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 [1] [2][3].

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 [4], [5].

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, contextual influences, cultural differences, and individual personality traits contribute to the variability of children’s facial expressions, adding complexity to emotion detection tasks [6], [7]. 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 [8].

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 [9]. Several studies have compared children’s and adults’ facial expressions to investigate the differences in their emotional displays. For example, [10] found that children’s facial expressions were more variable than adults’, and that they were less accurate at recognizing subtle emotions. Another study by [11] 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 two distinct groups: recognizing facial emotions via machine learning, and facial expression recognition via deep learning. Further, the literature that is specific for any age group is separately mentioned as FER in children and FER in adults.

1. *FER Using Machine Learning*

Recently, numerous machine learning algorithms for FER have been implemented by researchers. Geetha and Joseph [12] 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 [13] 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 [14] 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. [15] 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 [16] did research on automatic face emotion identification. For detecting seven distinct emotions, Verma et al. [17] 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. [18] 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. [19] 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. [20] 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. [21] 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 [22] 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 for Children*

Since there is a wealth of information on adult emotions, both conventional and deep learning techniques have been used to recognize facial emotions in databases of adult faces. However, there is little research on FER in kids for a variety of reasons, one of which is the scarcity of kid-emotional datasets. Modern techniques have not yet been trained using standardized databases that offer 100% "unbiased" or spontaneous and genuine emotions of youngsters [23]. The current databases additionally provide participants' changes in emotion, stance, and lighting across several geographic locations [24]. Similar problems with publicly accessible databases include the exclusion of children's facial expressions and the inclusion of solely adult facial expressions [25]. As a result, little research has been done on the identification of spontaneous infant facial expressions [26].

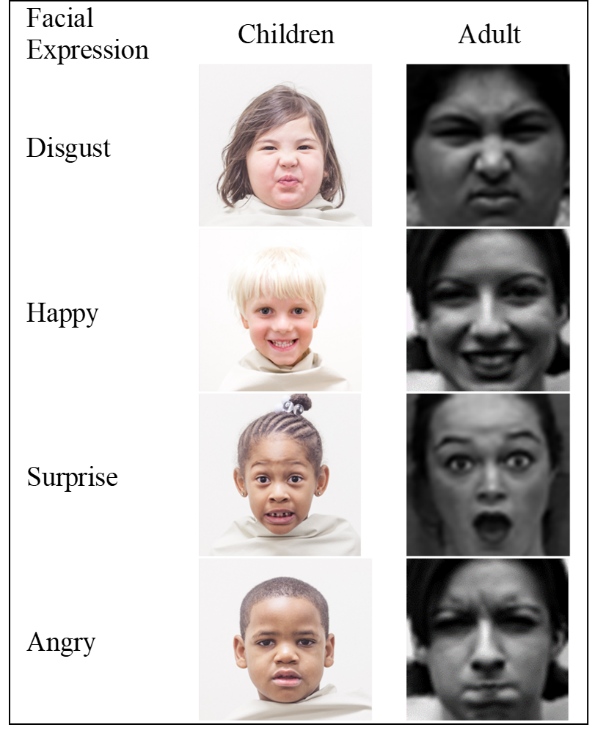
1. *FER for Children vs Adults*

Numerous studies have contrasted the facial expressions of adults and children to examine the variations in their emotional emotions. In their study, Witherow et al. [27], [28] choose the first strategy and suggest using the transfer learning method to their advantage. The CK+ [24] dataset of adult facial expressions is used by the authors to train a model based on CNN, while the CAFE [29] dataset of children facial expressions is used to fine-tune the proposed model. Similar questions about the adaptability of a trained model for adults to children are raised in different research by Zheng et al. [28]. A FER model trained on adult data, is more accurate when tested on adult data, while a model trained on children’s data, is more accurate when evaluated on children’s data, according to the authors. Similar findings from Qayyum et al. [30] support those from Zheng et al. They develop a progressive lightweight shallow network for analyzing spontaneous facial behaviour in children using the LIRIS-CSE [31] children dataset.

There is a covariate shift [32] between adults and children due to the differences in the morphology of their faces and the manner in which they exhibit their facial emotions. In particular, it is claimed that despite the fact that adults and children exhibit the same emotions, a number of facial features are different [28]. As demonstrated by a number of research [27], [28], models that are trained on adult facial expressions function poorly when evaluated on kid facial expressions because of these differences in the expression.

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 [24]. Each dataset offers valuable insights into emotion detection, catering to different age groups and emotional expressions.

The CAFE dataset [29] 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 and the CAFE (Child Affective Facial Expressions) dataset were utilized in this study to train and evaluate the proposed ensemble VGG-FE model for emotion detection.

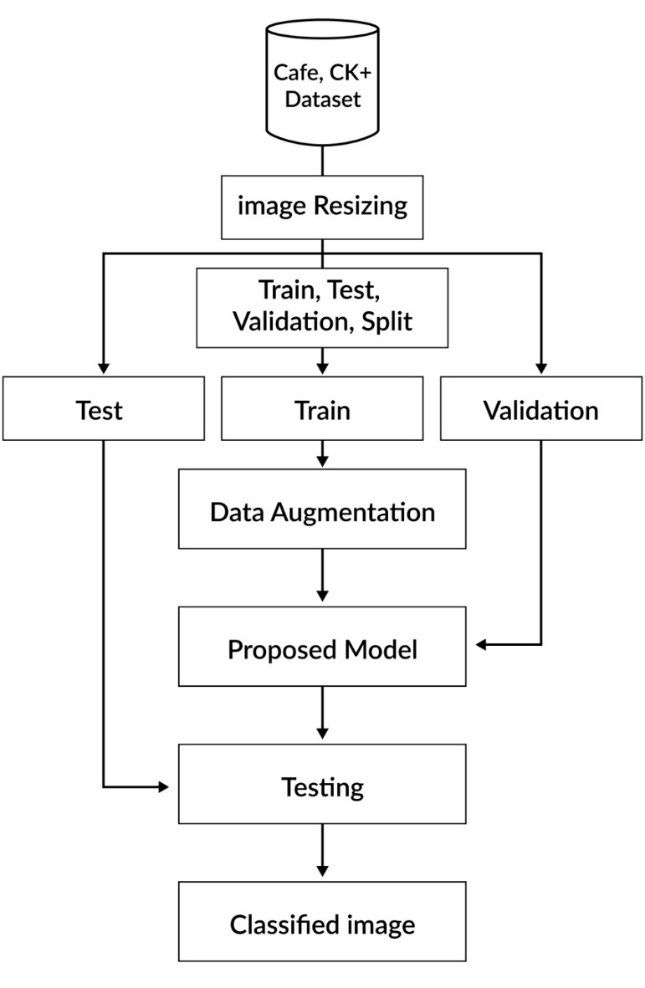
The facial images from both datasets were preprocessed to ensure consistency and optimal input for the model. The following preprocessing steps were applied: face detection, alignment, and normalization. Face detection was performed using established face detection algorithms to locate the facial region 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 [33].

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 ensemble VGG-FE 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 VGG-FE 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 network [34], 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 [35] 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 [34]. 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*

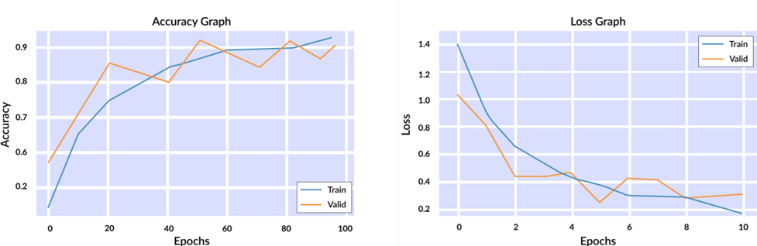
The deep ensemble VGG-FE model trained on the CAFE dataset achieved a high accuracy of 94% in detecting emotions in children as shown in Figure 13. This indicates that the model is effective in recognizing facial expressions and capturing the emotional states of children. The high accuracy can be attributed 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. The trained model showed a 92% accuracy score on the test samples of the CAFÉ dataset in Table 2.

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. The trained model also showed a 99% accuracy score on the test samples of the CK+ dataset in Table 3.

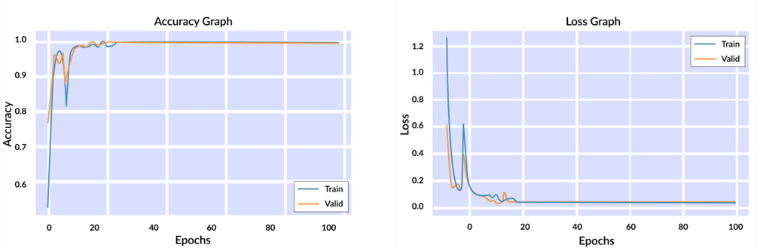


Fig. 13. Ensemble VGG-FE Model Performance on CK+ Dataset

|  |  |  |  |
| --- | --- | --- | --- |
| **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 of the CK+ dataset.

Table II demonstrates the performance of various emotion detection models on the CAFE dataset. The proposed VGG-FE model achieves an accuracy of 92%, with precision and recall scores of 90% and 91% 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 performance of the proposed ensemble VGG-FE model was significant on CAFÉ and CK+ datasets compared to the available state-of-the-art studies in Table 3. 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 85% 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 | 87 | 93 |
| Blurring | 0.01 | 85 | 92 |
| Brightening | 10% | 90 | 97 |
| Brightening | 20% | 90 | 96 |
| Darkening | 10% | 88 | 95 |
| Darkening | 20% | 85 | 94 |
| Salt & Paper | 0.005 | 91 | 97 |
| Salt & Paper | 0.01 | 90 | 97 |

|  |  |  |  |
| --- | --- | --- | --- |
| Study | Dataset | Method | Accuracy |
| Zheng et al. [28] | CAFÉ | Shape features + SVM | 77.40 |
| Witherow et al. [27] | CAFÉ | CNN | 76.03 |
| Nagpal et al. [36] | CAFE | msDBM + RF | 48.00 |
| Dias et al. [37] | cross-dataset (CK+, MUG, Oulu-CASIA) | Triplet Loss | 72.68 |
| Zheng et al. [28] | cross-dataset (CK+, CFEE, Multi-PIE) | Shape features + SVM | 64.70 |
| Witherow et al. [27] | cross-dataset (CK+) | CNN | 46.50 |
| Park et al. [38] | CK+ | CNN | 96.23 |
| Banerjee et al [39] | 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 [8]. Second, children may express emotions differently based on their developmental stage, cultural background, and individual differences. 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. 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, education, 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] J. Deng, J. Guo, N. Xue, and S. Zafeiriou, “ArcFace: Additive Angular Margin Loss for Deep Face Recognition”, Accessed: Jul. 14, 2023. [Online]. Available: https://github.com/

[2] P. Ekman and W. V. Friesen, “Constants across cultures in the face and emotion,” *J Pers Soc Psychol*, vol. 17, no. 2, pp. 124–129, Feb. 1971, doi: 10.1037/H0030377.

[3] A. Mollahosseini, B. Hasani, and M. H. Mahoor, “AffectNet: A Database for Facial Expression, Valence, and Arousal Computing in the Wild,” *IEEE Trans Affect Comput*, vol. 10, no. 1, pp. 18–31, Aug. 2017, doi: 10.1109/taffc.2017.2740923.

[4] L. A. Camras, “Differentiation, Dynamical Integration and Functional Emotional                Development,” *http://dx.doi.org/10.1177/1754073910387944*, vol. 3, no. 2, pp. 138–146, Apr. 2011, doi: 10.1177/1754073910387944.

[5] J. J. (Professor of psychology) Gross, “Handbook of emotion regulation,” p. 669.

[6] “The United Nations Convention on the Rights of the Child - UNICEF UK.” https://www.unicef.org.uk/rights-respecting-schools/the-rrsa/introducing-the-crc/ (accessed Jul. 14, 2023).

[7] M. W. Sullivan and M. Lewis, “Emotional expressions of young infants and children a practitioner’s primer,” *Infants Young Child*, vol. 16, no. 2, pp. 120–142, 2003, doi: 10.1097/00001163-200304000-00005.

[8] S. A. Denham, “Social and Emotional Learning, Early Childhood,” *Encyclopedia of Primary Prevention and Health Promotion*, pp. 1009–1018, 2003, doi: 10.1007/978-1-4615-0195-4\_147.

[9] M. Lyons, S. Akamatsu, M. Kamachi, and J. Gyoba, “Coding facial expressions with Gabor wavelets,” *Proceedings - 3rd IEEE International Conference on Automatic Face and Gesture Recognition, FG 1998*, pp. 200–205, 1998, doi: 10.1109/AFGR.1998.670949.

[10] J. Y. F. Lau *et al.*, “Individual Differences in Children’s Facial Expression Recognition Ability: The Role of Nature and Nurture,” *Dev Neuropsychol*, vol. 34, no. 1, p. 37, Jan. 2009, doi: 10.1080/87565640802564424.

[11] F. W. Paulus, S. Ohmann, E. Möhler, P. Plener, and C. Popow, “Emotional Dysregulation in Children and Adolescents With Psychiatric Disorders. A Narrative Review,” *Front Psychiatry*, vol. 12, p. 628252, Oct. 2021, doi: 10.3389/FPSYT.2021.628252/BIBTEX.

[12] A. Joseph and P. Geetha, “Facial emotion detection using modified eyemap–mouthmap algorithm on an enhanced image and classification with tensorflow,” *Visual Computer*, vol. 36, no. 3, pp. 529–539, Mar. 2020, doi: 10.1007/S00371-019-01628-3/FIGURES/14.

[13] M. Patil and S. Veni, “Driver emotion recognition for enhancement of human machine interface in vehicles,” *Proceedings of the 2019 IEEE International Conference on Communication and Signal Processing, ICCSP 2019*, pp. 420–424, Apr. 2019, doi: 10.1109/ICCSP.2019.8698045.

[14] M. Kolodziej, A. Majkowski, R. J. Rak, P. Tarnowski, and T. Pielaszkiewicz, “Analysis of facial features for the use of emotion recognition,” *Proceedings of 2018 19th International Conference Computational Problems of Electrical Engineering, CPEE 2018*, Oct. 2018, doi: 10.1109/CPEE.2018.8507137.

[15] A. M. Cadayona, N. M. S. Cerilla, D. M. M. Jurilla, A. K. D. Balan, and J. C. De Goma, “Emotional State Classification: An Additional Step in Emotion Classification through Face Detection,” *2019 IEEE 6th International Conference on Industrial Engineering and Applications, ICIEA 2019*, pp. 667–671, May 2019, doi: 10.1109/IEA.2019.8715171.

[16] W. Mellouk and W. Handouzi, “Facial emotion recognition using deep learning: review and insights,” *Procedia Comput Sci*, vol. 175, pp. 689–694, Jan. 2020, doi: 10.1016/J.PROCS.2020.07.101.

[17] A. Verma, P. Singh, and J. S. Rani Alex, “Modified Convolutional Neural Network Architecture Analysis for Facial Emotion Recognition,” *International Conference on Systems, Signals, and Image Processing*, vol. 2019-June, pp. 169–173, Jun. 2019, doi: 10.1109/IWSSIP.2019.8787215.

[18] K. C. Liu, C. C. Hsu, W. Y. Wang, and H. H. Chiang, “Real-Time facial expression recognition based on cnn,” *Proceedings of 2019 International Conference on System Science and Engineering, ICSSE 2019*, pp. 120–123, Jul. 2019, doi: 10.1109/ICSSE.2019.8823409.

[19] J. M. Guo, P. C. Huang, and L. Y. Chang, “A hybrid facial expression recognition system based on recurrent neural network,” *2019 16th IEEE International Conference on Advanced Video and Signal Based Surveillance, AVSS 2019*, Sep. 2019, doi: 10.1109/AVSS.2019.8909888.

[20] L. Zhong, C. Bai, J. Li, T. Chen, S. Li, and Y. Liu, “A graph-structured representation with BRNN for static-based facial expression recognition,” *Proceedings - 14th IEEE International Conference on Automatic Face and Gesture Recognition, FG 2019*, May 2019, doi: 10.1109/FG.2019.8756615.

[21] M. A. Ozdemir, B. Elagoz, A. Alaybeyoglu, R. Sadighzadeh, and A. Akan, “Real time emotion recognition from facial expressions using CNN architecture,” *TIPTEKNO 2019 - Tip Teknolojileri Kongresi*, Oct. 2019, doi: 10.1109/TIPTEKNO.2019.8895215.

[22] R. Pathar, A. Adivarekar, A. Mishra, and A. Deshmukh, “Human Emotion Recognition using Convolutional Neural Network in Real Time,” *Proceedings of 1st International Conference on Innovations in Information and Communication Technology, ICIICT 2019*, Apr. 2019, doi: 10.1109/ICIICT1.2019.8741491.

[23] P. Ekman and W. V. Friesen, “A new pan-cultural facial expression of emotion,” *Motiv Emot*, vol. 10, no. 2, pp. 159–168, Jun. 1986, doi: 10.1007/BF00992253/METRICS.

[24] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, “The extended Cohn-Kanade dataset (CK+): A complete dataset for action unit and emotion-specified expression,” *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops, CVPRW 2010*, pp. 94–101, 2010, doi: 10.1109/CVPRW.2010.5543262.

[25] M. Pantic, M. Valstar, R. Rademaker, and L. Maat, “Web-based database for facial expression analysis,” *IEEE International Conference on Multimedia and Expo, ICME 2005*, vol. 2005, pp. 317–321, 2005, doi: 10.1109/ICME.2005.1521424.

[26] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You Only Look Once: Unified, Real-Time Object Detection.” pp. 779–788, 2016. Accessed: Jul. 14, 2023. [Online]. Available: http://pjreddie.com/yolo/

[27] M. A. Witherow, M. D. Samad, and K. M. Iftekharuddin, “Transfer learning approach to multiclass classification of child facial expressions,” *https://doi.org/10.1117/12.2530397*, vol. 11139, pp. 275–282, Sep. 2019, doi: 10.1117/12.2530397.

[28] Z. Zheng, X. Li, J. Barnes, C. H. Park, and M. Jeon, “Facial Expression Recognition for Children: Can Existing Methods Tuned for Adults Be Adopted for Children?,” *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 11567 LNCS, pp. 201–211, 2019, doi: 10.1007/978-3-030-22643-5\_16/COVER.

[29] V. LoBue and C. Thrasher, “The Child Affective Facial Expression (CAFE) set: Validity and reliability from untrained adults,” *Front Psychol*, vol. 5, no. OCT, p. 127200, Jan. 2014, doi: 10.3389/FPSYG.2014.01532/BIBTEX.

[30] A. Qayyum, I. Razzak, N. Moustafa, and M. Mazher, “Progressive ShallowNet for large scale dynamic and spontaneous facial behaviour analysis in children,” *Image Vis Comput*, vol. 119, p. 104375, Mar. 2022, doi: 10.1016/J.IMAVIS.2022.104375.

[31] R. A. Khan, A. Crenn, A. Meyer, and S. Bouakaz, “A novel database of children’s spontaneous facial expressions (LIRIS-CSE),” *Image Vis Comput*, vol. 83–84, pp. 61–69, Mar. 2019, doi: 10.1016/J.IMAVIS.2019.02.004.

[32] H. Shimodaira, “Improving predictive inference under covariate shift by weighting the log-likelihood function,” *J Stat Plan Inference*, vol. 90, no. 2, pp. 227–244, Oct. 2000, doi: 10.1016/S0378-3758(00)00115-4.

[33] O. M. Parkhi, A. Vedaldi, and A. Zisserman, “Deep Face Recognition”.

[34] K. Simonyan and A. Zisserman, “Very Deep Convolutional Networks for Large-Scale Image Recognition,” *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, Sep. 2014, Accessed: Jul. 14, 2023. [Online]. Available: https://arxiv.org/abs/1409.1556v6

[35] Q. Cao, L. Shen, W. Xie, O. M. Parkhi, and A. Zisserman, “VGGFace2: A dataset for recognising faces across pose and age,” *Proceedings - 13th IEEE International Conference on Automatic Face and Gesture Recognition, FG 2018*, pp. 67–74, Oct. 2017, doi: 10.1109/FG.2018.00020.

[36] S. Nagpal, M. Singh, M. Vatsa, R. Singh, and A. Noore, “Expression Classification in Children Using Mean Supervised Deep Boltzmann Machine.” pp. 0–0, 2019.

[37] W. Dias *et al.*, “Cross-dataset emotion recognition from facial expressions through convolutional neural networks,” *J Vis Commun Image Represent*, vol. 82, p. 103395, Jan. 2022, doi: 10.1016/J.JVCIR.2021.103395.

[38] S. J. Park, B. G. Kim, and N. Chilamkurti, “A Robust Facial Expression Recognition Algorithm Based on Multi-Rate Feature Fusion Scheme,” *Sensors 2021, Vol. 21, Page 6954*, vol. 21, no. 21, p. 6954, Oct. 2021, doi: 10.3390/S21216954.

[39] A. Banerjee, O. C. Mutlu, A. Kline, S. Surabhi, P. Washington, and D. P. Wall, “Training and Profiling a Pediatric Facial Expression Classifier for Children on Mobile Devices: Machine Learning Study,” *JMIR Form Res 2023;7:e39917 https://formative.jmir.org/2023/1/e39917*, vol. 7, no. 1, p. e39917, Mar. 2023, doi: 10.2196/39917.