*Grenze International Journal of Engineering and Technology, June Issue*

# AI-Powered Emotion Interpretation from Facial Cues using Deep Learning Techniques

Ramya S[1], Jayasri B S[2], Kushi S[3], Neha Sharma[4], Naman Tiwari[5] and Pratham Sharma[6]
[1]Research Scholar, Department of CSE, NIE, Mysuru, Visvesvaraya Technological University, Belagavi, India
Email: ramyas@jssstuniv.in
[2-6]The National Institute of Engineering, Department of CSE, Mysuru, India
Email: jayasribs@nie.ac.in, khushiiis05@gmail.com, nehasharmaa268@gmail.com, namanstiwari@gmail.com, sharmapratham761@gmail.com

***Abstract*— Facial emotion detection plays a crucial role in understanding human behavior and has gained major consideration in several domains, including education. This research intends an automated method for facial emotion detection of students using deep learning techniques. This work leverages deep learning model, explicitly the model known as Convolutional Neural Networks (CNNs), which extract significant features from images of human face. The models are trained to classify emotions into commonly recognized kinds, such as happiness, anger, sadness, surprise, fear, and disgust. This work aims to provide a valuable tool for educators and researchers to objectively assess students' emotional states and tailor educational interventions accordingly. By understanding students' emotional responses, educators can identify areas where students may require additional support, personalize their teaching strategies, and create a positive and engaging learning environment.**

***Index Terms*— Emotion detection, Convolutional Neural Networks (CNNs), students, HAAR-Cascade, Deep learning.**

## I. INTRODUCTION

In this ultramodern period, where utmost of the work is anticipated to be done machines emotional discovery is also one of the most important, pivotal and grueling area. Emotion discovery is used in a wide range of operations like helping to estimate stress situations, blood pressure, somnolence and numerous further operations. The identification of facial features which in turn is used for emotional discovery include methods to identify emotions like happy, sad, neutral, disgusted, angry, fearful and surprised. Deep learning algorithms, specifically convolutional neural networks (CNNs) and intermittent neural networks (RNNs) have largely demonstrated remarkable success in capturing features and connections in emotion- related data. The main advantage of emotional recognition is that it helps to identify human (students) emotional status without asking them and also issuing cautions and displaying labels in readable format. The operation of emotion discovery using deep learning has various real- world applications to detect emotional wellbeing of person. It can enhance the accuracy of detecting emotions by automatically enabling system to extract features, understand and display the facial emotions by enforcing artificial intelligence. The human facial expression detection can be a complex and delicate task.

*Grenze ID: 01.GIJET.12.2.50_21*

High quality datasets are hard to set up and there are some crucial risks to avoid when designing similar system. Deep learning is one of the veritably well- known machine learning way that could be used for facial expression analysis and emotion recognition. Still, the quantum of data affects how well it performs. Also, the performance improves as the data volume increases. To enhance the frequency and the quantity of the data, several studies use addition ways during the preprocessing stage, similar as cropping, spanning, etc. These preprocessing styles are veritably good at enhancing the capabilities of deep literacy. Therefore, this study attempts to demonstrate use of deep literacy to describe facial expressions. One of the crucial areas in emotion recognition is facial expression analysis. Deep learning models can dissect facial images and directly identify emotions similar as happiness, sadness, surprise, etc. CNN infrastructures, similar as VGGNet, ResNet, or InceptionNet, are generally used for this purpose. These models are trained on expansive labeled datasets that contain facial images annotated with corresponding emotion markers. Other uses of deep learning can be, speech analysis that involves processing audio signals to extract features like pitch, intensity, and spectral content, which can be used to infer emotional status. RNN models, similar as Long ShortTerm Memory (LSTM) networks have been successfully used in speech emotion recognition tasks. Textual data, similar as social media posts, reviews, or converse exchanges, can also give precious information about people's feelings. Deep learning models, including Recurrent neural networks (RNNs) or motor- grounded infrastructures like BERT (Bidirectional Encoder Representations from Mills), can reuse textual data and prisoner contextual dependences to prognosticate feelings expressed in the textbook. The real objective of this paper is to make use of emotion recognition technology in classrooms so that the attendants or preceptors could cover a pupil's conditioning during the class. Real-time facial emotion recognition, in general, is divided by several phases. The first phase would be detecting general area of student face in the classroom and this process includes a shadowing system which requires the tackle element to cover the class. Alternate phase would be facial land marking, which points out more accurate facial point to be uprooted. This system will need a point, which is suitable to recognize the subject's facial expression and also further use the result as emotion. Though facial expressions obviously doesn't inescapably convey feelings, in computer vision community, the term of facial emotion detection frequently refers to the bracket of mortal facial features into one of the seven feelings happiness, sadness, fear, disgust, surprise, neutral and wrathfulness. With all this above in mind we aim to develop a socially intelligent software tool that can recognize and tell us the emotion that the human is trying to express non-verbally through print or videotape.

## II. Related Work

Researchers have used various methods to describe emotions in humans. In [1][3], the authors compiled a dataset of sleep gestures from numerous seminaries in Shanghai, encompassing real classes. They employed an undisclosed methodology referred to as LMT to analyze small-sized sleep gestures. Remarkably, their system demonstrated a substantial emotional improvement of 90 on a challenging dataset, while requiring minimal additional processing time [23]. The authors also utilized observance aspect rate and mouth aspect rate as general features and validated their approach using a pretrained ResNet50 model. This research paper presents a novel system aimed at detecting individuals who are sleeping in real classroom scenes. The system serves the purpose of monitoring scholars' attention levels. Sleep gesture detection poses several challenges, including occlusion, diverse sleep gestures, interference from similar features such as jotting, and the presence of small sleep gesture targets. To overcome these challenges, the authors first create a sleep gesture dataset by collecting data from numerous real classes in seminaries located in Shanghai. Furthermore, they propose a modified RFCN (Region-based Fully Convolutional Network) integrated with point total and deformable complications to effectively detect sleep gestures. Additionally, an efficient original multiscale testing algorithm is designed to address the detection of small sleep gestures.

In [2], a system is developed to assess the engagement level of scholars using only the standard built-in webcam of a laptop. The system is designed for realtime operation and integrates data from eye and head movements, as well as facial expressions, to generate an engagement score. This score is categorized into three levels: "truly engaged," "slightly engaged," and "not engaged at all." In [3], the authors have detected drowsiness using HAAR-CNN module among individuals which can be adopted to detect Drowsiness among students in classroom environment.

In [4][5], the authors use various ways to understand scholars attention throughout a lecture. The system incorporates methods such as tone estimation, face recognition, and the use of the MTCNN for the Face Discovery block. This block is responsible for supplying the Features and Pose extraction block with images of the scholars' faces, capturing their facial features and body posture. Additionally, it combines sensor technology with an artificial intelligence system. An Artificial Bee Colony (ABC) algorithm is employed to boost system performance, allowing instructors to evaluate students' attention levels and cognitive state. Based on these insights, teachers can deliver personalized support to learners who appear distracted. Using cameras to monitor classrooms offers a non-invasive method to digitize student behavior. Understanding students' attention span and recognizing behavioral cues that may signal a lack of attention is crucial for improving lecture dynamics and optimizing teaching effectiveness. This paper commences with an introduction to the existing research in the area, presenting a comprehensive overview of the most advanced and effective computer vision methods employed in classroom coverage. Following a thorough analysis of the relevant state-of-the-art approaches, we propose an agent designed to track the attention of scholars and capture relevant data. The main aim of this study is to advance the creation of an autonomous agent designed for this purpose. Additionally, we aim to provide valuable insights for both instructors and scholars and present preliminary results related to this topic.

In [6][7], arrangement of perceiving a human's feeling from a linked human's face. The visual features of mortal facial expression which are heavily associated with different articulation analogous as happy, grief, fright, restraint, annoyance, astonished and neutral are related to restraint geometric structures introductory dimension models of the conscious system. They use techniques like modern Support Vector Machine (SVM), perfected using Grid Search and also the generality of EC separation. Various approaches, including passing the training images through Gabor sludge, or Discrete Wavelet Transform (DWT) for better type of data are implemented.

In [8], the AV (Arousal-Valence) model is employed to assess positive, negative, and neutral values. Within the AV model, the absolute values of the seven corresponding emotions are determined by considering both extremes. In [9], facial expressions are considered for the recognition of passions in humans facial images from live image from webcam The need for real time discovery of emotion in real world surrounds has encouraged researchers to produce robotic emotion database. In [10], the authors utilize OpenCV (Open Source Computer Vision) as the foundation for their design, which is based on computer vision. The system requires a real-time webcam to record video and detect facial expressions. The system also integrates advanced Support Vector Machines (SVMs) [11], a supervised learning approach, to categorize eight distinct facial emotions. SVMs serve as powerful machine learning techniques that determine the best separating hyperplane for data analysis and pattern recognition, and they are frequently utilized in regression tasks.

Facial expression recognition [12] has gained significant relevance within the domain of pattern recognition. This study introduces a new system capable of detecting facial expressions by evaluating individuals' emotional states. The framework combines multiple methods, including the Viola–Jones face detection technique, histogram equalization for image enhancement, discrete wavelet transform (DWT), and a deep convolutional neural network (CNN). The facial features extracted via DWT are used as input to train the CNN model. In [13][15][24], the authors employed convolutional neural networks (CNNs) to train on grayscale images, testing various depths and max-pooling layers to optimize performance, ultimately reaching 89% accuracy. In[14], The paper focuses on developing a system that can automatically detect and confines to only 4 emotions (happiness, sadness, anger, and surprise) using Deep Learning techniques. The proposed approach proves to be highly effective for facial expression recognition while maintaining low computational costs, making it an ideal tool for image processing and computer vision.

## III. PROPOSED SYSTEM

The system is designed to capture students' faces in real-time during class using a webcam. With the proposed CNN architecture, the goal is to efficiently and quickly process pixel values in the regions containing facial expressions and enable rapid queries using the deep artificial neural network model developed. One of the tasks of the CNN is to transform or resize images into a format that is easier to process and analyze, while retaining key features necessary for accurate predictions. This step is crucial for making the algorithm scalable to large datasets. The system detects and crops the face during the image pre-processing stage, with the face boundary being identified using the HAAR Cascade classifier from the video. The cropped facial expressions are then resized to a uniform size, and the identified emotion is displayed to the examiner.

For this we divide our work in the following modules.

- Frame Capture Module
- Image Pre- recovering Module
- Training Module
- Facial Recognition Module
- Emotion Recognition Module

For the first module i.e frame prisoner, python provides colorful libraries for image and videotape processing. One of them is OpenCV, which is a vast library that helps in furnishing colorful functions for image and videotape operations. With this library, we can capture a videotape from the camera. It lets you produce a videotape prisoner object, which is helpful to capture vids through webcam, and also you may perform asked operations on that videotape. Following the prisoner of prints, we have the alternate module called image pre-processing module which will do image processing on the captured images. The slate scale prints will be created by converting the color original still frames to slate scale. In the upcoming module, the training process will involve working with a dataset of fingerprint data. The collected prints will be stored in a YAML file (YML) for convenient data serialization. YAML, similar to XML and JSON, is a human-readable language used for representing structured data.

Serialization is a process that allows one operation or service, which may use different data structures and technologies, to transfer data to another operation using a standard format. Two main approaches form the core of the design. The initial step in the facial recognition module involves training the host system using facial data that has been gathered. A webcam is utilized to capture the subject's face, recording 60 distinct images. In this process, face detection is carried out using the Local Binary Pattern (LBP) algorithm. The original LBP serves as a straightforward yet effective texture descriptor, where each pixel is labeled by comparing it with its neighboring pixels and encoding the outcome as a binary value. The Local Binary Pattern Histogram (LBPH) is then employed to perform the recognition task. Using the previously stored face ID and NAME, the system can identify faces within the database. The next step involves the emotion recognition module, which detects emotions on human faces through biometric indicators. The proposed system is built around a Convolutional Neural Network (CNN) as its fundamental component. The convolutional layers employ a series of trainable filters, also known as kernels, each defined by a particular width, height, and depth, and typically arranged in a square shape. Although these filters are small in spatial size, they extend across the full depth of the input volume. The CNN takes as input the image channels, such as three channels for RGB images, with one channel corresponding to each color component. Since the system processes and analyzes data from images, it can provide an unbiased and neutral emotional response or data. The facial expression corresponds to the likelihood of achieving the highest level of expression based on the individual's characteristics. One of seven possible facial expressions is identified.

Figure 1 shows the architecture of this proposed system, which clearly specifies phases as well as the output.

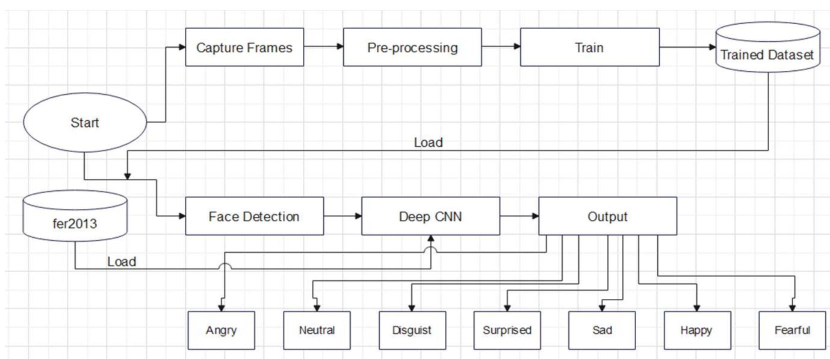



Figure 1. System Architecture

## IV. Implementation and Results

### *A. Dataset*

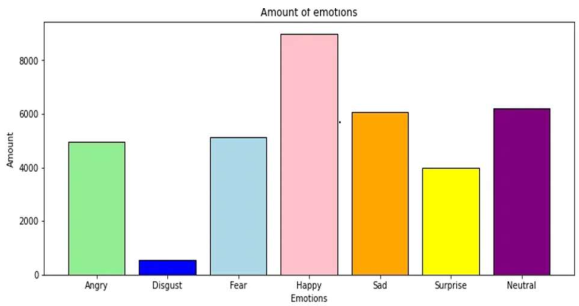
Fer2013 presently consists of roughly 28000 images with equal number of each of the 7 feelings in it. The dataset consists of grayscale facial images with a resolution of 48×48 pixels. Prior to use, the images are preprocessed to ensure that the faces are approximately centered and occupy a balanced portion of the frame.

The objective is to categorize each face based on the emotion reflected in its expression, assigning it to one of seven classes: 0 – Angry, 1 – Disgust, 2 – Fear, 3 – Happy, 4 – Sad, 5 – Surprise, and 6 – Neutral. The training dataset includes 28,709 samples, whereas the public test set comprises 3,589 samples.

The distribution of the FER2013 dataset is illustrated in Figure 2.

*Data Fields:*

- Images tensor containing the images.
- Markers tensor containing marker of images



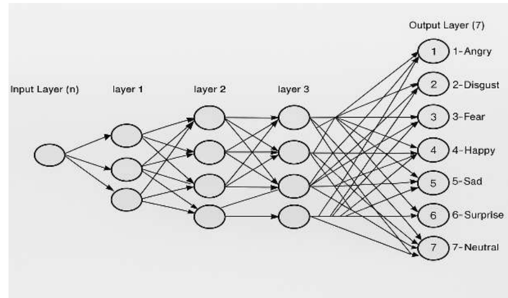




Figure 2. fer2013 dataset

Figure 3. Layered Neural Networks

*B. Algorithms used*

HAAR Cascade is a popular system for detecting faces in images. It uses a trained algorithm to identify features of a face, similar as the eyes, nose, and mouth, by analyzing the patterns of intensity in the image. Haar features are simple rectangular patterns that detect edges, lines, and differences in pixel intensity in an image. These features help in identifying the structure of objects like eyes, noses, and faces.

*A HAAR feature is defined as*

$$F = \sum I_{\text{white}} - \sum I_{\text{black}}$$

where: $I_{white}$ denotes total pixel intensity within the white region,
$I_{black}$ represents total pixel intensity within the black region
An integral image is used to calculate the features in HAAR cascade algorithm in constant time O(1).

$$I(x,y)=I(x,y-1)+I(x-1,y)-I(x-1,y-1)+f(x,y)$$

where I(x,y)) is the integral image, and f(x,y)) is the original pixel value.
To improve efficiency, HAAR Cascade uses a series of stages. If an image patch fails a stage, it is immediately rejected, reducing computation. Mathematically, a stage classifier is:

$$S_i(x) = \sum_{t=1}^{T_i} \alpha_t h_t(x)$$

where:

- Si(x) is the score for stage iii.
- Ti is the number of weak classifiers.

If
$S_i(x) \geq T_{\text{threshold},i}$ then the candidate region **passes** to the next stage. Otherwise, it is rejected. The object is detected when

$$\text{Object detected} \iff \forall i, S_i(x) \geq T_{\text{threshold},i}$$

Once the classifier is trained, it can be used to detect faces in images or vids by surveying the image with a sliding window. The classifier examines each window at different scales and exposures to identify regions of the image that match the learned patterns. The use of CNN algorithm with TensorFlow is the major part of the system which helps it to decide feelings and show them on the screen. Unlike traditional programming approaches, neural networks have the capability to learn patterns, form opinions, and make prognostications without unequivocal programming. analogous to mortal minds, a well- designed neural network can learn singly. A typical neural network consists of multitudinous artificial neurons arranged in successional layers, ranging from a many to millions of neurons. Each neuron is connected to the conterminous layers. In the proposed system, a three-layered convolutional network is employed, consisting of n input layers and seven hidden layers for each category, as shown in Figure 3. Convolutional Neural Networks (CNNs) have been demonstrated to provide higher accuracy rates compared to other neural network–based classifiers. Typically, a CNN is composed of two primary components: a convolutional layer and a subsampling (pooling) layer. These layers are designed to process two-dimensional image inputs. Keras is a library deduced from TensorFlow which is used heavily for neural networks related systems. We use MLP(Multi-Level Perceptron) to produce our model. The MLP(Multi-Layer Perceptron) model is composed of multiple layers of connected neurons. Each neuron is connected to all the neurons in the preceding and succeeding layers. The input layers admit the data, and also there may be one or further retired layers in between. The structure of the MLP model is a successional mound of layers, allowing for the accessible addition of new layers to the model. This inflexibility enables customization and trial with different infrastructures to ameliorate performance or address specific conditions. You can produce a consecutive model by calling the sequential () class and also adding layers to it using the add() functions as necessary for the system.

Convolutional layers in a neural network preserve the spatial relationship between pixels by applying filters to small squares of input data. The operations such as edge detection, blurring, and sharpening of the image is enabled by these filters. By convolving an image with different filters, the network learns to extract and detect various image features. The Rectified Linear Unit (ReLU) activation function is widely employed in Convolutional Neural Networks (CNNs) to introduce non-linearity into the model. ReLU guarantees that output values remain non-negative while maintaining linearity, a property that is beneficial for learning from real world data. By eliminating negative values and maintaining linearity, ReLU helps the network to model complex relationships and capture important features in the data.

In Convolutional Neural Networks (CNNs), pooling layers are commonly employed to reduce the spatial dimensions of input data while retaining essential information, thereby decreasing the number of parameters and mitigating overfitting. Pooling involves partitioning the input image or feature map into non-overlapping regions, referred to as pooling windows, and performing downsampling within each region. Common pooling strategies include max pooling, average pooling, and sum pooling. Max pooling selects the highest value from each pooling window, capturing the most salient features. Average pooling computes the mean value within each window, whereas sum pooling aggregates all values in the region. Following the pooling operation, the output is typically flattened into a vector, which can then be processed by a fully connected layer, analogous to a traditional neural network.

Frame augmentation involves applying various image transformations, such as flipping, rotation, saturation adjustment, and blurring, to enrich the dataset. These augmentations increase the diversity of training data, enhancing the model's ability to generalize. Following augmentation, the model is trained on the modified dataset to predict unseen data. In this study, training achieved a peak accuracy of 64.5% after 50 epochs, where an epoch represents a complete pass over the dataset. Increasing the number of epochs may further improve accuracy However, it is important to note that if the number of epochs is set to a value below 50, the accuracy may not reach the desired 64.5% mark, resulting in poorer recognition performance. Choosing an appropriate number of epochs is crucial, as too few epochs can result in underfitting, where the model fails to capture the complexities of the data, while too many epochs can lead to overfitting, where the model becomes overly specialized to the training data and performs poorly on new data.

*C. Model Training and Testing*

During the training phase, the system loads data from a designated training folder within the dataset. The dataset comprises grayscale facial images, each annotated with a corresponding emotion label. Throughout training, the neural network learns a set of optimized weights.

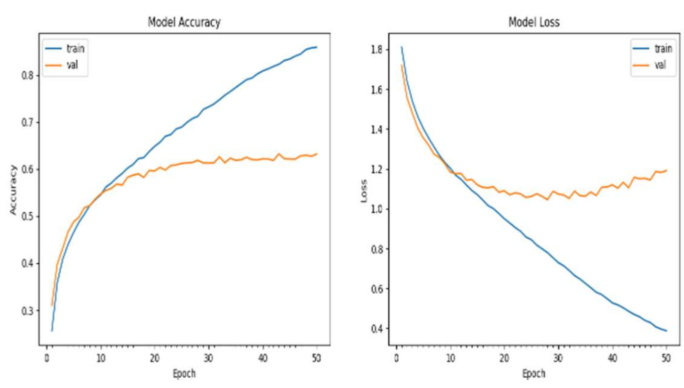



Figure 4. Accuracy Plot

The process involves taking preprocessed facial images and applying intensity normalization. These normalized images are subsequently used to train the Convolutional Neural Network (CNN), which updates its parameters to identify patterns and generate predictions based on the input data. In this study, 70% of the FER2013 dataset was allocated for training, with feature extraction performed prior to model training. To reduce the potential impact of sample ordering on training performance, a validation dataset is employed to determine the optimal set of model weights. This involves performing multiple training iterations with the samples presented in varying sequences. By analyzing the outcomes of these iterations, the final set of weights is selected based on validation performance. This strategy helps ensure that model performance is not biased by the order of training samples. The classifier utilizes a Convolutional Neural Network (CNN), which inherently detects emotions in all faces captured by the webcam. Using a simple five-class CNN, the model achieved a test accuracy of 64.5% after 50 epochs.

*D. Experimental Results*

Some of the results will be shown in figures 5-9

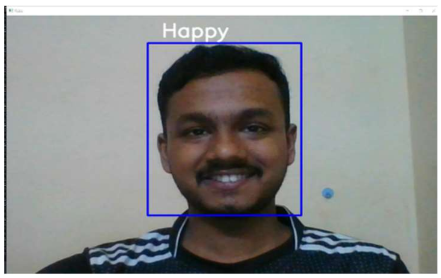



Figure 5. Happy Face identified

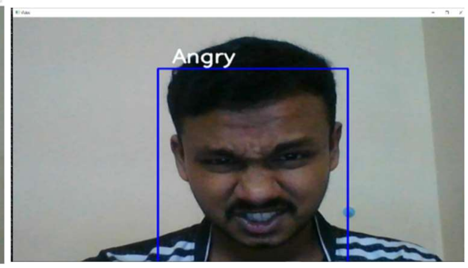



Figure 6. Angry Face Identified

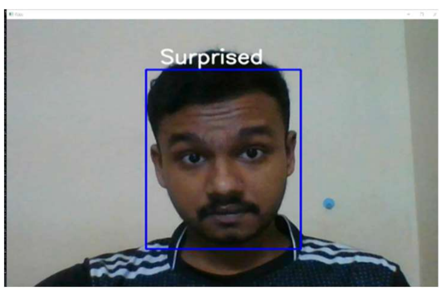



Figure 7. Surprised Face Identified

Figure 8. Multiple Faces Identified

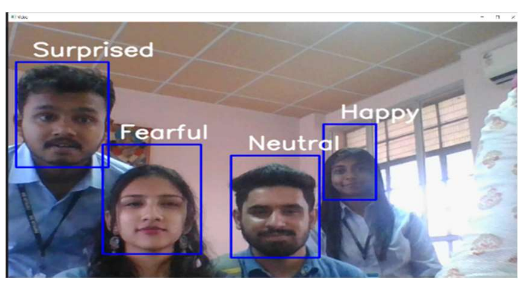



Figure 9. Multiple Faces Identified with corresponding emotions

In Figure 5. The out classifier identifies that a broad mouth length is observable in most of the training images. So when the classifier is run on this frame, it extracts all the facial features from the face and deduces a long mouth. This it predicts a happy face. In the Figure 6 the features that are common in the training dataset are generally the shortened eye, titled eyebrows, wide face and also particular cheek structure. With the detection of all these features, our system detects an angry face in Figure. 6. Similarly in Figure7, we can see a surprised face. Here the facial features detect wide opened eyes, forehead, etc. to determine which type of emotion is being displayed. The slightest of difference between surprised and fearful is the different lengths of mouth feature. Also for other features like neutral, the normal width and length of all features is taken into account by the classifier. In Figure 8 and Figure. 9 we can see the capability of this system to detect multiple faces from a single frame. It is easily able to detect all the faces present in the particular frame and also detect the variety of emotions being displayed in them.

## V. Conclusion and Future Work

In the proposed study, fusion features were utilized, and the input data were combined using Local Binary Patterns (LBP), Sphere, and Convolutional Neural Network (CNN) features. The study was conducted based on the CK dataset, and the proposed approach achieved a higher accuracy rate of 98.13% compared to other recent methods or techniques. In addition to the better performance of the proposed model in terms of accuracy, the FER2013 dataset also outperforms previous studies in terms of bracket delicacy. However, it is important to note that the proposed model is only considered average due to certain variables or factors. These variables could include the complexity of the dataset, limitations of the model architecture, or other factors specific to the analysis. Therefore, while the proposed model shows improved performance, it is still subject to certain limitations and may not achieve the highest possible accuracy or performance. For the future, the upcoming study focuses on investigating various human variables, such as personality characteristics, age, and gender, and their impact on the effectiveness of emotion detection. With the growing availability of extensive medical data, machine learning techniques are being employed to uncover emerging healthcare trends. In this study, face images will be captured in an unconstrained laboratory environment using a web camera. The goal is to analyze these images and extract meaningful insights regarding emotions, considering the influence of different personal variables. By utilizing machine learning, the study aims to enhance the understanding of emotions in healthcare environments and potentially improve healthcare outcomes. The detailed analysis results demonstrate that the proposed system, using CNN, achieves an accuracy of 85.71% in recognizing emotions through images of the mouth and eyes. In contrast, the transfer learning approach with pre-trained models using face images yields an accuracy of 64.5%. The proposed CNN-based system outperforms the pre-trained models in emotion recognition from real-time video.

## References


[1] “W. Li, F. Jiang and R. Shen, "Sleep Gesture Detection in Classroom Monitor System," ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2019, pp. 7640-7644,doi: 10.1109/ICASSP.2019.8683116.

[2] Sharma, P., Joshi, S., Gautam, S., Maharjan, S., Filipe, V., and Reis, M. J. C. S., “Student Engagement Detection Using Emotion Analysis, Eye Tracking and Head Movement with Machine Learning”, arXiv e-prints, 2019.

[3] Lahoti, Umang & Joshi, Rakshita & Vyas, Nupur & Deshpande, Kashmira & Jain, Sweta. (2020). Drowsiness Detection System for Online Courses. International Journal of Advanced Trends in Computer Science and Engineering. 1930-1934. 10.30534/ijatcse/2020/158922020.

[4] Canedo, Daniel & Trifan, Alina & Neves, António. (2018). Monitoring Students’ Attention in a Classroom Through Computer Vision. IEETA/DETI, University of Aveiro, 3810-193 Aveiro, Portugal 10.1007/978-3-319-94779-2_32.

[5] Su, Yen-Ning & Hsu, Chia-Cheng & Chen, Hsin-Chin & Huang, Kuo Kuang & Huang, YuehMin. (2014). Developing a sensor-based learning concentration detection system. Engineering Computations: Int J for Computer-Aided Engineering. 31. 10.1108/EC-01-2013- 0010. Ma

[6] Xiaoxi, Lin Weisi, Huang Dongyan, Dong Minghui, Haizhou Li, “Facial Emotion Recognition”, 2017.

[7] S. Bhardwaj, A. Dixit, K. Yadav and G. Rajan V, "A Study on Human Emotion Detection," 2021

[8] International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), 2021, pp. 909-913, doi: 10.1109/ICACITE51222.2021.9404755.

[9] C. Jain, K. Sawant, M. Rehman and R. Kumar, "Emotion Detection and Characterization using Facial Features," 2018 3rd International Conference and Workshops on Recent Advances and Innovations in Engineering (ICRAIE), 2018.

[10] H. -J. Lee and K. -S. Hong, "A study on emotion recognition method and its application using face image," 2017 International Conference on Information and Communication Technology Convergence (ICTC), 2017, pp. 370-372, doi: 10.1109/ICTC.2017.8191005.

[11] Renuka S. Deshmukh, Vandana Jagtap, Shilpa Paygude, “Facial Emotion Recognition System through Machine Learning approach” , 2017.

[12] Bendjillali Ridha Ilyas, Beladgham Mohammed, Merit Khaled, Abdelmalik Taleb Ahmed and Alouani Ihsen, “Facial Expression Recognition Based on DWT Feature for Deep CNN”, 2019 6th International Conference on Control, Decision and Information Technologies (CoDIT), 23- 26 April 2019, pp. 344 – 348, DOI: 10.1109/CoDIT.2019.8820410.

[13] Rohit Pathar, Abhishek Adivarekar, Arti Mishra and Anushree Deshmukh, “Human Emotion Recognition using Convolutional Neural Network in Real Time”, 2019 1st International Conference on Innovations in Information and Communication Technology (ICIICT), 25-26 April 2019, pp. 1 - 7, DOI: 10.1109/ICIICT1.2019.8741491.

[14] Phavish Babajee, Geerish Suddul, Sandhya Armoogum and Ravi Foogooa, “Identifying Human Emotions from Facial Expressions with Deep Learning”, 2020 Zooming Innovation in Consumer Technologies Conference (ZINC), 26-27 May 2020, pp. 36 - 39, DOI: 10.1109/ZINC50678.2020.9161445.

[15] P. Kaviya and T. Arumugaprakash, “Group Facial Emotion Analysis System Using Convolutional Neural Network”, 2020 4th International Conference on Trends in Electronics and Informatics (ICOEI), 15-17 June 2020, pp. 643 – 647, DOI: 10.1109/ICOEI48184.2020.9143037.

[16] M. I. N. P. Munasinghe, “Facial Expression Recognition Using Facial Landmarks and Random Forest Classifier”, 2018 IEEE/ACIS 17th International Conference on Computer and Information Science (ICIS), 6-8 June 2018, pp. 423 – 427, DOI: 10.1109/ICIS.2018.8466510.

[17] Rohit Pathar, Abhishek Adivarekar, Arti Mishra and Anushree Deshmukh, “Human Emotion Recognition using Convolutional Neural Network in Real Time”, 2019 1st International Conference on Innovations in Information and Communication Technology (ICIICT), 25-26 April 2019, pp. 1 - 7, DOI: 10.1109/ICIICT1.2019.8741491.

[18] I. Lasri, A. R. Solh and M. E. Belkacemi, "Facial Emotion Recognition of Students using Convolutional Neural Network," 2019 Third International Conference on Intelligent Computing in Data Sciences (ICDS), Marrakech, Morocco, 2019, pp. 1-6, doi: 10.1109/ICDS47004.2019.8942386

[19] U. Ayvaz, H. Gürüler, and M. O. Devrim, “USE OF FACIAL EMOTION RECOGNITION IN E-LEARNING SYSTEMS”, ITLT, vol. 60, no. 4, pp. 95–104, Sep. 2017

[20] A. Fathallah, L. Abdi and A. Douik, "Facial Expression Recognition via Deep Learning," 2017 IEEE/ACS 14th International Conference on Computer Systems and Applications (AICCSA), Hammamet, Tunisia, 2017, pp. 745-750, doi: 10.1109/AICCSA.2017.124

[21] A. Jaiswal, A. Krishnama Raju and S. Deb, "Facial Emotion Detection Using Deep Learning," 2020 International Conference for Emerging Technology (INCET), Belgaum, India, 2020, pp. 1-5, doi: 10.1109/INCET49848.2020.9154121

[22] E. Correa, A. Jonker, M. Ozo and R. Stolk, "Emotion recognition using deep convolutional neural networks", 2016

[23] Y. Lv, Z. Feng and C. Xu, "Facial expression recognition via deep learning," 2014 International Conference on Smart Computing, Hong Kong, China, 2014, pp. 303-308, doi: 10.1109/SMARTCOMP.2014.7043872

[24] Jayasri B S, Harshit Joshi, Prateek Kumar Choubey, Ritesh Bagati, Sartak Kumar Singh “Multi-facial Recognition Using Multi-task Cascaded Convolutional Neural Networks” pp 133–149 2025, https://link.springer.com/chapter/10.1007/978-981-96-4430-8_10”