International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

An Emotional Artificial Intelligence Based Attendance System for Classroom

Chetan K R

Professor and Head, Dept. of AIML, JNNCE, Shivamogga

Abstract
Attendance through a camera using Artificial Intelligence (AI) and Deep Learning (DL) is a modern approach to monitoring and tracking attendance in various settings such as schools, offices, and manufacturing facilities. It uses computer vision techniques and deep learning algorithms to automatically detect and identify individuals in an image or video captured by the camera. The process typically involves training the system on a dataset of labelled images of the individuals who will be taking attendance. This training data can include images of the individual's face, iris, or fingerprints, depending on the specific approach used. Once the system is trained, it can then use this knowledge to recognize these individuals in new images captured by the camera. When an individual is recognized, the system can log their attendance in a database or other record-keeping system. This can be done in real-time, allowing for immediate tracking of attendance, or it can be done at a later time for batch processing. However, AI has been expanding its horizon and face recognition with deep learning techniques can be augmented with emotion recognition as well. Sometimes, students feel really sad and overwhelmed by their school work and other responsibilities. They might feel like they can't keep up or that they are not good enough. When these feelings last for a long time, it's called depression. Depression can make it hard for students to do their school work, be with friends and family, or even get out of bed in the morning. When students are feeling really sad and hopeless, they might think about hurting themselves or ending their lives. This is called suicide. It's important to know that suicide is preventable and there are people who can help. It's important to take care of our mental health, just like we take care of our physical health. If you ever feel sad or overwhelmed, it's important to talk to someone you trust and get help. One way to prevent suicide is to detect if someone is feeling sad and hopeless, which can be a sign of depression. One way to do this is by using a CCTV camera and AI and DL technology to analyze the person's facial expression, body language, and speech patterns. Emotional AI and DL can be a powerful tool in detecting depression but it's not a substitute for a professional diagnosis. It's always advisable to consult with a mental health professional if you suspect that you or someone you know may be struggling with depression.

Keywords: Artificial Intelligence, Deep Learning, Depression Recognition

1. Introduction
Depression is a mental health condition that affects a significant portion of the population. It is characterized by feelings of sadness, hopelessness, and a lack of interest in activities. Emotional AI and ML can be used to detect depression by analyzing a person's speech, text, or image data to identify patterns that are indicative of the condition. This approach can be particularly useful in identifying depression in individuals who may be unwilling or unable to self-report their symptoms. Image analysis can also be used to

detect depression. For example, face based expression recognition can be used to identify changes in a person's facial expression that may indicate the presence of depression. Additionally, body language analysis can be used to identify changes in a person's posture, gait, and other physical cues that may indicate the presence of depression. It's important to note that depression is a complex condition and it's not always easy to detect. Emotional AI and DL can be a powerful tool in detecting depression but it's not a substitute for a professional diagnosis. It's always advisable to consult with a mental health professional if you suspect that you or someone you know may be struggling with depression.

Emotion Recognition is the leading academic challenge on emotion recognition and labeling. The work done in Image based Static Facial Expression Recognition Subchallenge used CNNs[1]. Yu and Zhang [2] proposes a CNN architecture specialized on emotion recognition performance. They put forth two novel constrained optimization frameworks to automatically learn the network ensemble weights by minimizing the loss of ensembled network output responses. Kim et al. [3] took a different approach by creating a committee of multiple deep CNNS. They also created a hierarchical architecture of the committee with exponentially-weighted decision-making process. Works are also reported on the use of Support Vector Machines (SVM) and Largest Margin Nearest Neighbor (LMNN) for classification [4]. The main difference between these were the feature descriptors. [5] used a system that extracts Pyramid of Histogram of Gradients (PHOG) and Local Phase Quantization (LPQ) features for encoding the shape and appearance information. [6] used Action Unit (AU) aware features that were generated after finding pairwise patches that are significant to discriminate emotion categories. The main insight was that previous research groups neglected to explore the significance of the latent relations among changing features resulted from facial muscle motions. Shan et al., [7] concentrated on person-independent facial expression recognition and used Local Binary Patterns (LBP) descriptors. Many of these algorithms that used feature based models aimed to mimic Ekman’s suggestions for human emotion at [8].

The following are the different objectives of the proposed system:

- Early detection of depression and mental health issues: By using advanced Artificial Intelligence algorithms, the system aims to identify signs of depression and other mental health issues in individuals, allowing for early intervention and treatment.
- De-stigmatizing depression and mental health: By using AI-based approaches to detect emotions, the system aims to remove the stigma surrounding mental health and promote open discussion about mental health issues.
- Improving treatment outcomes: By identifying the causes of depression and other mental health issues, the system aims to inform more effective and personalized treatment plans.
- Facial recognition attendance: By using AI and ML algorithms, the system aims to accurately and automatically recognize individuals' facial features and match them with their identity, in order to mark attendance.
- Automating and streamlining attendance systems: By using facial recognition technology, the system aims to automate attendance taking, reducing the need for manual intervention, and making the process more efficient and accurate.
- Improving productivity: By automating attendance taking and detecting mental health issues, the system aims to improve overall productivity and well-being of employees or students.
- Enhancing security: By using facial recognition technology, the system aims to enhance security by ensuring that only authorized personnel are able to access certain areas or resources.

2. System Architecture

Figure-1 shows the model of the proposed system. Images are captured with camera and normalization and other pre-processing techniques are applied. Important face and emotional parameters from image is extracted. The parameters are fed to deep CNN for creating a model for predicting emotional quotient. Based on the value, various analysis is made. The working system is deployed in the Azure/GCP cloud and a user interface is developed to check real time results.

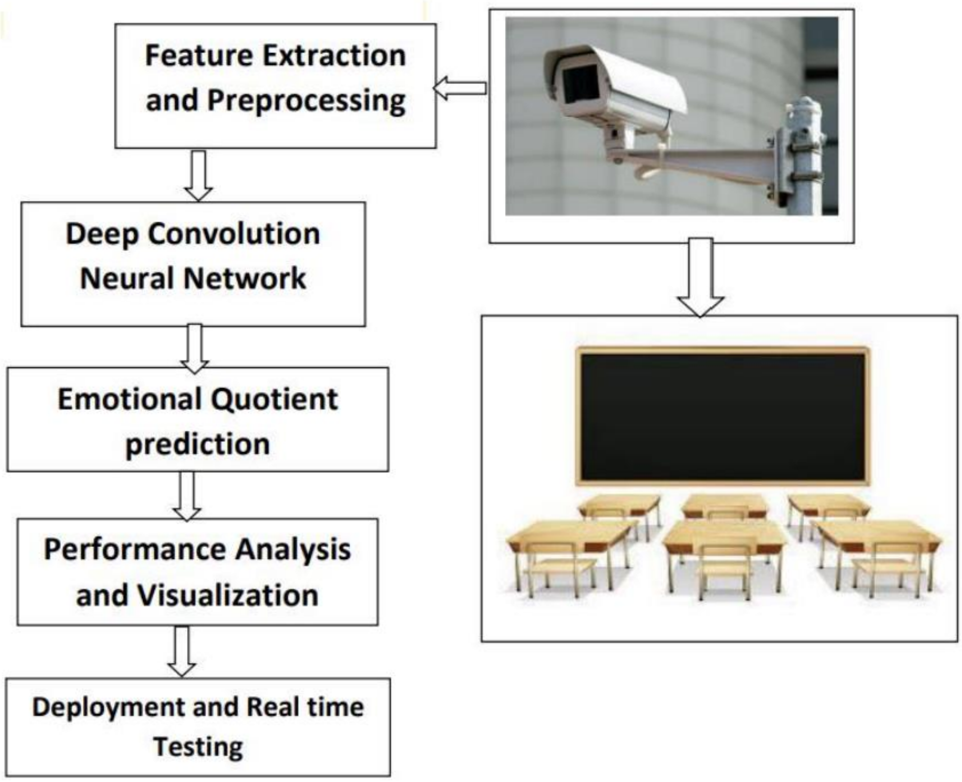

Figure 1: Proposed system architecture

3. Implementation

The system has been implemented as a modular architecture. Different modules implemented include: creation a deep learning model, facial recognition mechanism, emotion detection module, matching of test facial and emotion data. In the deep learning model, MultiTask Cascaded Convolutional Neural Network (MTCNN) is used, which leverages a 3-stage neural network detector. The image is resized multiple times to detect faces of different sizes. Then the P-network (Proposal) scans images, performing first detection. It has a low threshold for detection and therefore detects many false positives, even after NMS (Non-Maximum Suppression), but works like this on purpose. MTCNN is very accurate and robust. It properly detects faces even with different sizes, lighting and strong rotations. It's a bit slower than the Viola-Jones detector, but with GPU not very much. It also uses color information, since CNNs get RGB images as input. FaceNet is one of the state-of-the-art models for Face recognition. It creates embeddings of the image by mapping each image into a Euclidean space. These embeddings can then be used as feature vectors and fed to models like k-NN for face recognition. Clustering techniques can also be applied on the embeddings. FaceNet uses a loss function called as the triplet loss function. It makes sure that the distance between the positive image and the anchor image is as small as possible and the distance between the negative image and the anchor image is as large as possible. Flowcharts of various modules are depicted in Figures 2-5.

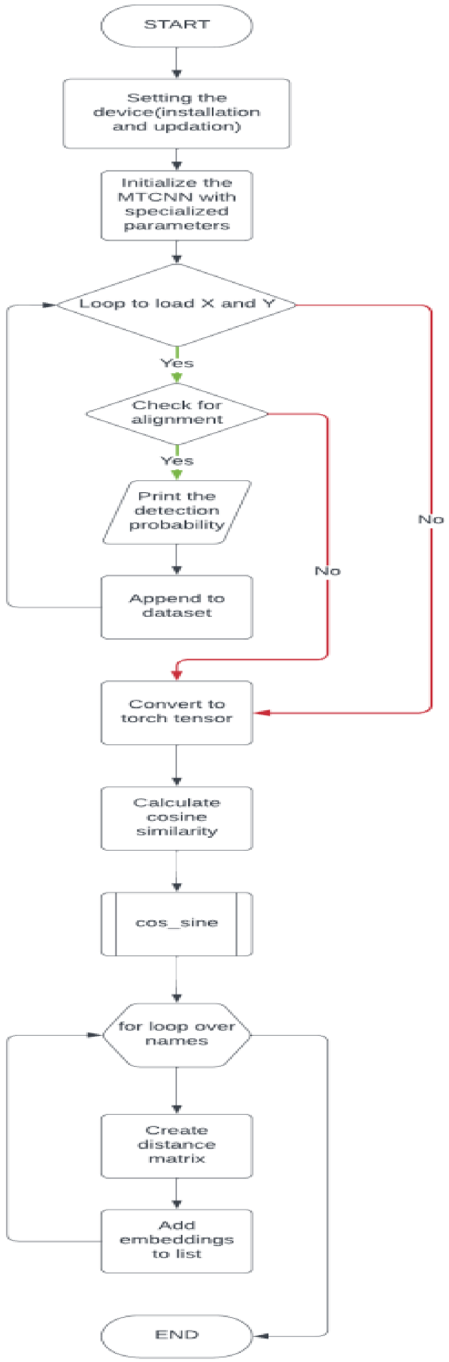

Figure 2: MTCNN based facial detection

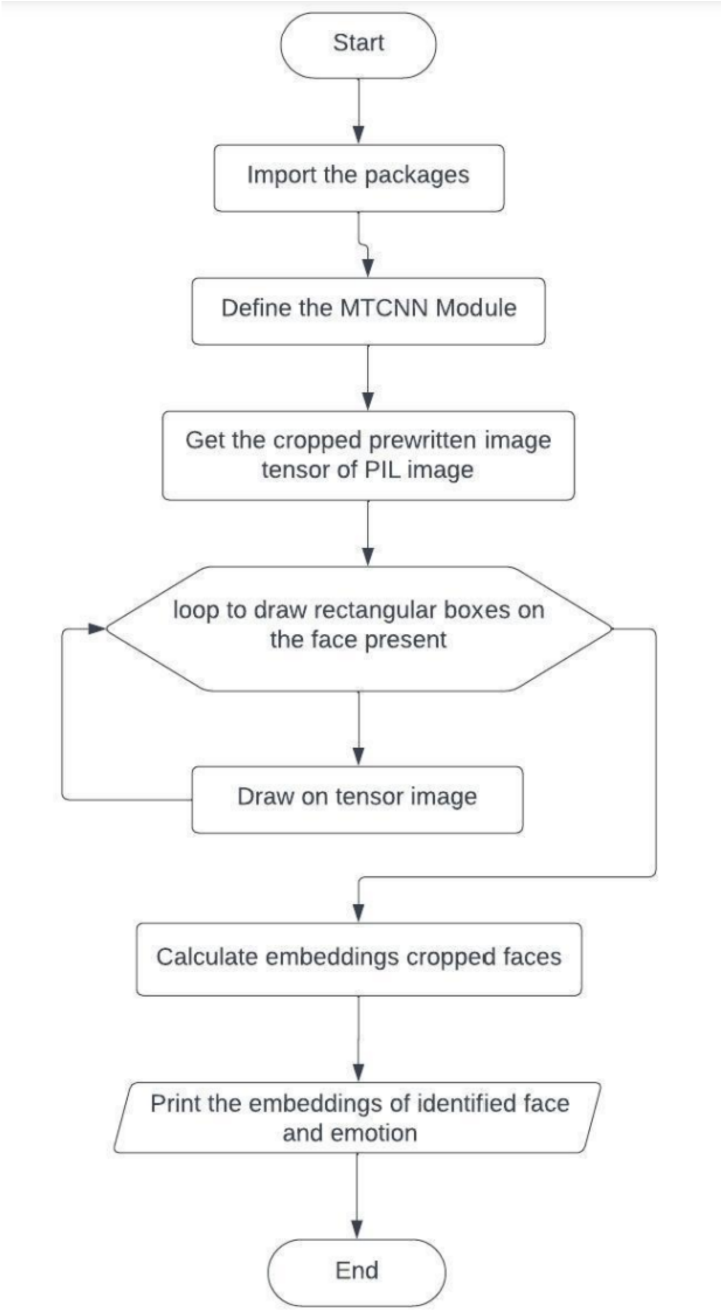

Figure 3: MTCNN based facial recognition

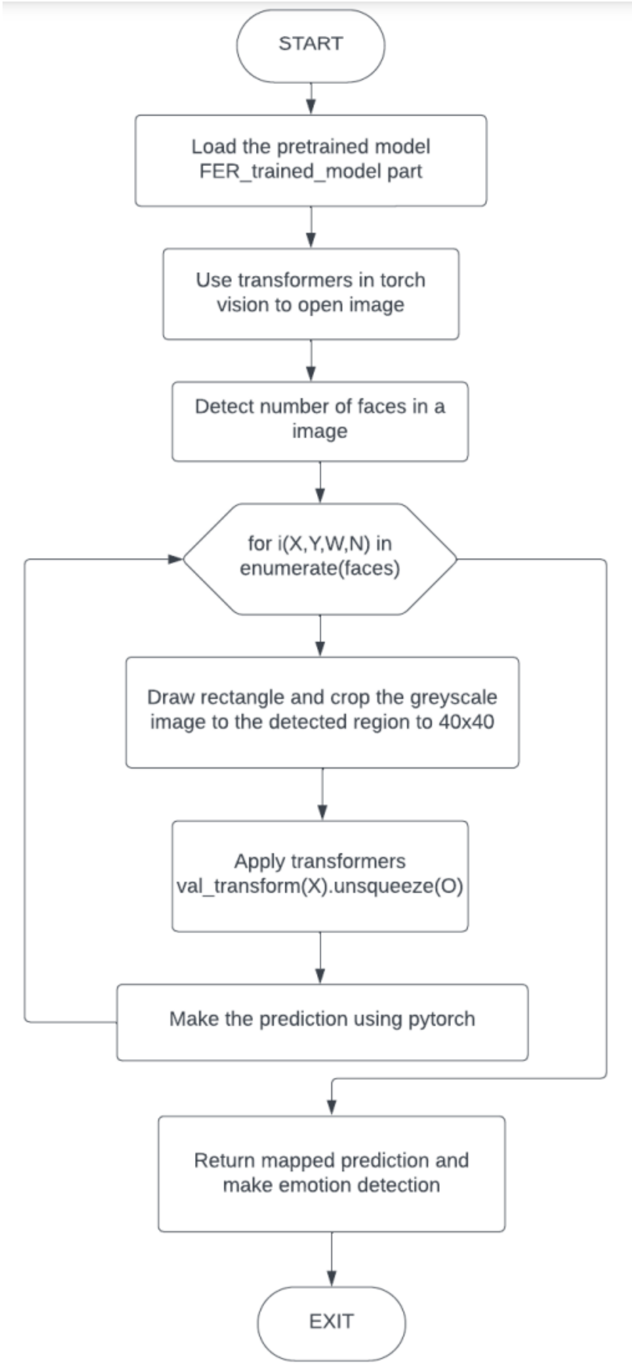

Figure 4: Facenet based emotion recognition

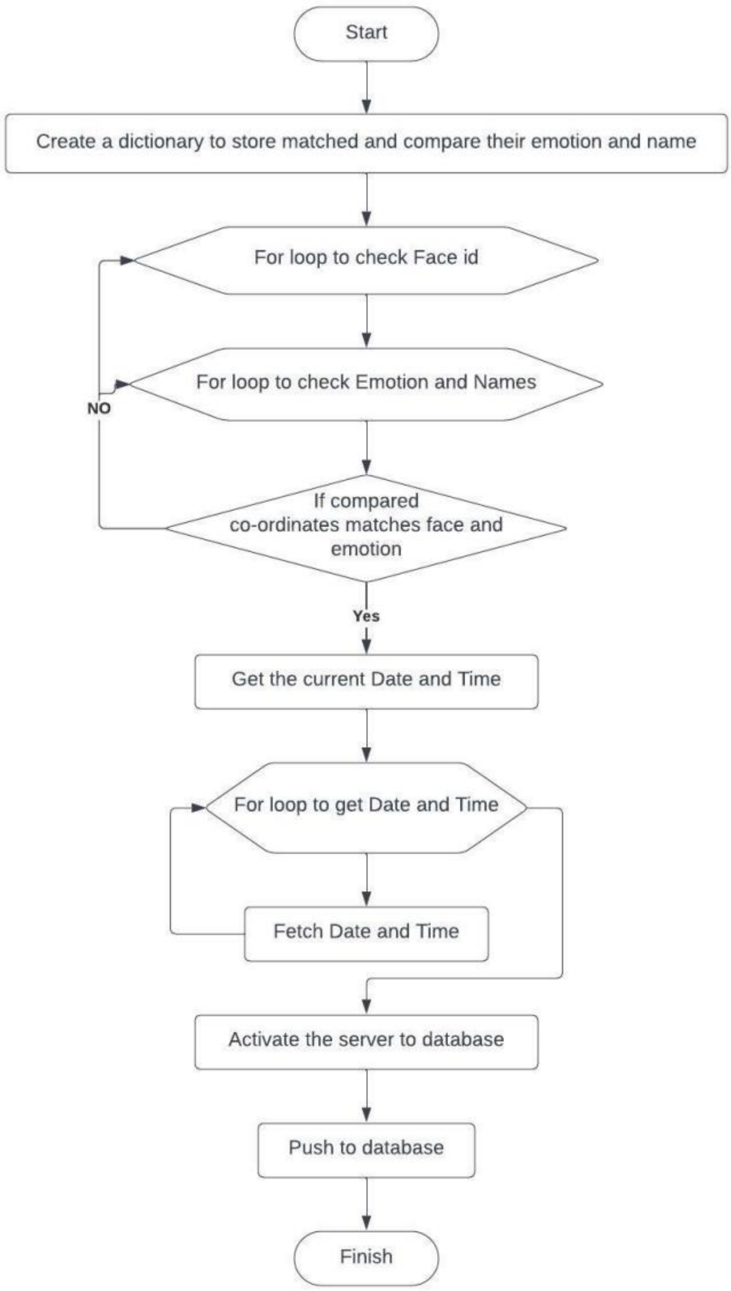

Figure 5: Matching of facing and emotion with database

4. Results and Analysis

Training Dataset : Training datasets used in AI for face and emotion detection serve as the foundation for teaching machine learning models to recognize facial features, expressions, and emotions accurately. Here's an explanation of how these datasets are utilized :

1. Facial Recognition Training Datasets: These datasets consist of a vast collection of facial images captured under different conditions, environments, and angles. They may include various Aspects to ensure the model's robustness and generalization.

International Journal for Multidisciplinary Research (IJFMR)

E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

2. Emotion Detection Training Datasets: Datasets for emotion detection contain images labeled with specific emotions (like happiness, sadness, anger, surprise, etc.). Each image is associated with a corresponding emotion label, providing ground truth for training the model.
3. Training Process : Before training, the datasets undergo preprocessing steps like normalization, resizing, and sometimes augmentation to enhance the model's ability to generalize across different scenarios.

Some training datasets are depicted in Figure 6.

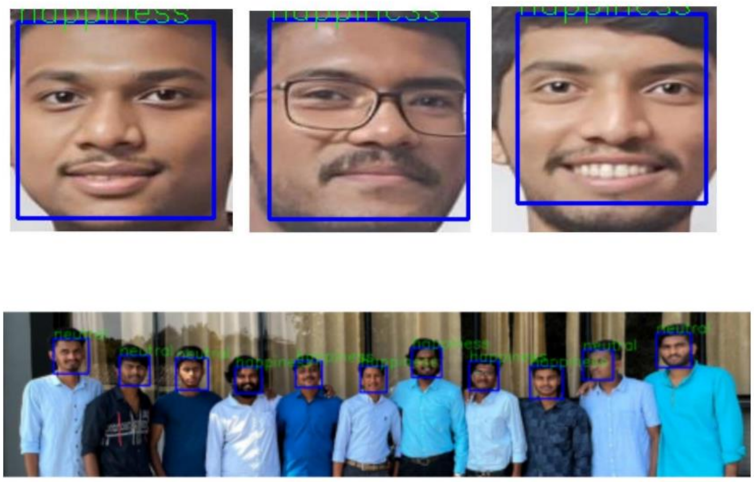

Figure 6: Training datasets

In this AI based emotion detection, the output includes the predicted emotions based on the input data, such as images, videos, or text. The output may provide a confidence score for each detected emotion, indicating the model's level of certainty. Additionally, the backend might generate statistical insights, such as the overall emotional tone or distribution of emotions within the analyzed content. Depending on the project's goals, the backend output could be integrated into various applications, such as emotion-aware interfaces, sentiment analysis tools, or emotion tracking systems. The screenshots are depicted in Figure 7.


```
Embedding: tensor([[ 0.0419,  0.0350,  0.0092,  ..., -0.0474, -0.0402,  0.0416],
        [ 0.0765, -0.0172, -0.0514,  ..., -0.0238, -0.0684,  0.0108],
        [ 0.0367,  0.0233, -0.0239,  ..., -0.0558, -0.0224,  0.0066]],
       grad_fn=<DivBackward0>)
Embedding Size: torch.Size([3, 512])
1 : {'name': 'H M Satwik', 'coordinates': (452, 343)}
2 : {'name': 'Krutharth E K', 'coordinates': (41, 406)}
3 : {'name': 'Sujan K M', 'coordinates': (716, 373)}
```

```
1 : {'name': 'H M Satwik', 'max emotion': 'happiness', 'execution_date': '2023-09-04', 'execution_time': '09:33:54'}
2 : {'name': 'Krutharth E K', 'max emotion': 'neutral', 'execution_date': '2023-09-04', 'execution_time': '09:33:54'}
3 : {'name': 'Sujan K M', 'max emotion': 'neutral', 'execution_date': '2023-09-04', 'execution_time': '09:33:54'}
```

Figure 7: Working results

The analysis of accuracy for various students' photos together is done and results are tabulated in Table 1. The corresponding graph is plot in Figure 8. It can be observed that an accuracy of more than 97% is obtained from the system and accuracy does not change much with number of students together. This is particularly required in the case of class rooms where students are seated together.

Table 1: Emotion Recognition Accuracy

Number of students together	**Accuracy (in %)**
30	98.7
40	98.4
50	98.2
60	98.2
70	97.9
80	97.8

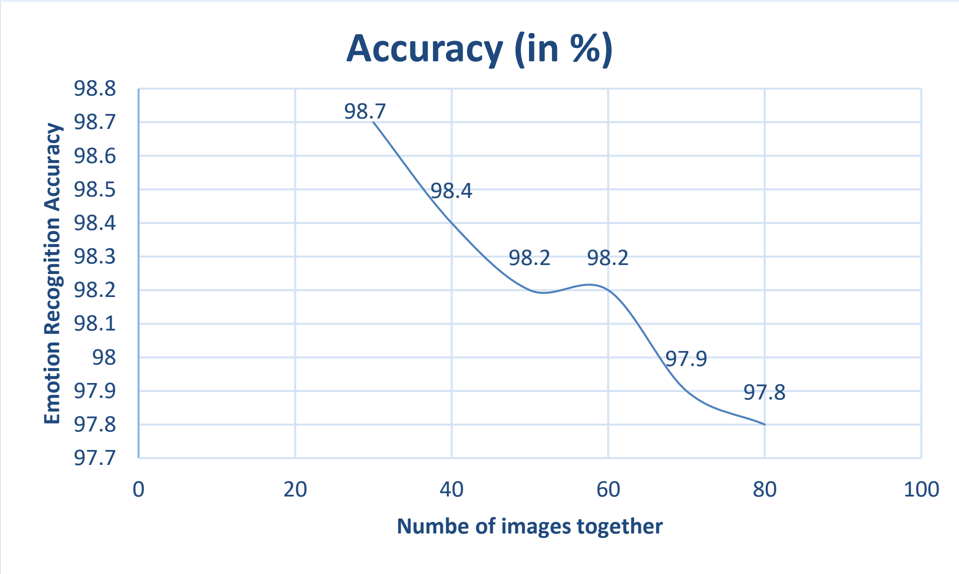

Figure 8: Plot of Emotion Recognition performance

5. Conclusion and Future Scope

In conclusion, this AI emotion detection project represents a significant stride in leveraging technology to understand and interpret human emotions. Through the integration of advanced algorithms and deep learning techniques, our system successfully analyzes input data, be it images, videos, or text, providing valuable insights into the emotional content. This project underscores the potential for AI to enhance human-computer interactions, opening avenues for emotion-aware applications and contributing to the evolving landscape of artificial intelligence. Looking ahead, the application and demand for AI emotion detection systems are poised for exponential growth. As technology continues to intertwine with daily life, the ability to accurately perceive and respond to human emotions becomes crucial. This project's implications extend to diverse sectors, including Human-Computer Interaction, Mental Health Monitoring, Customer Service and Experience, Education: Facilitating adaptive learning environments based on student emotional cues and Healthcare: Contributing to patient care by understanding emotional states. In the coming days, we anticipate an increased integration of AI emotion detection in various industries, from e-commerce to healthcare, enhancing user experiences and fostering more empathetic digital environments.

6. References

1. Bo-Kyeong Kim, Jihyeon Roh, Suh-Yeon Dong, and Soo-Young Lee. Hierarchical committee of deep convolutional neural networks for robust facial expression recognition. Journal on Multimodal User Interfaces, pp. 1–17, 2016
2. Zhiding Yu and Cha Zhang. Image based static facial expression recognition with multiple deep network learning. In Proceedings of the 2015 ACM on International Conference on Multimodal Interaction, pp. 435–442. ACM, 2015.
3. Bo-Kyeong Kim, Jihyeon Roh, Suh-Yeon Dong, and Soo-Young Lee. Hierarchical committee of deep convolutional neural networks for robust facial expression recognition. Journal on Multimodal User Interfaces, pp. 1–17, 2016.
4. L. A. Jeni, J. M. Girard, J. F. Cohn, and F. De La Torre. Continuous au intensity estimation using localized, sparse facial feature space. In Automatic Face and Gesture Recognition (FG), 2013 10th IEEE International Conference and Workshops on, pp. 1–7, April 2013.
5. F. B. Ilyasu, C. E. Erdem, A. A. Ahmad and Y. A. Badamasi, "Facial Age Estimation Using Geometric, Local Phase Quantization and Pyramid Histogram of Oriented Gradient Features," 2021 1st International Conference on Multidisciplinary Engineering and Applied Science (ICMEAS), Abuja, Nigeria, 2021, pp. 1-5.
6. Yao, L., Wan, Y., Ni, H. et al. Action unit classification for facial expression recognition using active learning and SVM. Multimed Tools Appl 80, pp. 24287–24301 (2021).
7. Caifeng Shan, Shaogang Gong, Peter W. McOwan, Facial expression recognition based on Local Binary Patterns: A comprehensive study,Image and Vision Computing,Volume 27, Issue 6, pp. 803-816, 2009.
8. Bennett, Casey & Sabanovic, S.. (2014). Deriving Minimal Features for Human-Like Facial expressions in Robotic Faces. International Journal of Social Robotics. pp. 367-381.