Proceedings of the International Conference on Sustainable Communication Networks and Application (ICSCNA 2023)
IEEE Xplore Part Number: CFP23DW8-ART; ISBN: 979-8-3503-1398-7

Semblance Unmasker: Hidden Emotion Recognition using Deep Learning Approach

Archana P. Lopes
Department of Eletcronics and Computer Science
Fr. Conceicao Rodrigues College of Engineering
Bandra, Mumbai-400050, India
archana_lopes@frcrce.ac.in

Sahil Bane
Department of Eletcronics and Computer Science
Fr. Conceicao Rodrigues College of Engineering
Bandra, Mumbai-400050, India
sahilbane02@gmail.com

Mohini Gautam
Department of Eletcronics and Computer Science
Fr. Conceicao Rodrigues College of Engineering
Bandra, Mumbai-400050, India
crce.9124.ecs@gmail.com

Shalaka Vengurlekar
Department of Eletcronics and Computer Science
Fr. Conceicao Rodrigues College of Engineering
Bandra, Mumbai-400050, India
shalaka.vengurlekar@gmail.com

2023 International Conference on Sustainable Communication Networks and Application (ICSCNA) | 979-8-3503-1398-7/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICSCNA58489.2023.10370099

***Abstract*— Emotion-aware systems have emerged as critical tools in the field of e-learning, aiming to address prevalent challenges such as low student self-esteem, high dropout rates, lack of motivation and engagement, self-regulation, and task performance. Customizing the educational experience for each learner requires the use of machine learning and deep learning techniques. The development of a machine learning model that can predict the mood or attitude of a group of students or a single student is recommended in order to assist teachers in responding properly. By adding emotional intelligence with educational technology, it is feasible to run cost-effective training programs while simultaneously significantly raising the quality of remote learning. It has been noted that emotion-aware systems have the potential to increase the effectiveness of online learning, and it has been recommended that further research and development be done on these systems.**

Keywords*— *Emotion detection, Machine Learning, Deep Learning, Face Recognition, Emotional Intelligence, Teaching strategies, Classroom Analysis, Attendance marking automation

I. Introduction

Teaching is a challenging profession that requires instructors to effectively contribute to students' learning and growth by considering into account their wants and needs, circumstances, and emotions. However, a student's mood can significantly impact their ability to learn, with positive moods facilitating better learning outcomes and negative moods hindering progress. Adapting teaching methods to match the mood of the class is crucial to ensure comprehensive understanding of the subject matter. Unfortunately, the challenges posed by large class sizes and limited time make it difficult for teachers to assess the moods of each student and adjust their approach accordingly. The exploration here focuses on the potential of machine learning and deep learning methodologies to create an emotion-aware system capable of real-time mood prediction for both individual students and groups. Such a system would provide teachers with valuable insights into their students' emotional states, enabling them to tailor their teaching methods and enhance learning outcomes. The paper emphasizes the significance of emotion-aware systems in improving teaching effectiveness and recommends further research for their development and implementation in educational settings.

Emotional intelligence in educational technology has the ability to significantly improve distant learning's quality and offer reasonably priced training options. It is meant to evaluate the issue statement by looking at facial recognition, experimenting with various machine learning models, and evaluating their effectiveness in real-time anticipating students' emotional states. By developing an emotion-aware system that can predict the mood or vibe of groups or individual students, valuable insights can be provided to teachers, enabling them to personalize their teaching methods and optimize learning outcomes. The significance of emotional intelligence in education is emphasized, and the potential of machine learning and deep learning methodologies in developing emotion-aware systems for educational settings is highlighted. The proposed research aims to contribute to the advancement of educational technology and provide new insights into the field of emotional intelligence in education.

II. Literature Survey

Emotion recognition, coupled with accurate face detection, holds immense potential in revolutionizing various domains such as human-computer interaction, affective computing, and social robotics. The ability to automatically detect and interpret facial expressions is crucial for understanding human emotions and improving the effectiveness of human-machine interfaces. In recent years, significant progress has been made in the field of emotion recognition and face detection, driven by advancements in deep learning, computer vision, and multimodal data processing. This literature survey aims to provide a comprehensive overview of the state-of-the-art techniques, datasets, and evaluation methodologies employed in the intersection of emotion recognition and face detection. By examining the existing body of research, identifying challenges and limitations, and exploring emerging trends, this survey seeks to contribute to a deeper understanding of the advancements in this field and offer insights into future research directions for achieving more robust and accurate emotion recognition in conjunction with face detection.

In the pursuit of advancing face recognition systems, researchers in [12] propose a novel and innovative bottleneck approach. Recognizing the need for more efficient and effective methods, their approach introduces a paradigm shift by incorporating a bottleneck layer into the face recognition pipeline.

The core principle behind this approach is to extract and utilize a compressed representation of facial features, referred to as the bottleneck feature. By strategically positioning a bottleneck layer within the architecture, the proposed method facilitates the extraction of the most salient and discriminative facial characteristics while significantly reducing the dimensionality of the feature space. By integrating this bottleneck layer, the researchers aim to overcome the computational challenges typically associated with face recognition systems. The compressed feature representation not only optimizes computational efficiency but also maintains a considerable level of accuracy in identifying and verifying faces. This breakthrough paves the way for real-time and large-scale face recognition applications that were previously hindered by computational constraints.

To implement this approach, a deep neural network, often based on convolutional neural network (CNN) architecture, is trained on a vast dataset. Through a supervised learning process, the network learns to extract highly informative features while ensuring that the bottleneck layer enforces dimensionality reduction. This training phase plays a crucial role in enabling the network to capture the essence of facial characteristics while discarding redundant or less essential information. Once the network is trained, the bottleneck layer can be leveraged to extract bottleneck features from new face images. These bottleneck features, being highly compact yet representative, can then be utilized for downstream face recognition tasks, such as classification or verification. The advantage of this approach is twofold: it significantly reduces computational complexity, allowing for real-time or resource-constrained scenarios, while preserving the discriminative power required for accurate face recognition. The proposed bottleneck approach introduced in [12] presents a paradigm shift in face recognition research, addressing the need for improved efficiency without sacrificing recognition accuracy. By compressing the feature space through the integration of a bottleneck layer, this method opens up new possibilities for faster and more scalable face recognition systems. Future research can explore the potential of this approach in various domains, driving advancements in human-computer interaction, biometric authentication, and security applications.

The bottleneck approach in face recognition, while offering advantages such as computational efficiency and compact feature representation, has some limitations. Information loss is a concern as dimensionality reduction may discard fine-grained details, affecting recognition accuracy. The approach can be sensitive to pose, lighting, and facial expression variations, potentially impacting performance. The compressed representation may have limited capacity to capture the full complexity of facial features, leading to reduced performance on diverse recognition tasks. The success of the approach relies heavily on diverse and representative training data. Lastly, there is a trade-off between efficiency and accuracy, where extreme compression can result in a significant loss of discriminatory information. Addressing these limitations will be crucial for enhancing the effectiveness of the bottleneck approach in face recognition. In their study [13], researchers investigate the effectiveness of different clustering algorithms, including K-means, agglomerative clustering, and DBSCAN, for the identification and categorization of individuals with similar biometric traits.

In the field of facial attribute detection and estimation, researchers have successfully applied transfer learning to neural network models, specifically convolutional neural networks (CNNs). Transfer learning involves utilizing pre-trained models that have been trained on a large dataset for a related facial recognition task, such as face identification or landmark detection. By leveraging the knowledge and learned features from these pre-trained models, the neural network models can improve the accuracy and efficiency of facial attribute detection and estimation. This approach is particularly beneficial when labelled data for the target facial attribute task is limited, as the pre-trained models provide a solid foundation. In parallel, deep features extracted from deep neural networks have proven to be effective in identifying distinguishing features for categorizing individuals with similar biometric traits, such as facial appearance. These deep features capture high-level representations and discriminative patterns within facial images. By passing face images through pre-trained layers of deep neural networks, such as CNNs, and extracting the activations from intermediate layers, deep features are obtained. These deep features are then used as input for clustering techniques, including k-means clustering or hierarchical clustering. The clustering algorithms analyse the similarity between deep features and assign individuals to different clusters based on shared biometric traits. This enables the identification of distinguishing facial features that contribute to the categorization of individuals with similar biometric traits.

By integrating transfer learning and clustering techniques based on deep features, researchers in [13] gained insights into the common characteristics among individuals and facilitate tasks such as face recognition, demographic analysis, or targeted advertising. This combined approach offers a comprehensive and powerful methodology for understanding facial attributes, identifying shared traits, and extracting valuable information from facial biometric data. While the approach of applying transfer learning and clustering techniques using deep features for facial attribute detection and categorization offers several advantages, it also presents several potential disadvantages. First, the reliance on pre-trained models introduces the risk of biases inherited from the training data. These biases can propagate and result in inaccurate or skewed categorization of facial attributes. Second, the success of transfer learning depends on the availability of labelled data for pre-training the models, which can be challenging to obtain in sufficient quantity and quality. Inadequate labelling may adversely affect the performance of the approach. Third, the generalization capability of the approach to unseen data is limited, as it heavily relies on the similarity of data patterns in the training dataset. The approach may struggle to adapt to novel or diverse facial attributes not adequately represented during training. Fourth, the inherent complexity and black-box nature of deep features and clustering techniques make it challenging to interpret and explain the decision-making process, which may hinder trust and transparency. Furthermore, the approach can be computationally intensive, requiring significant resources and time for training and inference, which may limit its practicality in resource-constrained or time-sensitive scenarios. Lastly, the approach may encounter difficulties in accurately detecting and categorizing highly variable or ambiguous facial attributes, affecting its reliability and robustness. Addressing these limitations necessitates careful consideration of biases, data labelling, generalization capabilities, interpretability, computational resources, and the challenges posed by variable and ambiguous facial attributes. Continued research and advancements in these areas are crucial to further enhance the effectiveness and mitigate the drawbacks of this approach.

In [15], the authors propose the utilization of Susan edge detection operation, facial geometry analysis, and edge projection analysis methods to extract facial features such as eyebrows, eyes, mouth, and nose. The Susan edge detection operation serves as an essential step in identifying the boundaries of these facial features by detecting local edge points using circular templates of varying sizes and orientations. This allows for precise edge detection based on intensity changes in facial images or video frames. Furthermore, the facial geometry analysis technique involves measuring and analysing the spatial relationships and proportions between different facial landmarks, enabling the extraction of important geometric information. Finally, the edge projection analysis methods project the extracted facial edges onto specific axes or planes, providing additional insights into the shapes, contours, and relative positions of the facial features. To evaluate the performance of these techniques, the researchers utilize the JAFFE database, a widely-used dataset containing facial images of Japanese female subjects exhibiting different facial expressions. By employing these methods and evaluating them on the JAFFE database, the researchers aim to enhance the accuracy and effectiveness of facial feature extraction, contributing to advancements in facial analysis, recognition, and understanding. In addition to the Susan edge detection operation, facial geometry analysis, and edge projection analysis methods mentioned in the previous paragraph, the approach presented in [15] incorporates the analysis of statistical features for emotion identification. This involves extracting and analysing various statistical measures from the facial feature data to infer and categorize the emotional states of individuals. Statistical features refer to quantitative measures derived from the extracted facial features. These features can capture important characteristics such as intensity, texture, shape, or symmetry of the facial expressions. Examples of statistical features commonly used for emotion identification include mean intensity, standard deviation, entropy, histogram-based features, and geometric ratios. By applying statistical analysis techniques to these features, the approach aims to identify patterns and correlations between the statistical measures and emotional states. This analysis can involve various statistical methods such as mean, variance, skewness, kurtosis, or machine learning algorithms that learn and generalize from the statistical properties of the facial feature data.

In reference to [17], a novel background subtraction algorithm is introduced. The algorithm addresses the task of isolating the foreground objects or individuals from the background in images or video frames. By leveraging advanced computer vision techniques, the proposed algorithm aims to accurately detect and extract the foreground elements, enabling various applications such as object tracking, activity recognition, and surveillance systems. The background subtraction algorithm operates by analysing the temporal changes in pixel values within a given image sequence. It starts by creating an initial background model that represents the static or stationary elements in the scene. This model is derived from a set of training images or by dynamically updating and adapting the background over time. Next, the algorithm compares each incoming frame with the background model and determines the pixel differences or discrepancies. Regions with significant differences are classified as foreground regions, indicating the presence of moving or dynamic objects in the scene. To refine the results, additional techniques such as morphological operations, noise removal, or shadow detection may be applied to further enhance the accuracy of foreground extraction. The proposed background subtraction algorithm introduces innovative strategies to handle challenges such as illumination changes, dynamic backgrounds, and noise. It may employ adaptive thresholding methods, statistical modelling, or machine learning techniques to adaptively adjust the background model and handle variations in the scene. By continuously updating the background model and accurately detecting foreground objects, the algorithm can effectively segment and isolate the desired objects of interest from the background.

The algorithm presented in [17] offers a valuable contribution to the field of computer vision by providing an efficient and robust solution for background subtraction. Its effectiveness in various applications relies on its ability to accurately distinguish between foreground and background elements, enabling precise object detection and tracking. The algorithm's performance is typically evaluated using benchmark datasets or custom datasets, and its results are compared against other state-of-the-art methods to assess its efficacy and efficiency.

The paper referenced as [18] introduces the utilization of the Audio-frame mean expression (AFME) algorithm for audio analysis. This algorithm aims to extract meaningful information from audio signals by analysing the average expression or characteristics of audio frames. The AFME algorithm operates by dividing an audio signal into small frames or segments, typically spanning a few milliseconds each. Within each frame, the algorithm calculates the mean expression by averaging the amplitude or energy values of the audio samples within that frame. This mean expression serves as a representative feature that encapsulates the overall energy or intensity level within the frame. By computing the AFME for each frame, the algorithm can capture variations in the audio signal over time. For instance, in the context of speech analysis, the AFME algorithm can reveal changes in speech volume, emphasize specific phonetic patterns, or detect emotional cues expressed through variations in vocal intensity. Similarly, in music analysis, the algorithm can help identify rhythm changes, distinguish between different musical passages, or highlight expressive dynamics. To further enhance the robustness and accuracy of the AFME algorithm, additional processing steps may be incorporated. These can include pre-processing techniques such as noise reduction, signal normalization, or feature selection to improve the reliability of the extracted mean expressions. Machine learning algorithms or statistical models may also be applied to classify or analyse the extracted AFME features for specific audio tasks, such as speech recognition, emotion detection, or music genre classification. The AFME algorithm presents several advantages in audio analysis. Firstly, by summarizing each frame's expression with a single mean value, it provides a compact representation that captures essential audio characteristics. Secondly, the algorithm is computationally efficient, enabling real-time analysis of audio streams. Lastly, the AFME algorithm can be easily integrated into existing audio processing pipelines or systems, making it applicable to a wide range of audio analysis applications.

In the context of [18], the AFME algorithm is proposed and evaluated on specific audio datasets or test cases. Its performance and effectiveness are typically measured using evaluation metrics such as accuracy, precision, recall, or mean squared error, comparing its results against other state-of-the-art audio analysis methods. By leveraging the AFME algorithm, researchers aim to enhance audio analysis capabilities and enable various applications in fields such as speech processing, music analysis, and audio-based affective computing.

In reference to [19], the adaptive background difference method is introduced as an alternative approach to traditional methods such as optical flow, background difference, and frame difference. The adaptive background difference method aims to accurately detect foreground objects or changes in a scene by adaptively updating the background model and considering dynamic variations in the environment. Traditionally, optical flow methods have been used to estimate the motion of objects by tracking the displacement of pixels between consecutive frames. However, optical flow methods can be sensitive to noise and often struggle to handle complex scenes with occlusions or rapid motions. Similarly, background difference methods compare each incoming frame with a static background model to detect foreground objects. However, these methods can be affected by gradual changes in lighting conditions, dynamic backgrounds, or camera noise, leading to false detections or unstable results. The frame difference method involves subtracting consecutive frames to highlight regions with significant pixel changes, indicating the presence of foreground objects. However, this approach can be sensitive to noise and may not effectively handle gradual variations or complex motion patterns. To address these limitations, the adaptive background difference method utilizes a more robust and adaptable approach. It dynamically updates the background model by continuously incorporating new incoming frames into the background estimation process. This allows the algorithm to adapt to changing environmental conditions and reduce the impact of noise or gradual variations. The adaptive background difference method typically involves techniques such as temporal filtering, statistical modelling, or pixel-wise adaptive learning. These techniques help refine the background model by selectively updating specific pixels based on their characteristics or considering the temporal relationships between frames. By adaptively adjusting the background model, the algorithm can better handle dynamic backgrounds, lighting changes, or camera motion, leading to more accurate foreground object detection. The proposed adaptive background difference method provides several advantages over traditional methods. It offers increased robustness and adaptability, allowing for improved performance in complex scenes with varying conditions. By adaptively updating the background model, the algorithm can effectively handle gradual changes, noise, and dynamic backgrounds, leading to more reliable and stable foreground object detection. In [19], the adaptive background difference method is likely presented along with experimental evaluations on benchmark datasets or real-world scenarios. The performance of the method is typically assessed using metrics such as precision, recall, F1-score, or receiver operating characteristic (ROC) curves to measure its effectiveness in detecting foreground objects while minimizing false alarms.

III. Proposed Methodology

The literature survey provides valuable insights into the intersection of emotion recognition, facial feature analysis, and background subtraction. It underscores the potential of these technologies to transform various fields, from human-computer interaction to security systems. The survey emphasizes the importance of accurate facial feature extraction for tasks like emotion detection and face recognition, shedding light on innovative techniques like the "bottleneck approach" that optimizes computational efficiency without compromising accuracy. Furthermore, it explores the integration of transfer learning and deep features for categorizing individuals with similar biometric traits, offering a powerful methodology for understanding facial attributes. Additionally, the survey discusses the adaptive background difference method, a robust approach to foreground object detection that adapts to changing environmental conditions. Overall, the survey paints a vivid picture of the cutting-edge techniques and challenges in these fields, providing a solid foundation for the proposed methodology's application in educational settings.

Adapting teaching strategies to the tone of the class becomes increasingly complex when dealing with a large number of students. The time restrictions faced by teachers make it impractical to individually assess the mood of each student and ensure comprehensive understanding of the subject for all. The challenge lies in finding a way to efficiently address the needs of multiple students simultaneously. To overcome this difficulty, the development of an emotion-aware system becomes crucial. By leveraging machine learning and deep learning methodologies, the system can analyse various data sources, including text, audio, and video inputs, to predict the emotional states of a group of students or individual students in real-time. This technology allows the teacher to receive valuable insights into the emotional atmosphere of the class without the need for individual assessments. The emotion-aware system can provide the teacher with actionable information, suggesting appropriate teaching strategies and interventions based on the predicted emotional states of the students. This way, the teacher can tailor their approach to address the overall mood of the class and optimize learning outcomes for the majority of students. While the system cannot replace the personalized attention that individual assessments would provide, it offers a practical solution for managing large class sizes and time constraints. By focusing on the collective emotional state of the class, the teacher can make informed decisions to create a supportive and engaging learning environment that benefits the majority of students. Implementing an emotion-aware system in the classroom would alleviate the burden on the teacher and enable them to allocate their time and resources more effectively. By addressing the emotional needs of the class as a whole, the teacher can ensure that each subject is understood and the syllabus is covered to the best extent possible within the given constraints.

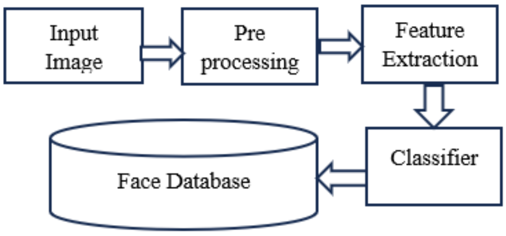

Fig. 1. Block Diagram for Face detection

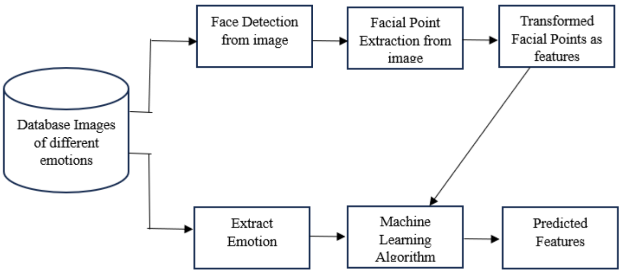

Fig. 2. Block diagram for Emotion recognition

For face detection, various approaches can be employed like PCA and Eigen face detection-based approach, CNN based approach, deep face analysis-based approach. The accuracy of PCA and Eigen face detection-based approach and CNN based approach is around 72% which is very low. In deep face analysis-based approach, first the class photo is given as input and individual faces are extracted. Extracted faces are cropped and saved with the time stamp. Then faces are compared with the student database. If match is found then attendance is marked, dominant emotion of the student is found using the DeepFace Library. DeepFace, an innovative Python library offers a compelling solution for emotion recognition. Its exceptional accuracy, powered by deep learning and pre-trained models, ensures reliable emotion detection across a spectrum of feelings. Its speed, scalability, and visualization tools further enhance its suitability making it a valuable asset for emotion analysis and interpretation.

A. *Integration of Face detection and emotion detection system*

The Emotion Detection Project involves combining a number of elements, such as face recognition, face cropping, and emotion detection. The initiative is also intended to record students' feelings and attendance in a class. In this context, the six phases that make up the integration process have been explored..

Step 1: Establishing a MongoDB DatabaseThe initial phase of the Emotion Detection Project is concentrated on building up a reliable MongoDB database. This database serves as the main repository for all the vital information needed to complete the project successfully. This well-designed MongoDB structure has been constructed to include two key collections.

a) Student Data Collection: This repository contains a wealth of data that was supplied by the students themselves. Every pertinent detail is completely captured not only names and addresses. This collection becomes a thorough library of student profiles, including basic biographical information, contact details, and even images.

b) Attendance Data Collection: The complete second collection is used to record attendance. The system's nerve center, this collection tracks student involvement. Each entry includes crucial attendance information, such as student identification numbers and timestamps. Additionally, based on the complexity of the project, this collection may be supplemented with information on the outcomes of emotion analysis or real-time attendance status.

Laying the foundations for the Emotion Detection Project, this basic phase also ensures data structure, accessibility, and security.

Step 2: Ingesting Student Data into the Student Data Collection: The crucial Step 2 is where the Emotion Detection Project really starts to take off. Now the emphasis is on adding data to the MongoDB database's Student Data collection. The rigorous processing of student data, spanning a wide range of data, is what distinguishes this approach.

a) Comprehensive Data Gathering: The vast majority of the data in this collection was contributed by the students themselves. This collection develops into a comprehensive library of student profiles, complete with basic biographical data, contact information, and even pictures.

b) Image Encoding Integration: The inclusion of picture encodings in data collecting is what makes this stage distinctive. As was already said, face comparison activities in projects benefit greatly from picture encoding. These encodings, which are kept in a binary file with minimal storage requirements, provide as the foundation for effective face comparisons. They are essential for identifying students and properly determining their emotional states.

The foundation for accurate face recognition is laid by incorporating picture encodings into data, which also helps to improve data storage. The ensuing phases are made possible by the rigorous data preparation, which guarantees that the Emotion Detection Project runs effectively and precisely.

Step 3: Extracting and Cropping Faces for Identification and Emotion Recognition: Identify individual students' faces in group images with precision during this crucial stage, taking into account numerous elements including lighting and emotions. Following face extraction, these extracted faces are meticulously cropped into distinct picture files to maintain clarity for upcoming face identification and emotion detection tasks. All processed photos are effectively saved in a designated directory, facilitating access for later stages of the project, and use a timestamp-based naming standard for each image to protect data integrity and prevent mix-ups. Accurate and trustworthy facial recognition and emotion analysis are enabled by this rigorous technique.

Step 4: Precise Face Recognition for Attendance: Step 4 involved carefully implementing facial recognition to correctly record attendance. When a match is found, attendance is recorded in the Attendance Data collection, which stores important attributes such as the student's identified name, date, the image encoding of the cropped face for reference, and the data directory path. This is done by methodically looping through the cropped student images and comparing them with stored images from the Student Data collection using precise face encodings. This process serves as the system's cornerstone and guarantees the dependability and accuracy of attendance tracking.

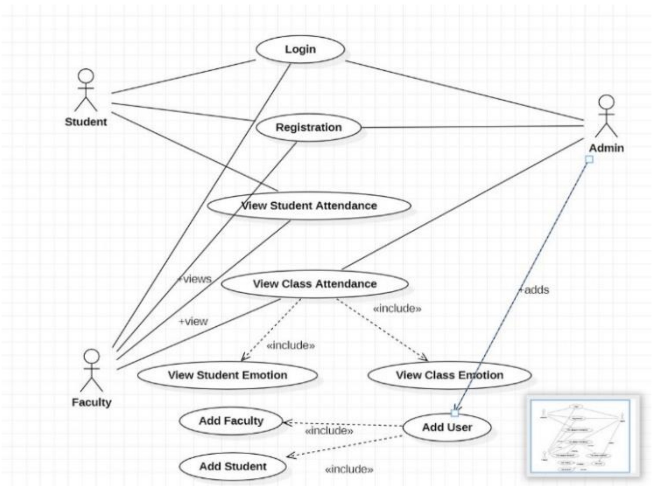

Fig. 3. Use Case diagram of the model

Step 5: Integration of Emotion Detection with Attendance System: In Step 5, the attendance system will be seamlessly integrated with emotion detection. The first step is to extract from the Attendance Data collection the picture encodings unique to a given day. Create NumPy arrays from the photos using these encodings, and then pass the arrays through the DeepFace.analyze() function of the DeepFace Library. The ethnicity, gender, age, and mood of the identified face are just a few of the characteristics that this potent library is excellent at identifying. The outcomes of the investigation with an emphasis on identifying the prevalent emotion. In order to complete the process, add this useful emotion detection data to the Attendance Data collection's "emotion" property. In this way, the attendance system efficiently combines the tracking of attendance with the comprehension of and monitoring of student emotions.

Step 6: Accessing Attendance and Emotion Data: Streamline access to attendance and mood data in the last step, Step 6. The system quickly locates and presents the pertinent data when asked for attendance records on a certain day. This well-organized method is excellent at accurately tracking student attendance as well as recording their feelings, providing information about the range of emotions present in the class as a whole. Figure 3's use case graphic, which depicts these features, gives a clear visual picture of how attendance management and emotion analysis are seamlessly integrated into the project.

IV. RESULTS AND DISCUSSIONS

A. PCA and Eigen Face Detection Model

Fig. 4 displays the results of the facial recognition algorithm, showcasing the faces that have been successfully identified and detected by the system. These recognized faces are subsequently saved and organized according to the predetermined size specifications outlined in the program.

In more detail, the system employs a facial recognition algorithm to analyze images or data, and in this case, it has successfully identified and located human faces within the input. These identified faces are then subjected to a process that ensures they conform to specific size requirements as defined by the program's parameters. This resizing and structuring of the recognized faces enable the system to standardize and prepare the data for further processing or storage, making it easier to work with and analyze as needed.

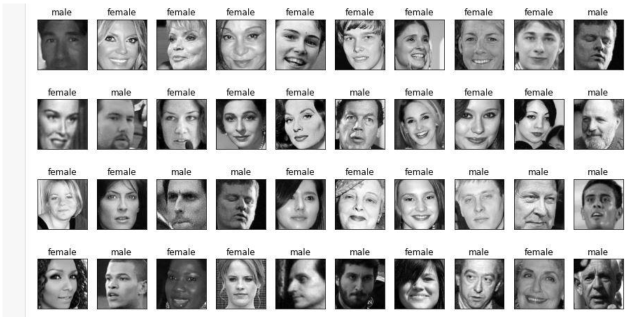

Fig. 4. Detected faces of group photograph

Fig. 5 provides us with valuable information about the emotions that have been detected from the cropped faces, accompanied by their respective eigen images. The eigen image is a crucial component as it contains all the essential elements necessary for predicting the emotion associated with each face.

To elaborate further, when the system detects and crops faces from the input data, it subsequently processes these cropped faces to generate eigen images. An eigen image is a representation that captures the key features and characteristics of a face related to its emotional expression. These eigen images serve as a condensed and informative representation of the facial features that play a significant role in determining the emotion being conveyed.

Once the eigen image for a given face is generated, it becomes a vital tool for predicting the associated emotion. This prediction is accomplished by comparing the eigen image of the detected face with the eigen images of faces stored in the system's dataset. By analyzing the similarities and differences between the eigen image of the detected face and those in the dataset, the system can make accurate predictions about the emotional state of the individual based on recognized patterns and correlations. This process enables the system to provide insights into the emotions expressed by the individuals in the images it analyzes, contributing to a better understanding of emotional states in various contexts.

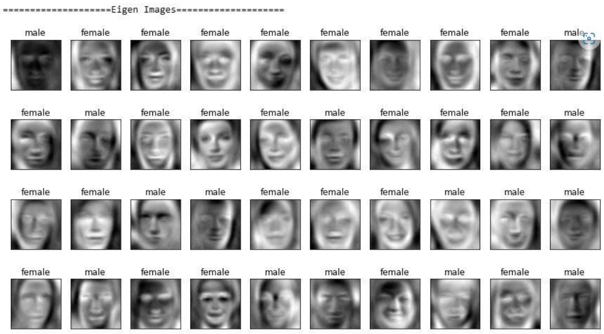

Fig. 5. . Eigen Images of the detected faces

B. CNN Based Approach

Fig. 6 shows a set of images that specifically convey the emotion of happiness. Similarly, if we opt for other emotions such as anger, disgust, fear, neutrality, sadness, and surprise, the corresponding images representing each of these emotions will be displayed as the output.

These images play a crucial role in the training and evaluation process of the emotion detection model. Once the images are displayed, they undergo a transformation into grayscale format. This conversion to grayscale is a necessary step as it simplifies the image data, making it easier for the classifier to detect and analyze facial features associated with different emotions. Grayscale images use varying shades of grey to represent the intensity of pixel values, which is particularly useful for capturing facial expressions and emotional nuances.

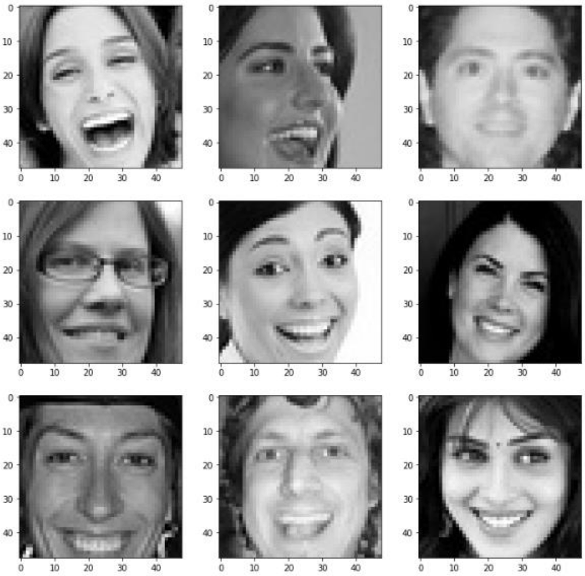

Fig. 6. Faces with happy emotions

In Fig. 7 and Fig. 8, the model's performance metrics, including accuracy and losses, are visualized and plotted. This visualization provides us with a comprehensive view of how well the emotion detection model is performing. Specifically, the figure reveals that the model achieves an accuracy rate of 72 percent in classifying emotions.

These accuracy and loss metrics are invaluable as they offer insights into the model's learning progress and overall performance. By closely examining the accuracy and loss trends depicted in Fig. 7, we can assess the model's effectiveness in accurately classifying emotions within the provided dataset. This evaluation process is critical for fine-tuning the model, making necessary adjustments, and enhancing its ability to correctly identify and classify a wide range of emotions, ultimately leading to improved emotion detection results.

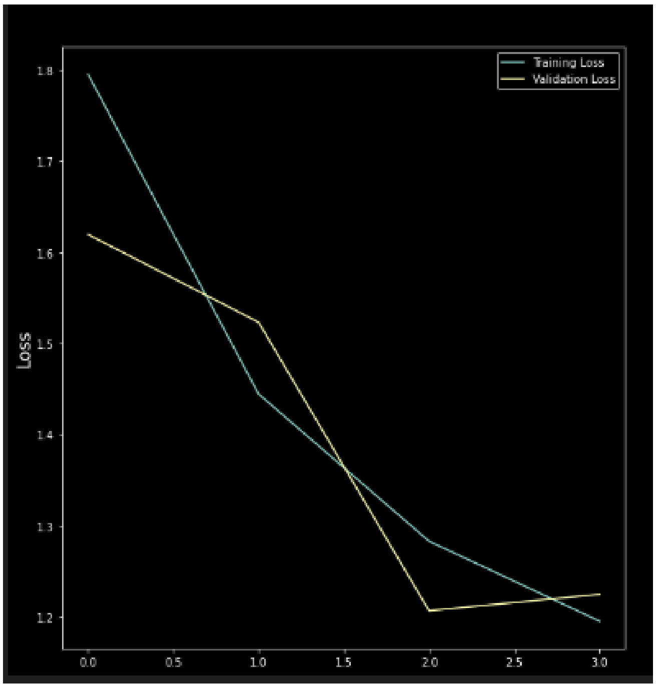

Fig. 7. Loss Graph

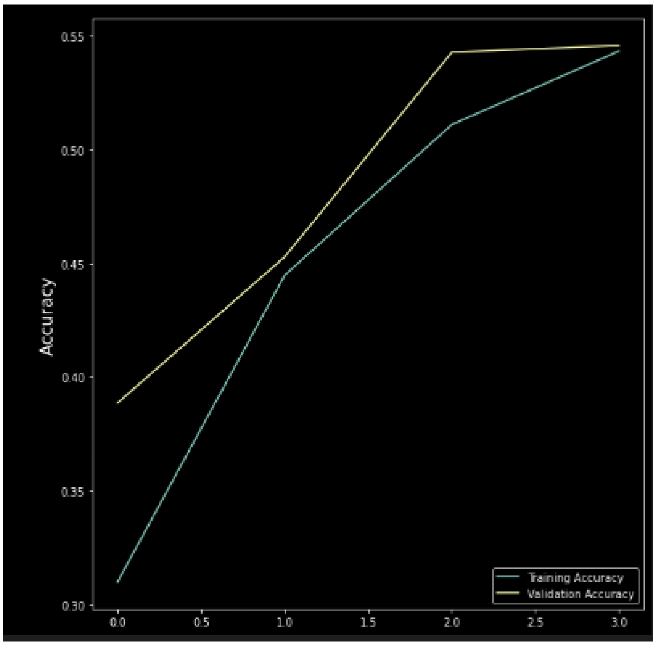

Fig. 8. Accuracy graph

C. Deepface Analysis Based Approach

In Fig. 9, we can observe an input image that is categorized as "unknown." This unknown image undergoes a series of processing steps within the "compare faces" function of the system. This function plays a critical role in determining whether the unknown image matches any of the known faces stored in the system's database.

The primary operation performed by the "compare faces" function is the calculation of the Euclidean distance between the features of the unknown image and the features of the known faces. The Euclidean distance is a mathematical measure used to quantify the similarity or dissimilarity between two sets of data points. In the context of face recognition, it serves as a crucial metric for determining how closely the unknown face aligns with the features of the known faces.

Fig. 9. Unknown image to be passed as input

In Fig. 10, we see the result of this comparison process. The function calculates the Euclidean distances and identifies the best match index. The best match index indicates which known face in the database is the closest match to the features extracted from the unknown image.

```
matches = face_recognition.api.compare_faces(known_face_encodings, np.array(face_encoding), tolerance=0.6)
# Instead, use face_distance to calculate similarities
face_distances = face_recognition.face_distance(known_face_encodings, np.array(face_encoding))
best_match_index = np.argmin(face_distances)
name = "Unknown"
if matches[best_match_index]:
    name = known_face_names[best_match_index]
print(name)

Lisa
```

Fig. 10. Comparison of the given face with the database

If the unknown image aligns sufficiently with the best match index, it signifies a successful face detection and recognition. This implies that the system has identified the individual in the unknown image as a known person whose face is stored in the system's database. Subsequently, the system saves the recognized image with the corresponding name or label, such as "lisa.jpg," as illustrated in Fig. 11.

Fig. 11. Detected face stored in the database

This entire process, from inputting an unknown image to calculating the Euclidean distance, determining the best match index, and saving the recognized image, enables accurate identification and storage of recognized faces. This contributes significantly to the overall effectiveness of the face recognition system, as it ensures that known individuals are correctly identified and their images are appropriately labelled and stored for future reference and use.

Fig. 12 serves as the starting point in our analysis or processing pipeline, presenting the original, unaltered image or input data. This image is rich in facial features and serves as the foundation for our subsequent steps. It provides us with the raw material from which we can extract valuable information.

Fig. 12. Original Image

Fig. 13 reveals the outcome of cropping the original image to zoom in on specific regions of interest, particularly facial characteristics essential for our emotion identification process. These cropped images are indispensable for singling out and manipulating precise elements within the original image.

Fig. 13. Cropped Photos

Moving on to Fig. 14, we witness the fruits of our identification process. This figure showcases the portions of the cropped images that have been successfully recognized and matched with well-known individuals or subjects. It represents a critical step in our pipeline, as it confirms the accuracy of our recognition efforts.

```
No match found for face_0_20230425_133102.jpg
No match found for face_10_20230425_133102.jpg
No match found for face_11_20230425_133102.jpg
No match found for face_12_20230425_133102.jpg
[0.74488791 0.63834476 0.80061417 0.70135485 0.71425941 0.8297155
 0.41832176]
Shriansh Jena is present
[0.73012655 0.64094024 0.8001857  0.68260027 0.39792564 0.76527105
 0.64187311]
Pratham Mahajan is present
No match found for face_15_20230425_133102.jpg
[0.68059042 0.66194772 0.75313295 0.49579607 0.66209443 0.84703631
 0.69969009]
Nimish Nandanvar is present
No match found for face_17_20230425_133102.jpg
[0.7312386  0.70450695 0.68448255 0.71734767 0.72136468 0.41264328
 0.73236248]
Lemmie Carvalho is present
```

Fig. 14. Identified Images

Finally, in Fig. 15, we delve into the emotional states or expressions identified for the recognized individuals. This figure offers us a window into the system's interpretation process, shedding light on how it understands and categorizes emotions based on the features it has identified.

```
Enter the date (yyyy-mm-dd): 2023-04-25
Attendance and Emotions for the date: 2023-04-25 00:00:00
Shriansh Jena : happy
Pratham Mahajan : neutral
Nimish Nandanvar : happy
Lemmie Carvalho : happy
Moeez Shaikh : happy
```

Fig. 15. Detected Emotions

Together, these figures constitute integral stages within our visual analysis and processing pipeline, illustrating the meticulous process of extracting, recognizing, and interpreting information from the original input data. Collectively, they contribute significantly to our comprehension and the overall functionality of the system or process detailed in our proposed methodology.

D. Observations and Comparisons

TABLE I. COMAPARATIVE ANALYSIS

Criteria	**Eigenfaces**	**DeepFace**	**CNN-based Models**
Accuracy	80%	99.93%	72%
Capture of Facial Features	Limited to overall variations in facial expressions	Captures complex patterns and subtle facial features	Captures complex patterns and subtle facial features
Capture of Facial Features	Limited to overall variations in facial expressions	Captures complex patterns and subtle facial features	Captures complex patterns and subtle facial features
Handling Pose and Lighting Variations	May struggle with large pose and lighting variations	Can handle to some extent	Can handle to some extent
Training Data Requirement	Requires limited training data	Requires a large amount of labeled data	Requires a large amount of labeled data
Computational Resources	Computationally efficient	Requires significant computational resources	Requires significant computational resources
Performance	Moderate accuracy	High accuracy	High accuracy with appropriate training

The Table1 provides a comparative analysis of three distinct face recognition models: Eigenfaces, DeepFace, and CNN-based models, with each model evaluated based on several criteria.

In terms of accuracy, Eigenfaces demonstrate a reasonable accuracy level of 80%, while DeepFace stands out with a remarkable accuracy rate of 99.93%. On the other hand, CNN-based models achieve an accuracy of 72%.

When considering the capture of facial features, Eigenfaces are somewhat limited, primarily capturing overall variations in facial expressions. In contrast, both DeepFace and CNN-based models excel in capturing intricate and subtle facial features, enabling them to recognize complex patterns effectively.

Regarding the handling of pose and lighting variations, Eigenfaces may encounter challenges, especially with large variations in pose and lighting. DeepFace exhibits a moderate capability to handle these variations, while CNN-based models also exhibit a similar capacity.

In terms of training data requirements, Eigenfaces demand relatively minimal training data, making them suitable for situations with limited data availability. DeepFace, however, necessitates a substantial amount of labeled data for effective training, a requirement shared by CNN-based models.

In the context of computational resources, Eigenfaces are computationally efficient, making them a practical choice for resource-constrained environments. Conversely, both DeepFace and CNN-based models demand significant computational resources to operate efficiently.

Performance-wise, Eigenfaces achieve moderate accuracy levels, DeepFace excels in delivering high accuracy, and CNN-based models, when appropriately trained, can also attain high accuracy levels.

These comprehensive evaluations underscore the distinct strengths and weaknesses of each face recognition model, making them suited for various applications based on specific requirements and available resources.

Eigenfaces, as a traditional approach, utilizes Principal Component Analysis (PCA) to extract eigenfaces, which represent the main variations in facial expressions. However, Eigenfaces may struggle with large pose and lighting variations and may not capture fine-grained details required for accurate emotion detection.

On the other hand, DeepFace, a deep learning-based approach, leverages deep neural networks to learn hierarchical representations from facial images, enabling it to capture complex patterns and subtle facial features relevant to emotion detection. DeepFace can handle variations in pose, lighting, and expressions to some extent and achieves high accuracy.

Similarly, CNN-based models, which also use deep neural networks, excel at capturing complex patterns and fine-grained details. However, both DeepFace and CNN-based models require a large amount of labeled training data and significant computational resources for optimal performance. Fine-tuning and optimization are necessary for these models to achieve high accuracy in emotion detection tasks.

In contrast, Eigenfaces are computationally efficient and require limited training data but may lack the capability to capture nuanced facial features necessary for accurate emotion detection.

V. CONCLUSION AND FUTURE WORK

Utilizing facial recognition technology, the system can analyse students' emotional expressions, detecting patterns and trends in their emotional states. This forms the basis for generating valuable insights and personalized recommendations for educators to improve learning outcomes based on students' emotional needs, either individually or collectively. This allows for efficient testing of various teaching methodologies, with the feedback guiding adjustments to content and delivery methods for more effective teaching and learning interactions. Automating attendance processes reduces the burden on the lecturer and eliminates the potential for recording errors. Developing an emotion-aware system capable of accurately predicting student moods and offering instructional advice is a complex undertaking. To tackle this challenge, we suggest leveraging the capabilities of machine learning and deep learning methodologies. These approaches enable the system to learn from historical data and effectively adapt to real-time inputs. Ensuring the system's accuracy and reliability across different scenarios requires rigorous testing and evaluation using a substantial dataset of student emotions. Additionally, addressing ethical concerns such as privacy and data security is paramount to ensure responsible system usage. Our objective in creating such an emotion-aware system is to contribute to the advancement of educational technology, ultimately enhancing the overall quality of students' learning experiences.

REFERENCES

[1] Nitisha Raut, San Jose State University, Facial Emotion Recognition Using Machine Learning, 2018.

[2] Emmeke Anna Veltmeijer, Charlotte Gerritsen, Koen Hindriks, Automatic emotion recognition for groups: a review, 2021. IEEE Transactions on Affective Computing.

[3] Emanuele Conti, Davide Salvi, Clara Borrelli, Brian Hosler, Paolo Bestagini, Fabio Antonacci, Augusto Sarti, Matthew C. Stamm, Deepfake Speech Detection Through Emotion Recognition: A Semantic Approach, 2022.International Conference on Acoustics, Speech, and Signal Processing (ICASSP).

[4] Aarohi Gupta,Emotion Detection: a Machine Learning Project, 2019. Towards Data Science.

[5] Rohit Dwivedi, My first CNN project – Emotion Detection Using Convolutional Neural Network With TPU,2020. Developers Corner.

[6] M. A. Mahima, N. C. Patel, S. Ravichandran, N. Aishwarya and S. Maradithaya, "A Text-Based Hybrid Approach for Multiple Emotion Detection Using Contextual and Semantic Analysis," 2021 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES), 2021.

[7] D. Joshi, A. Dhok, A. Khandelwal, S. Kulkarni and S. Mangrulkar, "Real Time Emotion Analysis (RTEA)," 2021 International Conference on Artificial Intelligence and Machine Vision (AIMV), 2021.

[8] S. Giri et al., "Emotion Detection with Facial Feature Recognition Using CNN & OpenCV," 2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), 2022.

[9] Ahong Xu, Jun Zhang, Jinwen Tian, Daimeng Zhang, Xiaomao Liu, "The improvement of VIBE foreground detection algorithm," Proc. SPIE 10608,MIPPR 2017: Automatic Target Recognition and Navigation.

[10] Ketan Vaidya, Emotion Recognition-Implementing a CNN model for facial emotion recognition with real-time implementation, 2020, Towards Data Science.

[11] D. Joshi, A. Dhok, A. Khandelwal, S. Kulkarni and S. Mangrulkar, "Real Time Emotion Analysis (RTEA)," 2021 International Conference on Artificial Intelligence and Machine Vision (AIMV), 2021.

[12] Rana Qarooni, Jonathan Prunty, Markus Bindemann, Rob Jenkins,"Capacity limits in face detection",Elsevier, Cognition Volume 228, November 2022, 105227.

[13] Andrea F. Abate, Paola Barra, Silvio Barra, Michele Nappi, Fabio Narducci, "Clustering Facial Attributes: Narrowing the Path From Soft to Hard Biometrics",IEEE Access, Digital Object Identifier 10.1109/ACCESS.2019.2962010, December 24, 2019.

[14] Jiangjin Gao,Tao Yang,"Face detection algorithm based on improved TinyYOLOv3 and attention mechanism",Elsevier, Cognition Volume 181, 1 January 2022, Pages 329-337.

[15] Ashok R. Shinde, Shriram D. Raut, Prashant P.Agnihotri,"Emotion Recognition by analyzing statistical features of cropped eyebrow portion as a Region of Interest of facial expression image",Advances in Computational Sciences and Technology ISSN 0973-6107 Volume 10, Number 8 (2017) pp. 2527-2532.

[16] J. -M. Guo, C. -C. Lin, M. -F. Wu, C. -H. Chang and H. Lee, "Complexity Reduced Face Detection Using Probability-Based Face Mask Prefiltering and Pixel-Based Hierarchical-Feature Adaboosting," in IEEE Signal Processing Letters, vol. 18, no. 8, pp. 447-450, Aug. 2011, doi: 10.1109/LSP.2011.2146772.

[17] O. Barnich and M. Van Droogenbroeck, "ViBe: A Universal Background Subtraction Algorithm for Video Sequences," in IEEE Transactions on Image Processing, vol. 20, no. 6, pp. 1709-1724, June 2011, doi: 10.1109/TIP.2010.2101613

[18] M. T. Teye, Y. M. Missah, E. Ahene and T. Frimpong, "Evaluation of Conversational Agents: Understanding Culture, Context and Environment in Emotion Detection," in IEEE Access, vol. 10, pp. 24976-24984, 2022, doi: 10.1109/ACCESS.2022.3153787.

[19] Zhang, Daimeng & xu, ahong & Zhang, Jun & Tian, Jinwen & Liu, Xiaomao. (2018). The improvement of VIBE foreground detection algorithm. 2. 10.1117/12.2281690.