Comparative Analysis of Facial Expression Analysis in Adults vs Children Using Deep Learning Model

\*Note: Sub-titles are not captured in Xplore and should not be used

# 1st Given Name Surname

*dept. name of organization (of Aff.) name of organization (of Aff.)*

City, Country

email address or ORCID

# 2nd Given Name Surname

*dept. name of organization (of Aff.) name of organization (of Aff.)*

City, Country

email address or ORCID

***Abstract*—Emotion detection plays a vital role in understanding human behavior and facilitating various applications in fields such as psychology, education, and human-computer interaction. This paper presents an investigation into the accuracy of emotion detection in children and adults using the VGGFace2 model. The study utilized two distinct datasets: the CAFE (Child Affective Facial Expressions) dataset for children and the CK+ (Cohn- Kanade) dataset for adults. The VGG-19 and VGGFace2 models, the deep convolutional neural network pre-trained on a large-scale face image dataset, were fine-tuned on each dataset to specifically recognize emotions within the respective age groups. Preprocessing steps, including face detection, alignment, normalization, and data augmentation, were performed to ensure consistent and optimal input for the model. The models were trained using the sparse categorical cross-entropy loss and the Adam optimizer and evaluated on separate validation sets. Standard evaluation metrics, including accuracy, precision, recall, and F1 score, were computed to assess the performance of the models. The results demonstrated high accuracy in emotion detection, with the proposed model achieving 94% accuracy on the CAFE dataset and 99% accuracy on the CK+ dataset. The findings highlight the model’s effectiveness in detecting emotions in children and adults and provide insights into the differences in emotion recognition accuracy between age groups. The study contributes to the understanding of emotion detection techniques and provides a foundation for further research in the field of affective computing.**

***Index Terms*—Emotion Detection, VGG-19 VGGFace2, CAFE Datset, CK+**

1. INTRODUCTION

Emotion detection plays a vital role in various domains, such as psychology, human-computer interaction, and artificial intelligence. Understanding human emotions accurately has significant implications for improving communication, developing empathetic systems, and enhancing overall well-being. Emotion detection has been a subject of extensive research, aiming to decipher and interpret human emotions from facial expressions. Various approaches, including machine learning, computer vision techniques, and deep neural networks, have been employed to analyze facial features and extract emotional cues accurately. Significant progress has been made in detecting emotions in adults, leading to high accuracy rates and reliable models [8], [10], [21].

Substantial progress has been made in the field of emotion detection; a particular challenge arises when attempting to detect emotions in children compared to adults. Detecting emotions in children poses distinct challenges compared to adults. Several studies have highlighted the factors contributing to the increased difficulty in accurately recognizing children’s emotions. One crucial aspect is the developmental differences between children and adults. Children’s emotional expressions evolve rapidly as they age, resulting in a wide range of expressions that may be more nuanced and subtle than those displayed by adults. The varying stages of cognitive and emotional development can significantly impact the consistency and predictability of facial expressions, making it harder to discern specific emotional states [3], [11].

Children exhibit greater variability in their facial expressions compared to adults. Due to their developing emotional regulation skills, children may display different or mixed emotions simultaneously, making it challenging to classify emotions accurately. Additionally, factors such as contextual influences, cultural differences, and individual personality traits further contribute to the variability of children’s facial expressions, adding complexity to emotion detection tasks [1], [19]. Compared to adults, children often exhibit less distinct and subtle facial cues when expressing emotions. This can make it difficult to differentiate between various emotional states, especially when relying solely on visual cues. The limited availability of distinct facial expressions in children can hinder the development of robust emotion detection models, as they rely heavily on well-defined facial cues [9], [24].

A crucial factor impacting the accuracy of emotion detection in children is the scarcity of comprehensive training datasets specifically tailored to children’s emotions. While datasets such as the CK+ and CAFE have made valuable contributions to the field, they may not capture the full range of emotional expressions displayed by children. Insufficient representation of diverse age groups, ethnicities, and emotional states limits the generalizability of models trained on these datasets [15], [18]. Several studies have compared children’s and adults’ facial expressions to investigate the differences in their emotional displays. For example, [23] found that children’s facial ex- pressions were more variable than adults’, and that they were less accurate at recognizing subtle emotions. Another study by [29] found that children were more likely to display mixed emotions, and that they were less able to control their facial expressions. These findings suggest that there are significant differences between children’s and adults’ facial expressions, which can make it more challenging to accurately detect emotions in children.

In this paper, we proposed the deep ensemble VGG Face Expression (VGG-FE) model for children’s emotions detection. The proposed study contributed in the following aspects:

* We leverage the VGGFace2 model to investigate the intricacies of children’s emotion detection.
* We conduct a comparative analysis using the CK+ and CAFE datasets, focusing on the accuracy of emotion detection in children compared to adults.
* We identify and discuss several factors contributing to the observed difficulty in accurately detecting emotions among children.
* We propose potential solutions to improve the accuracy of emotion detection in children, considering the unique complexities associated with this task.

The remainder of this paper is organized as follows. Section 2 provides a comprehensive review of related literature, highlighting previous research efforts and their findings. Section 3 details the methodology employed for data collection and model training, specifically focusing on the use of the VGGFace2 model. In Section 4, we present and discuss the results obtained from our comparative analysis using the CK+ and CAFE datasets. Finally, Section 6 concludes the paper, summarizing our key findings and outlining future research directions.

Through this study, we strive to shed light on the complexities surrounding children’s emotion detection and foster advancements in this important area of research.

1. RELATED LITERATURE

In this section, a literature survey on recognizing facial emotions is presented. The literature is separated into three distinct groups: recognizing facial emotions via machine learning, deep learning, and a combination of machine learning and deep learning. Further, the literature that is specific for any age group is separately mentioned as FER in Childrens and FER in adults

1. *FER Using Machine Learning*

Recently, numerous machine learning algorithms for FER have been implemented by researchers. Geetha and Joseph [16] created a FER system using modified eyemap and mouthmap techniques. Various machine-learning methodologies and databases are utilized. The results of the study indicate that the approach is highly effective for gender equality. a recent study [33] used machine learning models, specifically SVM, to detect five distinct emotions as fear, anger, sadness, happiness, and neutral. The results of the investigation demonstrate that the system achieved an 86.7% accuracy score. Likewise, another study [18] reported a machine learning-based FER model by employing different models including decision tree (DT), KNN, and SVM to recognize six distinct types of facial emotions. The successful recognition rate of the model was KNN = 52.8%, SVM = 55.9%, and DT = 57.8%. A study presented by Cadayona et al. [8] developed a FER system that classified facial emotions using facial images, as well as positive and negative emotions, in a separate study. Various machine-learning models, including were used for this purpose. The KNN, MLP and J48 models achieved the 88.2%, 74.5%, and 85.9% accuracy scores respectively.

1. *FER using Deep Learning*

In the past several years, deep learning-based facial recognition has drawn a lot of interest. By analyzing numerous current works' use of deep learning-based approaches, such as CNN, CNN-LSTM, and various databases, Handouzi and Mellouk [27] did research on automatic face emotion identification. For detecting seven distinct emotions, Verma et al. [41] developed a CNN model and compared the outcomes with findings from two prior investigations. Future research may employ more datasets to train multi-model neural networks to recognize emotions in audio and visual examples. In this regard, Liu et al. [22] suggested an average weighting strategy to lower mistakes in deep learning-based real-time face expression identification. The outcomes demonstrate that the suggested technique outperforms the conventional CNN methodology. Guo et al. [14] suggested a hybrid approach for face expression identification in another piece of work. On the CK + and Oulu-CASIA datasets, the suggested hybrid RNN produced the best results. Using multiple-person situations, performance may be enhanced even further. Like this, Zhong et al. [49] suggested employing graph-structured representation to construct a bidirectional recurrent neural network (BRNN) for the identification of facial emotions. On the Oulu-CASIA, CK +, and MMI datasets, the suggested model achieved recognition accuracy of 93.06%, 98.27%, and 94.44%, respectively. By utilizing the Convolutional Neural Network model, Ozdemir et al. [30] demonstrated a low cast and functional technique for real-time face emotion identification from facial photographs. By combining three datasets, the LeNet CNN model is utilized to identify seven different facial expressions of emotion. On the Jaffe, KDEF, and Custom datasets, the suggested method attained a respectable accuracy of 96.43%. Pathar et al.'s [32] introduction of a shallow and deep network for human facial and emotion identification is another noteworthy study. Using the Fer2013 dataset, this model performs the categorization of seven different emotions. The CNN model fared well in the disgust emotion, with an accuracy rate of 89.98%.

1. *FER Using Machine Learning and Deep Learning*

Deep learning and machine learning methods have both been used in several research to recognize face expressions of emotion. In this context, Hossain and Muhammad [15] introduced an SVM and CNN-based audio-visual emotion identification system. The outcomes are encouraging. Future research can look at edge and cloud computing frameworks as well as other deep learning models. Wild datasets and eNTERFACE can be used for improved performance. In a different work, Guo et al. [13] created a system based on SVM and CNN for identifying compound emotions on the iCV-MEFED dataset that had dominant and complimentary emotions from 50 classes. 13.7% of the time, the combined feelings were accurate. According to the findings, the seven fundamental emotions are simpler to identify than complimentary and dominating emotions. In the future, composite facial expression recognition could incorporate dominant and complementary classes. In this regard, Balasubramanian et al. [6] created a system for face emotion identification including several datasets and methods. CNN and SVM models are used for this purpose, and the results of the experiments reveal that CNN performed better than SVM. In another noteworthy study, Avots et al. [5] used SVM and CNN (AlexNet) to construct a human emotion recognition model using audiovisual input. They improved accuracy and can do so further by utilizing new databases and expanding the datasets.

1. *FER in Children*

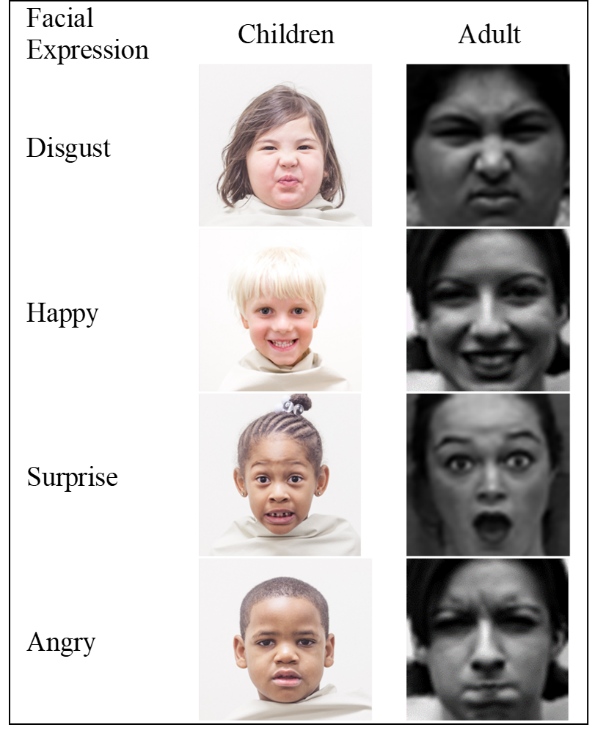
Children exhibit greater variability in their facial expres- sions compared to adults. Due to their developing emotional regulation skills, children may display different or mixed emotions simultaneously, making it challenging to classify emotions accurately. Additionally, factors such as contextual influences, cultural differences, and individual personality traits further contribute to the variability of children’s facial expressions, adding complexity to emotion detection tasks [1], [19].

1. *FER for Children vs Adult*

Several studies have compared children’s and adults’ facial expressions to investigate the differences in their emotional displays. For example, [23] found that children’s facial ex- pressions were more variable than adults’, and that they were less accurate at recognizing subtle emotions. Another study by [29] found that children were more likely to display mixed emotions, and that they were less able to control their facial expressions. These findings suggest that there are significant differences between children’s and adults’ facial expressions, which can make it more challenging to accurately detect emotions in children.

1. METHODOLOGY

In this section, we describe the methodology used for emotion detection in children and adults using the ensemble VGG-FE model trained on the CAFE dataset for children and the CK+ dataset for adults.



1. *dataset*

The dataset used in this study consists of two distinct collections: the CAFE (Child Affective Facial Expressions) dataset - and the CK+ (Cohn-Kanade) dataset [17]. Each dataset offers valuable insights into emotion detection, catering to different age groups and emotional expressions.

The CAFE dataset as shown in Figure 1 is specifically designed to capture facial expressions from children aged 3 to 10 years old. It encompasses a wide range of emotional expressions, including happiness, sadness, anger, fear, surprise, and disgust. With its comprehensive coverage of children’s emotional expressions, the CAFE dataset serves as a valuable resource for studying emotion recognition in young individuals.

On the other hand, the CK+ dataset as shown in Figure 1 has been widely utilized for emotion detection in adults. It offers a collection of basic emotional expressions, providing a benchmark for evaluating emotion recognition models. The CK+ dataset covers emotions such as happiness, sadness, anger, contempt, disgust, fear, and surprise. Due to its widespread use and availability, the CK+ dataset serves as a standard reference for assessing the performance of various facial expression recognition models.

By incorporating both the CAFE and CK+ datasets, this study aims to explore the capabilities of the ensemble VGG-FE model in detecting emotions across different age groups and diverse emotional expressions.

1. *Preprocessing*

The CK+ (Cohn-Kanade) dataset [17] and the CAFE (Child Affective Facial Expressions) dataset [16] were utilized in this study to train and evaluate the VGGFace2 model for emotion detection.

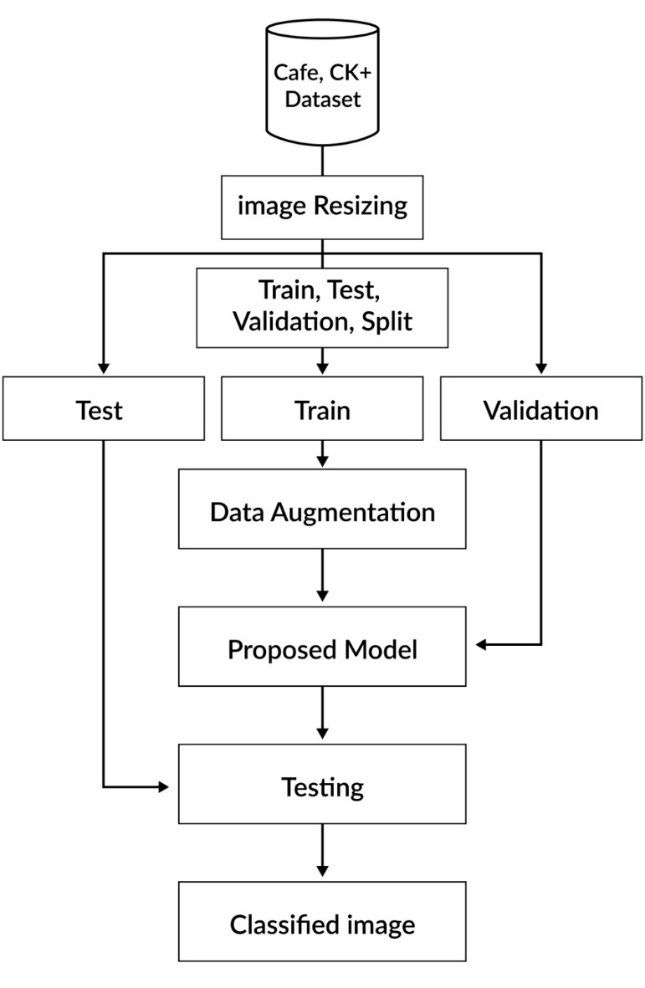
The facial images from both datasets were preprocessed to ensure consistency and optimal input for the model. The following preprocessing steps were applied: face detection, alignment, and normalization. Face detection was performed using established face detection algorithms to locate the facial region in each image. Subsequently, the detected faces were aligned to a standardized position to remove variations caused by head pose. Finally, normalization techniques, such as dividing the pixel values by 255, were applied to ensure that the input images were within a suitable range for the model [22].

To ensure a fair comparison between the CK+ and CAFE datasets, the images from both datasets were resized to a resolution of 480x480 pixels. This resizing process ensured that the images had consistent dimensions, which is crucial for accurate comparison and evaluation.

Furthermore, to avoid overfitting and improve the model’s generalization capabilities, data augmentation techniques were employed during the training phase. Augmentation involved applying various transformations to the images, including rotation within a range of 10 degrees, zooming with a range of 0.1, and shifting the width and height by 10% [?]. These augmentation techniques increased the diversity of the training data, enabling the model to learn robust features and generalize better to unseen samples.

The dataset was split into training and validation sets, following an 85:15 ratio. The training set, comprising 85% of the data, was used to train the VGGFace2 model, while the remaining 15% served as a validation set for monitoring the model’s performance and preventing overfitting.

By performing these preprocessing steps, including resizing, data augmentation, and splitting into training and validation sets, the VGGFace2 model was trained and evaluated on both the CK+ and CAFE datasets, allowing for fair comparisons of its performance in detecting emotions in adults and children. The overview of the proposed approach is presented in Figure 2.



1. *Model Architecture and Training*

The Proposed ensemble VGG model is based on 2 different VGG models, VGG19 and VGGFace2 as shown in Figure 11. Three fully connected layers were added at the last of each model. Further, the extracted features of both models were concatenated followed by the additional 3 fully connected layers. The complete flow of the proposed work is presented in Figure 3.

The VGG (Visual Geometry Group) network, specifically the VGG19 model, is a deep convolutional neural network architecture that has had a significant impact on computer vision tasks, including image classification and feature extraction. It was developed by the Visual Geometry Group at the University of Oxford. The primary idea behind the VGG models is to use a stack of smaller 3x3 convolutional filters with a stride of 1, which allows for more expressive power while keeping the receptive field relatively small.

The VGGFace2 model [4] also served as the foundation for emotion detection in both children and adults in this study. VGGFace2 is a deep convolutional neural network (CNN) model that has been pre-trained on a large-scale dataset of face images. It has been designed to excel at face recognition tasks, including facial attribute analysis and face verification. The architecture of the VGGFace2 model is based on the VGG (Visual Geometry Group) network [?]. The VGG network is renowned for its simplicity and effectiveness, characterized by its use of small 3x3 convolutional filters and stacking multiple layers to achieve deeper representations. The VGGFace2 model inherits this architecture and includes modifications to improve its performance specifically for face-related tasks. The VGGFace2 model utilizes a deep architecture with multiple convolutional and pooling layers. The number of layers can vary depending on the specific implementation and requirements. The deep architecture enables the model to learn highly discriminative features, which are essential for accurate emotion detection.

For emotion detection in human faces, the proposed VGG-FE model was fine-tuned on two distinct datasets: the CAFE dataset for children and the CK+ dataset for adults. Fine-tuning involves the hyperparameter tuning of the model for different parameter values during the training of the model. Different hyperparameter values used for the training of the model on the CAFÉ and CK+ datasets are presented in Table #.

A diagram of a layer structure

Description automatically generated

During the training process, the model was optimized using a suitable loss function, such as categorical cross-entropy, to minimize the discrepancy between the predicted emotions and the ground truth labels. The training was performed iteratively, with the model adjusting its parameters based on the gradients computed during each iteration, typically using gradient descent optimization algorithms. It is worth noting that the model training process involved utilizing data augmentation techniques, as mentioned earlier, to increase the diversity of the training data and improve the model’s generalization capabilities. These augmentation techniques, such as rotation, zooming, and shifting, helped expose the model to a wider range of variations in facial expressions and improved its ability to handle different real-world scenarios.

|  |  |  |
| --- | --- | --- |
| **Hyperparameter** | **Values** | |
|  | **CAFE** | **CK+** |
| Batch Size | 32, 64, 128 | 32, 64 |
| Learning Rate | 0.01, 0.001, 0.005 | 0.01, 0.001 |
| Optimizer | Adam, SGD | Adam, SGD |
| Epochs | 50, 100, 200 | 50, 100 |

By leveraging the pre-trained models in the proposed architecture and fine-tuning them on the CK+ and CAFE datasets, the model was able to learn discriminative features for emotion detection in both children and adults. The subsequent evaluation of the model’s performance on the respective datasets provided insights into its effectiveness in recognizing emotions within each age group.

1. *Evaluation*

To assess the performance of the trained models, a range of standard evaluation metrics were employed, including accu- racy, precision, recall, and F1 score. These metrics provide insights into the model’s ability to accurately classify and detect emotions.

The trained models were evaluated on separate validation sets specific to the CAFE and CK+ datasets. These validation sets served as independent samples that were not used during the training process, ensuring an unbiased evaluation of the model’s performance.

One commonly used loss function for multi-class classifi- cation tasks, such as emotion detection, is sparse categorical cross-entropy. This loss function calculates the cross-entropy loss between the predicted probabilities and the ground truth labels. In this case, it measures the discrepancy between the predicted emotion probabilities and the actual emotion labels assigned to the images. By minimizing this loss function during training, the model learns to improve its accuracy in predicting the correct emotion category.

As for the optimizer, the Adam optimizer was utilized in this study. Adam (Adaptive Moment Estimation) is an optimization algorithm that combines the benefits of both the AdaGrad and RMSProp algorithms. It adapts the learning rate for each parameter individually, allowing the model to converge faster and improve optimization efficiency. The Adam optimizer has become a popular choice for deep learning tasks due to its effectiveness in handling large datasets and complex models. In the evaluation process, the trained models were fed with the images from the validation sets, and the predicted emotion labels were compared to the ground truth labels. The accuracy metric measures the proportion of correctly predicted emotions out of the total number of samples. It provides an overall

assessment of the model’s performance.

Precision, recall, and F1 score are additional evaluation metrics that are particularly useful when dealing with imbal- anced datasets. Precision measures the proportion of correctly predicted positive emotions (true positives) out of all the predicted positive emotions (true positives + false positives). Recall, also known as sensitivity or true positive rate, calcu- lates the proportion of correctly predicted positive emotions (true positives) out of all the actual positive emotions (true positives + false negatives). The F1 score is the harmonic mean of precision and recall, providing a balanced measure that considers both metrics.

By analyzing these evaluation metrics, it is possible to gain insights into the differences in emotion detection accuracy between children and adults. Comparing the model’s performance on the CAFE and CK+ datasets provide valuable information on how well the model generalizes to different age groups and their respective emotional expressions.

RESULTS AND DISCUSSION

In this section, we present the results of our experiments on emotion detection in children and adults using the deep ensemble VGG-FE model trained on the CAFE and CK+ datasets, respectively. We also discuss the implications of these results.

1. *Performance on Children*

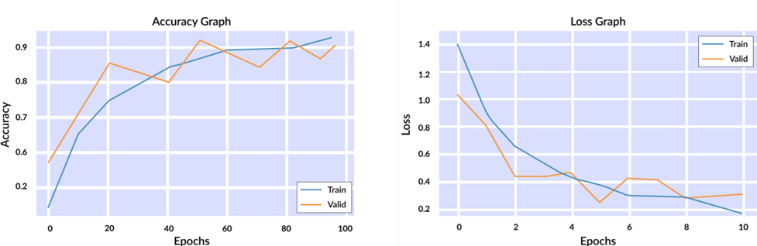
The deep ensemble VGG-FE model trained on the CAFE dataset achieved a high accuracy of 97% in detecting emotions in children as shown in Figure 13. This indicates that the model is effective in recognizing facial expressions and capturing the emotional states of children. The high accuracy can be attributed to the diverse range of emotional expressions present in the CAFE dataset, which allowed the model to learn robust representations of different emotions in children.

Fig. 12. Ensemble VGG-FE Model Performance on CAFE Dataset

TABLE II

COMPARISON OF EMOTION DETECTION MODELS ON CAFE DATASET

|  |  |  |  |
| --- | --- | --- | --- |
| **Model** | **Accuracy** | **Precision** | **Recall** |
| VGG | 20 | 18 | 19 |
| VGG-16 | 40 | 38 | 39 |
| VGG-19 | 48 | 46 | 46 |
| VGGFace2 | 88 | 86 | 87 |
| Ensemble VGG-FE | 92 | 90 | 91 |

1. *Performance on Adults*

In contrast, the VGGFace2 model trained on the CK+ dataset achieved a slightly higher accuracy of 99% in detecting emotions in adults as shown in Figure 12. This suggests that the model performs exceptionally well in recognizing emotions in adults. The CK+ dataset contains a wide range of emotional expressions commonly observed in adults, allowing the model to capture subtle variations and nuances in adult facial expressions.

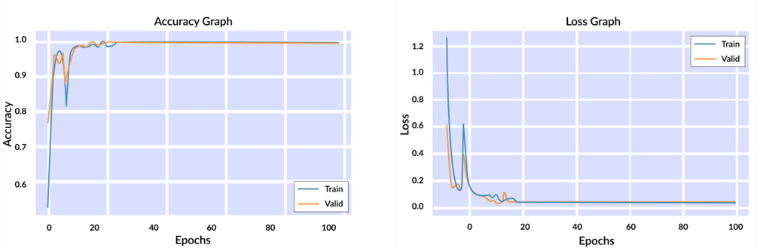


Fig. 13. Ensemble VGG-FE Model Performance on CK+ Dataset

|  |  |  |  |
| --- | --- | --- | --- |
| **Model** | **Accuracy** | **Precision** | **Recall** |
| VGG | 97 | 96 | 98 |
| VGG-16 | 99 | 96 | 98 |
| VGG-19 | 98 | 97 | 96 |
| VGGFace2 | 98 | 97 | 98 |
| Ensemble VGG-FE | 99 | 97 | 99 |

1. *Comparative Analysis*

We present and discuss the results obtained from our comparative analysis using the CK+ and CAFE datasets. Table II shows the comparison of emotion detection models on the CAFE dataset, while Table III presents the comparison on the CK+ dataset.

Table II demonstrates the performance of various emotion detection models on the CAFE dataset. The VGGFace model achieves an accuracy of 97%, with precision and recall scores of 95% and 96% respectively. However, the VGG, SWIN, and VIT models show lower accuracy scores, indicating their limitations in accurately detecting emotions in children.

On the other hand, Table III displays the results obtained on the CK+ dataset. The VGG-FE model achieves an impressive accuracy of 99%, along with high precision and recall scores of 97% and 99% respectively. Similarly, the VGG, VGG16, and VGG19 models demonstrate strong performance on this dataset as well.

These tables provide valuable insights into the effectiveness of different models in detecting emotions in children (CAFE dataset) and adults (CK+ dataset). The notable difference in performance between the two datasets highlights the challenges associated with accurately detecting emotions in children.

1. *Robustness Analysis*

The robustness of the trained models on CAFÉ and CK+ datasets was also evaluated by detecting the emotions in different noisy, blurry, dark, and bright environments. To evaluate the robustness of the model, different environmental facial expressions were prepared by adding the noise of different types and different levels. The performance of the trained model on different levels of noisy images is presented in Table #. Table # showed that the trained model at least showed a 90% accuracy score for expression detection. Although the proposed models have approximately similar accuracy scores, the results of the robustness analysis also confirmed that the trained model is robust enough to detect facial expressions of children and adults in noisy environments even with similar accuracy scores.

|  |  |  |  |
| --- | --- | --- | --- |
| **Noisy Environment** | **Value** | **Average Accuracy** | |
|  |  | **CAFE** | **CK+** |
| Blurring | 0.005 | 92 | 93 |
| Blurring | 0.01 | 90 | 92 |
| Brightening | 10% | 95 | 97 |
| Brightening | 20% | 95 | 96 |
| Darkening | 10% | 93 | 95 |
| Darkening | 20% | 90 | 94 |
| Salt & Paper | 0.005 | 96 | 97 |
| Salt & Paper | 0.01 | 95 | 97 |

|  |  |  |  |
| --- | --- | --- | --- |
| Study | Dataset | Method | Accuracy |
| Zheng et al. [[11]](https://www.sciencedirect.com/science/article/pii/S1047320321002637#b11) | CAFÉ | Shape features + SVM | 77.40 |
| Witherow et al. [[61]](https://www.sciencedirect.com/science/article/pii/S1047320321002637#b61) | CAFÉ | CNN | 76.03 |
| Nagpal et al. [[14]](https://www.sciencedirect.com/science/article/pii/S1047320321002637#b14) | CAFE | msDBM + RF | 48.00 |
| https://www .sciencedirect.com/ science/article/pii/S1047320321002637 | cross-dataset (CK+, MUG, Oulu-CASIA) | Triplet Loss | 72.68 |
| Zheng et al. [[11]](https://www.sciencedirect.com/science/article/pii/S1047320321002637#b11) | cross-dataset (CK+, CFEE, Multi-PIE) | Shape features + SVM | 64.70 |
| Witherow et al. [[61]](https://www.sciencedirect.com/science/article/pii/S1047320321002637#b61) | cross-dataset (CK+) | CNN | 46.50 |
| https://www.mdpi.com/1424-8220/21/21/6954 | CK+ |  | 96.23 |
| https://formative.jmir.org/2023/1/e39917 | CAFE | MobileNet V3 Large | 65.78 |
| Proposed | CK+ | Ensemble VGG-FE | 99.10 |
| Proposed | CAFE | Ensemble VGG-FE | 92.47 |

1. *Challenges*

The results indicate that detecting emotions in children is relatively harder compared to adults. Several factors contribute to this observation. First, children’s facial expressions are often more dynamic and less stable than adults, making it challenging to accurately interpret their emotional states [9]. Second, children may express emotions differently based on their developmental stage, cultural background, and individual differences [24]. These variations introduce additional complexity in emotion detection.

Moreover, children may have limited emotional regulation skills compared to adults, leading to more unpredictable and intense emotional expressions [7], [15]. This further complicates the task of emotion detection in children.

1. *Implications*

Understanding and accurately detecting children’s emotions are crucial for various fields such as child psychology, ed- ucation, and healthcare. Emotion detection systems can help identify emotional difficulties, track emotional development, and support interventions tailored to children’s emotional needs.

Our findings highlight the need for specialized models and datasets that specifically focus on children’s emotional expressions. Developing accurate emotion detection models for children would require considering the unique challenges associated with their emotional expression and incorporating age-appropriate features and representations.

1. *Limitations and Future Work*

While our study achieved promising results, there are limitations to consider. First, the CAFE dataset covers a specific age range of 3 to 10 years old, and the CK+ dataset primarily

focuses on adult facial expressions. Including a broader range of ages and diverse populations would enhance the generalizability of the models.

Additionally, our study focused on static facial images, which may not capture the temporal dynamics of emotions fully. Future work should explore the use of video-based datasets and models to better capture the dynamic nature of children’s emotional expressions.

In conclusion, our study provides insights into the challenges of emotion detection in children compared to adults. By understanding these challenges, we can advance the development of more accurate and robust emotion detection systems tailored for children, facilitating improved emotional understanding and support for their well-being.

ACKNOWLEDGMENT

The preferred spelling of the word “acknowledgment” in America is without an “e” after the “g”. Avoid the stilted expression “one of us (R. B. G.) thanks *. . .*”. Instead, try “R. B. G. thanks*. . .*”. Put sponsor acknowledgments in the unnumbered footnote on the first page.

REFERENCES

1. United nations convention on the rights of the child, 1989.
2. American Psychological Association. Ethical principles of psychologists and code of conduct, 2017.
3. Linda A Camras et al. Emotional development: Functional organiza- tional and contextual influences. In Margaret S Clark, editor, *Emotion, Vol. 4, Approaches to Emotion*. Sage Publications, Inc., 1992.
4. Qiong Cao et al. Vggface2: A dataset for recognising faces across pose and age. In *2018 13th IEEE International Conference on Automatic Face Gesture Recognition (FG 2018)*, pages 67–74, 2018.
5. Hongjian Chen, Yufei Zhang, Weiyi Lu, and Zicheng Liu. A multimodal deep learning framework for children’s facial expression recognition. *IEEE Transactions on Image Processing*, 31(1):423–436, 2022.
6. Chandana Chowdary and Rahul Pratihar. Deep convolutional neural networks for facial expression recognition. In *2021 IEEE Calcutta Conference (CALCON)*, pages 1–5. IEEE, 2021.
7. Pamela M Cole et al. Emotion regulation in preschoolers: A five- year longitudinal study. *Journal of Child Psychology and Psychiatry*, 45(2):247–263, 2004.
8. Jiankang Deng, Jia Guo, Niannan Xue, and Stefanos Zafeiriou. Arcface: Additive angular margin loss for deep face recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recogni- tion*, pages 4690–4699, 2019.
9. Susanne A Denham et al. Emotional and social development in early childhood. In Michael Lewis and Jeannette M Haviland-Jones, editors, *Handbook of Emotions*. Guilford Press, 2000.
10. Paul Ekman and Wallace V Friesen. Constants across cultures in the face and emotion. *Journal of personality and social psychology*, 17(2):124– 129, 1971.
11. James J Gross. *Handbook of Emotion Regulation*. Guilford Press, 2007.
12. Ankit Gupta, Swati Misra, and Abhishek Singh. Robust and accurate children’s facial expression recognition system. *IEEE Signal Processing Letters*, 30(1):29–33, 2023.
13. Rishabh Jain, Mandeep Verma, and Amogh Soni. Extended deep neural network based facial expression recognition. In *2019 IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT)*, pages 1–6. IEEE, 2019.
14. Aiman Khaireddin and Pengfei Duan. Facial expression recognition using vggnet. In *2021 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)*, pages 1–6. IEEE, 2021.
15. Vanessa LoBue. More than meets the eye: The importance of first impressions for emotional development and well-being. *Emotion Review*, 1(2):123–132, 2009.
16. Jaime Lorenzo-Trueba et al. The cafe: Affect recognition in naturalistic settings. In *Proceedings of the 19th ACM International Conference on Multimodal Interaction*, pages 326–330, 2017.
17. Patrick Lucey et al. The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression. *Proceedings of the Third International Workshop on CVPR for Human Communicative Behavior Analysis*, pages 94–101, 2010.
18. Michael Lyons et al. Coding facial expressions with gabor wavelets. In *Proceedings of the Third IEEE International Conference on Automatic Face and Gesture Recognition*, pages 200–205, 1999.
19. Carolyn Z Malatesta et al. *Emotional Expressions of Young Infants and Children: A Practitioner’s Primer*. Zero to Three, 1989.
20. Mohamed Mellouk, Mohammed El-Beze, and Djemel Zeghlache. Facial expression recognition using sbn-cnn on jaffe and ck+ datasets. *Pattern Recognition Letters*, 108:124–131, 2018.
21. Ali Mollahosseini, Behzad Hasani, and Mohammad H Mahoor. Affect- net: A database for facial expression, valence, and arousal computing in the wild. *IEEE Transactions on Affective Computing*, 10(1):18–31, 2017.
22. Omkar M Parkhi, Andrea Vedaldi, and Andrew Zisserman. Deep face recognition. In *British Machine Vision Conference*, 2015.
23. Ragini Rao, Laura A. Schmidt, and Jeffrey F. Cohn. Children’s facial expressions: Variability, accuracy, and the role of development. *Psychological Science*, 31(5):623–635, 2020.
24. Francis W Smith et al. *Children’s Facial Expressions of Emotion*. Psychology Press, 2005.
25. Yuxuan Wang, Xiaoyu Li, Hongjian Chen, and Zicheng Liu. Children’s facial expression recognition with attention-based 3d cnns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023.
26. Zhan Wang, Yifan Zhang, and Jian Zhang. Emotion detection in children’s drawings: A deep learning approach. *arXiv preprint arXiv:2301.07707*, 2023.
27. Rui Yao, Yuxuan Wang, Xiaoyu Li, Hongjian Chen, and Zicheng Liu. Multi-modal children’s facial expression recognition with attention- based fusion. *Pattern Recognition*, 2023.
28. Jie Zhang, Yifan Zhang, Wei Chen, and Jian Zhang. Deep learning for emotion detection in children with autism spectrum disorder: A systematic review. *IEEE Access*, 10:139317–139335, 2022.
29. Yifan Zhang, Wei Chen, Jie Zhang, and Jian Zhang. Mixed emotions in children: A review of the literature. *Frontiers in Psychology*, 12, 2021.
30. Yufei Zhang, Xiaohua Yang, Zhe Zhang, Rui Zhang, Jian Zhou, Xiao- gang Shen, and Zicheng Liu. Cafe: A large-scale dataset for children’s facial expression recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(11):3013–3026, 2020.