Journal Homepage: -www.journalijar.com

INTERNATIONAL JOURNAL OF ADVANCED RESEARCH (IJAR)

Article DOI: 10.21474/IJAR01/22408
DOI URL: http://dx.doi.org/10.21474/IJAR01/22408

ISSN NO. 2320-5407

RESEARCH ARTICLE

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

Marcella Peter[1,3], Jackie Ting Tiew Wei[1,3], Khairunnisa Ibrahim[2,3], Michael Tang Chi Seng[1,3] and Cody Cheng Jing Zhi[3]

1. DeTeC Research Centre, University of Technology Sarawak, Sibu, Malaysia.
2. ASSET Research Centre,University of Technology Sarawak, Sibu, Malaysia.
3. School of Computing and Creative Media, University of Technology Sarawak, Sibu, Malaysia.

Manuscript Info

Manuscript History
Received: 10 October 2025
Final Accepted: 12 November 2025
Published: December 2025

Key words:-
Facial emotion recognition, Deep Learning, DenseNet-201, Mobile Computing, Children's Affective Computing.

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

Corresponding Author:-Marcella Peter
Address**:-**DeTec Research Centre, School of Computing and Creative Media, University of Technology Sarawak, Sibu, Malaysia.

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

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

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

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

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

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

Methodology and Implementation:-

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

Dataset and Preprocessing:-

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

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

Model Selection and Training:-
The core process involved training and evaluating three well-known deep learning Convolutional Neural Network (CNN) models in MATLAB: DenseNet-201, ResNet-101, and Inception-V3.
DenseNet-201 utilizes dense block structures to enhance training by promoting direct feature reuse, resulting in high accuracy despite its intensive memory consumption [9]. ResNet-101 addresses the vanishing gradient problem through residual connections, enabling exceptionally deep networks [10], [11]. Inception-V3 employs the Inception module to efficiently capture features by processing inputs with multiple filter sizes simultaneously [12].

The performance comparison based on training results was as follows:
- DenseNet-201: 63.10% accuracy
- ResNet-101: 62.03% accuracy
- Inception-V3: 61.50% accuracy

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

System Implementation and Testing:-
The final system was coded using Visual Studio IDE and the .NET MAUI framework, chosen for its ability to build native cross-platform applications from a single codebase [13], targeting the Android platform.The interface design was adapted from a planned desktop view to a mobile view (Figure 2) to ensure portability. The user interface (UI) was designed to be intuitive, featuring controls to toggle the camera on/off and a “Run/Stop” button to initiate and halt the frame-by-frame recognition program. The analysis panel, displaying the detected emotion and corresponding emoji, is positioned at the bottom of the camera screen.Initial testing included generating confusion matrices for the trained models. User acceptance tests (UAT) were later conducted with seven respondents (4 adults, 3 children) to evaluate the UI’s intuitiveness, system usefulness, and responsiveness.

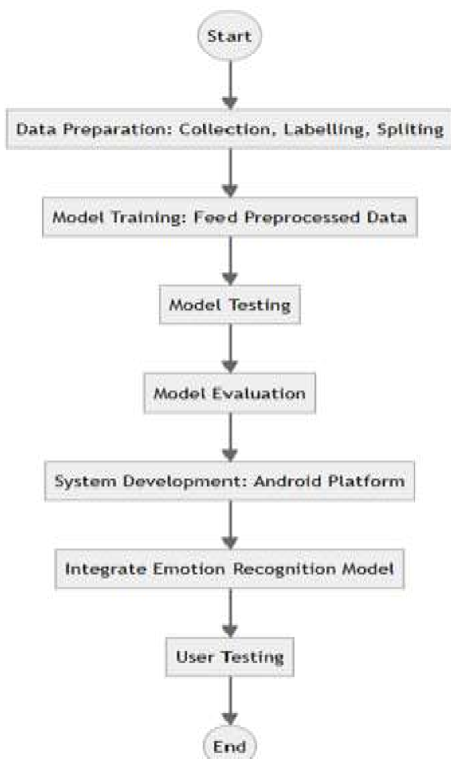

Figure 1: Project workflow

Figure 2: User Interface Design for Mobile View

Results and Discussion:-

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.

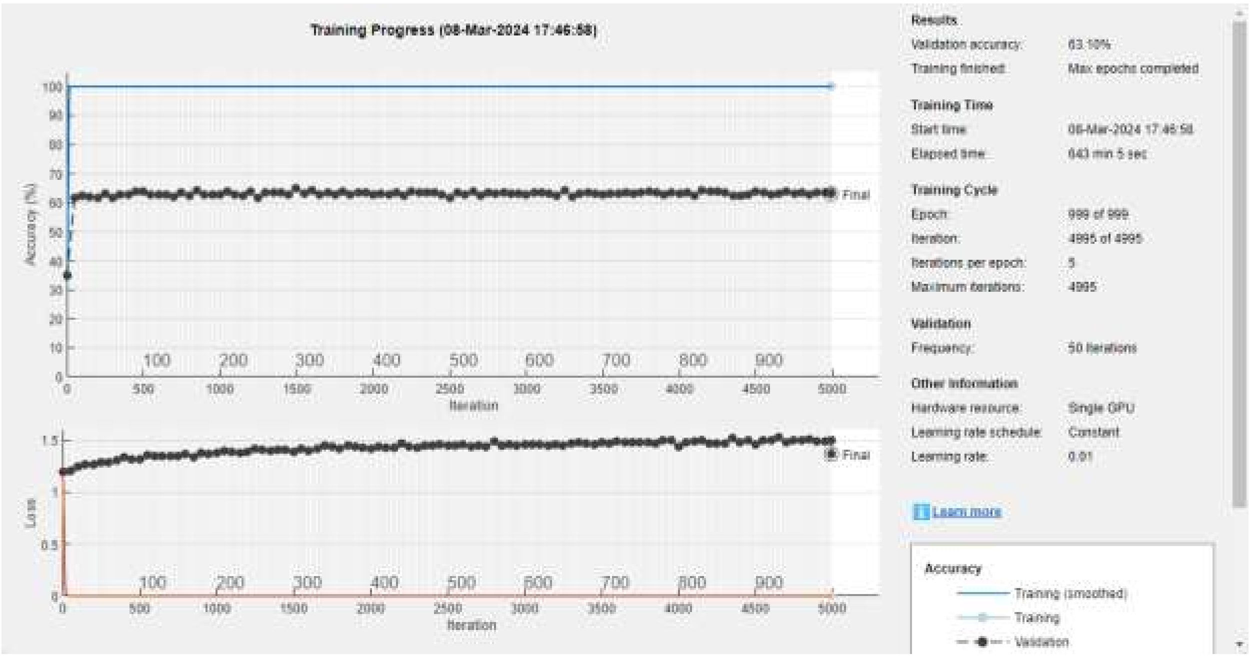

Figure 3: Training Accuracy and Loss History for DenseNet-201

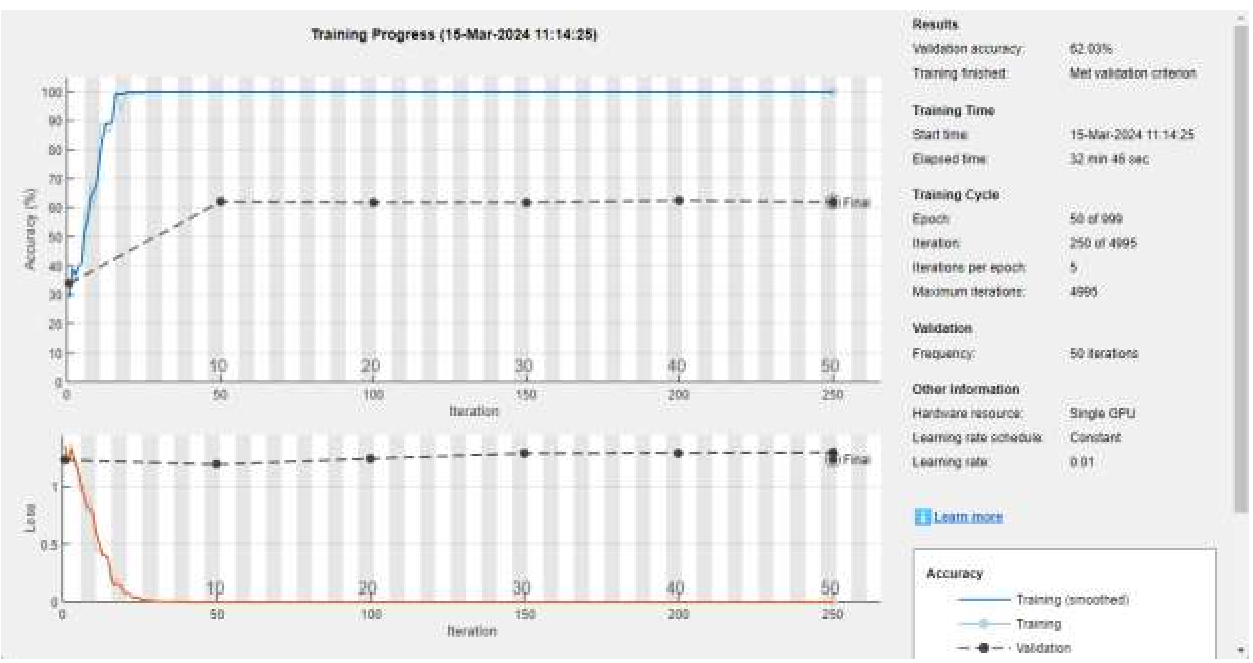

Figure 4: Training Accuracy and Loss History for ResNet-101

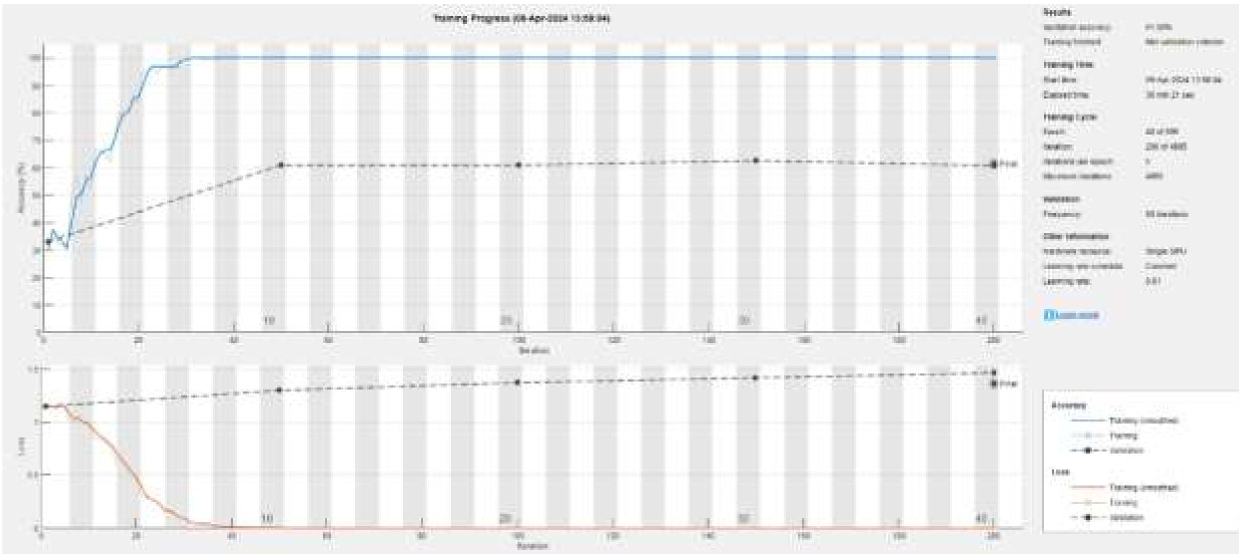

Figure 5: Training Accuracy and Loss History for Inception-V3

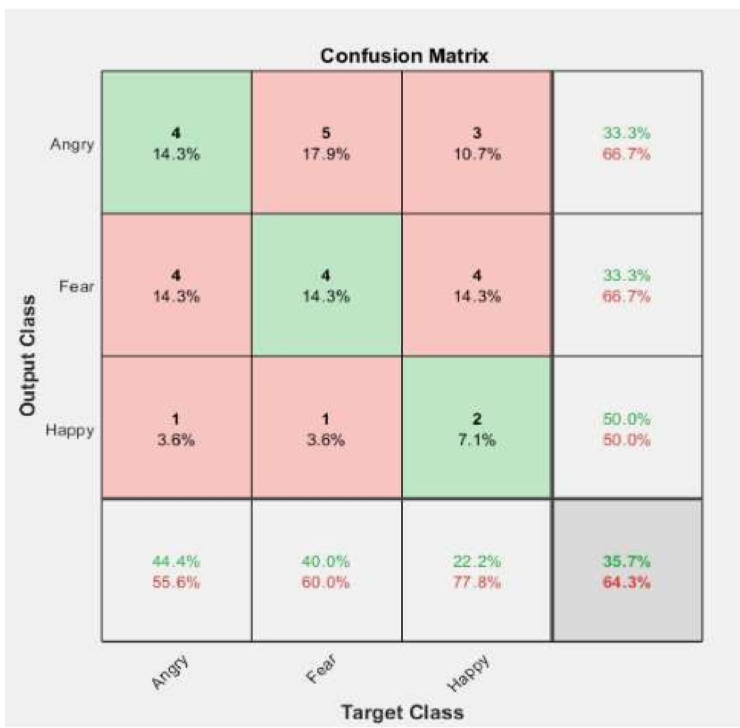

Figure 6: Confusion Matrix of DenseNet-201

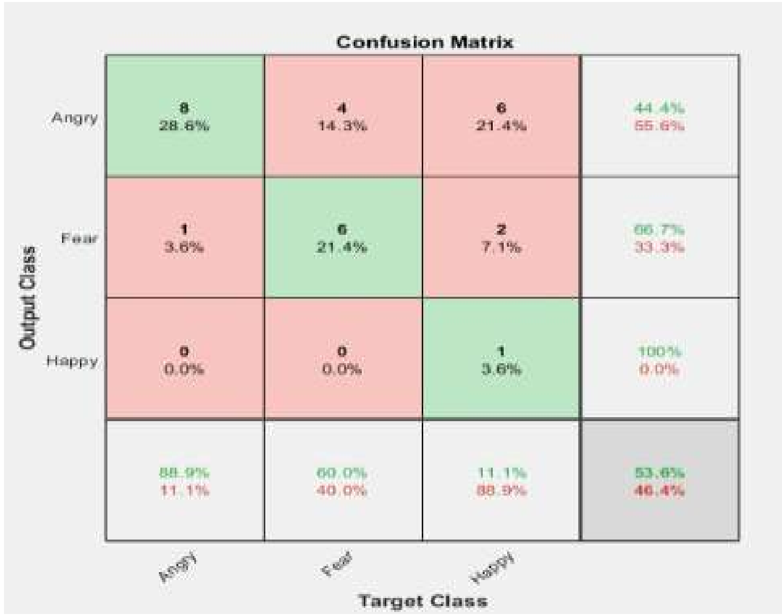

Figure 7: Confusion Matrix of ResNet-101

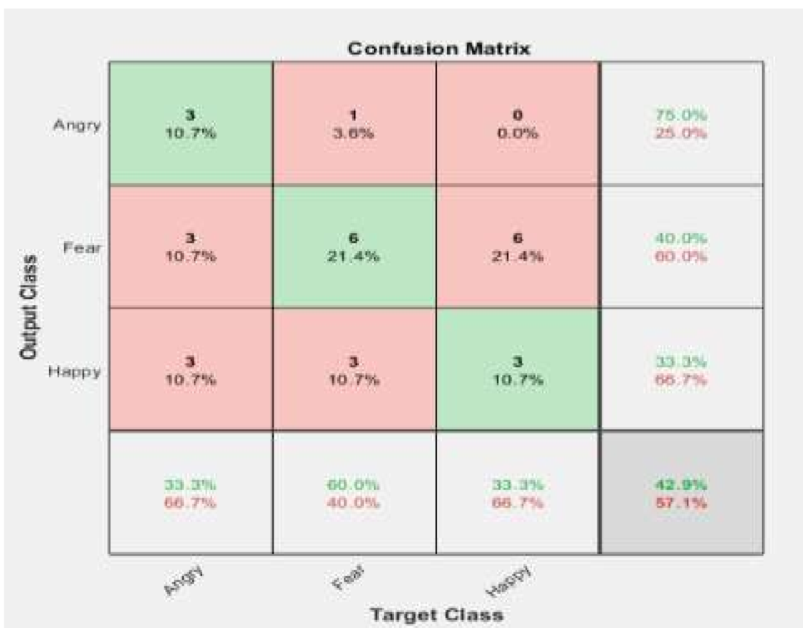

Figure 8: Confusion Matrix of Inception-V3

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

User Acceptance and System Latency:-
User acceptance testing results, compiled from feedback forms, indicated that most respondents found the developed system user-friendly, efficient, and easy to understand. Specifically, for ease of UI comprehension, 4 respondents rated the 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 hand, 5 of the respondents strongly agreed that the developed system is very helpful in recognizing facial emotions, and 2 more agreed on the statement (refer to Figure 11).

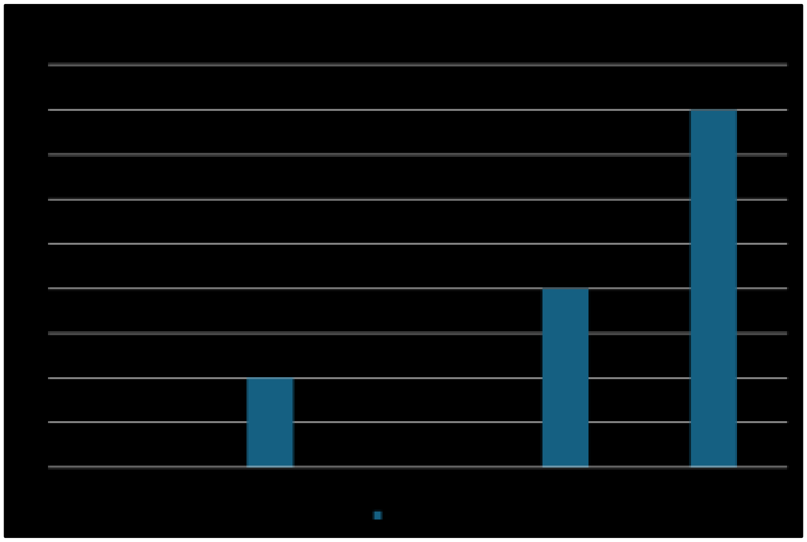

Figure 10: Statistics of Ratings for User Interface Satisfaction

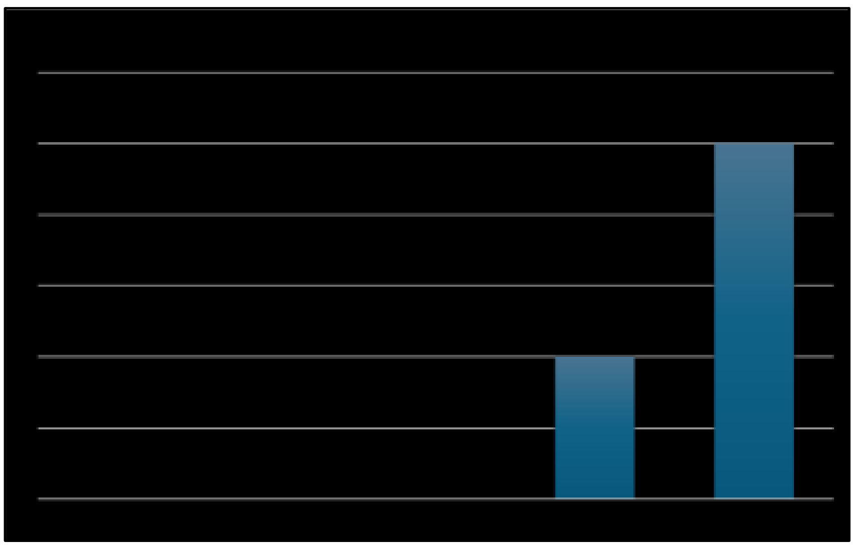

Figure 11: Statistics of Ratings for Usefulness of the Developed System in Recognizing Facial Emotions

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.

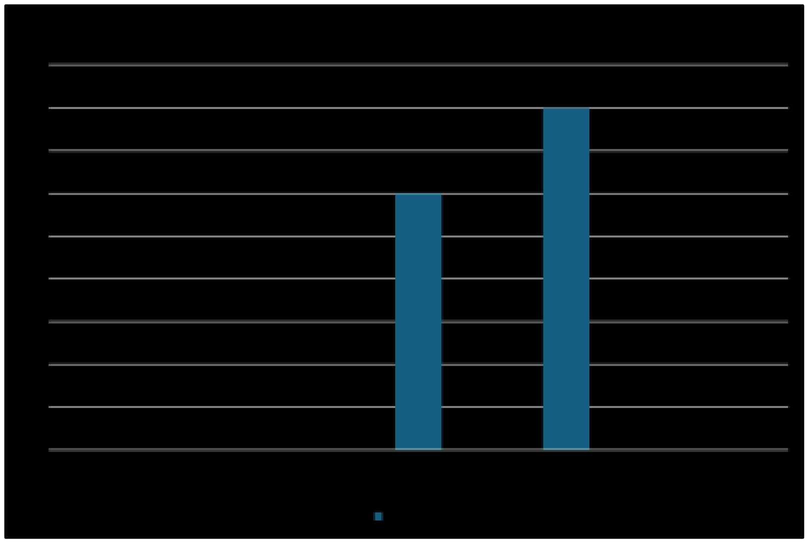

Figure 12: Statistics of Ratings for the Responsiveness/Delay in the System Operation

Scope Adjustment and Justification:-

Although the study initially aimed to recognize seven emotional states, the implemented system was refined to focus on three primary emotions: angry, fear, and happy. This reduction in scope was strategically motivated by the "Optimization Tradeoff" necessary for real-time mobile inference [13]. By narrowing the classification output, the system achieved a more stable inference speed (averaging within the 0.3–3s real-time threshold). It reduced the probability of false-positive detections in a pediatric context, where nuanced emotions like "Disgust" or "Surprise" are often physiologically similar to "Fear" or "Neutral" faces in children.

A notable limitation in current FER research is the scarcity of large-scale, publicly available datasets exclusively featuring children. Consequently, the use of a mixed-age dataset was necessary to provide sufficient training samples for deep learning. However, the moderate accuracy observed (63.10%) may be attributed to "age bias", where the model learns facial features more prevalent in adults, such as distinct bone structures or age-related skin markers, which are less pronounced in children [2]. Future iterations would benefit from fine-tuning on pediatric-only datasets to better capture the unique facial morphology of the 7 to 12-year-old age group.

The project addressed 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.

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.UAT results for UI comprehension and system usefulness were largely positive. 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:-

- Hardware-aware Modeling: Exploring pruning and quantization to reduce model size for mobile GPUs [13].
- Pediatric-specific Data: Incorporating targeted data to mitigate the age bias identified in this study.

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.

Acknowledgement:-
This project is supported by UTS internal grant (2/2021/02).

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. 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.
9. E. Hassan, S. A. Ghazalah, N. El-Rashidy, T. A. El-Hafeez, and M. Y. Shams, “DenseNet model with attention mechanisms for robust date fruit image classification,”International Journal of Computational Intelligence Systems, vol. 18, no. 1, pp. 1–33, 2025.
10. S. Meyyappan, “Leveraging attention mechanisms in ResNet101 for enhanced facial emotion recognition,” presented at the 3rd International Conference on Advancement in Computation & Computer Technologies (InCACCT), June 2025.
11. R. Kumari and J. Wasim, “A deep learning approach for human facial expression recognition using Residual Network–101,”Journal of Current Science and Technology, vol. 13, no. 3, pp. 517–532, 2023.
12. F. Chollet, “Xception: Deep learning with depthwise separable convolutions,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 1251–1258.
13. T. Zhao et al., “A survey of deep learning on mobile devices: Applications, optimizations, challenges, and research opportunities,”Proceedings of the IEEE, vol. 110, no. 3, pp. 334–354, 2022.