International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-8 Issue-6, August, 2019

Deep Learning based Student Emotion Recognition from Facial Expressions in Classrooms

Archana Sharma, Vibhakar Mansotra

Abstract**: *Classroom teaching assessments are intended to give valuable advice on the teaching-learning process as it happens. The finest schoolroom assessments furthermore assist as substantial foundations of information for teachers, serving them to recognize what they imparted fittingly and how they can improve their lecture content to keep the students attentive. In this paper, we have surveyed some of the recent paper works done on facial emotion recognition of students in a classroom arrangement and have proposed our deep learning approach to analyze emotions with improved emotion classification results and offers an optimized feedback to the instructor. A deep learning-based convolution neural network algorithm will be used in this paper to train FER2013 facial emotion images database and use transfer learning technique to pre-train the VGG16 architecture-based model with Cohn-Kanade (CK+) facial image database, with its own weights and basis. A trained model will capture the live steaming of students by using a high-resolution digital video camera that faces towards the students, capturing their live emotions through facial expressions, and classifying the emotions as sad, happy, neutral, angry, disgust, surprise, and fear, that can offer us an insight into the class group emotion that is reflective of the mood among the students in the classroom. This experimental approach can be used for video conferences, online classes etc. This proposition can improve the accuracy of emotion recognition and facilitate faster learning. We have presented the research methodologies and the achieved results on student emotions in a classroom atmosphere and have proposed an improved CNN model based on transfer learning that can suggestively improve the emotions classification accuracy.***

Keywords: **classification, convolutional neural network, deep learning, emotion recognition, face recognition**

I. INTRODUCTION

In the modern times where technology is forever evolving, the interaction between humans and machines is gaining importance and there is a growing demand for developing machines that can be intelligent and self-decisive and machines that can capture the gestures and emotions of the humans to automate tasks and handle the communication better. A machine that can understand the emotions of a human better can predict and respond to the human behavior better, which in turn can significantly improve the efficiency of the task that is meant to be done. This machine can act as an important tool for the behavioral science analysis of the social and communicative intricacies of the humans, thereby aiding in building complex and sophisticated software frameworks that can identify social and emotional behavior better in a way where they can be used in robots. Human emotions play a significant role in a way we respond to an action like making decisions, learning capabilities, motivating oneself, planning, reasoning, thinking, perception, etc., This potent hold that emotions wield in our daily life, dictating and affecting our reactions to the events, provides the impetus to analyze these better and build a model that can conclusively identify the right emotions in any situations has made emotion recognition, the trending topic today and has become a promising research field that can spring many surprises in the future about human emotions. Emotions are normally recognized by analyzing the speech signals, text content and facial expressions using intelligent software algorithms.

Of all the modes of emotion recognition, emotions expressed through facial reactions is the most powerful and universal way of conveying emotions. The expression of one's feelings through facial expressions is authentic and natural, finds out a psychological research. The human-machine interface (HMI) [1] framework aids us to construct devices capable of managing the interactions between human-machine. For instance, the facial emotions in our study can be deciphered as indicators for machines to investigate the fundamental sentimental state through its intelligent framework and can present the emotional state of the human being studied.

Facial expressions-based emotion recognition practices have a significant role in the modern fields of computer vision and artificial intelligence [2]. Although emotions can be captured and analyzed with wearable sensors, it is essential and more importantly adaptable to detect facial emotions with visual inputs without the need for a physical connection. This necessity to detect emotions in a real-world scenario, and the complexity and scalability of modern computer vision algorithms coupled with high performance hardware to process loads of data in real-time had boosted the development and application of artificial intelligence based facial emotion recognition a strong headstart among other contemporary methods used in facial emotions analysis.

Revised Manuscript Received on October 30, 2019.
* Correspondence Author
Dr. Archana Sharma, Department of Computer Science, Government M.A.M College, Cluster University of Jammu, Jammu, India. Email: archana.35188@yahoo.com
Dr. Vibhakar Mansotra, Department of Computer Science and IT, University of Jammu, Jammu, India. Email: vibhakar20@yahoo.co.in

However, this technology as poses its own challenges in facial analysis, detection, recognition and emotions classification. Simply mimicking the way by which humans recognize faces .

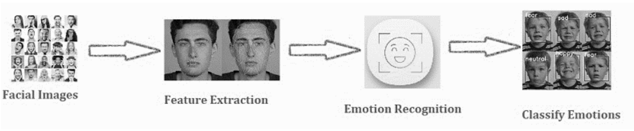

Fig. 1. Flow Representation of Facial Emotion Recognition System

instantly without much effort is a huge challenge for the machines. Several other factors like the broad range of individuals with own characteristic faces, further complicated by their age and ethnicity must also be considered. Furthermore, the head size variations, face orientations, facial hair, eyeglasses etc., make this task of identification more difficult. Also locating the face in the images where there are a group of people or group of objects

mixed with people can be more fundamental issue in facial emotional recognition. In an environment like classroom, meeting halls etc., the face poses may differ due to the camera angle, where they may be at an angle obscuring the facial features necessary for the analysis. This requires us to employ some good preprocessing methods to work on those input images that have a great deal of tactlessness to the scaling, and rotation of the head. Many feature-based techniques used in facial emotion analysis use local spatial analysis in identifying the facial features, and the automatic localization of those facial key points helps a lot in the analyzing the facial emotions in a robust manner. Recognizing the facial components like the eyes, nose, brows, lips, mouth etc., is also an important part in analyzing the facial emotions and there are many techniques that are employed today to identify those key facial points.

In our study and the review of the facial emotion recognition, we consider the classroom scenario where we analyze the mood of the students in the classroom to the lecture and analyze this moods or the emotions that they experience to understand the student psychology of what interests them and what makes them bored during a lecture. In the present classroom scenario, the teaching assessments are basic, non-graded, limited in the way they are designed to offer helpful feedback on the students' emotions in a classroom. This study is undertaken as the commitment and the involvement of the students during the lectures is lectures is vital for understanding the concepts of the topic being taught and can undoubtedly improve the academic credentials of the students. Even though direct supervision by the instructors is possible in a classroom environment, it cannot be used as a tool to measure the attentiveness of the students in the classroom. Also, there are quite a handful of students in every classroom who can lose their concentration and be inattentive to the lecture even under direct supervision. This presents a need for an approach to quantitatively assess and find out the lapses of concentration and attentiveness by the students in the classroom. Hence, a continuous assessment of the students' facial emotions in the class can act as an aid in predicting the complete class behavior like their attention to the class (neutral), laughing during the lecture (happy), sleeping during the lecture (bored/sad) and so on. This study has been presented to discuss on the previous works that were carried in the exact classroom scenario using the students' facial expression during a lecture to extract mood patterns. This way of analyzing the various emotions of the students within a classroom environment can offer more insights into their emotional states during the lecture can assist in designing teaching aids to improve their attentiveness and also improve the efficacy of the content delivery by the instructor in the classroom.

After analyzing the past works on the classroom student emotions recognition, we propose our own system using deep convolutional neural networks (CNN) technique for recognizing the students' classroom emotions. A video camera present in the classroom captures the live video of the students' in the classroom as a whole and this captured video frames are presented to the processing unit which extracts the key frames from the video and applies facial feature extraction techniques to detect the faces from the video frame. The CNN model has been trained on similar databases of people and faces and it can be used to predict the classes of emotions with highest accuracy and the most probable emotion is decided. An assessment can be carried out based on the emotion predictions, and the emotions classes like happy, sad, surprise, disgust, angry, fear and neutral emotions are predicted and analyzed to provide better teaching tools, improve lecture content and lead to a better classroom environment.

II. Related Works

Facial emotion identification and classification system has gathered a substantial amount of interest with a vast number of applications in the current era. In most of the applications, a user's face tracked by a digital high-resolution video camera and analyzed for emotions is growing exponentially. One can easily analyze the emotions to bring corresponding changes to the environment that surrounds it. To analyze the human facial expressions, face recognition must be the first step. The extraction and identification of the face from live images or live video-stream is one of the major challenges in this field.

By varying the physical attributes of a face causes a major change in the identification. Myunghoon et al.[3] proposed using an Active shape model (ASM) to take out the 77 facial features based on geometric-based model, where the detected image is iteratively distorted to fit and shape the model and filter out the facial features after comparing it with ASM. In 2014, Kamlesh et al. [4] used a hybrid approach based on Active Appearance Model (AAM) and Local Binary Patterns (LBP). This approach extracted 68 facial points, in which AAM is geometric- based approach and LBP is appearance-based approach. In 2016, Krithika L.B and Lakshmi Priya G.G [5] proposed a student face emotion identification system for e-learning environments. This model allows to capture the students' face and identifies the emotion of the students and captures the dynamically changing emotions in response to the part of the lecture being listened to in this e-learning environment. They had applied the Voila-Jones [6] method and Local Binary Patterns (LBP) to detect the face and classifying an

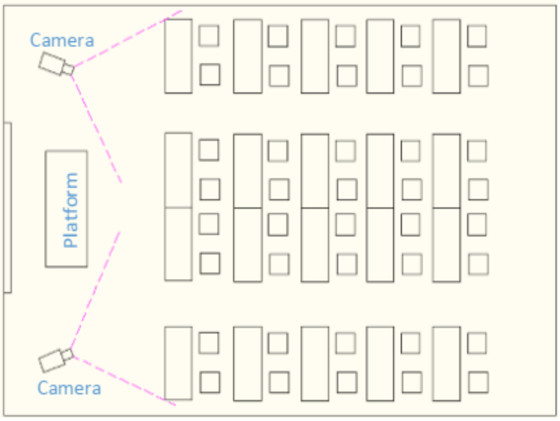

Fig. 2. Classroom Layout with the digital video cameras positioned on either side of the platform as proposed by Sheng Chen et al., [16].

expression or an emotion of a student.

In the other recent work, an intelligent classroom facial emotion recognition based on deep learning was presented by Chao M et al. [7].

They had used the FER 2013 database with fixed size of images such as 48 x 48 pixels and color-mode were grayscale. A three-layer CNN model was used in this paper. The results were not too precise to make any conclusive agreement on the emotion states, achieving 50-60% accuracy.

Recently, in 2019, Boonroungrut et al. proposed the student emotion identification in classroom environment by using cloud facial recognizer technology [8]. Surprisingly, this technology has been utilized in marketing to measure and explore the group satisfaction. Several emotion identification algorithms for facial expression analysis have been discussed there in the literature survey. Machine Learning methods [9] such as Support Vector Machine (SVM) [10], Random Forest Regression [11], Logistic Regression [12] and Neural Networks (NN) [13], Hidden Markov Model [14] and Adaboost [15] are different types of classifiers used for identification and classification of emotions. In a prior work, Chen et al. [16] proposed a classroom teaching feedback system based on machine learning based facial emotion recognition, with a perfect classroom setup for capturing the live streaming video to identify facial emotions and provide a feedback to the instructor. As we discussed above, machine learning models does not learn anything from old data. It just predicts or classifies certain things. Hence this model did not get better results compared to the previous work discussed above and layout can be shown in Fig 2. Models based on neural networks are very effective in capturing complex facial patterns from facial images. Both supervised and unsupervised learning approaches are used to train the neural network. Since finding an enough training dataset is questionable, unsupervised neural networks are preferable. Recent studies show that, Convolutional Neural Network (CNN) has become the most predominant technique used in the field of image processing. In 2018, another classroom-based environment designed by Tang et al. [17] used a classroom assessment and feedback system that utilizes the computer vision technology. With the help of FER 2013 dataset, CNN model was trained and used in predicting the emotion state of the students. An NVIDIA Jetson TX1, the first supercomputer to run a run a GPU Computing architecture was used in training the neural network, where 150K steps of the final trained model took around 4 hours to train it.

Another classroom teaching feedback-based system was proposed and implemented by Chen et al. [16]. Two GS3-U3 digital video cameras connected to HP Laptop (CPU: I7-855, 8GB RAM, Graphics: NVIDIA MX15) was used as the host computer to train and test the neural network model. A Multi-Layer Perceptron (MLP) Classifier was used in this study to classify the emotion states.

III. Existing System Frameworks

In this paper, we have intended on reviewing those research works and their system frameworks to analyze the research work done thus far in this field of emotion recognition in a classroom environment. For this purpose of investigation, we have selected four previous research works whose focus on the analysis on student emotions in a classroom arrangement. The selection of these papers were done based on the following criteria: year of publishing of the research work; we have primarily considered the most recent research paper, the methodology used in the work; we have selected papers that have used different methodologies for features extractions and classification of emotions, and finally the physical design and implementation; how the hardware setup consisting of cameras and the processing unit have been implemented in a classroom setup, the student population and the computing power of the processing unit used in facial image processing. of aim this survey for facial emotion identification of a student in a classroom. For our review, we have considered Abdulkareem Al-Alwani et al. [18] proposal on Mood Extraction Using Facial Features to Improve Learning Curves of Students in E-Learning Systems, where there is no direct supervision involved in gauging the emotions of the students. Another work, we have taken for our review was originally presented by Sahla K. S. and T. Senthil Kumar [19] for assessing the classroom teaching based on student emotions. This work is unique in the way that they had designed the system to capture the video of the teacher to predict his/her emotions along with the student’s emotions, hence providing a two-way dissection of the emotions experienced by both the instructor and the students, offering more insight into the classroom lectures. Our third paper for review was presented by Tang et al. [17] for the design of intelligent classroom based facial recognition using CNN and FER-2013 facial image dataset. This work is one of the few works that uses FER-2013 dataset for the classification of students’ emotions in a classroom. Since we have planned to use FER-2013 dataset for our proposed work, we found it appropriate to analyze their work better. The final research work considered here was by Chen et al. [16] that had proposed a real-time feedback system with a camera array to capture the facial reactions of the students in the classroom and identify their emotions, assisting the instructor to handle the learning process of the students dynamically.

A. *Dataset*

In the research works [18] [19], Cohn-Kanade AU-Coded (CK+) Facial Expression Database was used in the study. This is an extension of the former dataset, which was referred to as Cohn-Kanade (CK) dataset with increased number of subjects and image sequences. Every peak expression in this dataset is fully coded by FACS and a prototypic emotion label is assigned.

Fig. 3. Sample Expressions in a) CK+ (Extended Cohn-Kande) and b) JAFFE [23].

Fig. 4. Sample Images from the FER 2013 Dataset [24].

For 6-class expression recognition, the three or four most expressive images from each sequence was selected.

To build the 7-class dataset, the foremost image (neutral expression) from each of the 309 sequences was selected and added to the 6-class dataset. This is shown in Fig 3. (a) where a sample of seven prototypic expressions are taken from the CK+ dataset.

The research works [16] [17] have used the FER 2013 dataset which includes around 35,000 human facial expression images for seven emotion categories, including anger, disgust, fear, happiness, sadness, surprise, and calm. The image label breakdown shows a skew towards happiness and away from disgust, as can be clearly seen from the facial expression in the image (Fig 4).

B. *Face Detection*

In research work [18], the authors have not used any face detection algorithms, most likely because theirs was an e-learning system where, its most possible, that only one individual will be visible in the frame. Hence, there was no need to detect the face and facial expressions were directly analyzed which will be discussed in the next section.

However, a new method was used by the researchers in their paper [19], termed as, Key Frame Extraction, was used to avoid duplicate frames from the stored video frames due to slower processing. Each video frame is calculated by the intersection of color histograms and corresponding values are noted. The resulting values are compared with threshold values, a screen-cut is detected and removed from the training. Voila Jones Algorithm [6] was used to detect the faces from the selected frames

Tang et al. in [17] have used a convolutional model for detecting the face as well as predicting the facial expressions with machine learning schemes such as Support Vector Machines (SVM), Logistic Regression and Random Forest Regression. Once the face from the image gets located, it is cropped and resized into 48 x 48 pixels.

Chen et al. in [16] had used multiple methods for face detection like geometric feature-based methods, template matching methods, and subspace LDA, to detect the face in different angles from the two cameras positioned on both the sides of the classroom. But there can be some hindrances in capturing the facial features better due to low light conditions, background noise, hand postures etc., and normalization of the recognized faces can help eliminate this problem.

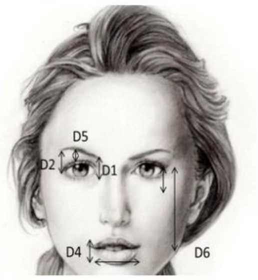

Fig. 5. Distance calculated between Facial Objects [18]

C. *Feature Extraction*

In the research work [18], the facial features were categorized into four different categories such as eyes, eyebrows, lips, and head. A Hidden Markov Model (HMM) was used to extract facial features and it differs from other approaches like template-based and appearance-based as it builds a series of observation vectors from the face pattern, represented in the Fig 5. The identified features are then labelled as D1, D2, D3… etc., and feed to the neural network. Any change in the distance metrics point to an instance of a facial feature and collection of these features can be used to classify those six facial emotions. Certain threshold values were provided for distance to make decisions related to facial emotions, which can then be used to classify an unknown pattern.

In their work [19], the researchers had used Haar Cascades based facial feature extraction, a well-known technique and implemented it in OpenCV. This cascade classifier consists of several stages, where each stage is an ensemble of weak action as shown in Fig 6. In each stage, the classifier labels the region defined by the current position of the sliding window as either positive or negative.

The authors [17] in their paper had proposed Regions with

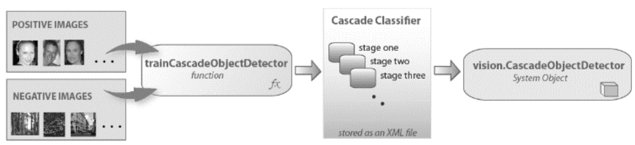

Fig. 6. Haar Cascade Feature Extraction model [25].

CNN (R-CNN) method to exactly identify the target position in the whole image, since regions suggest using precise information like color, texture and texture in an image. This overcomes the problems of missing features associated with the previous work that used sliding window method, by achieving a candidate window of better quality than the previous sliding window with fixed aspect ratio. This region proposal approach results in extracting the facial features well. In the research work [16], Gabor Filter and Discrete wavelet transform (DWT) were used for feature extraction.

Gabor wavelet transform, a potent image processing algorithm that mimics the perception of the human visual system, can improve the edge detection in the images A feature vector was calculated for facial features using Gaussian Kernel Function as seen in Fig 7. The image demonstrates precisely how the kernel output changes from low-level features like the eye corners, eyebrow edges to advance features like eyes, nose and mouth.

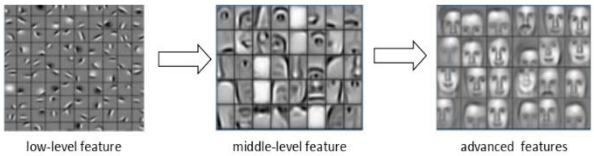

Fig. 7. Facial features extraction with Gabor Filter and Discrete Wavelet Transform [16].

D. *Training and Feature Learning*

Radial-basis function neural network was the preferred method used in [18] to classify the emotions such as happy, sad, surprised, confused and disturbed using the facial features that shows the distance points of eyes, mouth and lips. Data association mining technique was used to estimate the unknown relationships and decision rules in a dataset to improve decision making and predicting human facial expressions. Radial basis function for facial expression based on distance-based approach is shown in Fig.8.

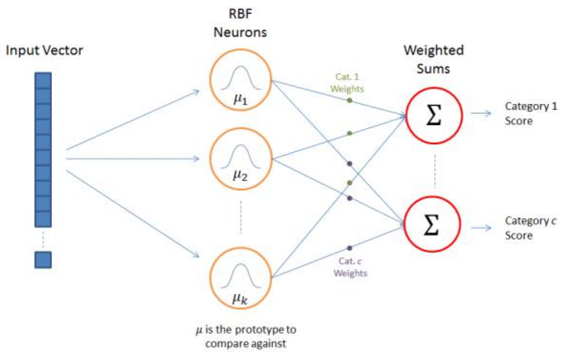

Fig. 8. Radial Basis Function Neural Network Model

The research work [19] had used CNN with Local Binary Patterns (LBP) Encoding, where the cropped faces are processed to obtain the inputs for CNN model. The classification of emotions is carried out as discussed below.

- LBP encoding method was applied to pixels.
- The metric space values were transformed from an unordered code values by using Multi-Dimensional Space (MDS).
- CNN Model was trained with RGB Cropped face images and input to the deep CNN model. The class with maximum average prediction be the final classification.

The researchers [17] had heavily indulged with three different machine learning algorithms like multi-class SVM, Random Forest and Logistic Regression for comparison with deep learning technique like CNN to classify the human emotions. The CNN architecture for face detection and emotion recognition is given below in Fig.9. In the preprocessing stage, the fixed size input layer (48x48) images is given as the input to the next layer. But first the faces are detected from each image using Adaboost to locate and crop faces, which in

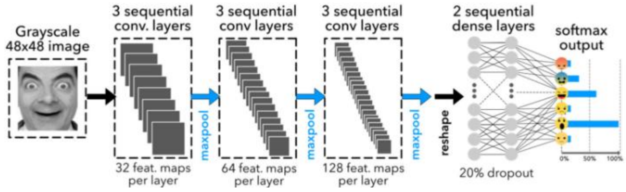

Fig. 9. Generic Convolutional Neural Network Model that takes 48x48 images as input and passes it through three layers of three sequential convolution layers, with the feature maps doubled in every layer to form two sequential dense layers with SoftMax activation function at the output [17].

turn reduces the size of the images. The input layer is then transferred to Convolution2D layer, where the number of filters is stated as a super-parameter. Each filter, such as a sliding window, moves through the whole image to create a feature graph with shared weights. The feature map constructed by the convolution layer show how the pixel values are elevated, for instance, like edge, light, and pattern detection. The CNN model produced better results

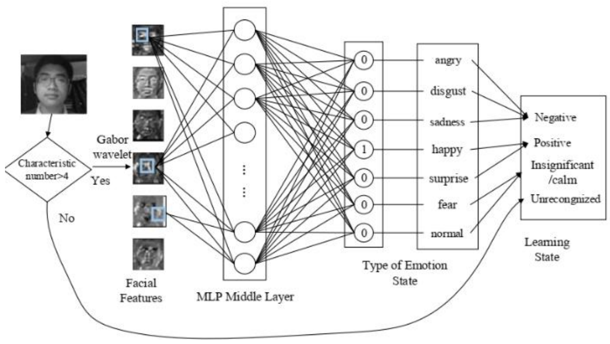

Fig. 10. Multi-Layer Perceptron based Emotion Classification Model [16]

than the machine learning techniques.

Sheng Chen, Jianbang Dai and Yizhe Yan in [16] had proposed an MLP (Multi-Layer Perceptron) Emotion Classifier that adopts the sigmoid function and reverses the back-propagation network for training, that can reduce the falling of the local optimum. Fig-10 shows the classification methodology of MLP where the perceptron is determined by multiplying weights and adding bias.

With this method, to express the student's emotion state like angry, disgust, sadness, happy, surprise, fear and normal, simple emotion mapping to the students learning state is realized by classifying them into positive, negative, insignificant and unrecognized.

IV. Proposed System Framework

The system architecture for proposed emotion recognition framework is represented by Fig.11.

A. *Dataset Description*

As there are certain publicly available databases which contain basic human expressions and widely used in emotion identification systems. Both these databases consist of 7 expressions. CK+ consists of 123 subjects and 593 samples, and FER2013 has 35,887 sample images. Our intention here is to use the FER2013 database a training set and transfer learning for a new dataset such as CK+.

B. *Face Detection*

There are few models that are frequently used for extracting the face from an image such as Voila-Jones, Haar Cascade and dlib face detectors. The efficiency of the face detection hinges on the effective uncovering of faces from the images that contains other objects and human body parts.

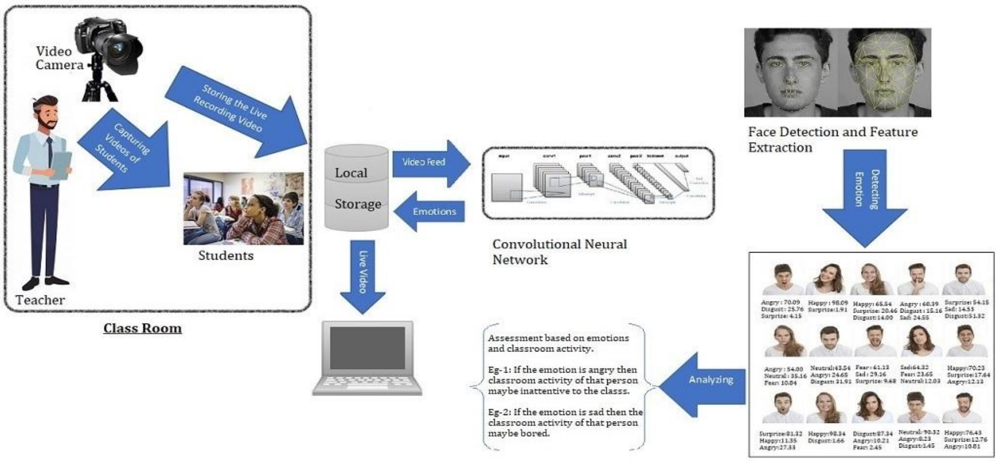

Fig. 11. Proposed Facial Emotion Classification Model – Process Flow Diagram

The proposed method for face detection is Voila-Jones [6]. In a multi-camera or dual camera assembly, it's not hard to identify the faces since either one or more cameras may have the proper alignment for easy face detection. Our proposed system uses a single digital camera pointing straight-ahead towards the students. This algorithm can be used to detect the face from a video frame and crop the faces detected as separate images of a fixed size, 150x150 pixels. The cropped images can be converted into grayscale if necessary and it will be highlighted by drawing a rectangular region around it. To remove the background and other edge-lying obscurities, the subject's face was must be cropped from the original image based on the positions of the eyes.

There will be some landmark locations which were provided for an image and each location represents some position on the face. By using those landmark locations, the distance between eyes were calculated once the midpoint of left and right eyes was identified. If you see the Fig.12, you can understand that the face was then cropped using empirically selected percentages of D with the center of the left eye as a reference point.

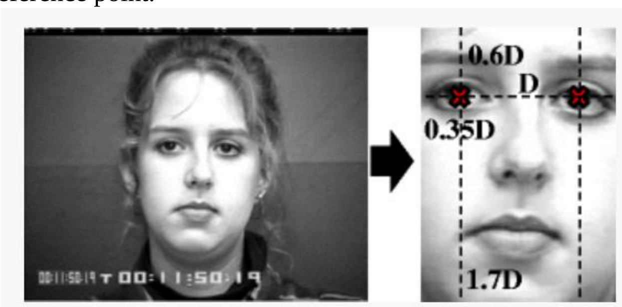

Fig. 12. Viola-Jones proposed Facial features extraction process [23].

C. *Feature Extraction*

After due consideration to all the prior discussed works for feature extraction, the Haar Cascades extraction method would be the most suitable technique for high performance images and high-resolution faces. It is because, as we have already referred [16], Gabor filter and wavelet transform extracts only certain facial features, missing out those when the face is not aligned to face the camera. In such a situation, it is hard to train the model and classifying the emotions with high accuracy.

D. *Training and Feature Learning*

The proposed method is to implement a CNN model with transfer learning. For image data, so called deep CNNs have proved to perform similar or even better than humans in some recognition tasks. One of the main assumptions of most CNN architectures is that the inputs to these networks are raw images; this aid them in encoding specific properties into the network structure. The CNN transforms the dimensions of the image layer by layer while the various activation functions in the different layers, until it finally reduces to a single vector of scores assigned to each class. These scores are arranged along the depth of the final layer.

Convolutional Networks (ConvNet) typically consist of three layers namely CONV, POOL and FC (fully connected). These layers are stacked to form a full ConvNet architecture. The activation function used for the network at hand is a Rectified Linear Unit (ReLU) activation function, which is R(z) = max (0, z). Further, Rectified Linear Unit (ReLU) apply an elementwise activation function. Symbolically, such a network can be described by [INPUT-CONV-RELU-POOLFC]. Once the features are learnt through ConvNet architectures, classification of emotions take place.

Deep Learning Algorithm 1*: CNN Training Model on FER 2013 dataset*

Stage 1. Pre-Processing:

1) **Load data:** database: FER2013, image size: 48x48 = 2304 vector. #classes=7 = [0=Angry, 1=Disgust, 2=Fear, =Happy, 4=Sad, 5=Surprise, and 6=Neutral].

2) **Split data:** (training: test) = (28273,7067).

3) **Augment data:** rotation, scaling, shift along X and Y axes.

Stage 2. Creating the Network.

Add layers sequentially [CONV-CONVNORM-RELU-POOL]x3 →[FC]

Stage 3. Training the Network:

TABLE I
EMOTION CLASSIFICATION RESULTS ACROSS REVIEWED RESEARCH WORKS

Authors	Dataset	Face Detection Methods	Feature Extraction Models	Classifier Models	Results
Abdulkareem Al-Alwani [18]	CK+	-NA-	Hidden Markov Model	Radial Basis Function	70%
Sahla K. S. and T. Senthil Kumar [19]	CK+	Haar Cascades	Haar Cascades	CNN with LBP Encoding	80.1%
Jielong Tang, Xiaotian Zhou, Jiawei Zheng [17]	FER 2013	CNN Model	R-CNN	SVM, Random Forest, Logistic Regression	59.3%, 55.1%, 54.0%
Sheng Chen, Jianbang Dai and Yizhe Yan [16]	FER 2013	Geometric Feature, Template matching, Subspace LDA	Gabor Filter Discrete Wavelet Transform	Multi-layer Perceptron (MLP)	85%

Num_epochs=100.
Fit model on batches with real-time augmentation.

Stage 4. Learning decision:
Determine loss on training and test sets over the training epochs.

Stage 5. Making Predictions:
Test on individual images.
Evaluate trained model on test set.

E. *Transfer Learning*

As a deep CNN is made up of many nodes, resulting in a high number of weights to be trained, requiring a large training data to train from scratch. Since our focus lies on on the recognition of the emotional classes, the training effort can be reduced with the use of pre-trained model from the devised deep learning algorithm already discussed. Since the weights are already known, small datasets like the Cohn-Kanade+ (CK) [20] datasets can be also be used. In doing so, training the network becomes rather a fine-tuning of the last 10 layers and the output layer's weights. This notion of representations learnt from pretrained networks trained on a different dataset (FER 2013) being transferred to a different dataset (CK+) for facial expression recognition is explored. The DL architecture that we have planned to deploy here is VGG16.

Deep Learning Algorithm 2 - Transfer Learning: *Pre-trained VGG16 on CK+ Database*

For the comparison, we have used VGG-16 network by Simonyan and Zisserman in Fig.13 [20] from the Keras library for Python with the TensorFlow backend. As described earlier the network was pre-trained given the FER2013 model weights. We then adapt the work in [22] for use of pretrained CNNs for learning and classifying samples from smaller databases of CK+. We choose to adapt the bottleneck features of a pretrained network to build a model
for the CK+ dataset. For the problem at hand, the output layer of the network was truncated and replaced by a SoftMax layer with seven output nodes. To avoid overfitting, we equipped a dropout layer before the output layer. The VGG16 architecture is represented as [[CONVx2-POOL]x3→[CONVx3-POOL]x2 →FCx3]. The features learnt from VGG16 only up to the convolutional model up to the fully connected layers is instantiated. This model is run on the training and validation data of CK+ once, thus recording the bottleneck features from VGG16 model. The model is then trained with a fully connected model on top of the stored features.

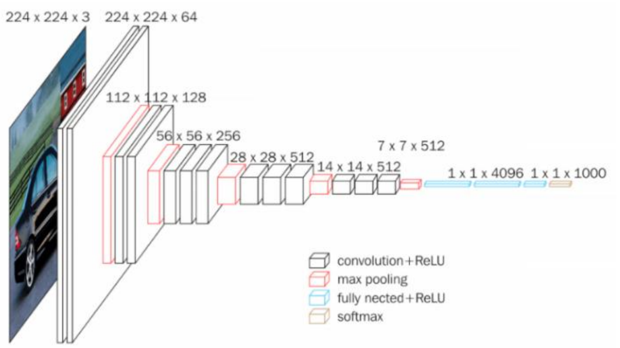

Fig. 13. VGG16 Architecture [26].

V. Result And Discussion

We have analyzed the previous research works done on facial emotion identification in a classroom environment. The best results were obtained by Al-Alwani [17] with 85% overall accuracy that was based on FER2013 Facial Emotion Dataset consisting of 7 Expressions.

Fig. 14: Proposed Classroom Emotion Recognition System

The results obtained by each of the reviewed methods is presented in the Table I. With our proposed system as seen in Fig.14, we have planned to even better classification accuracy in predicting the emotions, with the proposed CNN model and using transfer learning on our pre-trained VGG16 on the CK+ dataset to get higher accuracy.

TABLE II
PROPOSED EMOTION CLASSIFICATION APPROACH RESULTS

Emotion	Result	Emotion	Result
	Happy: 95% Surprise: 4.1% Neutral: 0.9%		Fear: 84.2% Sad:12. 4% Neutral: 3.4%
	Sad: 86.4% Neutral: 9.7% Fear: 3.9%		Happy: 96.6% Surprise : 3.4%

VI. CONCLUSION

Face detection and emotion recognition are one of the challenging problems in the field of computer vision. In this paper, we have presented a detailed survey about various research papers implemented in a classroom environment, as seen from the comparisons of databases, methodologies in Table I. As we have also inferred that, Chen et al., had achieved satisfactory results when compared to other methods used in this comparison as they had intensively concentrated on the pre-processing steps to extract the facial features which can be difficult considering the different angles and the sitting postures of the students. After due consideration and building on the techniques and methodologies from the discussed research works, we have proposed our paper with an improved methodology that can overcome the shortcomings of the above discussed frameworks by focusing on preprocessing and classification of emotions. For pre-processing, Haar Cascade, OpenCV and Voila-Jones algorithm performed well in extracting the facial features. For analyzing emotions, the transfer learning methodology that we have used can analyze emotions in real-time with faster processing than the other discussed works. This can especially work well for faces that are slightly blurred or at a different angles where it is hard to conclusively identify the emotions. Even so, there can be certain pitfalls in every approach and that actually have provided a scope for our future work that can use speech combined with facial features to analyze the emotions of the classroom as a whole, termed group emotions that can present the overall mood or sentiment of the classroom during the lecture.

REFERENCES

1. J.D. Bauer, H.F. II Kenneth, and R.N. Flores, “Intelligent human-machine interface,” U.S. Patent No. 7,966,269. 21 Jun. 2011.
2. S.J. Russell and P. Norvig, Artificial Intelligence: a modern approach. Malaysia: Pearson Education Limited, 2016.
3. M. Suk and B. Prabhakaran, “Real-time mobile facial expression recognition system-A case study,” Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2014, pp. 132-137.
4. K. Mistry, L. Zhang, S.C. Neoh, M. Jiang, A. Hossain and B. Lafon, “Intelligent appearance and shape based facial emotion recognition for a humanoid robot,” The 8th International Conference on Software, Knowledge, Information Management and Applications (SKIMA 2014), Dhaka, 2014, pp. 1-8.
5. L.B. Krithika, “Student Emotion Recognition System (SERS) for E-Learning improvement based on learner concentration metric,” Procedia Computer Science 85, 2016, pp. 767-776.
6. O.H. Jensen, “Implementing the Viola-Jones face detection algorithm,” MS thesis, Technical University of Denmark., DTU, Lyngby, Denmark, 2008.
7. C. Ma, C. Sun, D. Song, X. Li and H. Xu, "A Deep Learning Approach for Online Learning Emotion Recognition," 2018 13th International Conference on Computer Science & Education (ICCSE), Colombo, 2018, pp. 1-5.
8. C. Boonroungrut, T.T. Oo, and K. One, “Exploring classroom emotion with cloud-based facial recognizer in the Chinese beginning class: A preliminary study,” International Journal of Instruction 12.1, 2019, pp. 947-958.
9. M.S. Bartlett, G. Littlewort, M. Frank, C. Lainscsek, I. Fasel, and J. Movellan, “Recognizing facial expression: machine learning and application to spontaneous behavior,” IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), vol. 2, 2005, pp. 568-573.
10. M. Abdulrahman and A. Eleyan, “Facial expression recognition using support vector machines,” 23nd Signal Processing and Communications Applications Conference (SIU),2015, pp. 276-279.
11. E. Kremic and A. Subasi, “Performance of random forest and SVM in face recognition,” Int. Arab J. Inf. Technol. 13.2, 2016, pp. 287-293.
12. C. Xing, X. Geng, and H. Xue, “Logistic boosting regression for label distribution learning,” Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 4489-4497.
13. M. Ali, D. Chan, and M.H. Mahoor, “Going deeper in facial expression recognition using deep neural networks,” IEEE Winter conference on applications of computer vision (WACV), 2016,), pp. 1-10.
14. A.V. Nefian and M.H. Hayes, “Hidden markov models for face recognition,” Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP'98, vol.5, 1998, pp. 2721-2724.
15. P. Yang, S. Shan, W. Gao, S.Z. Li, and D. Zhang, “Face recognition using ada-boosted gabor features,” Proc. Sixth IEEE Inter-national Conference on Automatic Face and Gesture Recognition, 2004, pp. 356-361.
16. S. Chen, J. Dai, and Y. Yan, “Classroom teaching feedback system based on emotion detection,” 9th International Conference on Education and Social Science (ICESS 2019), 2019, pp. 940-946.
17. J. Tang, X. Zhou, and J. Zheng, “Design of intelligent classroom facial recognition based on deep learning,” Journal of Physics: Conference Series, vol. 1168, No. 2. IOP Publishing, 2019.
18. A. Al-Awni, “Mood extraction using facial features to improve learning curves of students in e-learning systems,” International Journal of Advanced Computer Science and Applications 7.11, 2016, pp. 444-453.
19. K.S. Sahla and T.S. Kumar, “Classroom teaching assessment based on student emotions,” The International Symposium on Intelligent Systems Technologies and Applications, Springer: Cham., 2016, pp. 475-486.
20. K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” arXiv preprint arXiv:1409.1556, 2014.
21. P. Lucey, J.F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, “The extended cohn-kanade dataset (ck+): a complete dataset for action unit and emotion-specified expression,” IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2010, pp. 94-101.
22. S. Ramalingam and F. Garzia, “Facial expression recognition using transfer learning,” International Carnahan Conference on Security Technology (ICCST), 2018, pp. 1-5.
23. Holder, Ross P., and Jules R. Tapamo. "Improved gradient local ternary patterns for facial expression recognition." EURASIP Journal on Image and Video Processing 2017.1 (2017): 42.
24. Pramerdorfer, Christopher & Kampel, Martin. (2016). Facial Expression Recognition using Convolutional Neural Networks: State of the Art.
25. J. Cuevas, A. Chua, E. Sybingco and E. A. Bakar, "Identification of river hydromorphological features using Viola-Jones Algorithm," 2016 IEEE Region 10 Conference (TENCON), Singapore, 2016, pp. 2300-2306.
26. Acharya, Debaditya & Yan, Weilin & Khoshelham, Kourosh. (2018). Real-time image-based parking occupancy detection using deep learning.

AUTHORS PROFILE

Dr. Archana Sharma received Doctorate from Jodhpur National University, Jodhpur, MCA from University of Jammu, India. She is currently working as Assistant Professor, Department of Computer Science, Govt.M.A.M College, Cluster University of Jammu. She has 12 years teaching experience at university of Jammu. Her research area is data mining and artificial intelligence. She has published several papers in National & International Journals.

Dr.Vibhakar Mansotra received Doctorate in computer science from University of Jammu, M.Sc, M.Phil.,(Physics), PGDCA, M.Tech.(IIT-Delhi),India. He is currently working as Professor, Former Head of Department of Computer Science and IT, Dean, Faculty of Mathematical Science & Director, CITES&M, Coordinator IGNOU(S.C-1201), University of Jammu and Chairperson Division-IV, Computer Society of India. He has 26 years teaching experience at university of Jammu. His research area is data mining, software engineering and information retrieval. He has published several papers in National & International Journals.