)        640
_________________________________________________________________
batch_normalization_1 (Batch (None, 46, 46, 64)        256
_________________________________________________________________
activation_1 (Activation)    (None, 46, 46, 64)        0
_________________________________________________________________
conv2d_2 (Conv2D)            (None, 44, 44, 64)        36928
_________________________________________________________________
batch_normalization_2 (Batch (None, 44, 44, 64)        256
_________________________________________________________________
activation_2 (Activation)    (None, 44, 44, 64)        0
_________________________________________________________________
max_pooling2d_1 (MaxPooling2 (None, 22, 22, 64)        0
_________________________________________________________________
dropout_1 (Dropout)          (None, 22, 22, 64)        0
_________________________________________________________________
conv2d_3 (Conv2D)            (None, 20, 20, 32)        18464
_________________________________________________________________
batch_normalization_3 (Batch (None, 20, 20, 32)        128
_________________________________________________________________
activation_3 (Activation)    (None, 20, 20, 32)        0
_________________________________________________________________
conv2d_4 (Conv2D)            (None, 18, 18, 32)        9248
_________________________________________________________________
batch_normalization_4 (Batch (None, 18, 18, 32)        128
_________________________________________________________________
activation_4 (Activation)    (None, 18, 18, 32)        0
_________________________________________________________________
conv2d_5 (Conv2D)            (None, 16, 16, 32)        9248
_________________________________________________________________
batch_normalization_5 (Batch (None, 16, 16, 32)        128
_________________________________________________________________
activation_5 (Activation)    (None, 16, 16, 32)        0
_________________________________________________________________
max_pooling2d_2 (MaxPooling2 (None, 8, 8, 32)          0
_________________________________________________________________
dropout_2 (Dropout)          (None, 8, 8, 32)          0
_________________________________________________________________
flatten_1 (Flatten)          (None, 2048)              0
_________________________________________________________________
dense_1 (Dense)              (None, 128)               262272
_________________________________________________________________
batch_normalization_6 (Batch (None, 128)               512
_________________________________________________________________
activation_6 (Activation)    (None, 128)               0
_________________________________________________________________
dropout_3 (Dropout)          (None, 128)               0
_________________________________________________________________
dense_2 (Dense)              (None, 7)                 903
_________________________________________________________________
activation_7 (Activation)    (None, 7)                 0
=================================================================
Total params: 339,111
Trainable params: 338,487
Non-trainable params: 704
_________________________________________________________________
```

Fig 6. CNN Model Configuration

C. Tools and Methodologies

For implementation of prototype for our proposed architecture, various tools and methods have been used which are as follows:

1. Open Computer Vision (OpenCV)

OpenCV is the immense open-source library for the computer vision, machine learning, image processing and now it is useful for detecting real-time facial expression of the human. All the OpenCV array are gets converted into multidimensional NumPy arrays. face recognition, count

number of people, counting number vehicle on highways, object recognition these are a few applications which are tackled using OpenCV. It has several inbuilt functions which are accessed via cv2 object such as cv2.imread() used for reading the image from the given path, cv2.imshow() used to display the image, cv2.imwrite() used to save an image. It is a computer vision framework and cv2.VideoCapture() function which is used for capturing continuous frames of the webcam and store it in the variable. Read () function which is used to check whether the frame read correctly or not and according to that it returns a bool (true/false).

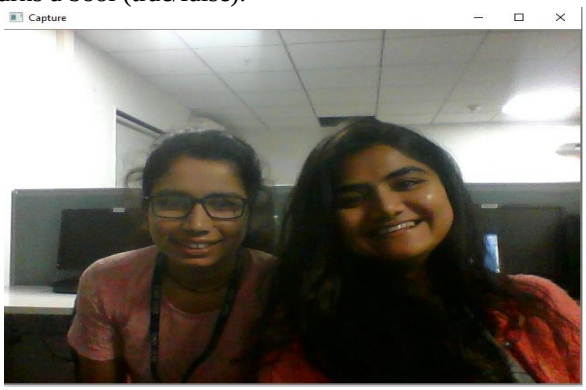

Fig 7. Screenshot of captured video

2. Haar-Cascade Frontal Face Detector Classifiers

The Haar-Cascade frontal face detector is used for detecting faces from webcam/images. It is based on the Viola and Jones algorithm mainly used for object detection in video or an image. OpenCV comes with the inbuilt trainer and a detector. It has the number of trained classifiers for face, eyes etc. We have to load the required XML classifiers file. At that point load input picture (or video) in grayscale mode. After that, we discover the faces in the picture. If faces are found, it returns the places of detected faces as Rectangle(x,y,w,h). When we get these locations, we can make an ROI for the face. The OpenCV has inbuilt function to load XML file that's cv2.CascadeClassifier().

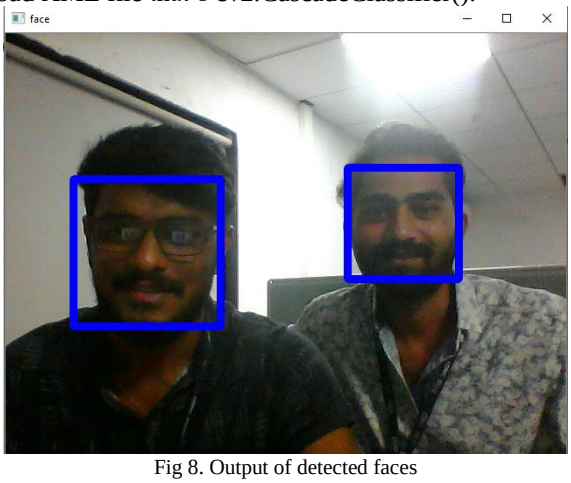

Fig 8. Output of detected faces

3. Keras

It's a python based deep learning framework.it is a high-level API for Tensorflow which runs on top of it. Useful for building simple classification model. It has a sequential model with the number of a different linear bunch of layers such as convolutional layers, max-pooling layers, dense layers etc. we can also import Batch Normalization, activation functions, dropout, optimizers such as Adam, loss function etc., from the Keras. Finally, the model gets complied, fit and evaluated.

4. NumPy

NumPy (Numerical Python) is mainly used for performing scientific operations in Python. It is a Python library that gives a multidimensional array object, different determined objects, (for example, masked arrays and matrices), and an arrangement of schedules for fast operations on arrays, including scientific, logical, shape manipulation, discrete Fourier transforms, selecting, sorting ,I/O, basic statistical operations, basic linear algebra, and much more.

5. Matplotlib

Matplotlib is a Python 2D/3D plotting library which can used in web application servers, python scripts, and GUI toolkits. It can generate bar charts, plots, histograms, scatterplots, error charts, area plot etc., with some few lines of code. Here we have used matplotlib to plot loss and accuracy graph.

6. Pandas

It is an opensource python library that will perform data preprocessing and data manipulation. It is used to create, manipulate, and wrangle the data and it can handle missing values, noisy values, unknown values etc. It's also a powerful solution for the time series data. It has a data structure named as DataFrame to store and handle multidimensional data.

D. Index of Learning Styles Questionnaire (ILS)

The Index of Learning Styles is an on-line review instrument used to survey inclinations on four measurements (active/reflective, sensing/intuitive, visual/verbal, and sequential/global) of a learning style model figured by Richard M. Felder and Linda K. Silverman. The instrument was created and approved by Richard M. Felder and Barbara A. Soloman. Clients answer 44 a-b questions and present the review, and their four inclinations are accounted for back to them quickly to be duplicated or printed out.[14]

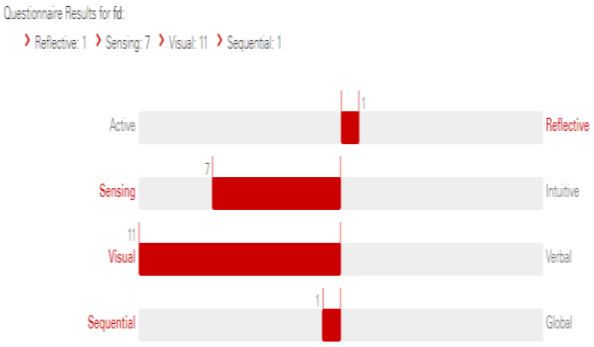

Fig 9. Result of the ILS Questionnaire [14]

V. RESULT AND ANALYSIS

We have analyzed an existing traditional approach such as LBP, HOG, LDP for extracting features from the facial expression of the human and classified into one of the categories of expression and found the issues/limitations such as they were taking too much time for extracting features and not effective when the image is small and quality is unclear. So here we have used the deep learning-based approach that's CNN for extract features with convolutional layers and later it classified in one of the expressions of categories.

A. Experimental results

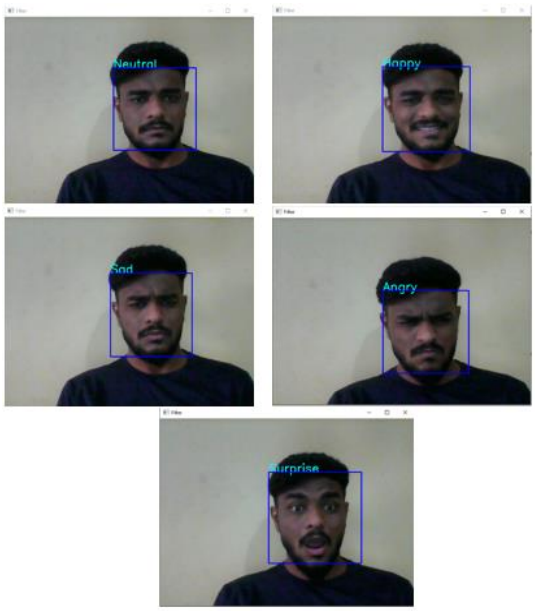

Fig 10. These are results with the different facial expressions such as Neutral, Happy, Sad, Angry, Surprise.

B. Evaluation Parameter Graph

Here we ran the model for 50 times with the Adam optimizer on training and validation datasets. we got the accuracy of 65% on a training set and 63% on a validation set.

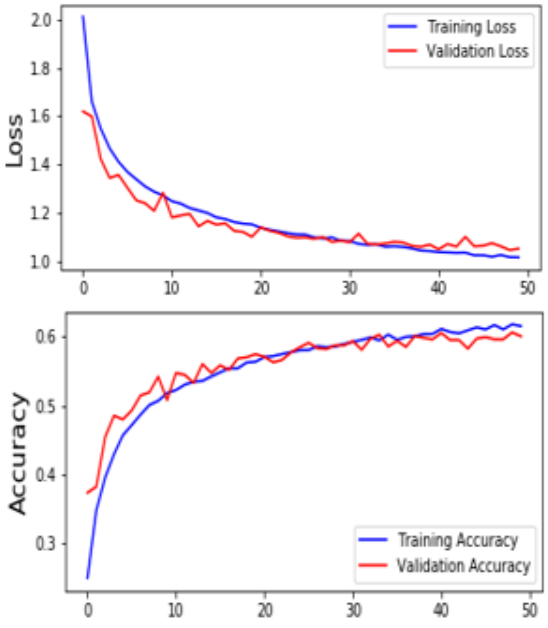

Fig 11. Evaluation Graph of Loss and Accuracy

CONCLUSION AND FUTURE SCOPE

From the above discussions, we can conclude that it is possible to detect and classify the facial emotions and behavior analysis of online learner with the help of Deep learning and machine learning technologies such as Convolutional Neural Networks (CNN) and libraries that includes OpenCV, Keras, NumPy, Pandas and Matplotlib. The learning style of the learner has been captured using the Felder-Silverman Index of Learning Style and stored it in a learner profile. So, we can adapt the content according to the facial expression and learning style of the online learner using content adaptation. So, the future scope is with the help of more convolutional layers and max-pooling layers and by removing overfitting with the help of regularization, dropout, data augmentation etc., we can achieve good accuracy.

REFERENCES

[1] Hingu, D.H., Khan, S.S. and Sinha, M.A.K., Facial Expression Analysis for Emotion and Behavior of Online Learner and Framework for Content Adaptation: A Survey.

[2] Duncan, D., Shine, G. and English, C., 2016. Facial emotion recognition in real time. Stanford University.

[3] 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.

[4] Kumari, J., Rajesh, R. and Pooja, K.M., 2015. Facial expression recognition: A survey. Procedia Computer Science, 58, pp.486-491.

[5] Dewan, M.A.A., Lin, F., Wen, D., Murshed, M. and Uddin, Z., 2018, October. A Deep Learning Approach to Detecting Engagement of Online Learners. In 2018 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI) (pp. 1895-1902) IEEE.

[6] Minaee, S. and Abdolrashidi, A., 2019. Deep-Emotion: Facial Expression Recognition Using Attentional Convolutional Network. arXiv preprint arXiv:1902.01019.

[7] Ayvaz, U., Gürüler, H. and Devrim, M.O., 2017. Use of facial emotion recognition in e-learning systems. Information Technologies and Learning Tools, 60(4), pp.95-104.

[8] Ciloglugil, B., 2016, November. Adaptivity based on felder-silverman learning styles model in e-learning systems. In 4th International Symposium on Innovative Technologies in Engineering and Science (ISITES2016) 3-5 Nov 2016 Alanya/Antalya-Turkey.

[9] Huang, H.C., Wang, T.Y. and Hsieh, F.M., 2012. Constructing an adaptive mobile learning system for the support of personalized learning and device adaptation. Procedia-Social and Behavioral Sciences, 64, pp.332-341.

[10] Liu, K.C., Hsu, C.C., Wang, W.Y. and Chiang, H.H., 2019, July. Real-Time facial expression recognition based on cnn. In 2019 International Conference on System Science and Engineering (ICSSE) (pp. 120-123). IEEE.

[11] Alkabbany, I., Ali, A., Farag, A., Bennett, I., Ghanoum, M. and Farag, A., 2019, September. Measuring Student Engagement Level Using Facial Information. In 2019 IEEE International Conference on Image Processing (ICIP) (pp. 3337-3341). IEEE.

[12] Haider, M., Sinha, A. and Chaudhary, B., 2010. An Investigation of relationship between learning styles and performance of learners. International Journal of Engineering Science and Technology, 2(7), pp.2813-2819.

[13] https://medium.com/@RaghavPrabhu/understanding-of-convolutional-neural-network-cnn-deep-learning-99760835f148

[14] https://educationdesignsinc.com/index-of-learning-styles/