

Received 10 June 2024, accepted 24 July 2024, date of publication 1 August 2024, date of current version 12 August 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3436531

RESEARCH ARTICLE

Decoding Student Emotions: An Advanced CNN Approach for Behavior Analysis Application Using Uniform Local Binary Pattern

UPPANAPALLI LAKSHMI SOWJANYA AND R. KRITHIGA ^{id}

School of Computer Science and Engineering, Vellore Institute of Technology, Chennai 600127, India

Corresponding author: R. Krithiga (krithiga.r@vit.ac.in)

ABSTRACT In this paper, Facial Emotion detection model, built using Uniform Local Binary Pattern, is explained using an application i.e., Student behavior detection. Student engagement is a critical factor in effective teaching and learning. Conventional approaches to evaluating student participation frequently depend on assertions made by students and observations made by teachers, which can be laborious and unreliable. The use of computer vision techniques to automatically identify the emotions and behaviors of students in real time has gained popularity in recent years. This research provides a novel computer vision-based method for detecting student behavior, emphasizing emotion recognition. We propose a system that utilizes Convolutional Neural Networks (CNNs) for emotion classification, data augmentation to enhance dataset, and Uniform Local Binary pattern (uLBP) feature extraction for improved texture analysis. The system aims to provide educators with real-time insights into student emotions, enabling them to adapt their teaching strategies accordingly. In order to expand the variety and breadth of the training dataset, we enhanced the facial expression data that we extracted from a dataset that's well suited for this work. We then applied uLBP feature extraction to capture local texture patterns in the facial images, which were then used as input to the CNN for emotion classification. According to our findings, the suggested approach successfully classified student emotions with a high accuracy of 94% and also completed the ablation study, which indicates its usefulness in practical settings.

INDEX TERMS Computer vision, deep learning, emotion detection, student behavior, data augmentation, uniform local binary pattern (uLBP).

I. INTRODUCTION

There are various examples and applications where we use Facial Emotion recognition. Here in our work, we take one such application i.e., Student behavior detection to evaluate and explain briefly, how the model works. There are various reasons why student behavior matters [1]. Firstly, because student engagement varies greatly depending on the environment, it is imperative to monitor and understand it both within schools and inside classrooms. Second, the likelihood of graduating with a degree and academic achievement are directly correlated with student behavior. Teachers

The associate editor coordinating the review of this manuscript and approving it for publication was Varuna De Silva ^{id}.

can support students in reaching their goals and improving their chances of success by encouraging positive behaviors. [2] Due to varying problems requiring varying approaches or solutions, university students do not exhibit any one behavior. Every term, teachers struggle to conceptualize the particular behavioral problems that need to be addressed in the classroom. Sometimes they have a few issues that prevent them from acting appropriately in class. [3] Teachers can successfully identify and meet the different needs of their students by analyzing data on student behavior. Schools can learn more about their students' mental well-being, including any indications of stress, anxiety, or emotional distress, by looking at their behavioral patterns. With the use of this information, educators and other support personnel can make

timely interventions or recommend students to specialized services. Additionally, according to [4] the analysis of behavior data offers the chance to pinpoint pupils who might need extra help along the way. Teachers can adapt their teaching strategies to each student's needs by recognizing their behavior patterns and degree of participation. For example, in order to improve a student's academic success, teachers can investigate alternate teaching styles or create tailored learning plans if the student repeatedly demonstrates low involvement. Instructions and a breakdown of typical student conduct. It explains how a teacher can identify or understand a student's behavior. There are many works on analysing Student performance like in [5], [6], [7], and [8] along with papers on how to use Artificial Intelligence in [9] and [10]. With all these existing research papers in the educational field or in student performance analysis, this research paper focuses on building a computer vision model on Student behavior and deriving the outcomes of the session.

Contribution to work:

- To conduct an extensive review of literature on utilizing Computer Vision for detecting student behavior, examining existing methodologies and the field's challenges.
- To apply data augmentation to expand the training dataset's diversity and size, improving the model's ability to generalize to novel data. Formulated and trained the Convolutional Neural Network (CNN) model for emotion classification, integrating advanced techniques like Local Binary Pattern (LBP) feature extraction to enhance accuracy.
- To perform comprehensive testing and evaluation of the model, scrutinizing performance metrics like accuracy, precision, recall, and F1-score to gauge its efficiency.
- Contribution to refining the model architecture based on evaluation outcomes, ensuring its effectiveness in practical scenarios.

II. RELATED WORK

This section focuses on the background study related to our topic. We have reviewed papers that worked on facial emotion recognition in the past to enhance our knowledge on this topic. Khanzada et al. in [11] used Ensemble of Deep Learning models for FER2013 and achieved up to 75.8% accuracy but there were considerable error analysis issues due to the subjective nature of emotions. Zahara et al. mentioned in their work [12] that they used CNN for real-time FER on Raspberry Pi with FER-2013 but they face problems with dataset constraints and hardware considerations. Ramis et al. [13] used a dataset approach with a CNN-based FER System, which gave up to 73.05% accuracy but it had limited adaptability to recent deep learning approaches. The work [14] used Deep neural networks with cross-database validation using AffectNet and RAF-DB but CNN network proposed is not consistently scalable to all Asian emotions. Huang et al. [15] used a Deep Neural Network with cross-database validation where

it was limited by the generalizability of the model to Asian faces, and cultural differences in expression interpretation. Mellouk et al. [16] mentioned Dynamic kernels for facial expression recognition but there's not much information on computational efficiency in real-time solutions. Perveen et al. [17] used a Robust FER System with a modular approach and 2D Taylor expansion. It needs further work to enhance automatic facial expression recognition in diverse conditions. Sujata et al. [18] amended the representation module for improved FER in [18] but there are limited insights into the scalability of the amended representation module across different datasets and conditions.

In work [19] global multi-scale and local attention network (MA-Net) was used which focused on improving challenging conditions and achieving state-of-the-art performance. It includes a feature pre-extractor, multi-scale module, and local attention module. Zhao et al. in [20] used Attentional CNN for facial expression recognition. In this, the network demonstrates notable advancements across diverse datasets. Visualization techniques pinpointed pivotal facial regions for emotion detection. Minaee et al. [21] used technology in early education assessments to investigate the role of technology in assessment practices within play-based kindergarten classrooms. It highlights positive views on technology integration for enhancing teaching and assessment practices. Danniels et al. [22] used observational assessment in early care and education, timing, and content/ format factors. Provides insights into the nuances of quality benchmarks in assessment practices. Thorpe et al. [23] used Spatio-Temporal Attention (STA) for 3d CNNs and introduced the STA module for enhancing 3D CNNs in action recognition and detection tasks. Addresses spatial and temporal variations in video frames. Achieves state-of-the-art performance on benchmark datasets. Residual Multi-Task Learning Framework (RMT-Net) [24] pioneering RMT_Net for simultaneous facial landmark localization and expression recognition. It utilizes a unique residual learning module.

Multi-Modal Facial Expression Recognition(MRAN) [25] uses color, depth, and thermal information for improved recognition. However, it has some specific challenges faced in multi-modal recognition that are not discussed. Dynamic Kernels with a Gaussian Mixture model are used in [26] presenting a robust facial expression recognition system with dynamic kernels and a universal Gaussian mixture model. Achieved effectiveness in accuracy and computational efficiency. However, it still has challenges with dataset constraints and hardware considerations. Recommendations for optimizing CNN architecture and incorporating diverse datasets. Themes Discovered in Review:

- Facial Expression Recognition (FER) Advancements
- Real-world Applicability and Deployment
- Cross-dataset Approaches and Challenges
- Innovations in Model Architectures
- Multi-modal Approaches for Enhanced Recognition
- Challenges in Interpretability and Error Analysis

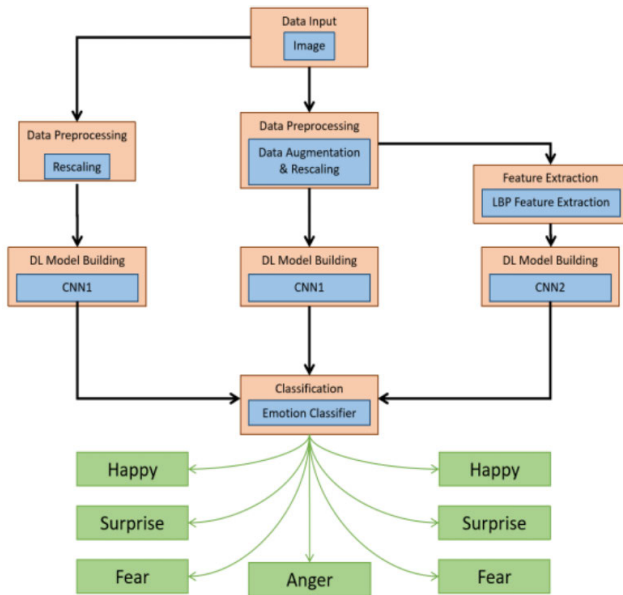


FIGURE 1. Architecture diagram.

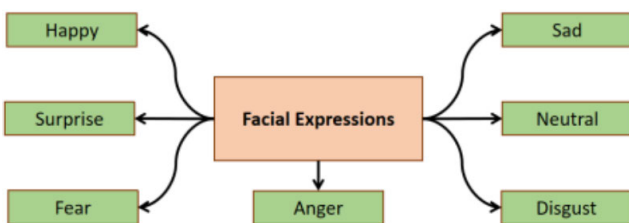


FIGURE 2. Classified facial emotions.

The overall architecture of the proposed work is depicted in Figure 1.

A. DATASETS

One of the most used datasets for Facial Expression Recognition is [27]. It has 48 by 48 grayscale pictures with labels for seven distinct emotions: disgust, fear, happy, neutral, sad, surprised, and angry (As shown in Figure 2 and Figure 3). Preprocessing includes shrinking the photos to a standard 48 × 48 pixel size and turning them to grayscale. The split folders module is used to split the dataset into training, validation, and test sets. This dataset is useful for training and testing emotion identification algorithms since it offers a wide variety of facial expressions.

B. CNN MODEL

A subclass of deep neural networks called convolutional neural networks, or CNNs, have demonstrated remarkable performance in image categorization tasks [28]. CNNs are particularly effective for image-based emotion classification because of their ability to automatically learn spatial hierarchies of features from input images [29]. Convolutional layers are utilized for the extraction of features, pooling layers are



FIGURE 3. FER2013 dataset.

used to compress spatial dimensions and fully connected layers are used for categorization in the CNN model employed in this work [30]. Because of its architecture, the model is ideally suited for problems like the recognition of emotions since it can derive intricate patterns and characteristics from the input images. The capability of CNNs to autonomously extract features on a wide scale sets them apart from traditional ML algorithms like SVMs and decision trees. This ability eliminates the need for manual feature engineering, which boosts efficiency [31]. CNNs can recognize and extract patterns and features from input regardless of changes in location, orientation, scale, or translation because of the translation-invariant properties that the convolutional layers confer on them. Numerous pre-trained CNN designs have proven to perform exceptionally well, such as VGG-16, ResNet50, Inceptionv3, and EfficientNet. Through a process called fine-tuning, these models can be made to work on new tasks with comparatively little data. In addition to image classification tasks, CNNs are flexible and can be used in many other fields, including speech recognition, time series analysis, and NLP. When compared to conventional machine learning methods, the accuracy and robustness of emotion recognition using CNNs have been substantially enhanced [32]. The ability of artificial intelligence to bridge the gap between human and computer capabilities has grown exponentially. Researchers and amateurs alike concentrate on numerous aspects of the field in order to do amazing things. Among these various disciplines is computer vision [33].

Convolutional Neural Networks (CNNs) offer a powerful tool for image classification tasks, allowing for the creation of models tailored to specific image categories by organizing images into folders. The post highlights CNN’s wide-ranging impact, citing Facebook’s DeepFace model with a remarkable 97% accuracy in facial recognition, surpassing human abilities [34]. Moreover, CNNs are revolutionizing the medical field, aiding in early cancer detection and diagnosing diseases like typhoid using X-ray images. The

post aims to inspire confidence in readers to explore CNN algorithms in their projects, showcasing the technology's versatility and rapid development. In order to address the issue of human face recognition on a small original dataset, this research [35] presents a novel technique that blends a convolutional neural network (CNN) with the augmented dataset. Several facial image alterations are used to expand the initial small dataset into a large one. By utilizing the clever CNN, it is possible to extract facial features more efficiently and obtain a greater level of face recognition accuracy based on the enhanced face picture dataset. Numerous tests and comparisons with a few popular facial recognition techniques may attest to the superiority and efficiency of the suggested method.

Deep Neural Networks show promise in Arabic OCR, achieving 98.46% accuracy on MNIST [36]. CNNs excel in face recognition, with a model achieving perfect classification on the UJ Face dataset after 80 epochs [37]. Regularization is crucial for deep network performance. Both tasks highlight the effectiveness of deep learning in complex recognition tasks.

C. DATA AUGMENTATION

One key strategy used in this work to artificially increase the quantity and diversity of the training dataset and improve the model's capacity for generalization is data augmentation [38]. Data augmentation creates variations by creating altered versions of the original photos, which help the model identify strong features and patterns linked to various emotional states [39]. More specifically, different instances of the original images are created by applying augmentation techniques such as rotation, flipping, scaling, and shearing [40]. Rotation involves rotating the image by a certain angle while flipping creates mirror images by flipping the original images horizontally or vertically. Scaling modifies the size of the images, and shearing changes the shape of the images by shifting one part of the image more than the other. These variations simulate different viewing angles, lighting conditions, and facial orientations, enabling the model to learn to recognize emotions under various circumstances [41].

The application of data augmentation is especially beneficial in scenarios where the size of the training dataset is limited, as it effectively increases the dataset's size without the need for additional data collection efforts [42]. Moreover, by applying the framework to a wider variety of instances during training, data augmentation helps reduce the danger of overfitting and enhances the model's capacity to generalize to new data [43].

D. UNIFORM LOCAL BINARY PATTERN

Uniform Local Binary Pattern (uLBP) feature extraction is a texture descriptor used for feature extraction from images. LBP compares each pixel in an image with its neighboring pixels and encodes the result into a binary pattern. In this study, uniform LBP is used to capture local

texture patterns in the facial images. This technique helps the model capture fine-grained texture details in the facial images, enhancing its ability to distinguish between different emotions. The extracted LBP features are then concatenated with the pixel intensities and fed into the CNN for emotion classification.

In order to extract features from facial photos, this work heavily relies on the extraction of Uniform Local Binary Pattern (uLBP) features, which serve as a texture descriptor [44]. Effective local texture pattern capturing is achieved by LBP by the comparison of individual pixels in a picture with their neighboring pixels, followed by the encoding of the resultant binary pattern [45]. To capture the fine-grained texture features in facial images which are crucial for differentiating between distinct emotional states, uniform LBP is specially used in the work [46].

By incorporating uniform LBP feature extraction, the model can better understand the nuanced variations in facial textures associated with different emotions, thus enhancing its ability to classify emotions accurately [47]. LBPs are far more potent than Haralick texture features; however, depending on the application (and the kind of LBP method employed), the enhanced discriminative power may be at the expense of computationally prohibitive and possibly exploding feature vector size [48].

Local Binary Pattern, or LBP for short, is a feature of an image's local representation, as its name implies. It is made up of relative values obtained by contrasting every pixel with its surrounding pixels [49]. LBP's two primary features are its low computation cost and resistance to variations in the grayscale values of images. Many advancements have been made since the 1994 first concept. It is commonly used for texture segmentation, facial image recognition, and other image analysis applications, especially in non-deep learning systems. Microstructures that the histogram can estimate, such as edges, lines, spots, and flat areas, are detected by the LBP. [50] discusses For face identification, a new discriminative face representation is suggested, which is obtained by the Linear Discriminant Analysis (LDA) of multi-scale local binary pattern histograms. First, the face image is divided into many non-overlapping areas. Regional features are created by concatenating multi-scale uniform local binary pattern histograms in each region. After that, the features are projected into the LDA space in order to utilize them as a distinguishing facial descriptor. With extremely encouraging results, the system is put into practice and evaluated for face verification on the XM2VTS database and for face identification on the standard Feret database.

Key steps in the process of implementing LBP for texture classification:

- Loading the Image: An image is loaded from the disk using OpenCV.
- Converting to Grayscale: The image is converted to grayscale since LBP operates on single-channel images.

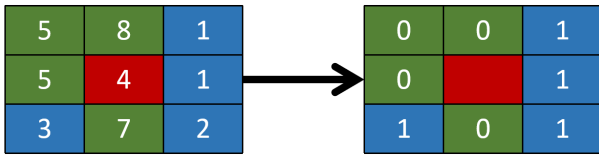


FIGURE 4. The first step in constructing an LBP is to take the 8-pixel neighborhood surrounding a center pixel and threshold it to construct a set of 8 binary digits.

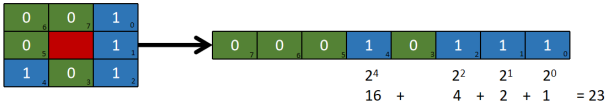


FIGURE 5. Taking the 8-bit binary neighborhood of the center pixel and converting it into a decimal representation.

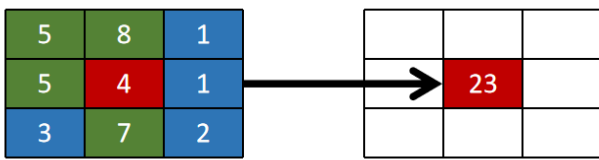


FIGURE 6. The calculated LBP value is then stored in an output array with the same width and height as the original image.

- Computing LBP: The LBP representation of the image is computed using the `feature.local_binary_pattern` function from the `scikit-image` library.
- Visualizing the LBP Image: The resulting LBP image is displayed to show the texture features.
- Histogram Calculation: A histogram of the LBP image is computed to serve as a texture descriptor. This histogram can then be used for classification purposes [51].

LBP Algorithm Local Binary Patterns is a simple and efficient texture descriptor. Here’s how it works:

- Pixel Comparison: For each pixel in the grayscale image, compare its intensity with its surrounding pixels. Typically, the surrounding pixels form a 3×3 grid centered on the pixel being processed.
- Binary Pattern: Generate a binary pattern by setting the center pixel as the threshold. If a surrounding pixel’s intensity is greater than or equal to the center pixel’s intensity, set the corresponding binary digit to 1; otherwise, set it to 0. As shown in **Figure 4**.
- Binary to Decimal Conversion: Convert the 8-bit binary pattern (formed by the 8 surrounding pixels) to a decimal value. As shown in **Figure 5**.
- LBP Image Creation: Replace the center pixel’s value with the corresponding decimal value, creating a new image where each pixel value represents a local binary pattern. As shown in **Figure 6**.
- Histogram Calculation: Compute a histogram of the LBP values across the image. Histogram captures the distribution of texture patterns and serves as a robust descriptor for texture classification.

Uniform Local Binary Patterns (uLBP) are outstanding for a variety of computer vision tasks because they provide a

Layer (type)	Output Shape	Parameters
Conv2D (32 filters)	(None, 46, 46, 32)	320
Conv2D (64 filters)	(None, 44, 44, 64)	18496
MaxPooling2D	(None, 22, 22, 64)	0
Dropout	(None, 22, 22, 64)	0
Conv2D (128 filters)	(None, 20, 20, 128)	73856
MaxPooling2D	(None, 10, 10, 128)	0
Conv2D (128 filters)	(None, 8, 8, 128)	147584
MaxPooling2D	(None, 4, 4, 128)	0
Flatten	(None, 2048)	0
Dense (1024 units)	(None, 1024)	2098176
Dropout	(None, 1024)	0
Dense (7 units)	(None, 7)	7175

FIGURE 7. Detailed description of CNN model.

strong and effective method for texture analysis and feature extraction. In order to encode the results into a binary pattern that represents local texture information, uLBP compares each pixel with its neighbors. This technique produces stable and discriminative representations by drastically reducing the dimensionality of the input while maintaining key features. The uniformity feature of uLBP ensures that only patterns with a small number of transitions—0 to 1 or 1 to 0—are taken into account, improving noise resistance and discriminative power. As shown in [52], uLBP captures small textural differences to deliver higher performance in facial recognition tests. Furthermore, [53] emphasizes the method’s capacity to identify drowsiness, where uLBP’s resilience to changes in illumination and occlusions is crucial. Moreover, [54] highlights uLBP’s adaptability in various applications and effectiveness in real-time processing scenarios because of its minimal yet efficient computational requirements.

III. PROPOSED METHODOLOGY

Following is a detailed explanation of the procedure followed in each method.

A. METHOD 1: CONVOLUTIONAL NEURAL NETWORKS

The architecture of the CNN model in this project is designed with multiple layers to extract features from input images and classify them into seven emotion categories. The initial layers consist of two convolutional layers, the first with 32 filters and the second with 64 filters, followed by a max pooling layer to reduce spatial dimensions. To prevent overfitting, a dropout layer is added, which randomly sets a fraction of input units to zero during training. The model then includes two additional convolutional layers, each with 128 filters, followed by another max pooling layer to further reduce spatial dimensions. After flattening the output, a dense layer with 1024 units is used for feature extraction, followed by another dropout layer for regularization. Finally, a dense layer with 7 units and a softmax activation function is used for multi-class classification into the seven emotion categories. The model has a total of 2,268,607 trainable parameters, which are adjusted during training to minimize the categorical cross-entropy loss function. The Adam optimizer is used for optimization during training. Refer **Figure 7** for a Detailed tabular description of the CNN Model used in this method.

B. METHOD 2: WITH DATA AUGMENTATION

Pseudocode:

Input: Facial Emotion Recognition Dataset.
 Output: Student behavior detection model.
Begin
 for img in the dataset:
 Data_augmentation(img)
 Remove duplicated(augmented_imgs)
 Rescale(augmented_imgs)
 Dataset.append(augmented_imgs)
 CNN_Model2(dataset) //Architecture of CNN_Model2 is same as CNN_Model1
 Train and compile (CNN_Model2)
 Test(CNN_Model2)
End

Data augmentation was implemented on each emotion class within the FER-2013 dataset using the Image Data Generator class from the Keras library. The augmentation parameters were meticulously selected to introduce realistic variations in the images, enhancing the model’s robustness and mitigating overfitting. Here’s a concise overview of the augmentation methods applied to each emotion: 1. **Angry**: 3500 images were augmented for the “angry” class. Augmentation included random rotation: 40 degrees, horizontal and vertical shifts: 20% of the image dimensions, shear transformations: 20 degrees, zooming: 20%, and horizontal flipping. These transformations aimed to simulate diverse facial expressions of anger. 2. **Disgust**: 6500 images were augmented for the “disgust” class. Similar to the “angry” class, augmentation included random rotations: 30 degrees, horizontal and vertical shifts: 10% of the image dimensions, shear transformations: 10 degrees, zooming: 10%, and horizontal flipping. These variations aimed to capture various facial expressions associated with disgust. 3. **Fear**: 3500 images were augmented for the “fear” class. Augmentation included random rotations: 40 degrees, horizontal and vertical shifts: 20% of the image dimensions, shear transformations: 20 degrees, zooming: 20%, and horizontal flipping. These transformations aimed to simulate diverse fearful expressions. 4. **Neutral**: 3500 images were augmented for the “neutral” class. Augmentation included random rotations up to 40 degrees, horizontal and vertical shifts: 20% of the image dimensions, shear transformations: 20 degrees, zooming: 20%, and horizontal flipping. These variations aimed to capture different neutral facial expressions. 5. **Sad**: 3500 images were augmented for the “sad” class. Augmentation included random rotations: 40 degrees, horizontal and vertical shifts: 20% of the image dimensions, shear transformations: 20 degrees, zooming: 20%, and horizontal flipping. These transformations aimed to simulate various sad facial expressions. 6. **Surprise**: 3500 images were augmented for the “surprise” class. Augmentation included random rotations: 40 degrees, horizontal and vertical shifts: 20% of the image dimensions, shear transformations: 20 degrees, zooming: 20%, and horizontal

TABLE 1. Dataset before and after data augmentation.

Emotions	Before Data Augmentation	After Data Augmentation
Happy	7215	7215
Sad	4830	9668
Neutral	4965	9888
Disgust	436	5196
Fear	4097	8560
Surprise	3171	7071
Angry	3995	8358
Total	28709	55956

TABLE 2. Dataset before and after Splitfolders.

Emotions	Before Split-folders	After Splitfolders TRAIN	After Splitfolders TEST	After Splitfolders VALIDATION
Angry	8358	5014	1673	1671
Disgust	5196	3117	1040	1039
Fear	8560	5136	1712	1712
Happy	7215	4329	1443	1443
Neutral	9888	5932	1979	1977
Sad	9668	5800	1935	1933
Surprise	7071	4242	1415	1414

flipping. These transformations aimed to capture different surprised facial expressions. Overall, data augmentation played a critical role in enhancing the dataset’s diversity, potentially leading to a more robust and accurate model for facial expression detection.

Table 1 shows the result of Data Augmentation on the dataset. The overall result of Data Augmentation shows the enlargement of the Dataset from 28709 images to 55956 images.

CNN Model used in Method 1 is also used in Method 2 (Data Augmentation being the only variation). Refer **Figure 7** for a detailed tabular description of the CNN Model used in this method.

C. METHOD 3: DATA AUGMENTATION WITH A COMBINATION OF LBP FEATURE EXTRACTION AND CNN

After Data Augmentation, the dataset increased and this caused overfitting. So, to avoid/ overcome this problem, split folders are used to further split the training dataset into train, test, and validation datasets for better results and to reduce the computation runtime. The results of Splitfolders are shown in Table 2.

Pseudocode:

Input: Facial Emotion Recognition Dataset.
 Output: Student behavior detection model.
Begin
 for img in the dataset:
 Data_Augmentation(img)
 Rescale(img)
 LBP_Feature_Extraction(dataset):
 feature_extraction_image = “LBP”

Covert(image,greyscale)

For a pixel in img:

Extract_LBP(pixel)

CNN_Model3(dataset)

Train and compile (CNN_Model3)

Test(CNN_Model3)

End

Here N is a number of LBP Feature vectors. i.e., 26.

- Input layers are created for images of size (48, 48, 1) and LBP features of length N.
- Two convolutional layers with 32 and 64 filters, respectively, are followed by max pooling and dropout layers.
- Add another convolutional layer with 128 filters, followed by max pooling and dropout layers.
- Flatten the output and pass it through a dense layer with 1024 units for image features.
- Process LBP features through a dense layer with 1024 units before concatenating them with the image features.
- Concatenated features go through additional dense layers with 512 and 256 units before the output layer with 7 units for softmax classification.
- Model developed for CNN with uniform LBP feature extraction can be briefly understood by referring to **Figure 8**.

IV. RESULTS AND DISCUSSION

Models built are evaluated and assessed using standard performance metrics. Here is some basic information about each of these metrics with relevant computational equations.

1) ACCURACY

The proportion of accurately predicted occurrences to all instances in the dataset is known as accuracy. It gauges how accurate the model is overall.

$$Accuracy = \frac{a + b}{a + b + c + d}$$

where a: True Positive

b: True Negative

c: False Positive

d: False Negative

2) PRECISION

The ratio of accurately anticipated positive instances to all expected positive instances is known as precision. It gauges how accurate positive forecasts are.

$$Precision = \frac{a}{a + c}$$

where a: True Positive

c: False Positive

3) RECALL

Recall is the ratio of correctly predicted positive instances to the actual positive instances in the dataset. It measures the

Layer Type	Output Shape	Parameters
Input	(48, 48, 1)	-
Conv2D	(46, 46, 32)	320
Conv2D	(44, 44, 64)	18496
MaxPooling2D	(22, 22, 64)	-
Dropout	(22, 22, 64)	-
Conv2D	(20, 20, 128)	73856
MaxPooling2D	(10, 10, 128)	-
Conv2D	(8, 8, 128)	147584
MaxPooling2D	(4, 4, 128)	-
Dropout	(4, 4, 128)	-
Flatten	2048	-
Dense (Image)	1024	2098176
Input (LBP)	N	-
Dense (LBP)	1024	N * 1024
Concatenation	2048 + N	-
Dense	512	(2048 + N) * 512
Dropout	512	-
Dense	256	512 * 256
Dropout	256	-
Output (Softmax)	7	256 * 7

FIGURE 8. Detailed description of CNN model with uniform local binary pattern.

ability of the model to identify all relevant instances. Recall is useful when the cost of false negatives is high. For example, in medical diagnosis, we want to minimize false negatives to ensure that all patients with a disease are correctly identified.

$$Recall = \frac{a}{a + d}$$

where a: True Positive

d: False Negative

4) F1-SCORE

The harmonic mean of recall and precision is known as the F1-score. It offers a harmony between recall and precision.

$$F1 - Score = \frac{2 * (p * r)}{p + r}$$

where p: Precision

r: Recall

5) CONFUSION MATRIX

A table known as a confusion matrix is widely employed to illustrate how well a classification model performs when applied to a set of data for which the true values are known. It makes it possible to illustrate how effectively an algorithm performs. The confusion matrix provides a summary of the predictions made by a classification model. It is particularly useful for understanding the types of errors the model is making, such as false positives and false negatives. It can potentially help in extracting performance metrics like *True Positives* (*Tp*), *True Negatives* (*Tn*), *False Positives* (*Fp*), *False Negatives* (*Fn*), *Accuracy*, *Precision*, *Recall (Sensitivity)*, *F1-Score*, *Misclassification Rate (Error rate)*. These points help in understanding how well a model is performing.

A. ANALYSIS OF RESULTS:

1) METHOD 1: CNN

The confusion matrix of the CNN Model built-in method 1 is shown in **Figure 9**. and the Corresponding Classification

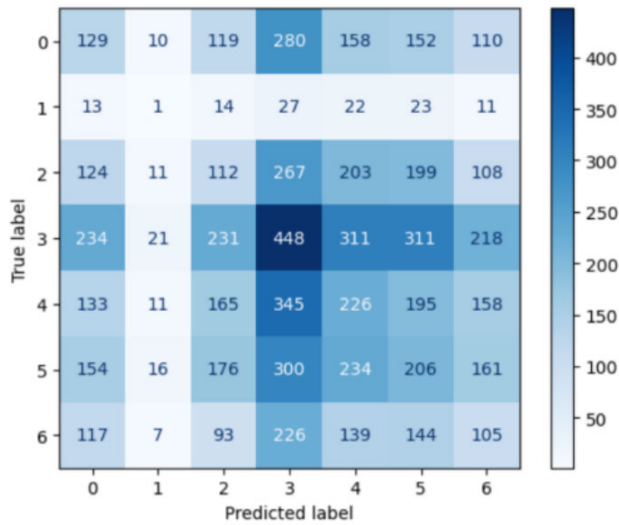


FIGURE 9. Method1:CNN confusion matrix.

TABLE 3. Method 1 classification report.

Emotions	Precision	Recall	f1-score	support
0	0.14	0.13	0.14	958
1	0.01	0.01	0.01	111
2	0.12	0.11	0.12	1024
3	0.24	0.25	0.24	1774
4	0.17	0.18	0.18	1233
5	0.17	0.17	0.17	1247
6	0.12	0.13	0.12	831
accuracy			0.17	7178
macro avg	0.14	0.14	0.14	7178
weighted avg	0.17	0.17	0.17	7178

Report is shown in Table 3. With an overall accuracy of only 17%, the model correctly identified only 17% of the 7178 photos. This emphasizes how urgently significant improvements are needed in order to discriminate between different emotions. Performance differed significantly between the emotion classes, with class 3 (happy) showing better results than class 1 (disgust), where all metrics were at 1% (precision: 24%, recall: 25%, F1-score: 24%). This shows that the model has trouble telling one emotion from another, such as disgust. The model showed low precision and recall across most emotion classes, suggesting that it missed many real instances of specific emotions (low recall) in addition to misclassifying a large number of images (low precision). Despite the support values indicating a balanced dataset with a similar number of images per class, the consistently poor performance across most classes suggests potential issues with the model’s architecture or training process. In conclusion, the basic CNN model performed inadequately in this student emotion recognition task, exhibiting low overall accuracy and struggling to precisely identify most emotion classes.

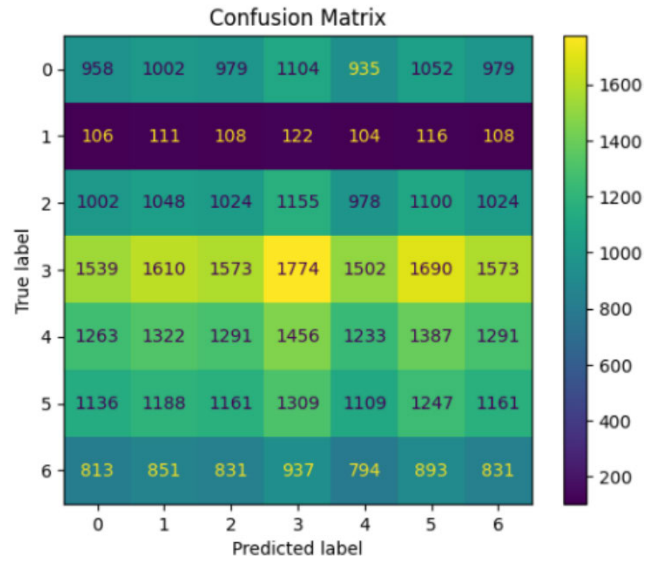


FIGURE 10. Method2 confusion matrix.

TABLE 4. Method 2 classification report.

Emotions	Precision	Recall	f1-score	support
0	0.83	0.90	0.86	958
1	0.85	0.84	0.85	111
2	0.87	0.84	0.85	1024
3	0.80	0.81	0.81	1774
4	0.89	0.86	0.87	1233
5	0.84	0.81	0.82	1247
6	0.86	0.85	0.85	831
accuracy			0.85	7178
macro avg	0.85	0.85	0.85	7178
weighted avg	0.85	0.85	0.85	7178

2) METHOD 2: CNN WITH DATA AUGMENTATION

The confusion matrix of the CNN Model with Data Augmentation built in method 2 is shown in Figure 10, and the Corresponding Classification Report is shown in Table 4.

Data augmentation improved the model’s accuracy greatly, raising it from an unknown low value to 85%. This indicates that 85% of the 7,178 photos in the test set are now accurately classified by the model. According to the report, performance was largely matched across all emotion classes (0-angry to 6-surprise), with 85% for precision, recall, and F1-score. This implies that the model can now accurately distinguish between various emotions. Performance does, however, differ slightly throughout classes. As an example, class 0 (angry) has an 83% precision, 90% recall, and an 86% F1-score; class 3 (happy) has an 89% accuracy, 86% recall, and an 87% F1-score. This implies that the model may be marginally more adept at identifying joyful emotions than furious ones. Finally, data augmentation successfully addressed the shortcomings of the prior model, yielding a high degree of overall accuracy and well-balanced performance in the majority of emotion classes.

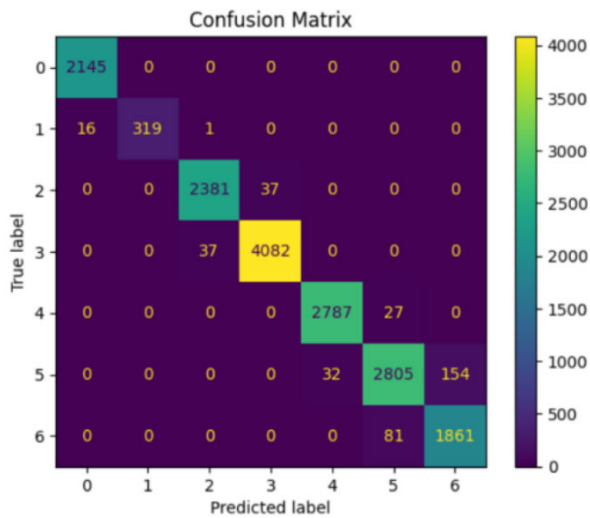


FIGURE 11. Method3 confusion matrix.

TABLE 5. Method 3 classification report.

Emotions	Precision	Recall	f1-score	support
0	0.95	0.95	0.95	2259
1	0.95	0.95	0.95	336
2	0.94	0.96	0.95	2480
3	0.94	0.95	0.95	4292
4	0.95	0.94	0.95	2969
5	0.96	0.94	0.95	2991
6	0.94	0.95	0.95	1970
accuracy			0.95	11197
macro avg	0.95	0.95	0.95	11197
weighted avg	0.95	0.95	0.95	11197

3) METHOD 3: CNN WITH DATA AUGMENTATION AND UNIFORM LBP FEATURE EXTRACTION

The confusion matrix of CNN Model built-in method 1 is shown in Figure 11. and the Corresponding Classification Report is shown in Table 5. With an overall accuracy of 95%, the model performs exceptionally well, correctly categorizing 95% of the 11,197 photos in the test set. This is a big step up from earlier techniques that only used a rudimentary CNN model or data augmentation. All emotion classes (0-angry to 6-surprise) had consistently high precision, recall, and F1-scores (around 95%), suggesting balanced accuracy across emotions and avoiding bias towards any particular class. Class 2 (fear) had a precision of 94%, recall of 96%, and an F1-score of 95%. Class 5 (sad) had a precision of 96%, recall of 94%, and an F1-score of 95%. These slight differences in performance between the classes are insignificant. All in all, the model’s consistency in identifying every emotion experienced by students is impressive. In summary, the combination of data augmentation with uniform LBP feature extraction is a powerful method for accurately identifying student emotions through the successful extraction of pertinent features from students’ facial expressions.

The overall result or comparison between all three models used in this project procedure is projected in Tables 6 and 7

TABLE 6. Summary of results.

Model	Training Accuracy	Testing Accuracy
Simple CNN Model	0.8989	0.6258
CNN with Data Augmentation	0.9224	0.8587
CNN with Data Augmentation and LBP Feature Extraction	0.9655	0.9482

TABLE 7. Results comparison with other state-of-art methods.

Methods	Accuracy
Ensemble for Deep Learning [11]	0.7580
CNN for Real-Time FER on Raspberry Pi [12]	0.6597
Cross-Database approach with CNN-based FER System [13]	0.7305
DenseNet121 [51]	0.59
ResNet-50 [51]	0.60
Proposed Method (CNN with Data Augmentation and LBP Feature Extraction)	0.9482

show the results compared with other state-of-the-art methods. When we go from a basic CNN model to more sophisticated methods, the accuracy of emotion recognition clearly improves, as Table 3 demonstrates. At 17%, the accuracy of the most basic CNN model is the lowest. This is greatly enhanced with data augmentation, as the model achieves an accuracy of 85%. The most sophisticated approach obtains the highest accuracy of 95% by combining LBP feature extraction with data augmentation. This implies that utilizing LBP in conjunction with data augmentation to extract local pattern information from facial photos is a potent method for teaching students to recognize emotions.

Three distinct models created for the project on student behavior detection using computer vision are compared in the findings summary table. A CNN with a Data Augmentation Model, a CNN with Data Augmentation and LBP Feature Extraction Model, and a Simple CNN Model for Emotion Classification are among the models. The training and testing accuracy attained by each model is displayed in the table.

Training accuracy of 0.8989 and testing accuracy of 0.6258 were attained by the Simple CNN Model. Given that the test accuracy is much lower than the training accuracy, this model most likely suffered from over fitting. This suggests that the model might not have generalized well to previously encountered data, instead memorizing the training set. Next, data augmentation was used to solve this problem.

Training accuracy was increased to 0.9224 and testing accuracy to 0.8587 using the CNN with Data Augmentation Model. In order to increase the model’s capacity to generalize, data augmentation is a popular approach used to artificially increase the amount of a dataset by producing altered versions of photographs. Testing accuracy significantly improved, indicating that data augmentation reduced over fitting and enhanced model performance.

Ultimately, the CNN with Data Augmentation and LBP Feature Extraction Model had the best testing and training accuracy, coming in at 0.9482 and 0.9655, respectively. The

performance of the model was significantly improved by the use of LBP feature extraction. Texture patterns are vital for tasks like emotion recognition, and LBP is a texture descriptor that captures the local structure of an image. The excellent testing accuracy shows that the LBP feature-based model caught key elements for emotion classification and adapted well to new data.

Ensemble Deep Learning models for FER2013 and achieved unprecedented 75.8% accuracy, practical applicability demonstrated through a mobile web app with rapid recognition speed. Interpretability methods shed light on facial features. Successfully surpassed human-level accuracy. In [12], the authors used the FER-2013 dataset to develop a prediction system for micro-expressions using a Convolutional Neural Network (CNN) algorithm implemented on a Raspberry Pi. Real-time prediction on Raspberry Pi achieved 65.97% accuracy [11]. This model considers facial position, distance, and photo recognition well. However, this paper lags in optimizing the model and dataset. The paper [13] uses a cross-dataset approach to enhance generality and performance. Their system achieved accuracies ranging from 31.56% to 61.78% on single-dataset tests, which improved up to 73.05% using the cross-dataset approach. Additionally, they compared CNN's performance with human participants, who achieved 83.53% accuracy, highlighting a correlation between human and system results. The paper [51] utilizes transfer learning with pre-trained models (ResNet-50, Xception, EfficientNet-B0, Inception, DenseNet-121) to prevent overfitting by leveraging large-scale datasets and fine-tuning them for emotion recognition. Uses a combination of freezing layers of pre-trained models and fine-tuning them, which is effective for most models. Reports significant accuracy improvements with transfer learning models. All models averaged 59% accuracy when trained using AffectNet, Xception was the most accurate (61%), followed sequentially by Inception (60%), EfficientNet-B0 (59%), ResNet-50 (58%), and DenseNet-121 (57%). The key difference here lies in the reliance on transfer learning and pre-trained models versus the development of a custom model with specific feature extraction and augmentation techniques.

V. CONCLUSION AND FUTURE WORK

Our research presents an innovative approach to leveraging computer vision for detecting student behavior, particularly focusing on emotion recognition. By employing Convolutional Neural Networks (CNNs) for emotion classification, implementing data augmentation to diversify the dataset, and integrating Uniform Local Binary Pattern (LBP) feature extraction for enhanced texture analysis, we have developed a system capable of automatically identifying student emotions in real-time. This system provides educators with valuable insights into student engagement, allowing them to adapt their teaching methods accordingly. Our findings indicate that our approach achieves high accuracy in classifying student emotions. While the Simple CNN Model initially showed effectiveness, it displayed signs of overfitting, which

we mitigated through data augmentation. The CNN with Data Augmentation Model notably improved performance, highlighting the efficacy of data augmentation in reducing overfitting and improving model generalization. Additionally, the CNN with Data Augmentation and LBP Feature Extraction Model further enhanced performance, underscoring the importance of texture analysis in emotion recognition tasks. Our study contributes to educational technology by offering a reliable, automated method for monitoring student behaviors and emotions, potentially fostering more dynamic and engaging learning environments. By providing real-time feedback on student emotions, our system enables educators to personalize their teaching strategies to better meet students' needs. Overall, our research demonstrates the potential of computer vision techniques in enhancing the teaching and learning process, suggesting avenues for future exploration to enhance emotion recognition accuracy and broaden the capabilities of educational technology in supporting student engagement and learning.

The approach in this research paper differs from other papers that utilized Convolutional Neural Networks (CNN) and Local Binary Pattern (LBP) in the following ways: An innovative fusion of techniques: The research paper introduces a novel method that combines Convolutional Neural Networks (CNNs) for emotion classification, data augmentation to enhance dataset diversity and size, and Uniform Local Binary Patterns (uLBP) feature extraction for improved texture analysis. This fusion of techniques represents a fresh and unique approach to FER, leveraging the strengths of each method to achieve high accuracy in emotion classification. Incorporation of Data Augmentation: The research addresses the challenge of model overfitting by implementing data augmentation to diversify the dataset, improving the model's ability to generalize to novel data. This effectively mitigates overfitting, which is a common issue in CNN-based models, setting this approach apart from others that may not have explicitly addressed this challenge. Integration of Uniform LBP: The paper specifically incorporates Uniform Local Binary Pattern (uLBP) feature extraction for enhanced texture analysis. This represents a departure from traditional LBP feature extraction, showcasing a more refined and nuanced approach to texture pattern capture. Emphasis on real-time insights into student emotions: The research paper places a significant focus on providing educators with real-time insights into student emotions, enabling them to adapt their teaching methods accordingly. This practical emphasis distinguishes it from other papers that may focus solely on technical aspects without a direct application to educational settings.

REFERENCES

- [1] T. D. Nguyen, M. Cannata, and J. Miller, "Understanding student behavioral engagement: Importance of student interaction with peers and teachers," *J. Educ. Res.*, 2016, doi: [10.1080/00220671.2016.1220359](https://doi.org/10.1080/00220671.2016.1220359).
- [2] M. C. Heitink, F. M. Van der Kleij, B. P. Veldkamp, K. Schildkamp, and W. B. Kippers, "A systematic review of prerequisites for implementing assessment for learning in classroom practice," *Educ. Res. Rev.*, vol. 17, pp. 50–62, Feb. 2016.

- [3] M. Chimienti, T. Cornulier, E. Owen, M. Bolton, I. M. Davies, J. M. J. Travis, and B. E. Scott, "The use of an unsupervised learning approach for characterizing latent behaviors in accelerometer data," *Ecol. Evol.*, vol. 6, no. 3, pp. 727–741, Jan. 2016.
- [4] [Online]. Available: <https://www.interventioncentral.org>
- [5] S. Mulay and S. Poddar, "Data mining techniques to predict student academic performance in higher education: Literature review," *J. Adv. Zool.*, vol. 45, no. S4, pp. 167–173, 2024.
- [6] N. S. R. Najjar and K. Al-Jammali, "Prediction and evaluation of students' performance in E-learning using data mining algorithm," *Int. J. Adv. Multidisc. Res. Stud.*, vol. 3, no. 5, pp. 488–491, 2023.
- [7] A. Kubegenova, Z. Abuova, A. Gumarova, and G. Kamalova, "Using the data mining tool to analyze student performance," in *Proc. Int. Conf. Internet Mod. Soc. Cham, Switzerland: Springer, Jun. 2022*, pp. 39–51.
- [8] N. A. B. M. Zahruddin, N. D. Kamarudin, R. M. Jusoh, N. A. A. Fataf, and R. Hidayat, "Case study: Using data mining to predict student performance based on demographic attributes," *JOIV Int. J. Informat. Visualizat.*, vol. 7, no. 4, pp. 2460–2468, Dec. 2023.
- [9] V. Harish, R. Sharma, G. Rana, and A. Nayyar, "Artificial intelligence in sustainable education: Benefits, applications, framework, and potential barriers," in *The Role of Sustainability and Artificial Intelligence in Education Improvement*. Boca Raton, FL, USA: CRC Press, 2023, pp. 219–236.
- [10] Z. S. Manukyan, "Opportunities for using artificial intelligence in the higher education system (in the field of international relations)," *J. Digit. Economy Res.*, vol. 1, no. 4, pp. 85–101, Feb. 2024.
- [11] A. Khanzada, C. Bai, and F. T. Celepcikay, "Facial expression recognition with deep learning," 2020, *arXiv:2004.11823*.
- [12] L. Zahara, P. Musa, E. P. Wibowo, I. Karim, and S. B. Musa, "The facial emotion recognition (FER-2013) dataset for prediction system of micro-expressions face using the convolutional neural network (CNN) algorithm based raspberry pi," in *Proc. 5th Int. Conf. Informat. Comput. (ICIC)*, Nov. 2020, pp. 1–9.
- [13] S. Ramis, J. M. Buades, F. J. Perales, and C. Manresa-Yee, "A novel approach to cross dataset studies in facial expression recognition," *Multimedia Tools Appl.*, vol. 81, no. 27, pp. 39507–39544, Nov. 2022.
- [14] J. Li, Y. Mi, J. Yu, and Z. Ju, "A novel convolutional neural network for facial expression recognition," in *Proc. 4th Int. Conf. Cogn. Syst. Signal Process.*, Beijing, China. Singapore: Springer, Nov. 2019, pp. 310–320.
- [15] X. Huang, S.-J. Wang, X. Liu, G. Zhao, X. Feng, and M. Pietikainen, "Discriminative spatiotemporal local binary pattern with revisited integral projection for spontaneous facial micro-expression recognition," *IEEE Trans. Affect. Comput.*, vol. 10, no. 1, pp. 32–47, Jan. 2019.
- [16] W. Mellouk and W. Handouzi, "Facial emotion recognition using deep learning: Review and insights," *Proc. Comput. Sci.*, vol. 175, pp. 689–694, Jan. 2020.
- [17] N. Perveen, D. Roy, and K. M. Chalavadi, "Facial expression recognition in videos using dynamic kernels," *IEEE Trans. Image Process.*, vol. 29, pp. 8316–8325, 2020.
- [18] Sujata and S. K. Mitra, "Modular FER: A modular facial expression recognition from image sequence based on two dimensional (2D) Taylor expansion," *Social Netw. Comput. Sci.*, vol. 2, no. 3, pp. 1–15, May 2021.
- [19] J. Shi, S. Zhu, and Z. Liang, "Learning to amend facial expression representation via de-albino and affinity," 2021, *arXiv:2103.10189*.
- [20] Z. Zhao, Q. Liu, and S. Wang, "Learning deep global multi-scale and local attention features for facial expression recognition in the wild," *IEEE Trans. Image Process.*, vol. 30, pp. 6544–6556, 2021.
- [21] S. Minaee, M. Minaei, and A. Abdolrashidi, "Deep-emotion: Facial expression recognition using attentional convolutional network," *Sensors*, vol. 21, no. 9, p. 3046, Apr. 2021.
- [22] E. Danniels, A. Pyle, and C. DeLuca, "The role of technology in supporting classroom assessment in play-based kindergarten," *Teaching Teacher Educ.*, vol. 88, Feb. 2020, Art. no. 102966.
- [23] K. Thorpe, P. Rankin, T. Beaton, S. Houen, M. Sandi, I. Siraj, and S. Staton, "The when and what of measuring ECE quality: Analysis of variation in the classroom assessment scoring system (CLASS) across the ECE day," *Early Childhood Res. Quart.*, vol. 53, pp. 274–286, 4th Quart., 2020.
- [24] J. Li, X. Liu, W. Zhang, M. Zhang, J. Song, and N. Sebe, "Spatio-temporal attention networks for action recognition and detection," *IEEE Trans. Multimedia*, vol. 22, no. 11, pp. 2990–3001, Nov. 2020.
- [25] B. Chen, W. Guan, P. Li, N. Ikeda, K. Hirasawa, and H. Lu, "Residual multi-task learning for facial landmark localization and expression recognition," *Pattern Recognit.*, vol. 115, Jul. 2021, Art. no. 107893.
- [26] J. Lee, S. Kim, S. Kim, and K. Sohn, "Multi-modal recurrent attention networks for facial expression recognition," *IEEE Trans. Image Process.*, vol. 29, pp. 6977–6991, 2020.
- [27] Kaggle. (2013). *FER2013: Facial Expression Recognition Challenge*. [Online]. Available: <https://www.kaggle.com/datasets/msambare/fer2013?resource=download>
- [28] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.
- [29] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 25, Jan. 2012, pp. 4–12.
- [30] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.
- [31] J. Wu, "Introduction to convolutional neural networks," Nat. Key Lab Novel Softw. Technol., Nanjing Univ., Nanjing, China, vol. 5, no. 23, p. 495, 2017.
- [32] D. Zhu, Y. Fu, X. Zhao, X. Wang, and H. Yi, "Facial emotion recognition using a novel fusion of convolutional neural network and local binary pattern in crime investigation," *Comput. Intell. Neurosci.*, vol. 2022, no. 1, 2022, Art. no. 2249417.
- [33] P. G. O. Lobaton, J. E. C. Magtiay, V. G. E. Mendez, G. J. G. Moster, M. M. Casabuena, M. M. Gaspar, and P. V. Mojares, "LEAFLET: A web-based leaf classification system using convolutional neural networks," *J. Elect. Syst.*, vol. 20, no. 1, 2024.
- [34] R. T. B. M. D. L. Duvva, S. Bethu, and Y. Garapati, "Deep learning feature extraction architectures for real-time face detection," *SN Comput. Sci.*, vol. 4, no. 5, 2023, Art. no. 645.
- [35] P. Lu, B. Song, and L. Xu, "Human face recognition based on convolutional neural network and augmented dataset," *Syst. Sci. Control Eng.*, vol. 9, no. sup2, pp. 29–37, May 2021.
- [36] K. S. Younis and A. A. Alkhateeb, "A new implementation of deep neural networks for optical character recognition and face recognition," in *Proc. New Trends Inf. Technol.*, 2017, pp. 157–162.
- [37] [Online]. Available: <https://vitalflux.com/cnn-basic-architecture-for-classification-segmentation/>
- [38] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *J. Big Data*, vol. 6, no. 1, pp. 1–48, Dec. 2019.
- [39] L. Perez and J. Wang, "The effectiveness of data augmentation in image classification using deep learning," 2017, *arXiv:1712.04621*.
- [40] M. Kang, M. Kang, and S. Kim, "Catch-up Mix: Catch-up class for struggling filters in CNN," in *Proc. AAAI Conf. Artif. Intell.*, vol. 38, no. 3, Mar. 2024, pp. 2705–2713.
- [41] E. Sadeghnezhad and S. Salem, "InceptionCapsule: Inception-ResNet and CapsuleNet with self-attention for medical image classification," 2024, *arXiv:2402.02274*.
- [42] S. H. Kim, J. S. Park, H. S. Lee, S. H. Yoo, and K. J. Oh, "Combining CNN and Grad-CAM for profitability and explainability of investment strategy: Application to the KOSPI 200 futures," *Expert Syst. Appl.*, vol. 225, Sep. 2023, Art. no. 120086.
- [43] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, 2014.
- [44] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971–987, Jul. 2002.
- [45] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
- [46] G. Zhao and M. Pietikainen, "Dynamic texture recognition using local binary patterns with an application to facial expressions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 6, pp. 915–928, Jun. 2007.
- [47] T. Ahonen, A. Hadid, and M. Pietikainen, "Face recognition with local binary patterns," in *Proc. 8th Eur. Conf. Comput. Vis.*, Prague, Czech Republic. Berlin, Germany: Springer, May 2004, pp. 469–481.
- [48] CV Explained. (2020). *Local Binary Patterns (LBP)*. [Online]. Available: <https://cvexplained.wordpress.com/2020/07/22/10-7-local-binary-patterns/>
- [49] [Online]. Available: <https://aihalapathirana.medium.com/understanding-the-local-binary-pattern-lbp-a-powerful-method-for-texture-analysis-in-computer-4fb55b3ed8b8>

- [50] C. H. Chan, J. Kittler, and K. Messer, "Multi-scale local binary pattern histograms for face recognition," in *Proc. Int. Conf. Adv. Biometrics*, Seoul, South Korea. Berlin, Germany: Springer, Aug. 2007, pp. 809–818.
- [51] C.-T. Yen and K.-H. Li, "Discussions of different deep transfer learning models for emotion recognitions," *IEEE Access*, vol. 10, pp. 102860–102875, 2022.
- [52] A. Rosebrock. (2015). *Local Binary Patterns With Python & OpenCV*. PyImageSearch. [Online]. Available: <https://pyimagesearch.com/2015/12/07/local-binary-patterns-with-python-opencv/>
- [53] N. A. M. Hayat, Z. M. Noh, N. M. Yatim, and S. A. Radzi, "Analysis of local binary pattern using uniform bins as palm vein pattern descriptor," *J. Phys., Conf. Ser.*, vol. 1502, no. 1, Mar. 2020, Art. no. 012043.
- [54] D. A. Navastara, W. Y. M. Putra, and C. Fatichah, "Drowsiness detection is based on facial landmarks and uniform local binary patterns," *J. Phys., Conf. Ser.*, vol. 1529, no. 5, 2020, Art. no. 052015.
- [55] F. Abbas, A. Gattal, and R. Menassel, "Local binary pattern and its derivatives to handwriting-based gender classification," *Bull. Electr. Eng. Informat.*, vol. 12, no. 6, pp. 3571–3583, Dec. 2023.



UPPANAPALLI LAKSHMI SOWJANYA is currently pursuing the degree in computer science and engineering specializing in artificial intelligence and machine learning with Vellore Institute of Technology, Chennai, Tamil Nadu, India. Her ongoing projects, such as network intrusion detection systems and driver drowsiness are mainly based on deep learning methodologies and their applications. She has applied many innovative approaches to tackling the problems encountered during the course of the projects. Her research interests include artificial intelligence and machine learning and their applications.



R. KRITHIGA received the Ph.D. degree in machine language, medical image processing, and natural language processing from the College of Engineering, Guindy, Anna University, Chennai. She is currently an Assistant Professor (Sr.G.) with the School of Computer Science Engineering, Vellore Institute of Technology, Chennai Campus. She has participated in and presented papers at many International conferences. Her research achievements and activities include publishing articles in refereed journals.

...