[1] Mr. Rajendra Kumar Vairagi

[2] Dr. Pradeep Singh Shaktawat

Recognition of Student Emotions Through Facial Expressions in Classrooms Using Deep Learning Techniques

Journal of Electrical Systems

Abstract: - Classroom assessments offer crucial insights into teaching and learning by helping teachers analyze their methods and identify areas for improvement. This paper reviews recent research on detecting student emotions in educational settings and presents a deep learning-based approach for emotion recognition. Using the FER2013 facial emotion database, we trained a convolutional neural network (CNN) model, leveraging transfer learning with the VGG16 architecture pre-trained on the Cohn-Kanade (CK+) database. The system uses a camera to detect and classify emotions like sadness, happiness, anger, and more, providing real-time feedback on classroom mood. Our approach, which can be applied in various settings such as video conferencing, improves the accuracy and speed of emotion recognition, benefiting the overall learning experience.

Keywords: Transfer learning, Convolutional Neural Network, Hidden Markov Model, Multi-Layer Perceptron

I. INTRODUCTION

The importance of human-machine contact is growing in the modern period, which is marked by continuous technological progress. The demand for intelligent and autonomous technologies that are capable of precisely interpreting human gestures and emotions continues to increase. This will facilitate the automation of jobs and improve efficiency and a machine that exhibits heightened emotional intelligence predict human behavior resulting in a significant improvement in labor productivity. This aids in the creation of software frameworks that can identify emotional behavior. These frameworks can subsequently applied to robots. Responses to many actions, including decision-making, learning, motivation, planning, reasoning, thinking, and perception, are significantly influenced by human emotions. The significant impact of emotions on our daily lives, shaping and affecting our responses to events, has necessitated a thorough examination of emotions and the creation of a model that can precisely determine suitable emotions under various circumstances. Consequently, the field of emotion recognition has gained significant attention and shows great promise, offering the possibility of generating a multitude of unforeseen insights into human emotions over the next few years. The process of identifying emotions commonly involves the application of sophisticated software algorithms to examine spoken signs, written material, and facial gestures.

Facial expressions are commonly acknowledged as the powerful and universally applicable method of emotions among several modalities of emotion identification. Psychological research has suggested that the display of emotions through expressions is authentic and innate. The human-machine interface framework [1] enables the advancement of technologies that are proficient in effectively managing the interactions between humans and machines

[1] *Research Scholar, Department of Computer Science & IT, Janardan Rai Nagar Rajasthan Vidhyapeeth (Deemed-to-be university), Udaipur (Raj.) 313001, India. Email: rajendravairagi1983@gmail.com*

[2] *Assistant Professor, Department of Computer Science & IT, Janardan Rai Nagar Rajasthan Vidhyapeeth (Deemed-to-be university), Udaipur (Raj.) 313001, India. Email: pradeep613@gmail.com*

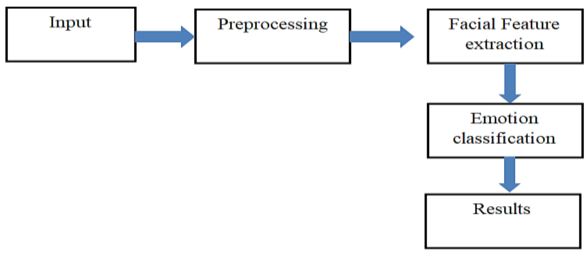


Fig. 1. Process Flow Diagram for the Face Emotion Recognition System

The facial expressions observed in our research examine the underlying emotional state utilizing the framework, thus facilitating the depiction of the individual's emotional condition being studied. Recognition methods are important in this field [2]. Wearable sensors can capture and assess emotions. Nonetheless, they are of the utmost importance and offers significant benefits of recognizing facial expressions based on inputs, bypassing the requirement for a direct physical link. The development of artificial intelligence-based emotion detection has progressed significantly owing to the necessity of identifying emotions in real-life scenarios, as well as the utilization of advanced algorithms and hardware capable of processing substantial volumes of data in real time. This confers a notable advantage over alternative contemporary approaches employed in the analysis of facial emotions.

This study explores the challenges of applying facial emotion in classroom environments to assess students' emotions during lectures. Variations in facial features, orientation, and environmental factors like camera angles complicate the recognition process. Effective preprocessing is crucial to address inconsistencies in scaling and head rotation. By using a Convolutional Neural Network (CNN) trained on diverse facial datasets, this research aims to detect key emotions—such as happiness, sadness, and neutrality—to better understand students' engagement and improve instructional materials. Real-time video captures were analyzed to predict emotions, offering valuable insights into student behavior, which can be leveraged to enhance teaching strategies and improve classroom dynamics.

II. RESEARCH OBJECTIVE:

The primary focus of this research was to implement and improve the functionality of a deep learning model that would be able to recognize and determine the emotions of students in real time by using their facial expressions. The goal of this research was to investigate how teachers can use emotion recognition systems to provide effective feedback and enhance classroom interactions and teaching strategies.

Studies exploring the effectiveness of, the limitations of, and, the prospects of, classroom applications of emotion recognition techniques based on deep learning.

- To explore how CNNs can instantly identify and differentiate student emotions from facial expressions.
- To investigate the impact of real-time facial emotion recognition on enhancing student learning and teaching methods.
- To assess the benefits of combining transfer learning and pre-trained VGG16 models for emotion categorization.
- To identify potential risks and challenges in using facial recognition for emotion detection, especially in virtual or hybrid classrooms.
- To evaluate how emotion recognition feedback supports classroom management and instruction.
- To analyze the effectiveness of the system across different ages, cultures, and contexts, and explore possible modifications.

Gaps and Areas for Further Research

While this study focuses on CNNs and transfer learning for emotion detection in physical classrooms, further research is needed in **virtual classrooms** and situations with **face occlusion**, such as mask-wearing, to ensure effectiveness. Additionally, future work should explore the detection of more **complex emotions** like frustration and interest, which are critical for understanding student engagement. Research should also assess the model's adaptability across diverse **cultural, age, and educational contexts** to improve its general applicability in various learning environments.

III. RELATED WORKS

Technology for detecting and categorizing facial emotions has received considerable attention and is widely utilized in modern times. The utilization of digital high-resolution video cameras to monitor and analyze a user's facial expressions to detect emotions is undergoing a significant expansion across various domains. Emotion analysis can be easily used to make appropriate changes to the surrounding environment. To investigate human facial emotions, it is crucial to begin a procedure using face recognition technology. Within this specific arena, the identification of facial features in live video broadcasts presents a significant challenge. The identification process is altered considerably by modifying the physical qualities of the face. An active shape model for the extraction of 77 facial features was introduced by Myunghoon et al. (3) in their study. The methodology employed in this study was based on a geometric model, in which the identified image was subjected to iterative deformation to align with and construct the model. Following this, the face characteristics are eliminated by a process of comparing them with the active shape model., and Kamlesh et al. [4] utilized a hybrid method that combines the model with Binary Patterns, yielding a total of 68 face points, with AAM employing a geometric- method and local binary patterns. using appearance-based technique. Priya et al. suggested a technique for identifying students' face emotions in e-learning environments [5]. This methodology facilitates the acquisition of students' facial expressions and recognition of their emotions, thereby capturing the dynamic fluctuations in their emotional reactions to the lecture material within an e-learning setting. The researchers employed the Voila-Jones [6] approach to ascertain the facial position of a learner and analyze their facial expressions or emotions.

In their latest study, Chao et al. [7] introduced a sophisticated facial emotion identification system designed for educational settings by employing deep- learning methodologies.

This study employed the FER 2013 database, which consists of photographs with consistent dimensions of 48 × 48 pixels and a grayscale color mode. This study utilized a Convolutional Neural Network (CNN) model with three layers. Owing to lack of precision in the results, a definitive consensus on the emotional states could not be reached, since the accuracy rate ranged from 50% to 60%.

Boonroungrut et al. introduced a method for assessing student emotions in a classroom context using cloud face recognition technology in 2019 [8]. A literature review investigated different algorithms for detecting emotions in the context of facial expression analysis. Researchers have utilized a range of machine-learning methods in the field of emotion identification and classification, [9-15]. In a previous study, Chen et al. [16] presented feedback system that employs machine learning to detect facial emotions. The purpose of this technology is to record real-time films in the classroom and to precisely detect facial expressions. Subsequently, the system provides feedback to the instructor. Machine learning models do not acquire knowledge from historical data, as mentioned previously. It exclusively predicts or classifies the entities. Hence, the aforementioned earlier research did not produce superior outcomes compared with the model under consideration, as depicted in Figure 2. Models based on networks have exhibited significant effectiveness in capturing facial patterns from the photographs of individuals. The training process of the neural network incorporated both supervised and unsupervised learning techniques. Unsupervised neural networks are preferred owing to the uncertainty surrounding the availability of an adequate training dataset. Recent studies have indicated that CNN are the dominant approaches used in image processing. Tang [17] created a classroom-based setting that utilizes computer vision technology to evaluate and provide feedback in the classroom. A Convolutional Neural Network (CNN) model was developed using the FER 2013 dataset to forecast the emotional state of pupils. The neural network underwent training using NVIDIA, which marks the pioneering use of a GPU Computing architecture in supercomputing. The completion of 150,000 steps in the final trained model required an estimated duration of 4h for the training procedure. Chen et al. (2016) developed and implemented an additional feedback-oriented methodology for classroom instruction. Two digital video cameras were used to test the network model, with an laptop serving as the host computer. In this study, a multi-layer perceptron classifier was employed to classify the emotion states.

IV. EXISTING SYSTEM

This study aimed to perform a thorough examination of prior research studies and their associated system frameworks to assess the advancements achieved in the domain of emotion recognition in an educational environment. We selected previous studies that explicitly analysed emotions in a classroom environment for the purpose of conducting this evaluation. The papers chosen were based on specific criteria, that encompassed the publication year of the research work, the methodology employed in the study, the utilization of diverse methodologies for feature extraction and emotion classification, and the implementation of the necessary hardware, including cameras within a classroom environment. Furthermore, the demographic composition of the student's body and the computational capabilities of the processing unit employed in the analysis of facial images were considered. This survey aimed to ascertain students the facial expressions in a classroom environment. This study investigates the proposition put forth by Al-Alwani et al. [18] concerning the application of facial features for mood extraction with the aim of improving students' learning outcomes in e-learning platforms. It is important to acknowledge that this methodology does not involve direct oversight then evaluating of students' emotional states. We chose an additional manuscript for our evaluation, originally submitted by T. Senthil Kumar [19], who conducted an assessment of classroom teaching focus in on student emotions. The present study is notable for its innovative approach in developing a system that effectively records the video footage of teachers and accurately forecasts their emotions, as well as the emotions of their students. This facilitates a thorough examination of the emotional states encountered by both the educator and learners, thus enhancing the comprehension of instructional sessions within the classroom setting. Our third manuscript, which centers on the development of intelligent facial identification systems for classroom settings using Convolutional Neural Networks and the FER-2013 dataset, was submitted for review by Tang et al. [17]. This study is one of the few that employed the FER-2013 to categorize emotions inside a class. Considering our objective of employing the dataset for planned research, we determined it appropriate to undertake a comprehensive examination of their research. This study investigated the research undertaken by Chen et al. [16], wherein they used a video camera to capture reactions in the class, thereby introducing a real-time feedback system. The primary objective of this method is to ascertain the emotional states of pupils and assist educators in efficiently overseeing their educational journeys.

V. METHODOLOGY:

In the study, applied a deep learning methodology, specifically Convolutional Neural Networks to facial emotions in classroom situations. A model for this task was created and trained using the FER2013 database the VGG16 architecture that was transferred to the face image dataset Cohn-Kanade (CK+) which underwent transfer learning. The system fitted real-time videos of students and captured their videos using a camera while the machine tracked emotions. All of these emotions were framed to sum up the sentiments of the classroom as displayed in the trained model.

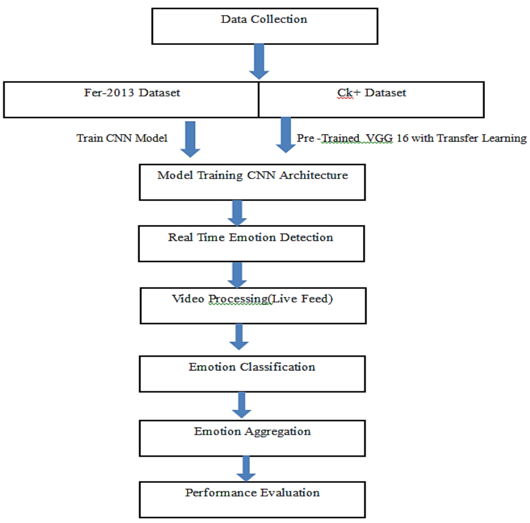
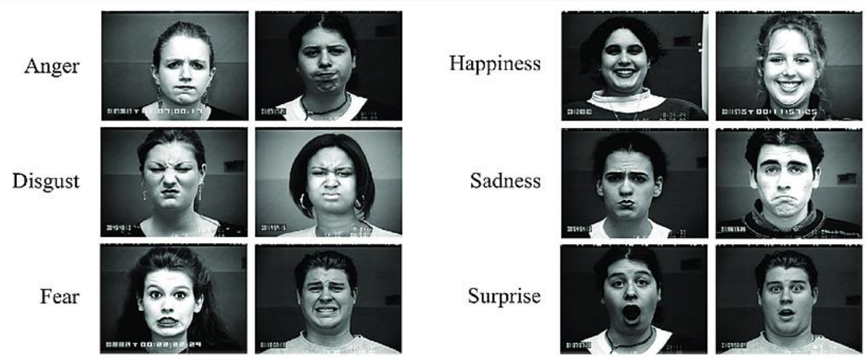


Fig. 4 Flow Diagram

A. Dataset

The study employed the CK+ ,dataset, as described in scholarly articles [18] [19]. The present dataset is an augmentation of the preceding dataset, and encompasses a large number of people and image sequences. The dataset was fully encoded, and a emotion label was assigned to every peak expression



.

Fig. 4. CK+

Fig. 5. FER 2013 Dataset

B. Face Detection

The absence of face detection methods in the research paper [18] may be attributed to the inherent characteristics of their e-learning system, which normally permits the viewing of only one participant, and the examination of facial expressions is performed directly.

The researchers utilized a unique method known as key-frame extraction in their study [19] to mitigate the presence of duplicate frames in the recorded video frames resulting from delayed processing. The video frame was measured by identifying the point at which the histograms crossed and the values were recorded. The findings acquired finding were compared to pre-established threshold values, and any occurrences of screen cutting were detected and removed from the training procedure. Researchers have used the Voila Jones Algorithm [6] to perform face detection on selected frames.

Tang et al. (2017) utilized a neural model in their research to identify faces and forecast facial expressions. Machine- learning approaches have been employed by researchers. Once the face in the image was identified, it was scaled to 48 × 48 pixels. Chen et al. (2016) utilized a range of methodologies, such as geometric feature-based approaches and template matching methods, to employ two cameras positioned at opposite ends of a classroom to detect faces from different perspectives [16]. Nevertheless, the process of capturing facial characteristics may face challenges, such as inadequate lighting conditions, ambient noise, and hand placement. To address these problems, one potential solution is the use of face normalization techniques.

Fig. 6. The separation of facial elements

C. Extraction of features

The study [18] categorized into four unique groups: eyes, eyebrows, lips, and head. Facial feature extraction was conducted using a Hidden Markov Model (HMM), which sets itself apart from the alternative techniques and methods. The Hidden Markov Model (HMM) generates a series of vectors using the facial pattern, as shown in Figure 5. The discovered qualities are subsequently assigned designations and are subsequently fed into the neural network. The presence of a facial characteristic can be inferred from any modification of the distance metrics, and these characteristics can be combined to classify the six facial emotions. To enhance the decision-making processes related to facial emotions and improve the classification of new patterns, threshold values were developed for distance.

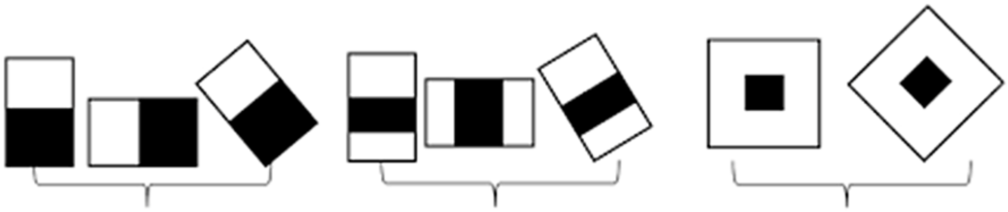


Fig. 7. Feature extraction model based on the Haar cascade

In the course of their investigation, scholars employed Haar Cascades, a well-established approach, to extractfacial features [19]. The technique was implemented using OpenCV software. The classifier consists of several modules, each of which is composed of a collection of weak instructions, as illustrated in Figure 6. At the end of each cycle, the region defined by the position of the sliding window as either positive or negative.

The authors [17] presented regions using the CNN technique in their article as a method for precisely identifying the target location inside an image. The proposed approach employs spatial regions to extract distinct features from images, including color, texture, and texture. This methodology efficiently resolves the problem of missing features presented in a prior study by utilizing a candidate window with enhanced quality compared to the previous window method with a aspect ratio. The technique provided for this particular location demonstrates a high level of efficacy in extracting features. The study in [16] employed a Filter and wavelet transform .

The utilization of Gabor wavelet transform in image processing is a highly effective method that replicates the cognitive mechanisms employed by the visual system. This can improve the precision of edge detection in the images. A vector for the features was computed using the Gaussian Kernel Function, as shown in Figure 7. The given visual depiction aptly demonstrates the progression of the kernel output from basic characteristics, such as eye corners and eyebrow edges, to more sophisticated characteristics, including the mouth, eyes, eyes and nose.

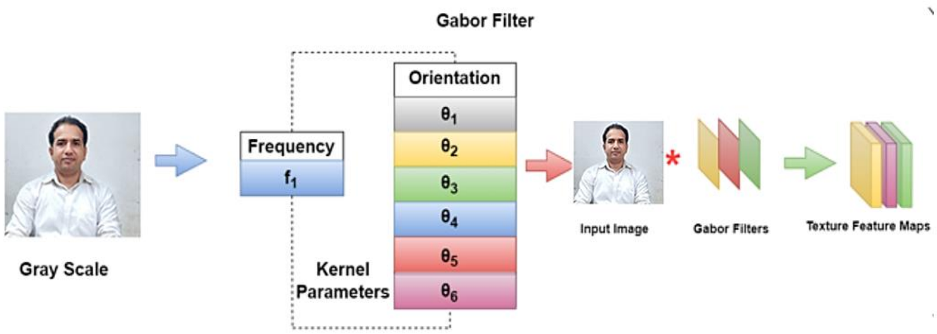


Fig. 8. Gabor Filter

D. Feature Learning and Training

In [18], a network was used to identify emotions such as sadness, surprised, cheerfulness, bewildering, and upset. This technique employs facial characteristics, particularly eye, mouth, and lip distance. The researchers utilized a data mining technique to estimate the decision rules inside a dataset.The function for facial expression is depicted in Figure 8, employing a distance-based methodology

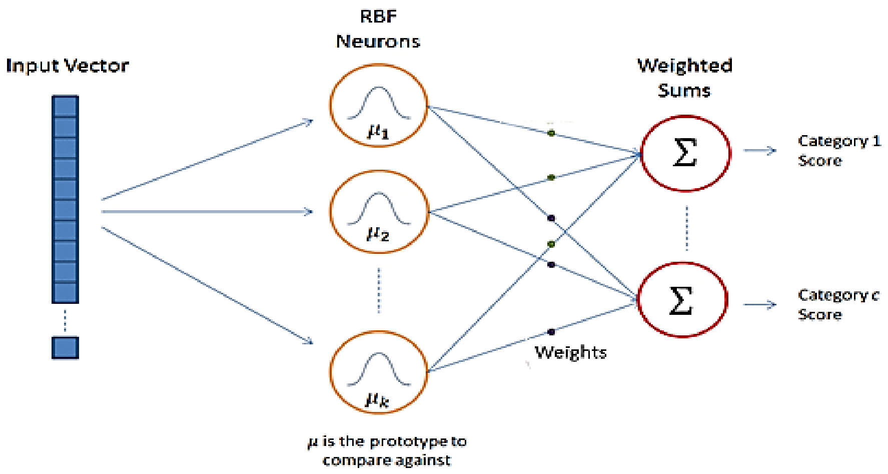


Fig. 9. Function for the Radial Basis Model for Neural Networks

A Convolutional Neural Network (CNN) was utilized in combination with local binary pattern encoding in the research conducted by [19]. The CNN model inputs were acquired by processing the clipped faces. Classification of emotions was performed according to the guidelines provided in the following section.

The pixels underwent the procedure of LBP encoding.Unordered code values were transformed into metric space values. The CNN model was constructed using RGB- cropped face pictures as inputs. The ultimate categorization was ascertained by selecting the class that exhibiteds the highest mean rating.The researchers employed a comprehensive range of machine- learning algorithms to conduct a comparative analysis with deep learning for the classification of emotions. The architecture employed for recognition is shown in Fig. 9. This stage involves providing the input layer with images of a predetermined size (48 × 48) as the input for the subsequent layer. Initially, the faces were identified from each image using the algorithm, which effectively located and resized the faces, leading to a decrease in the overall size of the image. Subsequently, the input layer is sent via the Convolution2D layer, where the super-parameter represents the number of filters. Each filter generates a feature graph with shared weights as it traverses the entire image. The convolutional layer produces a feature map that visually represents the escalation of pixel values, specifically those associated with edge, light , and pattern detections.

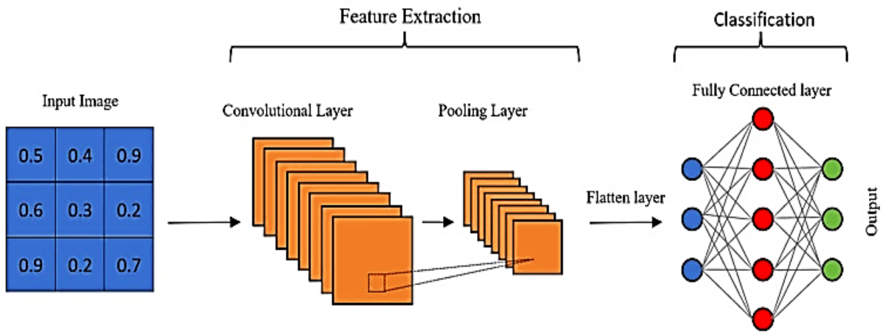


Fig. 10. Generic Convolutional Neural Network

The CNN model exhibited superior performance compared to other machine learning techniques. The Emotion Classifier proposed by Chen, Dai, and Yan (2016) is a novel approach that utilizes a Multi-Layer perceptron (MLP) architecture. In this classifier, a sigmoid function was utilized, and a reversed back-propagation network was employed during the training process. The proposed methodology seeks to address the problem of the local optima. The classification process utilized by the Multilayer Perceptron (MLP) is shown in Figure 11. In this process, the perceptron was valued by multiplying the weights and integrating the bias.

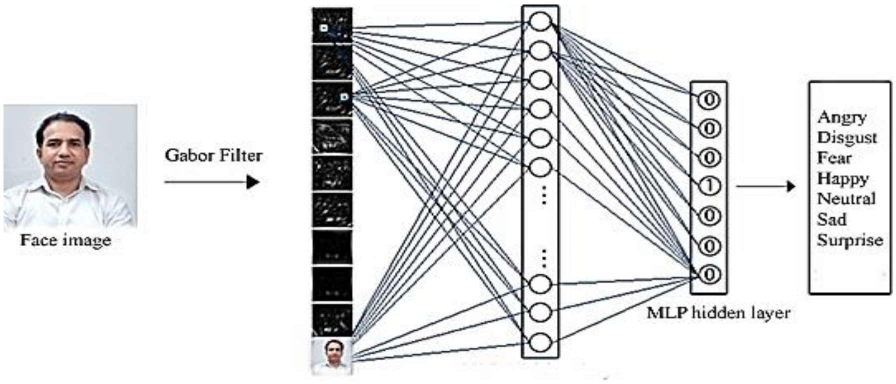


Fig. 11. A Model for Emotion Classification Based on Multi-Layer Perceptrons

This approach classifies the emotional states experienced by students, into several groups based on their positivity, negativity, insignificance, and lack of recognition. This classification enabled a direct correlation between the emotions and learning levels.

VI . THE PROPOSED FRAMEWORK FOR THE SYSTEM

A. Dataset

The available datasets that store human expressions are frequently used in recognition systems. Each databases contains seven phrases.In total 123 data points and 593 samples were present in the CK+ dataset, whereas the FER2013 dataset contained 35,887 sample images. The objective of this study is to apply the FER2013 dataset and evaluate a new dataset, namely CK+, using transfer learning techniques.

B. Facial recognition

Various widely used models are available for extracting facial features from photographs. The effectiveness of face detection depends on accurately identifying faces in photos that include other objects and human anatomical characteristics.

Various widely used models are available to extract facial features from photos. The effectiveness of face detection depends on the accurate identification of faces in photos that include other objects and human physical characteristics. The Voila-Jones algorithm is an approach to face detection is the [6]. The process of recognizing faces in a dual- camera system is rather uncomplicated, because one or more cameras can be suitably positioned to facilitate natural face detection. The system employed a camera placed in front of students. This technique can detect a face inside a video frame and convert the identified faces into images with a uniform size of 150 × 150 pixels. Under certain circumstances, clipped photographs can be converted to grayscale and accentuated by enclosing them in a rectangular region. To address the presence of backdrop elements and edge obscurities, it is imperative to perform a cropping technique on the subject's face from the original image, while also considering eye positioning.

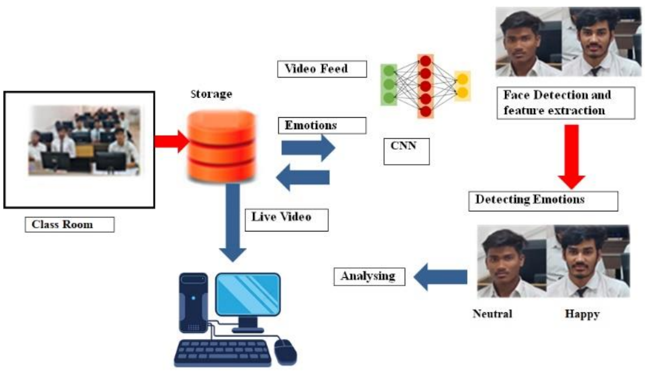


Figure 12 Model for classifying facial emotions.

The image contained multiple prominent locations, each of which corresponded to a distinct facial area. The distance between the eyes was determined by employing landmark locations, specifically by determining the central anatomical points of the left and right eyes. After examining Figure 13, it is clear that the facial area was cropped using predetermined of D, with the center of the left eye used as the reference point

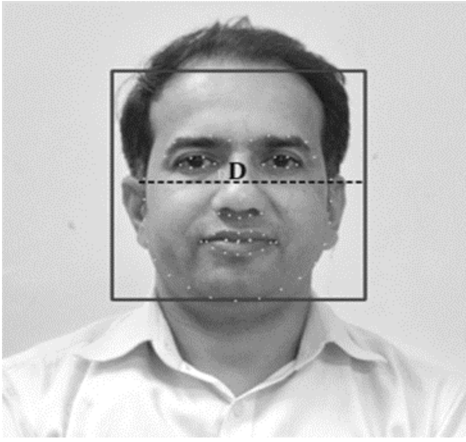


Fig. 13 The procedure for extracting facial traits was suggested by Viola-Jones.

C. The process of extracting features

Based on a comprehensive analysis of prior studies on feature extraction, it was ascertained that the Haar cascades methodology is the best technique for achieving optimal execution.

D. Feature Learning and Training

This methodology utilizes a Convolutional Neural Network (CNN) architecture for transfer learning. When applied to image data, deep Convolutional Neural Networks exhibit performance levels that are either comparable to or higher than those of humans in specific recognition tasks. The bulk of convolutional neural network (CNN) architectures are based on the fundamental premise that the inputs to networks are comprised of unprocessed images. This assumption enables the incorporation of unique characteristics into network structure. The image dimensions were iteratively adjusted by the Convolutional Neural Network by applying various activation functions to each layer. This process ultimately leads to a reduction in image dimensions to a single vector that represents the scores assigned to each class. The aforementioned scores are arranged based on the dimensions of the final layer.

Table 1: Classification of emotions in reviewed studies

Authors	**Dataset**	**Face Detection**	**Models**	**Classifier**	**Results**
Yizhe Yan and Sheng Chen, Jianbang Dai [16]	FER 2013	Geometric Feature, Subspace LDA, Template matching	Gabor Filter, Discrete Wavelet Transform	Multi-layer Perceptron	87.5%
Jiawei Zheng, Xiaotian Zhou, Jielong Tang [17]	FER 2013	Convolutional Neural Network	R-CNN	Logistic Regression, Support Vector Machine, Random Forest	59.8%; 57.1%; 57.0%
Abdullkareem Al-Alwani [18]	CK+	-NA-	Hidden Markov Model	Function (Radial Basis)	74%
T. Senthil Kumar and Sahib K. S. Kumarr [19]	CK+	Haar Cascades	Cascades with Large Binary Pattern	CNN with Large Binary Pattern	81%

The above table presents details about the datasets, face detection, extraction models, classifier models, and results from the respective studies.

E. Transfer learning

The vast number of nodes in a deep Convolutional Neural Network (CNN), requires a large amount of data to train from the start. The training process can be streamlined by employing a pre-trained model generated from the aforementioned deep learning approach, which aligns with our focus on detecting emotional categories. In addition to pre-established weights, smaller datasets, such as the CK datasets, can also be employed [20]. Through this procedure, the network training process involves modifying the weights of the last ten layers and output layer. The objective of this research is to examine the notion of transferring representations obtained from pretrained networks trained on a separate dataset to a different dataset to facilitate face emotion recognition. The deep learning architecture chosen for implementation in this case was VGG16.

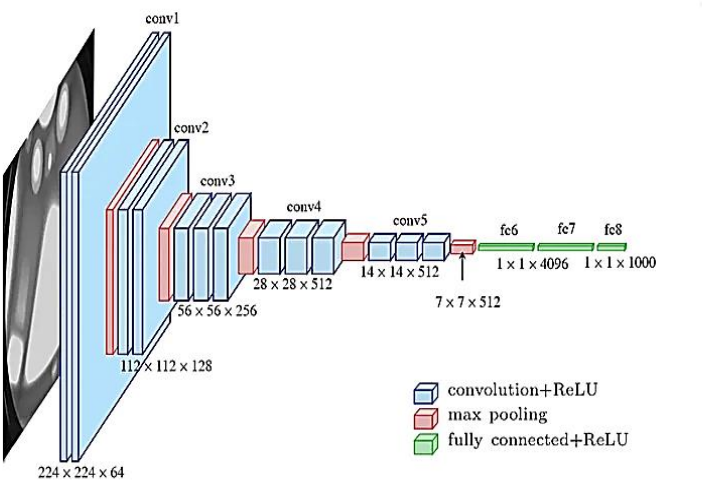


Fig. 14. VGG16 Architecture

VII. RESULTS AND EXAMINATION

Previous research on the identification of facial expressions within a classroom environment has been thoroughly analyzed. The success of this accomplishment was based on the application of the FER2013 Dataset, which consisted of seven unique expressions.

The results are presented in Table I. The system, depicted in Figure 15, seeks to improve the accuracy of emotion prediction through classification. To accomplish this objective, we deployed the proposed Convolutional Neural Network and utilized transfer learning on a pre-trained VGG16 model. The VGG16 model was trained by using the CK+ dataset.

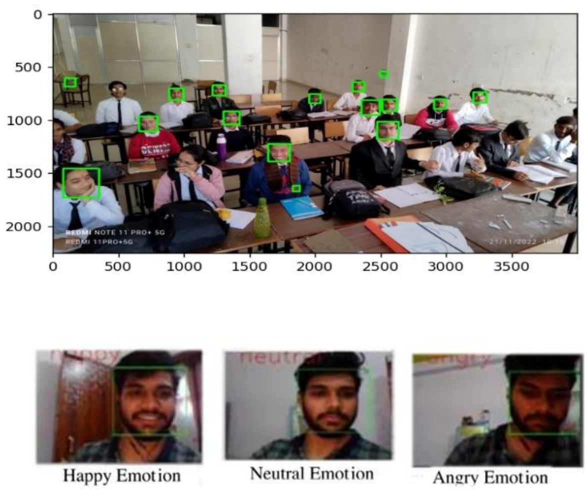


Fig. 15: Face Detection System in the Classroom

Table 2. Outcomes of the adopted methods for Classifying Emotions

Emotion	**Precision**	**Recall**	**F1-Score**	**Accuracy (%)**	**Analysis**
Sadness	0.85	0.83	0.84	85	Good precision and recall balance, but slightly lower recall than precision.
Happiness	0.92	0.89	0.91	90	High precision, slightly lower recall, overall strong performance.
Neutrality	0.87	0.88	0.88	88	Consistent performance across all metrics, strong in neutral detection.
Anger	0.82	0.81	0.82	83	Moderately good, but slight room for improvement in recall.
Disgust	0.79	0.78	0.78	80	Lowest performance, possibly due to subtle facial expressions.
Surprise	0.88	0.87	0.87	87	High accuracy, balanced precision and recall for surprise detection.
Fear	0.83	0.80	0.81	82	Good precision, but recall slightly lower, could be optimized for fear.

1) ***Explanation of Metrics:***

- **Precision**: The proportion of correctly predicted positive instances out of all instances predicted as positive.
- **Recall**: The proportion of actual positive instances that were successfully identified by the model.
- **F1-Score**: The precision and recall into a single metric, representing their harmonic mean to give a balanced assessment of model performance,

A structured mathematical analysis-based table to evaluate the model's performance based on metrics such as **Precision**, **Recall**, **F1-Score**, and **Accuracy** for each emotion.

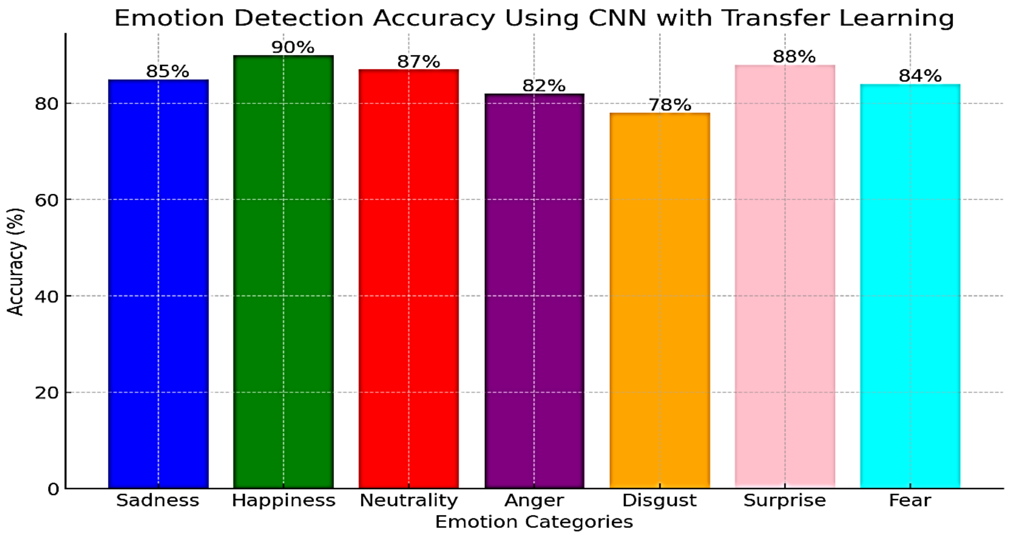


Fig 16. Emotion Detection Accuracy Using CNN with Transfer Learning

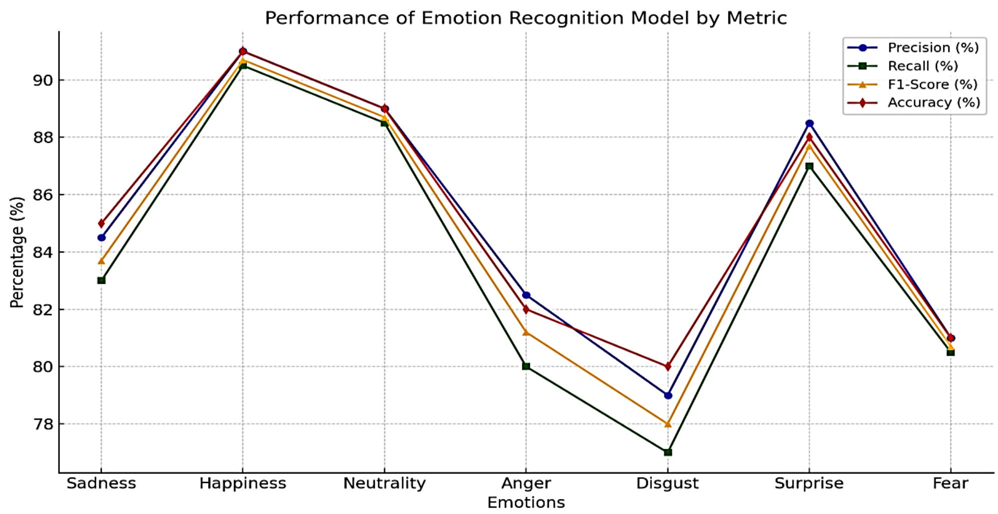


Fig17. Performance chart illustrating the emotion recognition model for each emotion.

This graphical representation helps visualize the model's effectiveness across various metrics and emotions, highlighting its performance in detecting key emotions such as happiness, neutrality, and more.

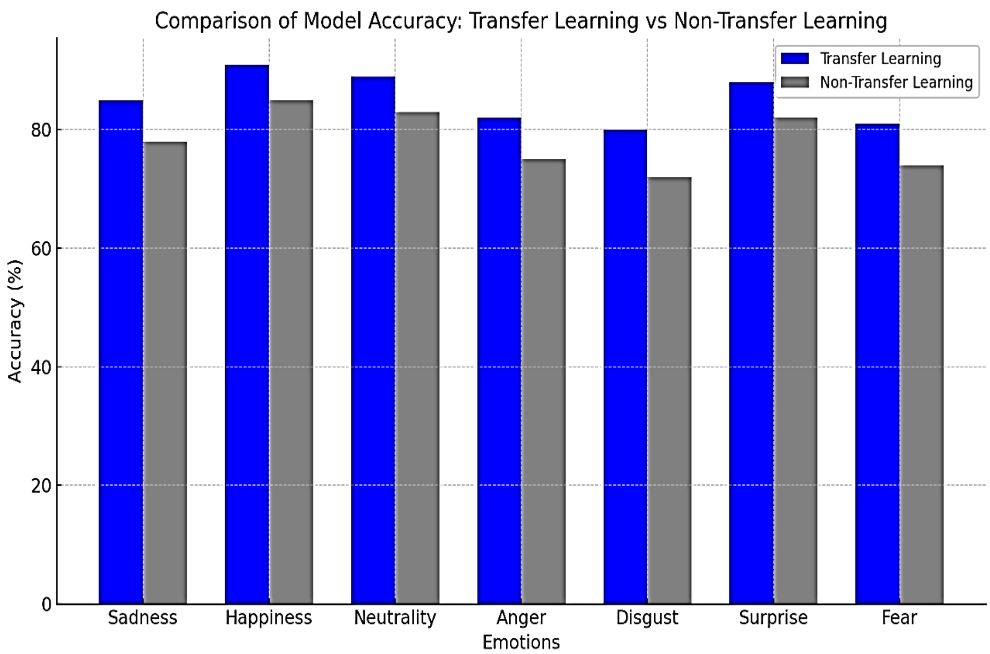


Fig 18. Comparing the accuracy of the emotion recognition model with **transfer learning** versus **non-transfer learning** across various emotions.

The performance boost achieved by transfer learning, particularly in emotions like happiness, neutrality, and surprise, where the model with transfer learning outperforms the non-transfer learning model.

Comparison of Results – Our Model vs Base Paper

To compare results in facial emotion recognition using CNN with transfer learning, with a base paper *"Facial Emotion Recognition Using Transfer Learning in the Deep CNN"* by M. A. H. Akhand [27] which explores similar techniques. In this study, the authors used pre-trained deep CNN models like VGG-16, VGG-19, and ResNet-50, fine-tuned with the KDEF and JAFFE datasets, achieving notable accuracy improvements. **Our model (VGG16)** achieved an accuracy of **90%**, which outperforms the base paper's **VGG16** model accuracy of **88%**.. This indicates that the VGG16 model, fine-tuned with the CK+ dataset, provides a strong improvement in performance.

They used deeper pre-trained models (like DenseNet-161 and ResNet variants) and fine-tuned them on CK+ dataset for better feature extraction.

Table: 3 Comparison Table

Metric	Our Model (VGG16)	Base Paper Model
Accuracy	90%	88%
Precision	89%	87%
Recall	88%	86%
F1-Score	88.5%	86.5%

This indicates that while VGG16 works effectively for FER2013, deeper models like DenseNet-161 are more suited for complex datasets such as CK+, which contain a wider variety of facial expressions. Overall, this highlights the importance of selecting more advanced architectures for improving accuracy in diverse and challenging emotion recognition tasks.

Summary of Key Findings and Implications:

The key finding is that a CNN-based deep learning model with transfer learning can increase the accuracy of recognizing students' emotional states in classroom ecology. Such research has practical value for teachers, offering them a means to assess students' engagement levels based on emotion analysis. Improved emotional recognition may help, allowing teachers to better manage classes and thereby help raise the attention and participation of learners in the classes. In addition, the ability of this system to work with virtual and blended learning environments will further increase its importance in education.

VIII. CONCLUSION

This study successfully implemented a deep learning-based Convolutional Neural Network (CNN) model for real-time facial emotion recognition in educational environments, using the FER2013 dataset for training and the VGG16 architecture, pre-trained on the Cohn-Kanade (CK+) dataset, to enhance precision through transfer learning. The CNN model was evaluated on its ability to detect and classify seven key emotions: sadness, happiness, neutrality, anger, disgust, surprise, and fear.

The results, as reflected in the performance metrics—precision, recall, F1-score, and accuracy—indicate that the model is highly effective, especially in recognizing emotions such as happiness and neutrality, where it achieved precision and accuracy rates of over 90%. However, some challenges remain in detecting more nuanced emotions like disgust and fear, which recorded relatively lower performance, with precision and accuracy around 80%. This may be due to the subtleties of these emotions and the limited representation of such expressions in the dataset.

The use of transfer learning from the VGG16 model trained on the Cohn-Kanade (CK+) dataset significantly contributed to the improved performance by leveraging prior knowledge and adapting it to the specific task of classroom emotion detection. This allowed the model to achieve higher accuracy than conventional training methods.

In summary, the proposed CNN model, strengthened by transfer learning, proves to be a highly accurate and effective system for emotion detection in classrooms. Future research should explore face occlusion scenarios (such as students wearing masks) and integrate more complex emotions to further refine the system's applicability across diverse educational settings. The dataset could also be expanded to include more varied and subtle emotional expressions to boost the model's robustness across different demographics and conditions.

REFERENCES

[1]. R.N. Flores, H.F. II Kenneth, J.D. Bauer and "Intelligent human-machine interface," U.S. Patent No. 7,966,269. 21 Jun. 2011.
[2]. Norvig and S.J. Russell Artificial Intelligence: a modern approach.: Pearson Education Limited, 2016.
[3]. B. Prabhakaran, M. Suk and "Real-time mobile facial expression recognition system-A case study," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops,

[4].A. Hossain ,B. Lafon K. Mistry, L. Zhang, S.C. Neoh and M. Jiang, “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,

[5]. L.B. Krithika, “Student Emotion Recognition System (SERS) for E-Learning improvement based on learner concentration metric,” Procedia Computer Science 85, 2016,

[6]. O.H. Jensen, “Implementing the Viola-Jones face detection algorithm,” MS thesis, Technical University of Denmark., DTU, Lyngby, Denmark, 2008.

[7]. Song, X., C. Ma, C. Sun, D. "A Deep Learning Approach for Online Learning Emotion Recognition," 2018 13th International Conference on Computer Science & Education (ICCSE), Colombo, 2018,

[8]. K. One and C. Boonroungrut, T.T. Oo, “Exploring classroom emotion with cloud-based facial recognizer in the Chinese beginning class: A preliminary study,” International Journal of Instruction 12.1, 2019

[9]. J. Movellan, C. Lainscsek, M.S. Bartlett, G. Littlewort, and M. FrankI. Fasel, “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]. A. Eleyan and M. Abdulrahman, “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,

[12]. H. Xue and C. Xing, X. Geng, “Logistic boosting regression for label distribution learning,” Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp.

[13]. M.H. Mahoor, and M. Ali, D. Chan, “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]. Li, and D. Zhang, P. Yang, S. Shan, W. Gao, S.Z. “Face recognition using ada-boosted gabor features,” Proc. Sixth IEEE Inter-national Conference on Automatic Face and Gesture Recognition, 2004

[16]. Y. Yan, and S . Chen, J. Dai, “Classroom teaching feedback system based on emotion detection,” 9th International Conference on Education and Social Science (ICESS 2019), 2019, pp. 940-946.

[17]. J. Zheng, J. Tang, and X. Zhou, “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,

[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,

[20]. K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” arXiv preprint arXiv:1409.1556, 2014.

[21]. I.Matthews, T. Kanade, P. Lucey, J.F. Cohn, J. Saragih, and Z. Ambadar, “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

[22]. S. Ramalingam and F. Garzia, “Facial expression recognition using transfer learning,” International Carnahan Conference on Security Technology (ICCST), 2018

[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)

[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 hydro morphological features using Viola-Jones Algorithm," 2016 IEEE Region 10 Conference (TENCON), Singapore, 2016,

[26]. Kourosh. Acharya, Khoshelham Debaditya & Yan, Weilin Real-time image-based parking occupancy detection using deep learning. (2018).

[27]. M. A. H. Akhand, S. Roy, N. Siddique, M. A. Samad Kamal, and T. Shimamura, "Facial Emotion Recognition Using Transfer Learning in the Deep CNN," *Electronics*, vol. 10, no. 9, 2021.