A PROJECT REPORT ON

STUDENT CONCENTRATION ANALYSER

BY

Amit Ghogre	**Student ID. 8833038**
Aashi Aggarwal	**Student ID. 8920299**
Akash Chandrasekar	**Student ID. 8947100**

UNDER THE GUIDANCE OF

Prof. Islam Mahmud

SUBMITTED TO

CONESTOGA

Connect Life and Learning

Conestoga College
299 Doon Valley Dr. Kitchner

2022– 2023

Abstract

In the digital age of education, online learning has become increasingly prevalent, bringing with it unique challenges in maintaining student engagement and focus. Our project aims to address this by leveraging technology to accurately assess student concentration levels during live online sessions. By analyzing facial expressions, we seek to develop a sophisticated model capable of discerning whether a student is attentive or distracted. Moreover, we propose to enhance the learning experience by integrating adaptive engagement mechanisms, such as timed quizzes for re-engagement and rewards for sustained focus. Through this endeavor, we aim to create a more dynamic and effective online learning environment for students worldwide. This report presents a comprehensive study of learner's facial emotion recognition using the FER dataset, providing an overview of its structure, size and different emotion categories. There are several deep learning architectures for facial emotion recognition, the method that is used for this project is Convolutional Neural Network (CNN). The model is implemented and trained on the FER dataset, and their performance are evaluated using various metrics like precision, recall, f1-score.
Keywords: Emotion recognition, CNN, Deep learning, dataset, precision, re- call, f1-score.

Contents

ERA OF ONLINE EDUCATION, INITIATIVE TO TRANSFORM AI EDUCATORS TO DETECT THEIR LEARNER'S CONCENTRATION LEVEL AND EMOTIONS.

Introduction

Emotion recognition refers to the ability to identify and understand human emotions based on various cues, such as facial expressions, voice tone, body language, and physiological signals. Facial emotion recognition also known as facial expression recognition, is a significant area of research within computer vision and affective computing. Student Concentration Analyzer specifically focuses on analyzing and interpreting emotions and concentration level of the learner through facial expressions. It involves detecting and interpreting different facial cues, such as changes in muscle movements, eyebrow position, eye gaze, mouth shape, and other facial features. It uses the automated recognition and categorization of human emotions based on facial expressions seen in pictures or videos. In order to cultivate a more dynamic and effective online learning environment for learners worldwide, this technology is essential for enabling machines to understand and react to learner's emotions [1].
Online Education is a fast-evolving field with the center focus of online classes and our project's concentration level analyzer focuses on computer vision, machine learning, and affective computing. It could be used in numerous ways in fields like marketing, human-robot interaction, virtual reality, and healthcare, among others, emotion recognition from facial expressions has drawn a lot of interest.
Over last decade, advancements in deep learning algorithms and the increase in availability of large-scale data-sets has increased significantly in facial emotion recognition. Convolutional neural networks (CNNs), have showed impressive capabilities in extracting distinguishing features from facial images, which results in improved accuracy of emotion classification [3].
The dataset of this project is FER, it is a widely used dataset for emotion recognition, mainly used for training and evaluating deep learning models, especially Convolutional neural networks (CNNs). This dataset is an addition to the original FER dataset [4] that has been enhanced with more samples and better annotations to make it better suited for modern deep learning techniques [5]. The dataset contains a diverse collection of facial images, captured from various individuals, portraying seven different emotions: neutral, happy, sad, anger, disgust, fearful and surprised [6]. The dataset contains total 65520 images belonging to 7 classes. Each image is gray-scaled and resized to a resolution of 48x48 pixels for efficient processing.
In this report, the aim is to develop a website in which teacher can detect emotions and concentration level of the learner in a LIVE class environment from their live cameras. The project thoroughly investigates the methodologies and advancements in facial emotion recognition, with a focus on the use of the FER dataset. Modern deep learning approaches will be examined and contrasted, along with data pre-processing methods. The ethical issues of privacy, bias, and the responsible use of emotion recognition systems will also be covered.

Motivation

Online Education is an evolving market. Concentration analyzer using both emotion recognition and facial expression recognition have applications in various fields, including psychology, human-computer interaction, market research, entertainment, and healthcare. Facial emotion recognition improves human-computer interaction by adjusting responses based on users' emotional states, enabling personalized and intuitive interactions. Utilizing the facial emotion recognition model to analyze learners' concentration level in LIVE classes aids educators in understanding student engagement and responses, enabling effective teaching methods and learning environments. This could be utilized in various field:

- IT aids in diagnosing and treating mental health conditions, monitoring emotional well-being, and early detection of mood disorders, improving mental health support.
- IT helps businesses understand customer reactions to products, advertisements, and services, enabling tailored marketing strategies and better customer engagement.
- In robotics enhances human interactions, making them more natural and intuitive in caregiving, companionship, and customer service roles.
- IT aids security systems by analyzing real-time emotional states, detecting potential threats and suspicious behavior.
- IT improves entertainment industry by enhancing video games, virtual reality, and animated characters' responsiveness to players' emotions.
- IT enhances autonomous driving by monitoring drivers' emotional states, enabling vehicle systems to respond to agitation or distraction for safety.
- This technology enhances accessibility for disabled individuals by enabling non- verbal communication through facial expressions and eye movements.

The primary objective is to create a comprehensive platform capable of accurately detecting and interpreting learners' emotional states from visual data they provide. To achieve this, a sophisticated web application has been developed using streamlit, leveraging advanced facial emotion recognition technology. By analyzing the intricate nuances of learners' facial expressions captured through visual input, this application aims to discern a wide spectrum of emotions, providing a deeper understanding of their feelings and reactions. This endeavor not only strives to enhance human-computer interaction by tailoring responses to learner's emotional cues but also holds potential in diverse domains such as mental health assessment, personalized marketing, and immersive virtual experiences.

This project aims at engaging learner in a virtual Education environment by participating in quizzes prompted based on their emotion and concentration level.

Literature review

According to various studies, nonverbal components convey two-thirds of human communication and verbal components one-third, with people generally inferring the emotional states of others, such as joy, sadness, and anger, using their facial expressions and vocal tones. [7], [8] In a study by [9], they proposed an approach to learn identity and emotion jointly. They used deep convolutional neural networks (CNNs) to increase the sensitivity of facial expressions and their better recognition. This approach can be utilized in designing concentration level analyzer for learners in an effective manner. From their study, they concluded that emotions and identifications are different and separate features, which are being used by CNNs for Facial expression recognition (FER). They deduced a statement that expression and identity can be both used to deep learned tandem facial expression (TFE) feature and can be used to form a new model. Experimental results from this study presented the fact that this model approach achieved 84.2 percentage accuracy on FER+ Dataset. Identity and emotion combined model were experimented using different methods including ResNet18, ResNet18+FC, and TFE Joint Learning. They gave an accuracy of 83, 83 and 84 percentage respectively as seen in table 1. From different studies, different models were studied for Old FER2013, and New FER+ database models. Results of different models on old FER and FER+ are given in table below.

Table 1: Accuracy of Previous Systems

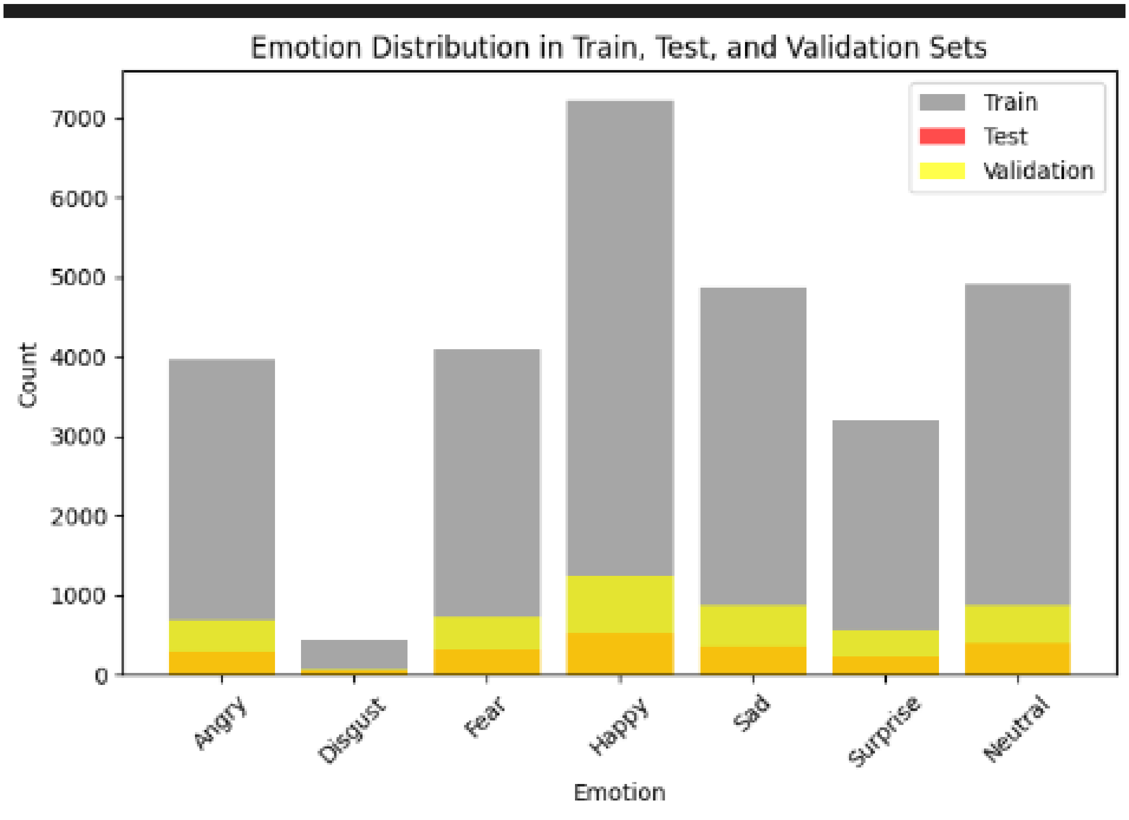
Dataset	Methods	Accuracy
FER 2013	DLSVM-L2[8]	71
	Zhou et al.[8]	69
	Maxim Milakov[8]	68
	Radu+Marius+Cristi	67
New FER+	This Study	71
	VGG13(MV)[8]	83
	TFE-JL[8]	84

In another study [10] Based on the features of the human face, the database was able to distinguish five human emotions—happiness, anger, grief, surprise, and neutral—with an average recognition accuracy of 81.6 percentage. [11] In another study, Eigen spaces and a dimensionality reduction technique were used to identify the fundamental emotions of sadness, anger, contempt, fear, happiness, and surprise in people's facial expressions. [12] The system that was created had an accuracy rate for recognition of 83 percent. A different article's research extracts local face features using principal component analysis, classifies facial expressions using an artificial neural network, and uses a unique method dubbed Canny. According to research, the method average level of facial emotion categorization accuracy is 85.7 percent on FER+.

Dataset

In 2013 a dataset was created for facial emotion recognition named as FER-2013[4]. This dataset was small and had numerous flaws. The FER13+ dataset was introduced in 2016 [13]in this paper, the authors describe the process of creating the FER+ dataset, which involved refining the emotion labels present in the FER 2013 dataset and obtaining probability distributions to capture the uncertainty associated with each label. The authors conducted a study in which human annotators adjusted the emotion labels of the images, resulting in more accurate annotations.
The project uses FER 2013+ [15] as its dataset. In total 65520 images are in the dataset. Data pre-processing techniques are used for increasing the accuracy of the model [16]. Data pre-processing involves resizing the image to 48x48 pixels so that when the images from the dataset is taken as input it reduces memory usage and increases the speed of training [17]. Gray scaling is also done on the dataset which results in images having single channel which further leads to faster training and gray-scaling also increases memory efficiency [18].
The dataset is divided into two parts training set and validation set, training set. The dataset includes 7 different emotion classes: Anger, Disgust, Fear, Happy, Sad, Surprise, Neutral. It can be seen that happiness had the highest number of images followed by neutral, sad, fear, angry, surprise and lastly disgust having least number of images. This difference in number of images as input will also result in variance in emotion accuracies.

Figure 1: Data distribution

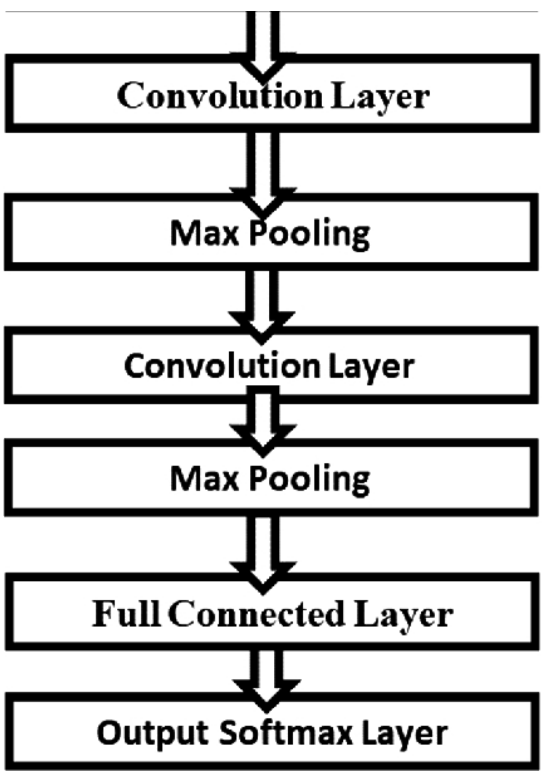


Project Design

Convolutional neural network (CNN) is used as the model for Student Concentration Analyzer in this project. CNNs has spatial hierarchical feature means it automatically detects patterns on the face provided if it within the screen [19]. CNNs provides Translation Invariance meaning emotion can be recognized irrespective of their location in the image by using shared weights in convolutional layers [20]. CNNs also provides Feature Hierarchies in which the model learns from the corners/edges first and progressively learn higher-level features in deeper layers. This helps in enhancing the model's ability to discriminate emotions effectively [22].

The CNN model architecture, as exemplified by the provided Deep Emotion class, consists of two convolutional layers (conv1 and conv2). Each convolutional layer is followed by Rectified Linear Unit (ReLU) activation functions, fostering non-linearity in the model's feature extraction process. MaxPooling layers (pool1 and pool2) are strategically employed after each convolutional layer to down sample spatial dimensions, effectively reducing computational complexity and averting overfitting. Dropout layers (self.dropout) are judiciously added following the MaxPooling operations, serving to randomly deactivate neurons during training, thereby bolstering the model's resilience against overfitting. Subsequently, the model transitions into three fully connected layers (fc1, fc2, and fc3), each accompanied by Batch Normalization, ReLU activation, and Dropout. These fully connected layers play a pivotal role in discerning higher-level patterns from the extracted features, essential for accurate emotion classification. The output of the last MaxPooling layer is meticulously flattened into a 1D vector, facilitating seamless integration with the fully connected layers. Finally, the model culminates with a dense layer (fc3) equipped with a SoftMax activation function, comprising seven neurons, one for each emotion class. This final layer adeptly produces a probability distribution over the emotion classes for each input image, encapsulating the model's predictive capabilities.

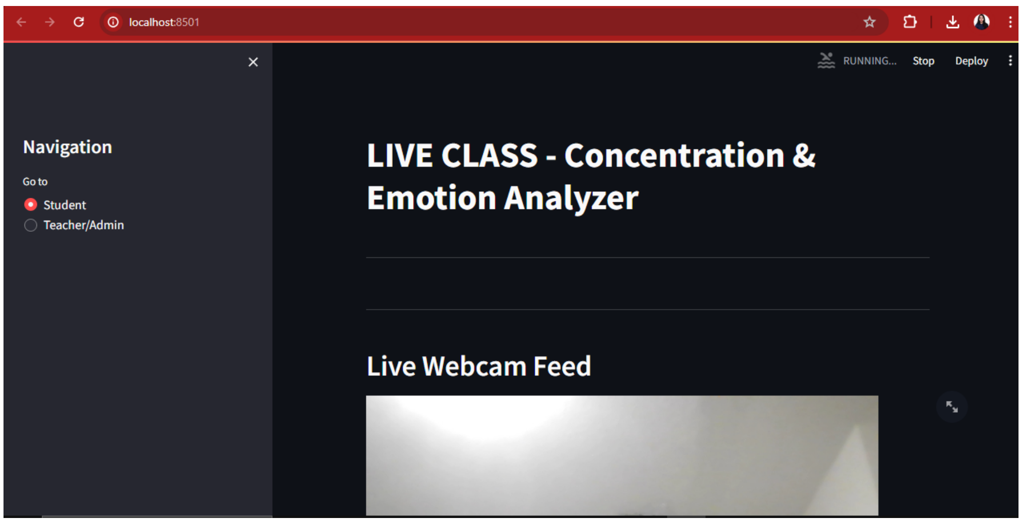
Figure 2: CNN model architecture



The CNN model's final layer is a dense layer featuring a SoftMax activation function, comprising seven neurons, each corresponding to an emotion class. This layer efficiently generates a probability distribution across the classes for each input image, facilitating accurate emotion classification. The model leverages the Adam optimizer, a variant of stochastic gradient descent (SGD), which dynamically adjusts the learning rate for each parameter during training. The specified learning rate is set to 0.0001, optimizing the model's training process. To effectively address the multi-class classification problem inherent in emotion recognition, the chosen loss function is categorical cross-entropy, ensuring optimal model performance.

Furthermore, to display the emotion recognition system a web application using Streamlit is developed. Streamlit is a user-friendly Python library that allows us to build interactive web applications easily. Streamlit supports python libraries like OpenCV(cv2), TensorFlow and WebRTC.
Below fig. 3, represents the user interface of the system through which the user can operate to detect facial emotions either from image/videos or live cameras.

Figure 3: User Interface



Experiments

The code is experimented in two phases to increase the accuracy of the model.

Phase 1:

In phase 1, the model undergoes on all 7 facial features batches combined together.

The Python function train_model serves as a pivotal component in the development of a Convolutional Neural Network (CNN)-based facial recognition model, specifically tailored to recognize seven distinct facial emotion features. This function orchestrates the training process over 30 epochs, overseeing both the training and validation phases essential for model optimization and performance evaluation.
During each epoch, the model undergoes rigorous training on batches of all 7 facial feature images, meticulously extracted from the dataset and fed through the CNN architecture. Within the training phase, the model's parameters are iteratively adjusted to minimize the training loss, achieved through forward and backward propagation of gradients facilitated by the Adam optimizer. Concurrently, the function meticulously monitors training accuracy, ensuring the model's ability to accurately predict facial emotion labels.
Following the training phase, the model transitions into a critical evaluation stage, wherein its performance is rigorously assessed on a separate validation dataset. This phase is crucial for gauging the model's generalization capabilities and guarding against overfitting. Through meticulous calculation of validation loss and accuracy, the function offers insights into the model's performance beyond the training data, enabling informed decisions regarding model refinement and optimization.
To mitigate the risk of overfitting and promote model generalization, the function incorporates an early stopping mechanism, halting training if the validation loss fails to improve over a specified number of epochs. This strategic decision ensures that the model achieves optimal performance while mitigating the risk of overfitting to the training data.
Furthermore, the function provides invaluable visualization capabilities, offering a comprehensive overview of the training and validation losses over epochs through intuitive Matplotlib plots. These visualizations empower stakeholders with actionable insights into the model's training dynamics, facilitating informed decisions regarding model refinement and optimization strategies.

Phase 2:

In phase 2, the model undergoes each facial feature batches individually.

The model encapsulates a novel approach to facial recognition model training by evaluating each facial feature individually. This granular analysis provides a deeper understanding of the model's performance nuances, enabling targeted optimizations tailored to specific facial characteristics. By monitoring the accuracy of predictions for each of the seven facial features, the function facilitates fine-grained adjustments to the model architecture and training strategies, optimizing performance across diverse facial expressions and attributes.
Incorporating early stopping mechanisms into the training process underscores the function's commitment to achieving optimal model performance while mitigating the risk of overfitting. By dynamically halting training when validation loss fails to improve over consecutive epochs, the function ensures that the model generalizes well to unseen data, a critical requirement for real-world deployment scenarios. This strategic decision not only safeguards against overfitting but also expedites the model development lifecycle by

preventing unnecessary iterations and resource expenditures.
Moreover, the function's visualization capabilities offer stakeholders a comprehensive overview of the training and validation dynamics, empowering informed decision-making and facilitating collaboration among multidisciplinary teams. By visualizing the temporal evolution of training and validation losses, the function enables stakeholders to identify trends, diagnose performance bottlenecks, and prioritize optimization efforts effectively.
In conclusion, the train_model function represents a pivotal milestone in the development of a robust and scalable CNN-based facial recognition model. By embracing a granular approach to performance evaluation, incorporating early stopping mechanisms, and providing insightful visualizations, the function equips stakeholders with the tools and insights necessary to navigate the complexities of facial recognition model development effectively. As the project progresses into subsequent phases, the function's contributions will continue to underpin the model's evolution, driving advancements in accuracy, scalability, and real-world applicability.

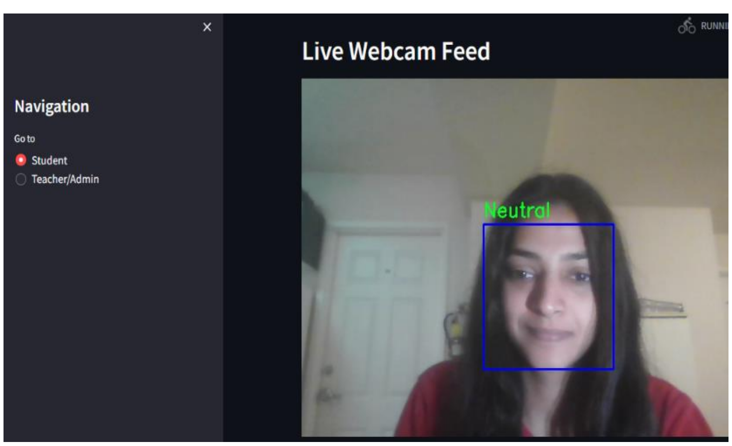
Based on these experiments, we observed that Phase 2, turns out to be an efficient model. Even though its accuracy is low comparatively but phase 1 shows overfitting. We will discuss its evaluation parameters in Evaluation section of the report.

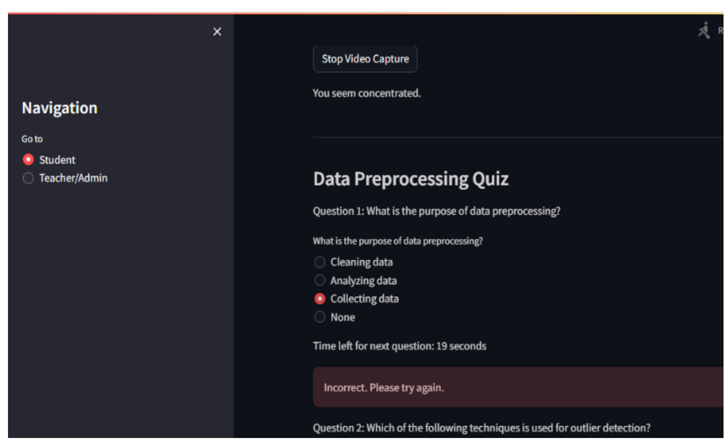
Results

Fig. 4 the results of the system. These images are taken from within the dataset. Hence, the accuracy of the emotions is higher. All the emotions like Happy, Fear, Sad, Neutral, Surprise were correctly detected.

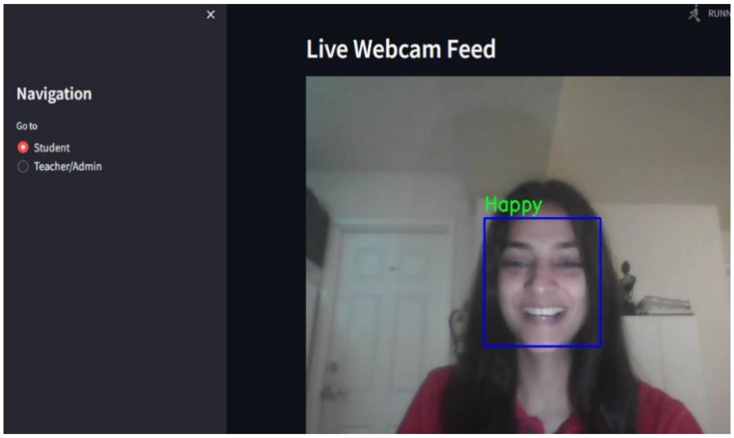
Face Detection Fig 4:

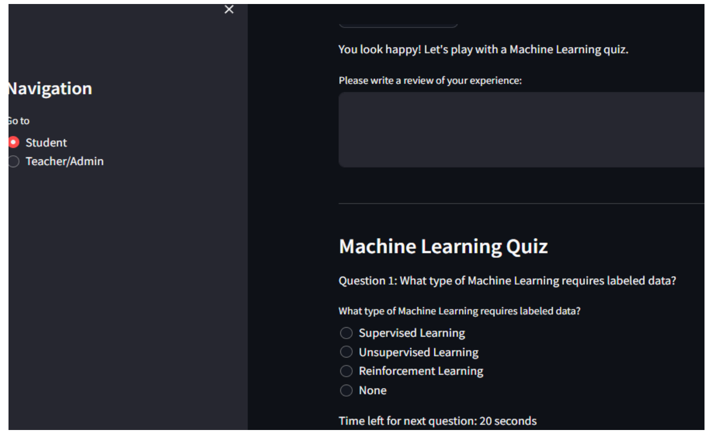
1. NEUTRAL Fig 4.1



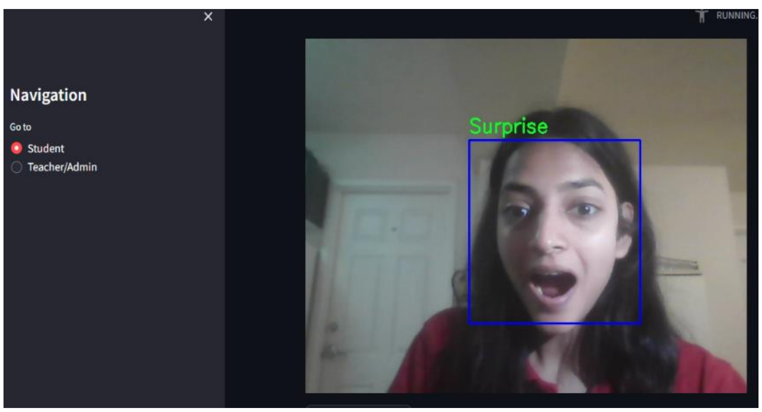


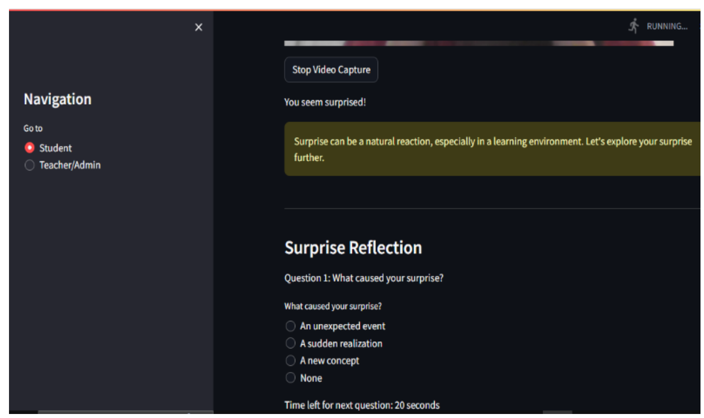
2. HAPPY Fig 4.2



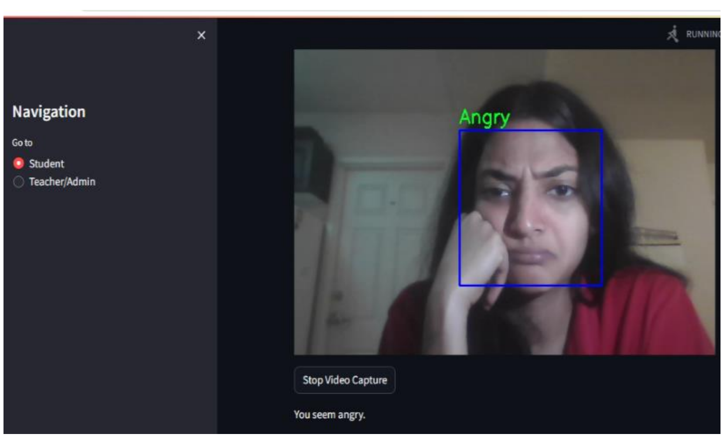


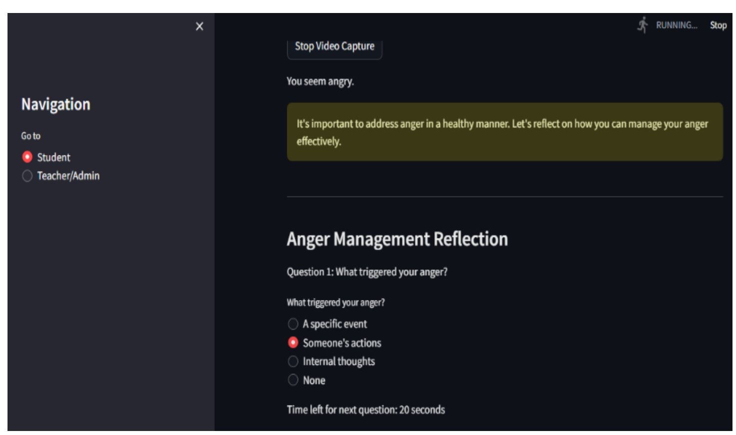
3. SURPRISE Fig 4.3



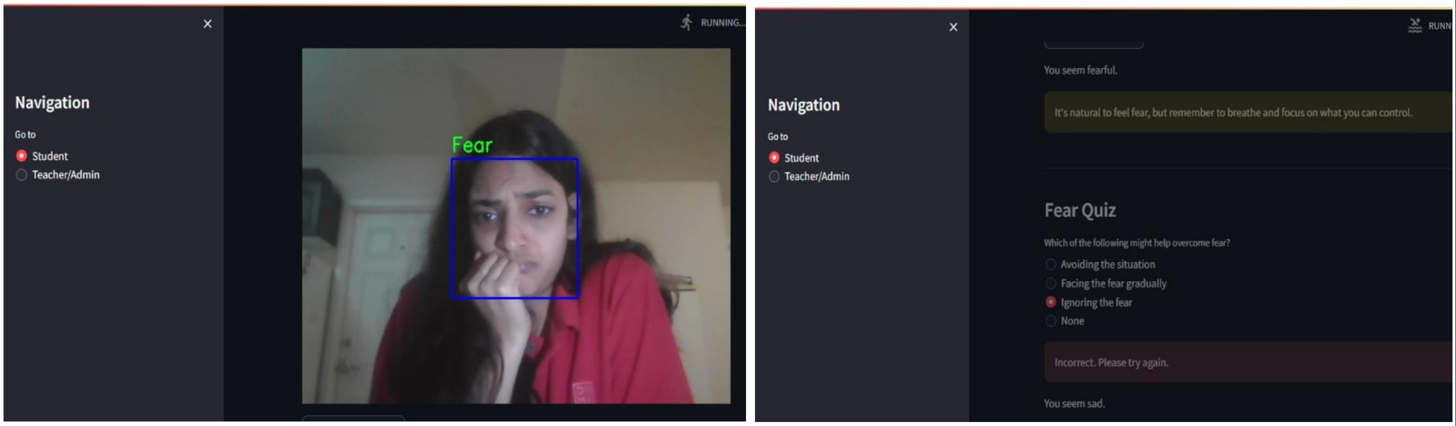


4. ANGRY Fig 4.4





5. FEAR Fig 4.5



6. SAD Fig 6.6

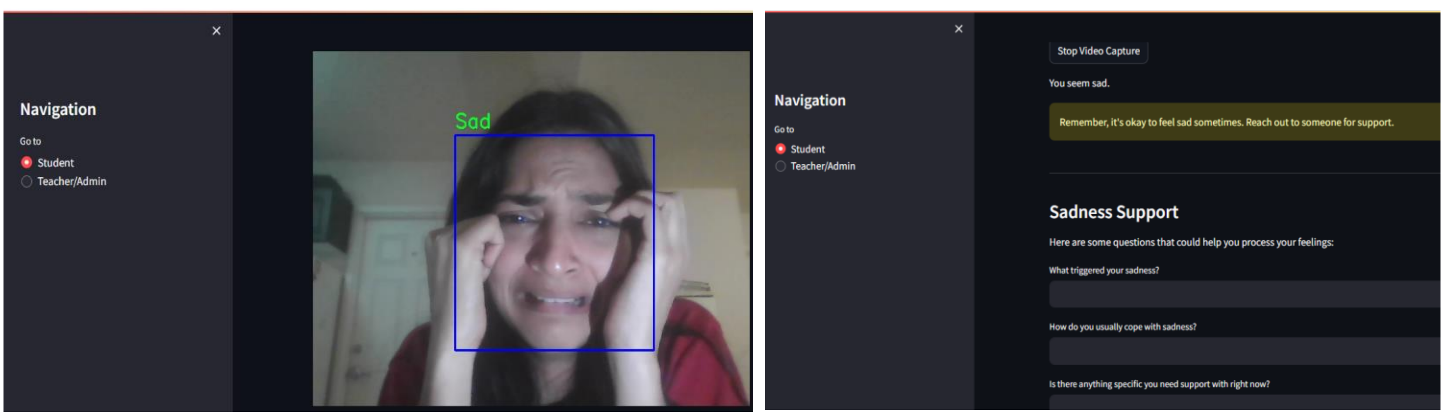


Fig 5 is image with low lighting. The emotion is detected accurately even though half face in not visible properly, the system catches the parameters such as eyebrows, nose, eyes, mouth shape and accurately detects the emotion.

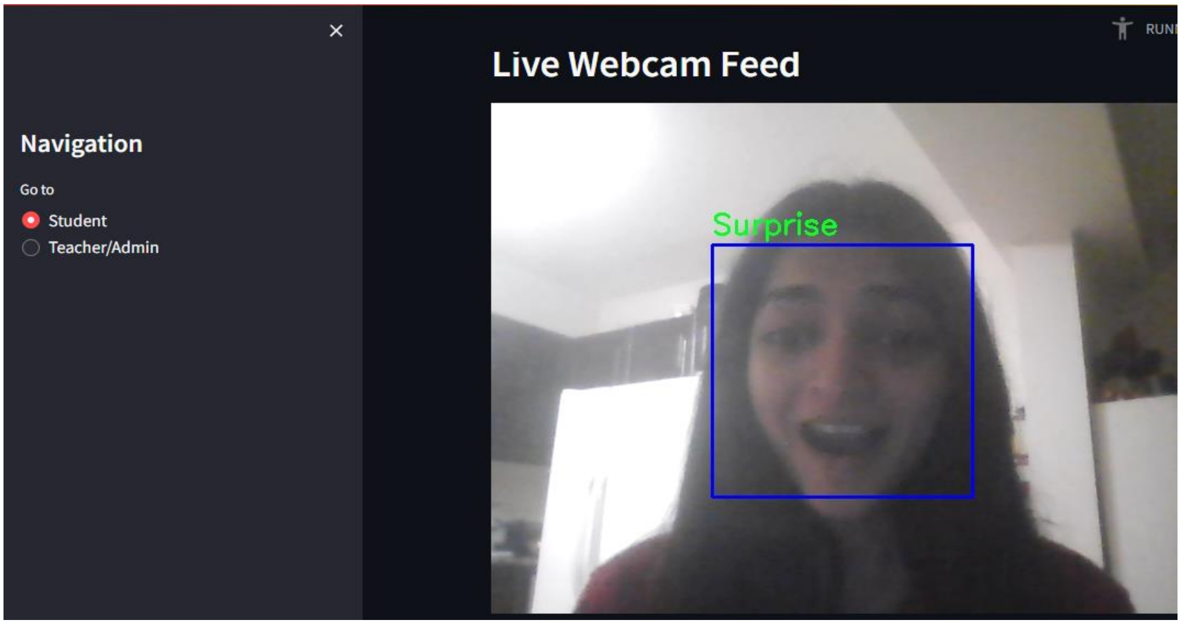


Fig 5 Low Lightning

The Fig. 6 is a photo of covering mouth, even though the mouth is covered the system was able to detect the emotion accurately using the other features such as raised eyebrows, eyes.

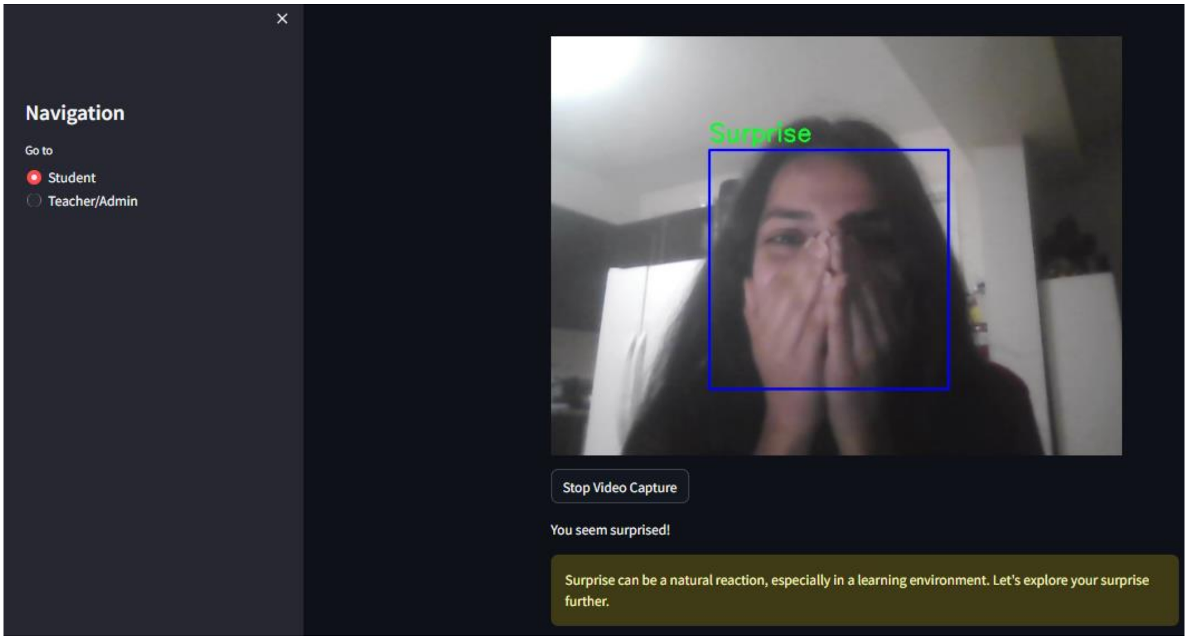


Fig 6 Face Covered

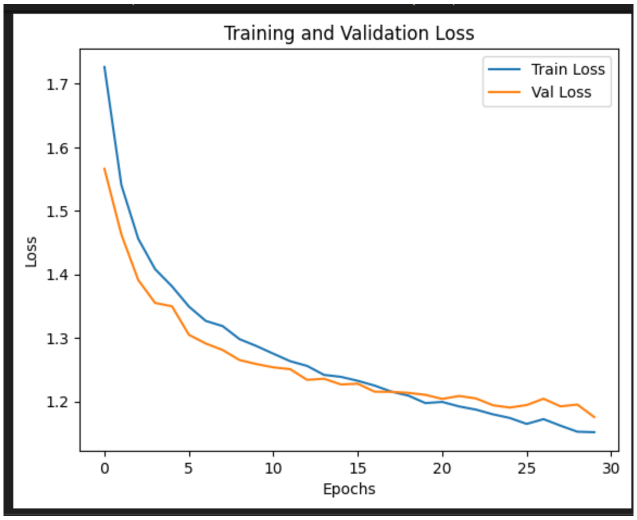
Evaluation

Model evaluation is an essential component of any machine learning project, and properly reporting the evaluation results is critical for effectively communicating the model's performance. The evaluation used to assess the model's performance are Accuracy, Precision, Recall, F1-score and Confusion Matrix.

Training and Validation: Accuracy and Loss

Training and Validation Loss: In deep learning, training loss and validation loss are frequently used metrics. They depict the difference between the model's predicted output and the actual target output [25].

PHASE 1:



P H A S E 2 :

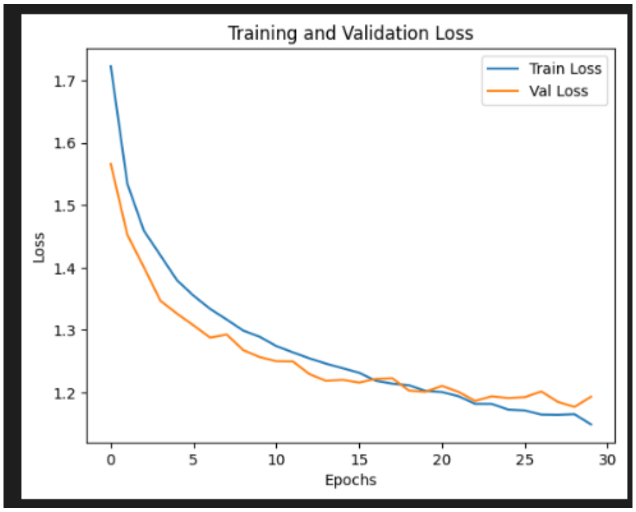


Fig 7: **Training and Validation Loss**

This clearly shows, Phase 2 has better performance as it seems comparatively smoother with less spikes stating less overfitting.

Confusion Matrix

The effectiveness of a classifier on a multi-class classification task is represented by a confusion matrix fig. 5. The true classes are represented by each row in the matrix, while the predicted classes are shown by each column. The top-left to bottom-right diagonal elements of the matrix represent correctly classified samples for each class, whereas the off-diagonal elements represent incorrect classifications [27].

PHASE 1:

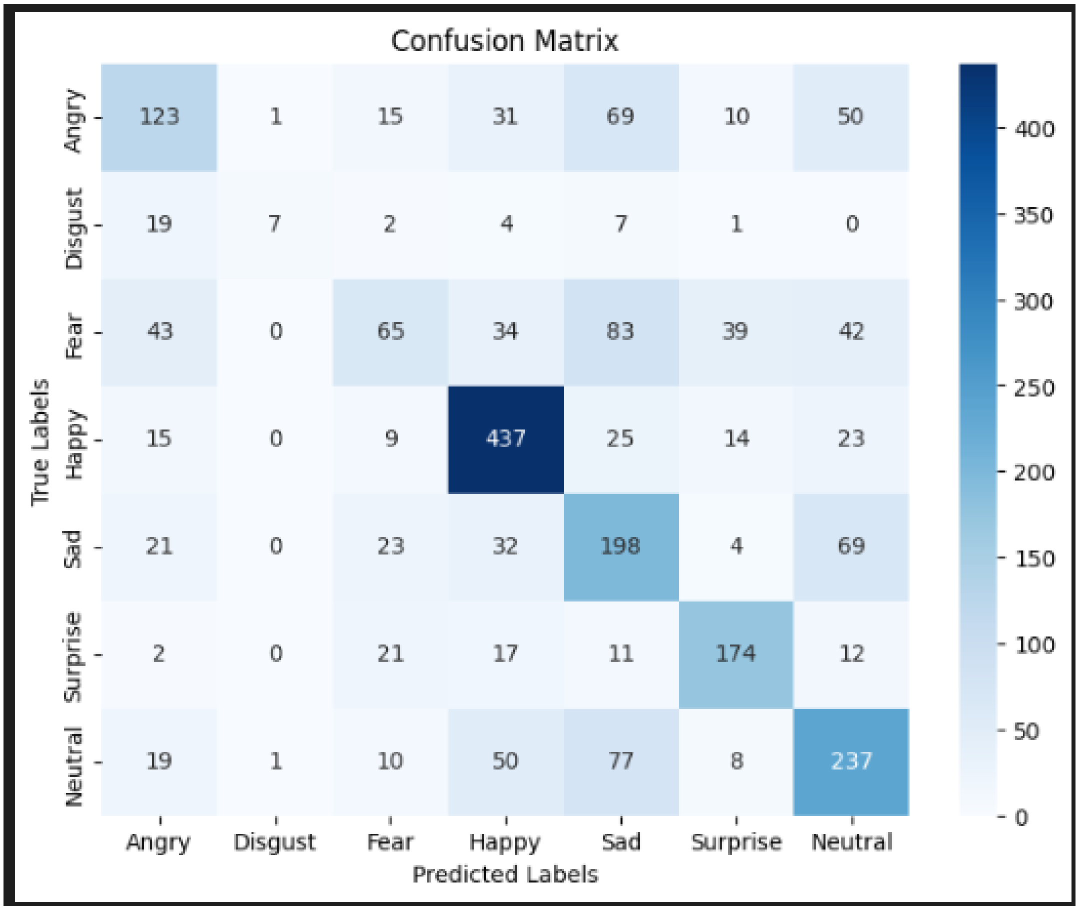


Fig 8 Confusion Matrix – Phase 1

The diagonal elements represent correct predictions by the model. The largest numbers on the diagonal are for 'Happy' (437) and 'Neutral' (237), indicating that the model is most accurate in predicting these emotions.

Least Accurately Predicted , 'Fear', ‘Disgust’ seems to be the emotion that the model struggles with the most, as it has a relatively low number of correct predictions (65) (7) compared to the number of times it was mistaken for other emotions, particularly 'Happy'.

The model seems to perform best when predicting 'Happy' and 'Neutral' emotions and struggles with 'Fear'. The model's ability to distinguish between 'Angry' and 'Neutral' could be improved given the relatively high number of misclassifications.

PHASE 2:

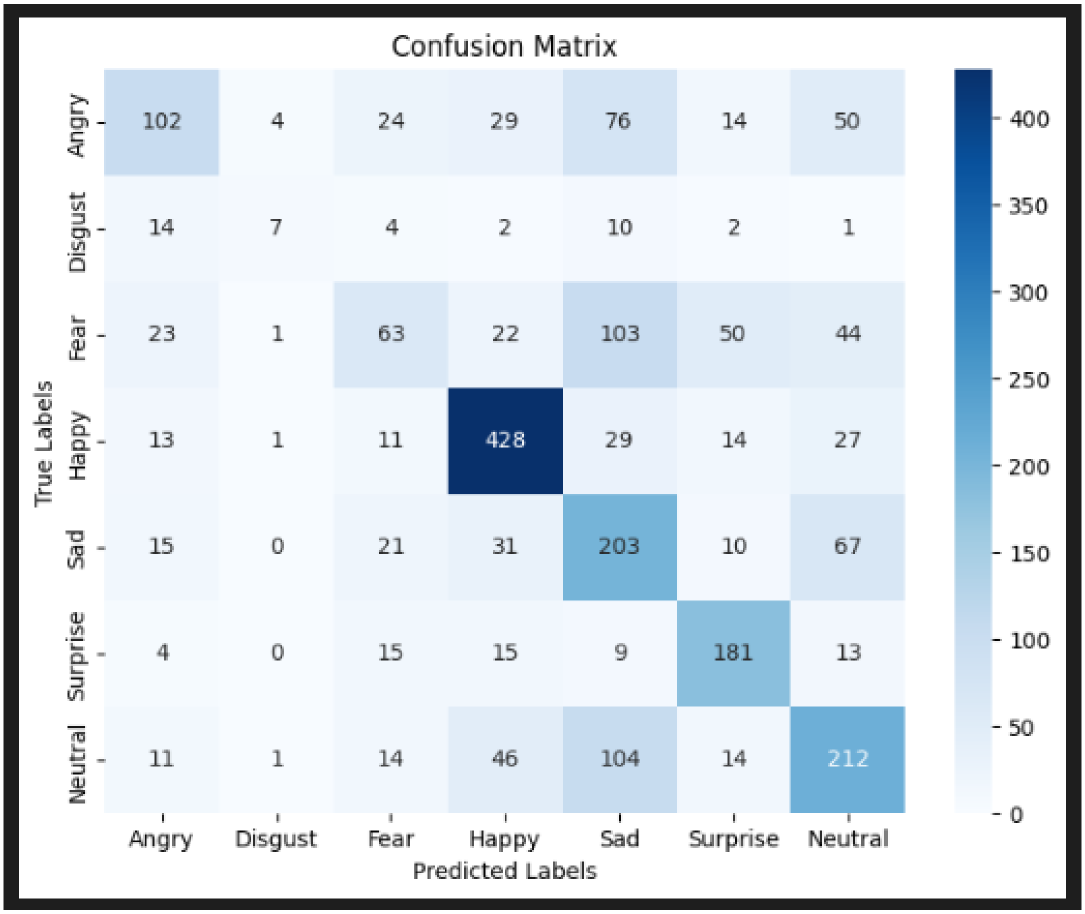


Figure 9: Confusion Matrix – Phase 2

The diagonal elements represent correct predictions by the model. The largest numbers on the diagonal are for 'Happy' (428) and 'Neutral' (212), indicating that the model is most accurate in predicting these emotions.
Least Accurately Predicted: 'Fear' and ‘Disgust’ seems to be the emotion that the model struggles with the most, as it has a relatively low number of correct predictions (63) (7) compared to the number of times it was mistaken for other emotions, particularly 'Happy'.
The model seems to perform best when predicting 'Happy' and 'Neutral' emotions and struggles with 'Fear'. The model's ability to distinguish between 'Angry' and 'Neutral' could be improved given the relatively high number of misclassifications.

Classification Report

The classification report offers a thorough summary of different evaluation metrics for each class in the dataset. After the CNN model has been trained, it is typically used to evaluate how the model performed on a validation set. For each class, the classification report includes the following crucial metrics given in the table 2

Performance Metrics Report:

Emotion	Precision	Recall	F1-Score	Support
Angry	0.51	0.41	0.45	299
Disgust	0.78	0.17	0.29	40
Fear	0.45	0.21	0.29	306
Happy	0.72	0.84	0.77	523
Sad	0.42	0.57	0.48	347
Surprise	0.70	0.73	0.71	237
Neutral	0.55	0.59	0.57	402
Average Weighted	0.57	0.58	0.56	-
Macro	0.59	0.50	0.51	-

Accuracy: 0.58

Figure 10: Performance Metric – Phase 1

Performance Metrics Report:

Emotion	Precision	Recall	F1-Score	Support
Angry	0.56	0.34	0.42	299
Disgust	0.50	0.17	0.26	40
Fear	0.41	0.21	0.28	306
Happy	0.75	0.82	0.78	523
Sad	0.38	0.59	0.46	347
Surprise	0.64	0.76	0.69	237
Neutral	0.51	0.53	0.52	402
Average Weighted	0.55	0.56	0.54	-
Macro	0.54	0.49	0.49	-

Accuracy: 0.56

Figure 11: Performance Metric – Phase 2

Precision: Precision measures the accuracy of positive predictions for each class [28]. For instance, for emotion Angry in phase 1, the model's predictions are 0.51, meaning that 51 percentage of the predicted positive samples for Angry emotion are correct.

Recall: Recall, also known as sensitivity, quantifies the model's ability to correctly identify positive samples for each class [26]. For example, for emotion Happy in phase 1, the model can recall 84 percentage of the actual positive samples in the dataset.

F1-score: The F1-score is the harmonic mean of precision and recall. It provides a balanced measure of a model's performance by considering both precision and recall. Higher F1-scores indicate better overall performance [28]. For example, emotion Surprise in phase 1 has an F1-score of 0.71, which is quite good.

Support: Support represents the number of actual samples in the test set for each class [29]. It shows the distribution of samples across the different classes.

Accuracy: The accuracy is the overall performance metric that measures the ratio of correctly predicted samples to the total number of samples in the test set [26]. In this case phase 1, the model achieved an accuracy of 58 percentage, indicating that it correctly predicted 58 percentage of the samples in the test set.

Macro Avg: The unweighted mean of precision, recall, and F1-score for all classes is calculated by the macro average. It contributes equally to every class, regardless of size [26]. The macro-averaged precision, recall, and F1-score in phase 1 are variant as 59,51,50 which shows misbalancing and overfitting of model. Due to this we are considering Phase 2 to be more efficient model. all close to 56 percentage.

Weighted Avg: The precision, recall, and F1-score across all classes are calculated using a weighted average, where the weights are determined by the support (number of samples) for each class. It takes into account the dataset's class imbalance [28]. The weighted average precision, recall, and F1-score in phase 1 are all close to 56 percentage.

On Keenly observing each parameter for both the phases of model, it is observed that despite having better parameters, phase 1 seems to be overfitting as for some features its very high and its train and validation plot also shows its overfitting.
Considering the phase 2 model proves to be the efficient model to analyses student emotions and concentration accurately.

The plot below shows the accuracy of each emotion, it is seen that happiness had highest accuracy among all other emotions

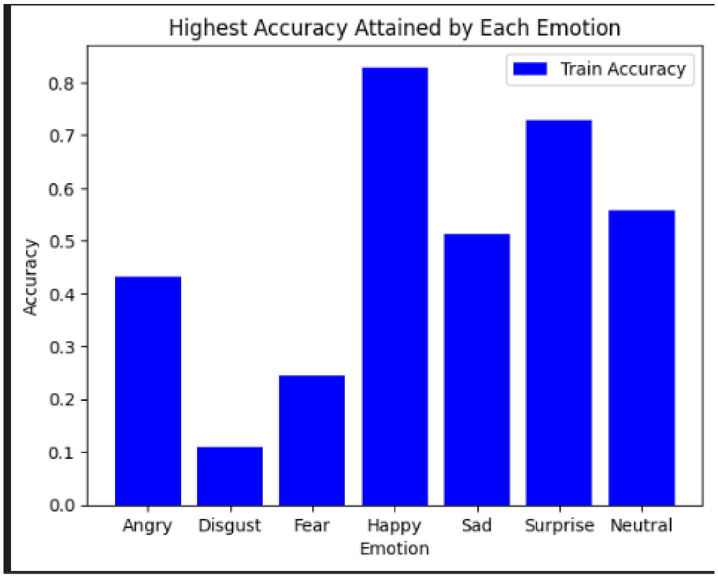


Figure 12: Highest Accuracy Plot

Limitations or Challenges

There are numerous challenges faced by this system. The challenges are listed below:

1. One of the challenges faced was distinguishing emotion between anger and dis- gust. Disgust has the least number of image input as a result its accuracy is least, but also the facial expression of anger and disgust is similar like eyebrows lowering.
2. Lighting on the face during testing can be a major factor it can cause inaccuracies.
3. Detecting emotion for multiple faces at the same time can be done but it can result in delay.
4. Covering face while testing can also result in low accuracy. System gives result on the basis of facial expressions like eyebrows, eyes, mouth shape, nose. If the face is covered, the system will only be able to read from the features available during testing.
5. Privacy concern is also a major challenge in a system where users face is taken as input.

6. Model must be capable of generalizing to novel, unobserved face expressions. It is difficult to achieve this level of generalization because facial expressions vary widely depending on the person, their cultural background, and the situation.
7. Future of Emotion detection technologies can be enhanced by introducing new methodologies in which face recognition can be done in dim-light and also better bifurcations between different emotions.

Conclusions

In conclusion, student emotion analyzer using face emotion detection CNN is a complex technology that has the potential to revolutionize various fields such as healthcare, marketing, and security. By accurately identifying emotions, it can help doctors diagnose mental health disorders, marketers tailor their advertising campaigns, and security personnel detect potential threats. It can be a potential game changer in Ed-tech sector for introducing AI Educators and analyzing learner's emotions during LIVE sessions.

However, there are also significant challenges and ethical implications associated with this technology. Developers must overcome technical difficulties such as lighting and pose variations, while also addressing privacy concerns and potential biases in the data used to train these systems.

Despite these challenges, the potential benefits of facial expression recognition are too great to ignore. As we continue to develop and refine this technology, it is important for us to approach it with thoughtfulness and consideration for its impact on society.

References

1. D'Mello, S. K., Kory, J. (2015). "A Review and Meta-Analysis of Multimodal Affect Detection Systems." ACM Computing Surveys (CSUR), 47(3), 1-36.
2. Abadi, M., et al. (2016). "TensorFlow: A System for Large-Scale Machine Learning." 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI).
3. Ekman, P. Friesen, W. (1978). "Facial Action Coding System." Consulting Psychologists Press.
4. Yan, W., Garcia, C. (2013). "Facial expression recognition using deep learning." in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops.
5. Barsoum, E., Zhang, C., Ferrer, C. C., Zhang, Z. (2016). "Training deep networks for facial expression recognition with crowd-sourced label distribution." Proceedings of the 18th ACM International Conference on Multimodal Interaction (ICMI).
6. Goodfellow, I., Bengio, Y., Courville, A. (2016). "Deep Learning." MIT Press.
7. B. E. Boser, I. M. Guyon, and V. N. Vapnik, "Training algorithm for optimal margin classifiers," Proc. Fifth Annu. ACM Work. Compute. Learn. Theory, pp. 144–152, 1992, Doi: 10.1145/130385.130401.
8. S. J. Wang, H. L. Chen, W. J. Yan, Y. H. Chen, and X. Fu, "Face recognition and micro-expression recognition based on discriminant tensor subspace analysis plus extreme learning machine," Neural Process. Lett., vol. 39, no. 1, pp. 25–43, Feb. 2014, Doi: 10.1007/S11063-013-9288-7.
9. M. Li, H. Xu, X. Huang, Z. Song, X. Liu, and X. Li, "Facial Expression Recognition with Identity and Emotion Joint Learning," IEEE Trans. Affect. Compute., vol. 12, no. 2, pp. 544–550, Apr. 2021, Doi: 10.1109/TAFFC.2018.2880201.
10. P. Giannopoulos, I. Perikos, and I. Hatzilygeroudis, "Deep learning approaches for facial emotion recognition: A case study on FER-2013," Smart Innov. Syst. Technol., vol. 85, pp. 1–16, 2018.
11. K. Bahreini, R. Nadolski, and W. Westera, "Towards multimodal emotion recognition in e-learning environments," Interact. Learn. Environ., vol. 24, no. 3, pp. 590–605, Apr. 2016,
12. C. Clavel, "Surprise and human-agent interactions," Rev. Cogn. Linguist., vol. 13, no. 2, pp. 461–477, Dec. 2015,
13. M. Li, H. Xu, X. Huang, Z. Song, X. Liu, and X. Li, "Facial Expression Recognition with Identity and Emotion Joint Learning," IEEE Trans. Affect. Compute., vol. 12, no. 2, pp. 544–550, Apr. 2021, Doi: 10.1109/TAFFC.2018.2880201.
14. P. Giannopoulos, I. Perikos, and I. Hatzilygeroudis, "Deep learning approaches for facial emotion recognition: A case study on FER-2013," Smart Innov. Syst. Technol., vol. 85, pp. 1–16, 2018.
15. K. Bahreini, R. Nadolski, and W. Westera, "Towards multimodal emotion recognition in e-learning environments," Interact. Learn. Environ., vol. 24, no. 3, pp. 590–605, Apr. 2016,
16. C. Clavel, "Surprise and human-agent interactions," Rev. Cogn. Linguist., vol. 13, no. 2, pp. 461–477, Dec. 2015,

17. M. Li, H. Xu, X. Huang, Z. Song, X. Liu, and X. Li, “Facial Expression Recognition with Identity and Emotion Joint Learning,” IEEE Trans. Affect. Compute., vol. 12, no. 2, pp. 544–550, Apr. 2021, Doi: 10.1109/TAFFC.2018.2880201.

18. P. Giannopoulos, I. Perikos, and I. Hatzilygeroudis, “Deep learning approaches for facial emotion recognition: A case study on FER-2013,” Smart Innov. Syst. Technol., vol. 85, pp. 1–16, 2018.

19. K. Bahreini, R. Nadolski, and W. Westera, “Towards multimodal emotion recognition in e-learning environments,” Interact. Learn. Environ., vol. 24, no. 3, pp. 590–605, Apr. 2016,

20. C. Clavel, “Surprise and human-agent interactions,” Rev. Cogn. Linguist., vol. 13, no. 2, pp. 461–477, Dec. 2015,

21. "Adaptive Face Region Representation for Affect Recognition" by Barsoum et al. (2016)

22. "Facial Expression Recognition with Intra-Sample Label Switching: A Unified Approach" by E. Barsoum et al. (ICLR 2017)

23. IMANO00,"Dataset3Modified."Kaggle.[Online].Available: https://www.kaggle.com/datasets/imano00/dataset3modified

24. S. Dari, A. A. Hussein, and M. R. El-Sakka, "The Importance of Data Preprocessing in Machine Learning: A Review," in Procedia Computer Science, vol. 65, 2015, pp. 1042-1051. Doi: 10.1016/j.procs.2015.09.037.

25. R. K. Adu-Gyamfi, M. Zhang, S. M. Anthony, "The Benefits of Resizing Datasets for Machine Learning Tasks," in 2019 IEEE International Conference on Big Data (Big Data), 2019, pp. 2272-2275. Doi: 10.1109/BigData47090.2019.9006418.

26. J. Li, Q. Fan, J. Wang, “ Benefits of Gray scaling the Dataset Im- ages for Deep Learning," in 2021 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 2021, pp. 1566-1570. Doi: 10.1109/ICASSP39728.2021.9413639.

27. Yamashita, R., Nishio, M., Do, R.K.G. et al. Convolutional neural networks: an overview and application in radiology. Insights Imaging 9, 611–629 (2018).

28. Divyanshu Soni., "Translation Invariance in Convolutional Neural Networks",

29. Nov. 2019

30. Harsh Yadav., "Dropout in Neural Networks", Jul. 2022

31. LeCun, Y., Bengio, Y., Hinton, G. (2015). "Deep learning." Nature, 521(7553), 436-444.

32. Jason Brownlee., "Gentle Introduction to the Adam Optimization Algorithm for Deep Learning", July. 2017

33. S. Ioffe and C. Szegedy, "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift," in Proceedings of the 32nd Inter- national Conference on Machine Learning (ICML), 2015, pp. 448-456.

34. Ian Goodfellow, Yoshua Bengio, and Aaron Courville. Deep Learning. MIT Press, 2016.

35. M. Li, H. Xu, X. Huang, Z. Song, X. Liu, and X. Li, “Facial Expression Recognition with Identity and Emotion Joint Learning,” IEEE Trans. Affect. Compute., vol. 12, no. 2, pp. 544–550, Apr. 2021, Doi: 10.1109/TAFFC.2018.2880201.

36. P. Giannopoulos, I. Perikos, and I. Hatzilygeroudis, “Deep learning approaches for facial emotion recognition: A case study on FER-2013,” Smart Innov. Syst. Technol., vol. 85, pp. 1–16, 2018.

37. K. Bahreini, R. Nadolski, and W. Westera, “Towards multimodal emotion recognition in e-learning environments,” Interact. Learn. Environ., vol. 24, no. 3, pp. 590–605, Apr. 2016,

38. C. Clavel, “Surprise and human-agent interactions,” Rev. Cogn. Linguist., vol. 13, no. 2, pp. 461–477, Dec. 2015,

39. "Adaptive Face Region Representation for Affect Recognition" by Barsoum et al. (2016)

40. "Facial Expression Recognition with Intra-Sample Label Switching: A Unified Approach" by E. Barsoum et al. (ICLR 2017)

41. IMANO00,"Dataset3Modified."Kaggle.[Online].Available: https://www.kaggle.com/datasets/imano00/dataset3modified

42. S. Dari, A. A. Hussein, and M. R. El-Sakka, "The Importance of Data Preprocessing in Machine Learning: A Review," in Procedia Computer Science, vol. 65, 2015, pp. 1042-1051. Doi: 10.1016/j.procs.2015.09.037.

43. R. K. Adu-Gyamfi, M. Zhang, S. M. Anthony, "The Benefits of Resizing Datasets for Machine Learning Tasks," in 2019 IEEE International Conference on Big Data (Big Data), 2019, pp. 2272-2275. Doi: 10.1109/BigData47090.2019.9006418.

44. J. Li, Q. Fan, J. Wang, "Benefits of Gray scaling the Dataset Im- ages for Deep Learning," in 2021 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 2021, pp. 1566-1570. Doi: 10.1109/ICASSP39728.2021.9413639.

45. Yamashita, R., Nishio, M., Do, R.K.G. et al. Convolutional neural networks: an overview and application in radiology. Insights Imaging 9, 611–629 (2018).

46. Divyanshu Soni., "Translation Invariance in Convolutional Neural Networks",

47. Nov. 2019

48. Harsh Yadav., "Dropout in Neural Networks", Jul. 2022

49. LeCun, Y., Bengio, Y., Hinton, G. (2015). "Deep learning." Nature, 521(7553), 436-444.

50. Jason Brownlee., "Gentle Introduction to the Adam Optimization Algorithm for Deep Learning", July. 2017

51. S. Ioffe and C. Szegedy, "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift," in Proceedings of the 32nd Inter- national Conference on Machine Learning (ICML), 2015, pp. 448-456.

52. Ian Goodfellow, Yoshua Bengio, and Aaron Courville. Deep Learning. MIT Press, 2016.

53. Christopher M. Bishop. Pattern Recognition and Machine Learning. Springer, 2006.

54. Géron, Aurélien. Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow. O’Reilly Media, 2019. (Chapter 3: Classification)

55. S. Ioffe and C. Szegedy, "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift," in Proceedings of the 32nd Inter- national Conference on Machine Learning (ICML), 2015, pp. 448-456.

56. Ian Goodfellow, Yoshua Bengio, and Aaron Courville. Deep Learning. MIT Press, 2016.

57. Christopher M. Bishop. Pattern Recognition and Machine Learning. Springer, 2006.

58. Géron, Aurélien. Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow. O'Reilly Media, 2019. (Chapter 3: Classification)

59. Kevin P. Murphy. Machine Learning: A Probabilistic Perspective. MIT Press, 2012. (Chapter 28: Classifier Evaluation)

60. Sebastian Raschka and Vahid Mirjalili. Python Machine Learning, 3rd Edition. Packt Publishing, 2019. (Chapter 6: Learning Best Practices for Model Evaluation and Hyperparameter Tuning)

61. Lawrence, Kate Campbell, Ruth Skuse, David. (2015). Age, gender, and puberty influence the development of facial emotion recognition. Frontiers in Psychology. 6. 10.3389/fpsyg.2015.00761.

62. Justin M., Low Light Portrait[Photograph]., Attribution-NonCommercial (CC BY-NC 2.0)

63. "Image: Freepik.com". This cover has been designed using assets from Freepik.com