FEASIBILITY OF A REAL-TIME MOBILE EMOTION RECOGNITION SYSTEM FOR CHILDREN

Abstract

This feasibility study focused on developing a robust Facial Emotion Recognition (FER) system specifically tailored for children aged 7 to 12, aiming to address the limited availability of specialized tools in this field. The system development followed the standard FER pipeline, including face detection, feature extraction, emotion classification, and evaluation. To determine the most effective classification method, three deep learning models, DenseNet-201, ResNet-101, and Inception-v3, were trained using MATLAB. The results showed that DenseNet-201 yielded the highest overall accuracy at 63.10% and was subsequently chosen for integration into the mobile system. The developed system was able to analyze facial expressions frame by frame, offering accurate insights into the emotional states of children. It also supports a user-initiated video processing workflow through an intuitive user interface (UI), minimizing user effort while strictly adhering to privacy protocols.

Introduction:-

Facial Emotion Recognition (FER) is a pivotal research topic in computer vision and artificial intelligence, leveraging visual cues to interpret complex human emotional and physiological states. Emotions influence and regulate cognitive activities such as attention, memory, and thought processes, significantly impacting learning efficiency [1]. Therefore, emotion recognition is particularly crucial for children, especially those who may struggle with verbal communication, as facial expression serves as a primary way to reflect their inner feelings. Identifying emotional difficulties early on allows for timely interventions and support, preventing potential long-term emotional and mental health challenges.

Despite the importance of FER, there is a lack of well-established, comprehensive tools available specifically for accurately and reliably recognizing emotions in children [2]. Existing tools are often optimized for adults, failing to account for the unique characteristics of children's emotional development, such as rapidly changing expressions. This study was initiated to address these gaps and to develop an effective real-time emotion recognition system that accurately analyzes children's facial expressions.

The system targets the recognition of seven emotional variations: happiness, surprise, neutral, disgust, anger, fear, and sadness. Ethically, strict adherence to privacy protocols is maintained, especially concerning children's personal data, which is safeguarded by ensuring the system does not record or save facial data.

Related Work:-

The field of FER is grounded in recognizing that facial expressions are rich sources of emotional data, categorized through psychological approaches, and analyzed via artificial intelligence. Historically, the foundation for analyzing facial expressions stems from the Facial Action Coding System (FACS), which taxonomizes facial movements into action units (AUs). This categorical approach relies on seven universal "Basic Emotions", such as joy, anger, fear, disgust, sadness, surprise, and contempt.

However, this Basic Emotions view is highly controversial, challenged by arguments that emotions are socio-culturally constructed and context-dependent [3]. Contemporary research finds that expressions are more

25 contextually contingent than the Basic view suggests, and that single labels often fail to capture complex emotional
26 behavior [4].

27 Modern FER systems, particularly those relying on affective computing and machine learning (often labeled as
28 Emotional AI), typically employ a pipeline involving face detection, feature extraction, and emotion classification:

29 • **Face Detection** methods like the Viola-Jones algorithm, often utilizing Haar-like features and AdaBoost,
30   rapidly identify and localize faces within images or video frames. These methods are essential for
31   initializing the face-related recognition pipeline in real-time applications [5].

32 • **Feature Extraction** heavily relies on Convolutional Neural Networks (CNNs), which automatically learn
33   and hierarchically abstract complex spatial features specialized for image pattern recognition. Traditional
34   approaches also include techniques like the Active Shape Model (ASM), which defines the geometric
35   structure of the face.

36 • **Classification** is accomplished using algorithms such as Support Vector Machines (SVM) or K-Nearest
37   Neighbors (KNN), which categorize emotions based on the extracted features. In contemporary FER, the
38   final dense layers of the pre-trained CNN often serve as the classifier, mapping high-level features directly
39   to emotion labels [6].

40 In the context of Education Technology, the deployment of emotional AI based on facial coding raises significant
41 ethical and methodological concerns. Critics note the debatable methodology behind facial coding, the
42 commodification of children's emotional data for broader commercial interests, and the resulting tension between
43 private profit motives and serving the public good [7], [8].

44

Methodology and Implementation:-
46 This research developed a cross-platform mobile application for facial emotion recognition by training and
47 evaluating three deep learning CNN models on a preprocessed, grayscale dataset of faces. DenseNet-201 was
48 selected for integration into the final system after achieving the highest accuracy of 63.10%. The implemented
49 application features an intuitive camera interface and was validated through initial model testing and user acceptance
50 trials.

51 **I. Dataset and Preprocessing**

52 The project utilized image datasets sourced from the Kaggle website (https://www.kaggle.com), containing a variety
53 of human face data, including children, teenagers, adults, and elders, spanning an age range from approximately 1 to
54 99 years old. This 2D image dataset was used to train models for the recognition of seven emotions: angry, fear,
55 happy, disgust, neutral, sad, and surprise. No data labeling was required, as the images were pre-normalized to
56 48x48 pixels in grayscale, ensuring faces were more or less centered and occupied a consistent amount of space
57 within each image.

58 The initial plan was to collect data from a specific child development centre. However, in the end, the diverse
59 Kaggle dataset was used with the purpose of broadening the target demographic from children at a single center to
60 all children, and avoiding the complexity of processing large BAG files.

61 **II. Model Selection and Training**

62 The core process involved training and evaluating three well-known deep learning Convolutional Neural Network
63 (CNN) models in MATLAB: DenseNet-201, ResNet-101, and Inception-V3.

64 DenseNet-201 utilizes dense block structures to enhance training by promoting direct feature reuse, resulting in high
65 accuracy despite its intensive memory consumption [9]. ResNet-101 addresses the vanishing gradient problem
66 through residual connections, enabling exceptionally deep networks [10], [11]. Inception-V3 employs the Inception
67 module to efficiently capture features by processing inputs with multiple filter sizes simultaneously [12].

68 The performance comparison based on training results was as follows:

69 • DenseNet-201: 63.10% accuracy

70 • ResNet-101: 62.03% accuracy

71 • Inception-V3: 61.50% accuracy

72 As DenseNet-201 demonstrated the highest accuracy, it was selected for integration into the final system for facial
73 emotion recognition.

74 **III. System Implementation and Testing**

75 The final system was coded using Visual Studio IDE and the .NET MAUI framework, chosen for its ability to build
76 native cross-platform applications from a single codebase [13], targeting the Android platform.

77 The interface design was adapted from a planned desktop view to a mobile view (Figure 2) to ensure portability. The
78 user interface (UI) was designed to be intuitive, featuring controls to toggle the camera on/off and a "Run/Stop"
79 button to initiate and halt the frame-by-frame recognition program. The analysis panel, displaying the detected
80 emotion and corresponding emoji, is positioned at the bottom of the camera screen.

81 Initial testing included generating confusion matrices for the trained models. User acceptance tests (UAT) were later
82 conducted with seven respondents (4 adults, 3 children) to evaluate the UI's intuitiveness, system usefulness, and
83 responsiveness.

84

85

86

87
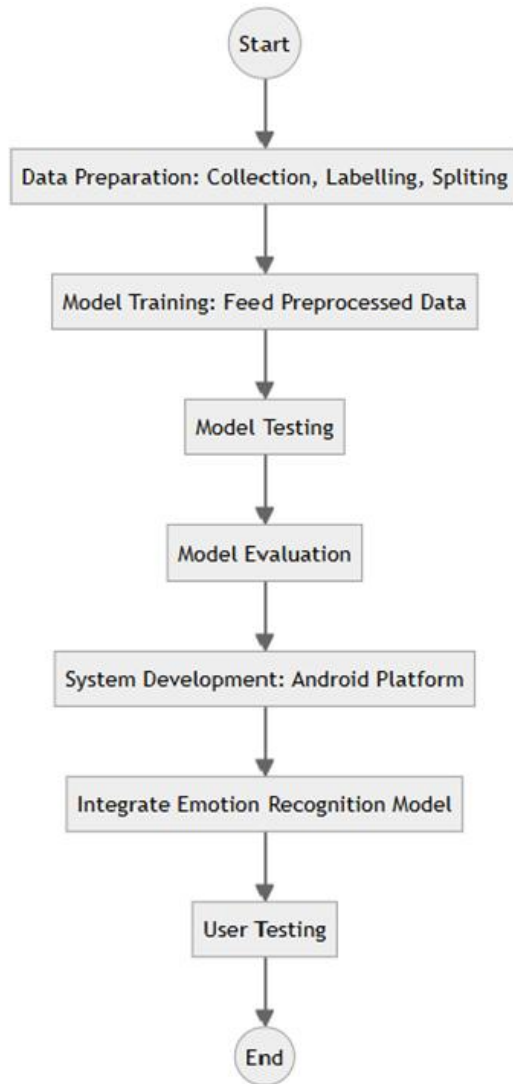
**Figure 1:** Project workflow
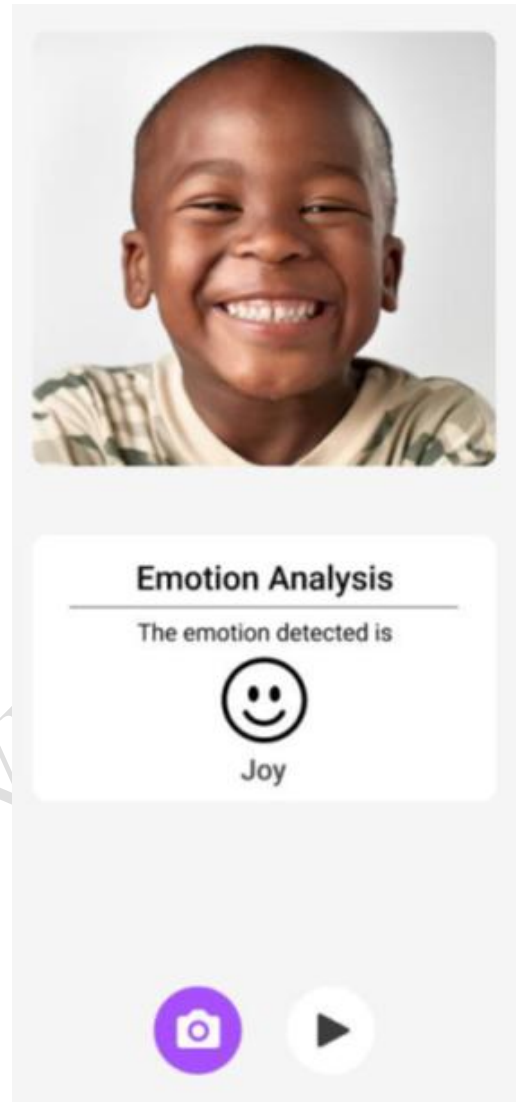

Figure 2: User Interface Design for Mobile View

88

Results and Discussion:-
### I. Model Performance and Evaluation

To comprehensively assess model performance, confusion matrices were generated based on testing images (10 prepared for each emotion). Figures 3 through 5 show the training results, and Figures 6 through 8 present the confusion matrices.
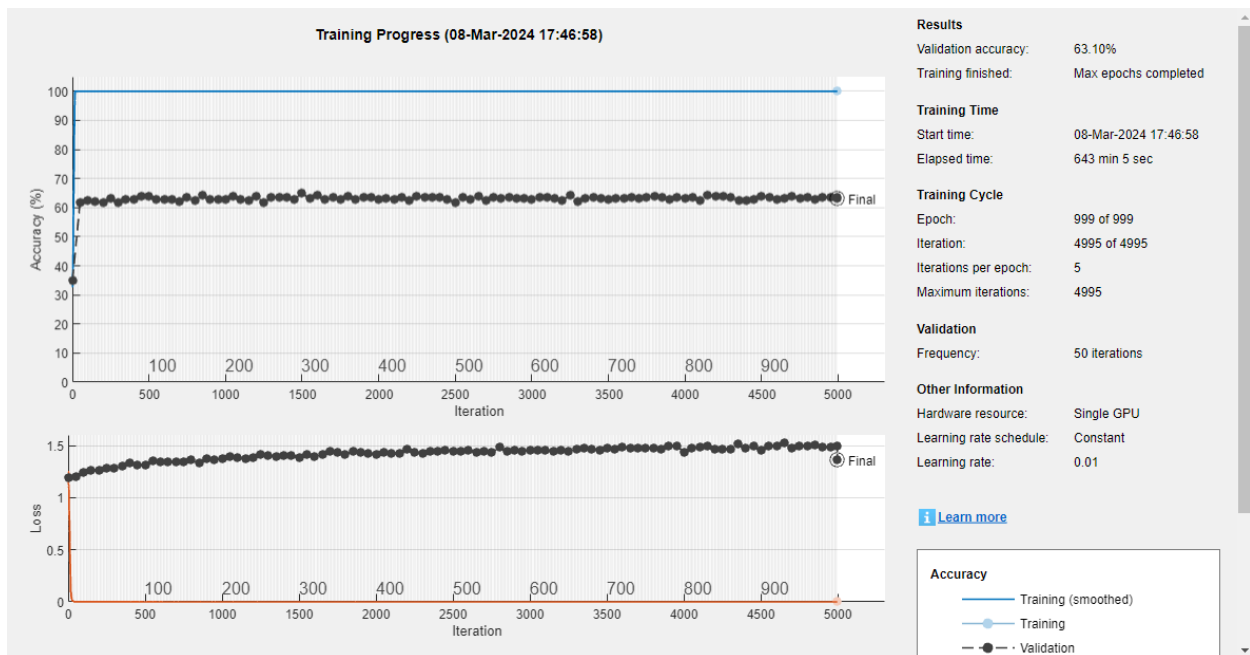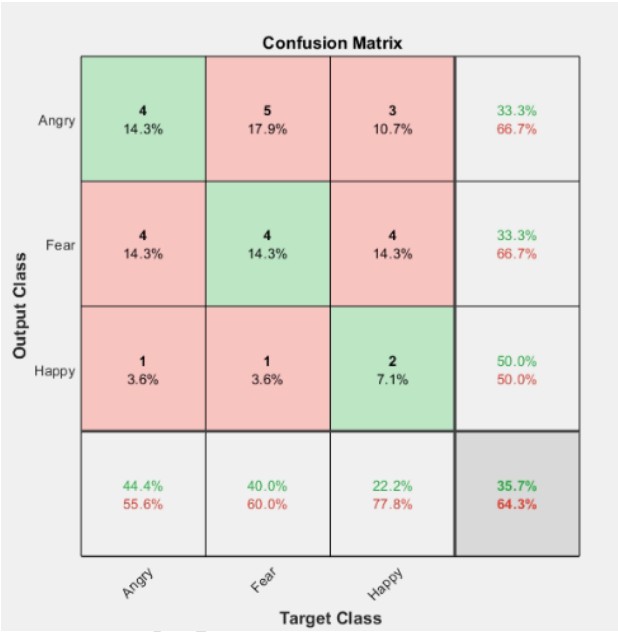
Figure 3: Training Accuracy and Loss History for DenseNet-201
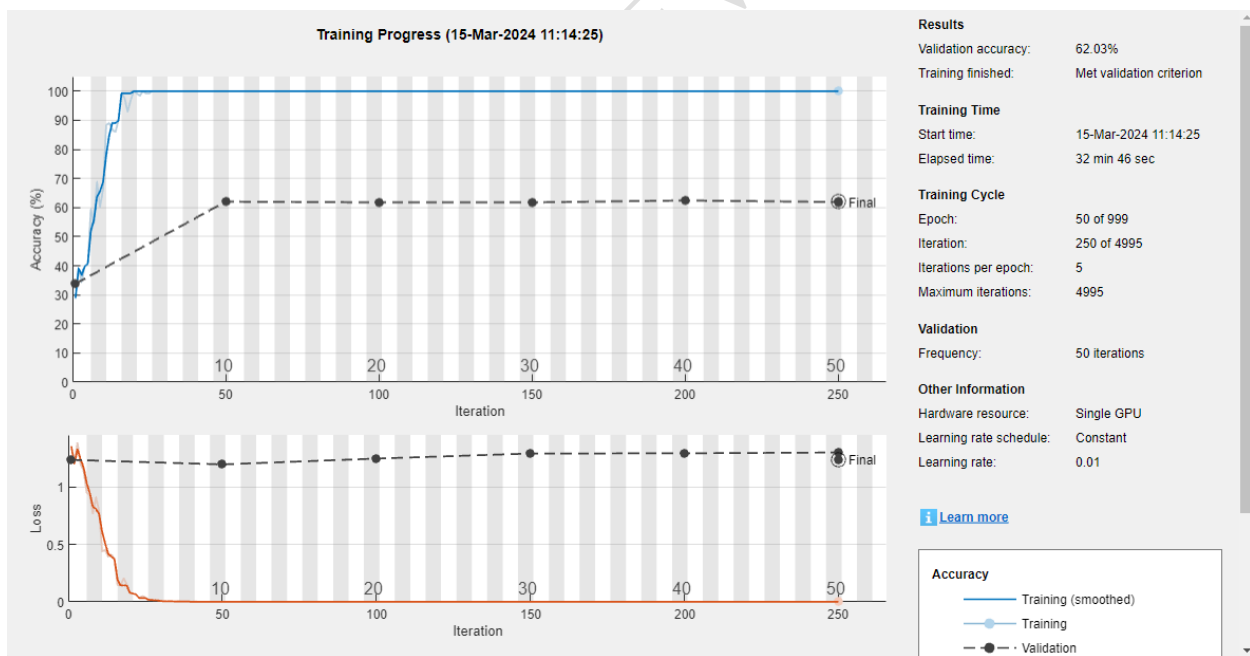
Figure 4: Training Accuracy and Loss History for ResNet-101
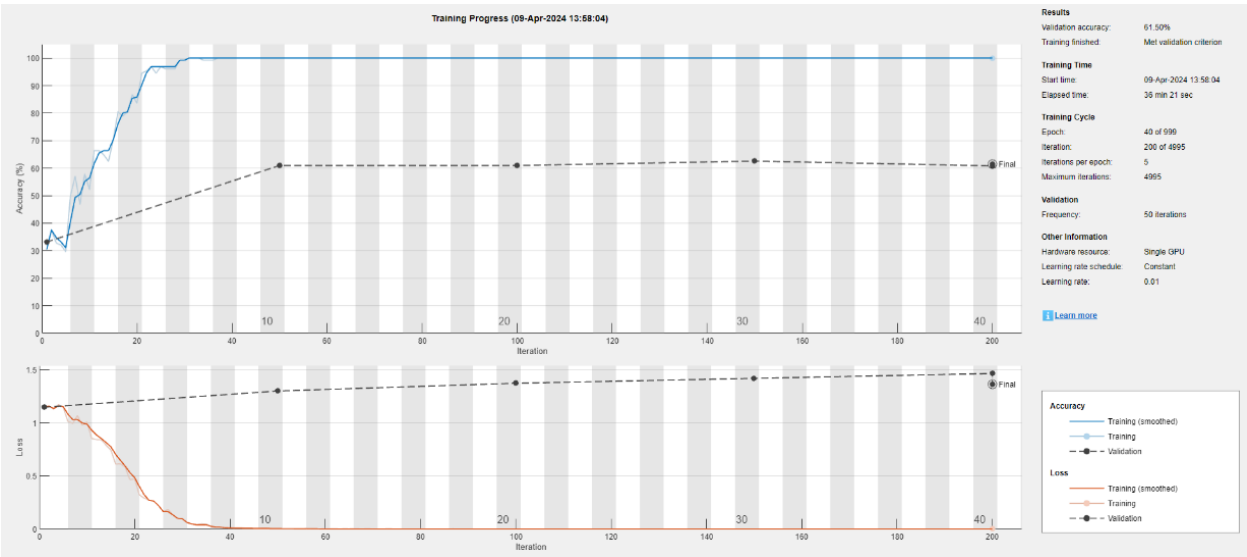
Figure 5: Training Accuracy and Loss History for Inception-V3

Figure 6: Confusion Matrix of DenseNet-201

Figure 7: Confusion Matrix of ResNet-101

Figure 8: Confusion Matrix of Inception-V3

While DenseNet-201 was expected to be the strongest model due to its 201 layers, its confusion matrix accuracy was lower than anticipated. This outcome highlights the challenges of training instability and hyperparameter optimization in complex architectures. Simpler models are frequently easier to optimize effectively, a sentiment echoed by [13]. Their paper emphasizes that deeper models, while theoretically more accurate, pose "Optimization

7

116 Trade-offs" for mobile use. Specifically, they lead to increased energy consumption and inference lag, which can
117 negatively impact real-time performance.

118 **II. User Acceptance and System Latency**

119 User testing results, compiled from feedback forms, indicated that most respondents found the developed system
120 user-friendly, efficient, and easy to understand. Specifically, for ease of UI comprehension, 4 respondents rated the
121 UI as easy (4), 2 rated it as very easy (5), and 1 found it somewhat difficult (2) (refer to Figure 10). On the other
122 hand, 5 of the respondents strongly agreed that the developed system is very helpful in recognizing facial emotions,
123 and 2 more agree on the statement (refer to Figure 11).

124
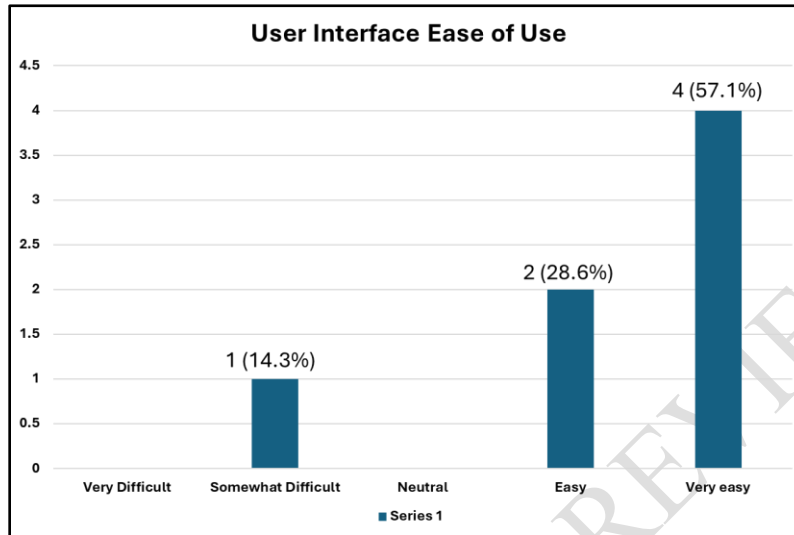
125 **Figure 10:** Statistics of Ratings for User Interface Satisfaction

126

127
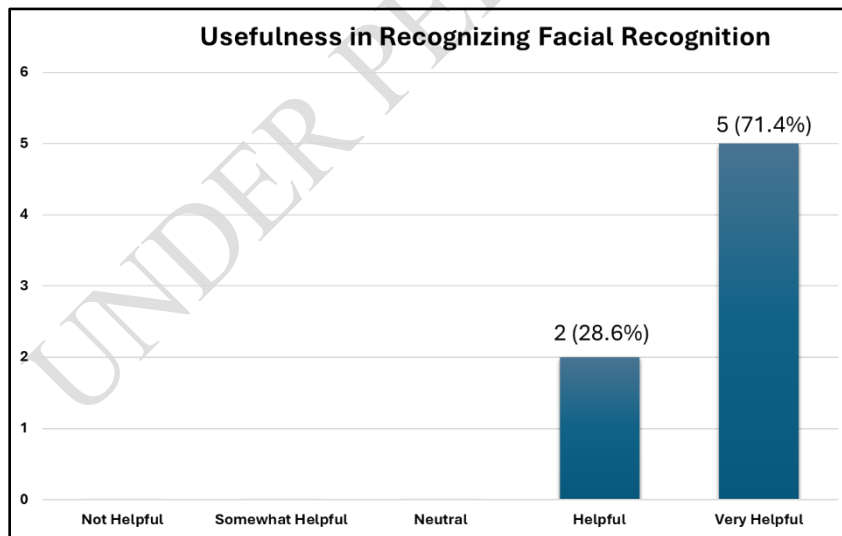
128 **Figure 11:** Statistics of Ratings for Usefulness of the Developed System in Recognizing Facial Emotions

129

130

A notable finding was the relatively low rating for system responsiveness (refer to Figure 12). Three respondents rated performance as average (3), and four rated it as running smoothly (4). This latency was attributed to the mobile phone's insufficient processing power, which acts as a bottleneck when simultaneously capturing video, processing each frame for emotion recognition using the DenseNet-201 model, and updating the UI.

135
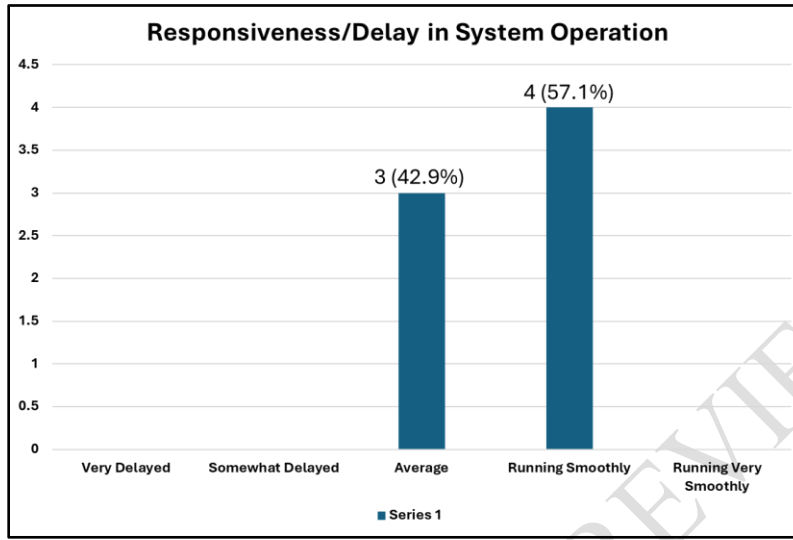
Figure 12: Statistics of Ratings for the Responsiveness/Delay in the System Operation

138

System Limitations. Limitations were encountered during the project implementation:

1. **Limited Number of Emotions Recognized:** Due to the complexity of distinguishing nuanced facial expressions, the initial scope of recognizing seven emotions was restricted. To improve accuracy and effectiveness in the field, the system's focus was restricted to recognizing only the three most significant emotions: angry, fear, and happy.

2. **Images Latency Issue:** As discussed, real-time operation experienced delays when the program started running. This was attributed to insufficient mobile processing power.

The project successfully addressed the primary problem statement, the lack of comprehensive tools for real-time emotion recognition in children, by developing a dedicated, accurate system that includes built-in privacy measures, such as not recording or saving users' facial data.

149

Conclusion:-
This project successfully developed a real-time Facial Emotion Recognition system, fulfilling the objectives of creating an effective recognition tool and supporting a user-initiated workflow. The successful training and deployment of the DenseNet-201 model, selected for its superior performance (63.10% accuracy) over ResNet-101

and Inception-V3, yielded a reliable model suitable for practitioners aiding in the early detection and management of emotional issues in children.

Despite these achievements, challenges remain, notably the limitation on the number of emotions the system can reliably recognize, and the image latency caused by insufficient processing power in mobile devices.

Future research should address these limitations through several countermeasures:

- **Progressive Enhancement and Feature Engineering:** Focus should be placed on improving the accuracy for the most prevalent emotions before expanding the scope and utilizing additional facial features like landmarks or texture descriptors to enhance the model's ability to discriminate between emotions.

- **Optimization of Algorithm and Hardware Upgrades:** To counter latency, optimizing the recognition algorithm to reduce computational load is recommended. Additionally, considering hardware upgrades would allow the recognition program to run more smoothly.

Overall, this system represents an important step in supporting the psychological development of children by providing a dedicated tool for accurately assessing their emotional states.

References:-

[1] E. M. Younis, S. Mohsen, E. H. Houssein, and O. A. S. Ibrahim, "Machine learning for human emotion recognition: A comprehensive review," *Neural Computing and Applications*, vol. 36, no. 16, pp. 8901–8947, 2024.

[2] E. Kim, D. Bryant, D. Srikanth, and A. Howard, "Age bias in emotion detection: An analysis of facial emotion recognition performance on young, middle-aged, and older adults," in *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*, 2021, p. 62.

[3] A. Katirai, "Ethical considerations in emotion recognition technologies: A review of the literature," *AI and Ethics*, vol. 4, no. 4, pp. 927–948, 2024.

[4] M. Lecker and H. Aviezer, "More than words? Semantic emotion labels boost context effects on faces," *Affective Science*, vol. 2, no. 2, pp. 163–170, 2021.

[5] Y. Guo and B. C. Wünsche, "Comparison of face detection algorithms on mobile devices," in *2020 35th International Conference on Image and Vision Computing New Zealand (IVCNZ)*, IEEE, 2020.

[6] G. S. Kumar and J. Cheriyan, "Unleashing facial expression recognition for stress detection using deep CNN model," *Procedia Computer Science*, vol. 259, pp. 306–315, 2025.

[7] V. Bakir, A. Laffer, A. McStay, D. Miranda, and L. Urquhart, "On manipulation by emotional AI: UK adults' views and governance implications," *Frontiers in Sociology*, vol. 9, Art. no. 1339834, 2024.

[8] L. F. Barrett, R. Adolphs, S. Marsella, A. M. Martinez, and S. D. Pollak, "Emotional expressions reconsidered: Challenges to inferring emotion from human facial movements," *Psychological Science in the Public Interest*, vol. 20, no. 1, pp. 1–68, 2019.

191   [9] E. Hassan, S. A. Ghazalah, N. El-Rashidy, T. A. El-Hafeez, and M. Y. Shams, "DenseNet model with attention
192   mechanisms for robust date fruit image classification," *International Journal of Computational Intelligence Systems*,
193   vol. 18, no. 1, pp. 1–33, 2025.

194   [10] S. Meyyappan, "Leveraging attention mechanisms in ResNet101 for enhanced facial emotion recognition,"
195   presented at the *3rd International Conference on Advancement in Computation & Computer Technologies
196   (InCACCT)*, June 2025.

197   [11] R. Kumari and J. Wasim, "A deep learning approach for human facial expression recognition using Residual
198   Network–101," *Journal of Current Science and Technology*, vol. 13, no. 3, pp. 517–532, 2023.

199   [12] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proceedings of the IEEE
200   Conference on Computer Vision and Pattern Recognition*, 2017, pp. 1251–1258.

201   [13] T. Zhao et al., "A survey of deep learning on mobile devices: Applications, optimizations, challenges, and
202   research opportunities," *Proceedings of the IEEE*, vol. 110, no. 3, pp. 334–354, 2022.

203