International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
IJISAE
ISSN:2147-6799 www.ijisae.org Original Research Paper

Classifying Feelings Using Facial Expression Recognition

Reem Khalid Sahi[1], M. M. Elgayar*[1], Noha A. Hikal[1]

Submitted:15/03/2024 **Revised**: 24/04/2024 **Accepted**: 07/05/2024

Abstract: This paper challenges the recent methodologies of identify the identical facial expressions in this study we use a dataset containing 48x48 pixel grayscale face images. The importance lies in classifying seven emotions (Angry, Disgust, Fear, Happy, Sad, Surprise, Neutral) because their applications in human-computer interaction. We implemented three deep learning models (Regular Convolutional Neural Network (CNN), Attention-Aided CNN, and ResNet50), to evaluate their performance. The results show that the regular deep CNN achieved the highest accuracy at 79% for identifying the identical faces. These findings underscore the potential of networks depth in capturing most of facial expression patterns. This study contributes to the field by illustrating the effectiveness of different models in emotion recognition and provides insights for future research to improve classification accuracy and practical applications.

Keywords: *Facial Expression, Classification, CNN, Attention, Deep Learning, ResNet50.*

1. Introduction

Recognition of human facial expressions forms the most crucial task for the development of advanced human-computer interactive systems, psychological research, and security systems. Thus, the automated recognition of human facial expressions is one area with increasing interest in vast numbers of research areas. This paper studies the use of deep learning models for the classification of facial expressions into the seven classes of emotion: angry, disgust, fear, happy, sad, surprise, and neutral, with the use of a dataset that consists of faces with a size of 48x48 pixels of grayscale as shown in Figure 1. Training and testing databases are collections of 28,709 and 3,589 examples, respectively, from automatically aligned faces to make them consistent in images [1]. In the paper, three deep learning models have considered important: standard Convolutional Neural Network (CNN), CNN with improved attention mechanisms, and ResNet50, for comparing in emotion classification [2,3]. The differences in each model are very broad, providing a broad-spectrum analysis of feature extraction and the impact of its architectural complexities on the classification accuracy [4, 5]. Our findings, therefore, underpin that the superior performance of ResNet50, using the deep residual learning framework, accomplished a significant classification accuracy of 72% and hence opened great possibilities in capturing subtleties associated with facial expressions [6].

This work gives a very important contribution to the field of computer vision, more precisely to the research on emotion recognition, pointing at the effectiveness of ResNet50 in the problem of facial expression classification tasks and drawing parallels with the influence of the network architecture on emotion recognition [7, 8]. The outcomes of this study would yield more than just an insight into automatic emotion recognition but further give an impetus to the research for improving classification accuracy and real-world investigations on practical applications. This would also be valuable for providing an input toward the formation of an effective framework for practical applications.

[1] *Department of information Technology, Faculty of Computers and Information– 35516, Mansoura University, Egypt*

Corresponding Author: M.M.Elgayar
ORCID ID: 0000-0002-4644-1835
Email: Mostafa_elgayar@mans.edu.eg

1.1. Background and importance of facial expression recognition.

Facial expression recognition (FER) bears a great impact on human health, human-computer interaction (HCI), robotics, and security, as it requires empowering machines to understand human expressions and feel their emotions genuinely [1,2]. For every individual, the most important form of non-verbal language is the recognition and interpretation of facial expressions, by which a person is able to recognize the emotional state or intentions of another person [4]. In healthcare, the technology is useful in-patient monitoring, which includes pain assessment and the review of mental health, providing a clear, non-invasive way of measuring emotional and psychological well-being [5]. In the domain of HCI and robotics, emotion-aware interfaces or robots would help in making the user experience better, where responses can be adapted according to the perceived emotions, thus making the interaction more natural and, therefore, more effective.

The great refinement and improvement in convolutional neural network (CNN) models have carried the field of FER to another high, making it capable of creating tools that can capture and analyze even the most delicate and sensitive subtleties of facial expressions. For example, ResNet50, based on its powerful capabilities in deep feature extraction and large potential to learn from big datasets, is one of those deep learning models that has shown success in many FER tasks [9,10]. Besides, the incorporation of attention mechanisms in CNNs will further fine tune model performance with focus on salient features that are relevant to recognition in the emotional context, hence increasing accuracy and reliability for emotion classification.

Figure 1. Facial Emotions Classes

1.2. Objective of the study

Therefore, this study, by getting into the details, seeks to explore, affirm, and re-establish the performance of some of the most recent state-of-the-art deep learning models for better detection and classification into seven discreetly different emotions: Angry, Disgust, Fear, Happy, Sad, Surprise, and Neutral. The main aim of this research study is to examine the potential of a 48 × 48 pixels grayscale image dataset in two aspects: firstly, to further reveal the most promising methods from the wide field of automatic techniques for emotion recognition and, secondly, to apply these methods within the wide range of domains this research includes, like domains dealing with human-computer interaction, psychological research, and security systems. Specifically, the study aims to:

This work will consider three deep learning models: a basic Convolutional Neural Network (CNN), a CNN with attention mechanisms, and ResNet50 against recognition of facial expression. Its heterogeneity allows a general analysis of the considered architectures in feature extraction and in the impact their architectural complexity causes to the classification accuracy. From the classification accuracies, therefore, the selected models were compared to conclude which architecture could be the most effective at capturing the nuanced patterns associated with facial expressions. The most interesting one is ResNet50, which is of focus since it has a deep residual learning framework and is expected to perform much better in the context of emotion recognition.

These results provide insight into how network architecture factors into emotion recognition and help to drive the field of computer vision and emotion recognition forward. This paves the way for future work increasing the classification accuracy in the quest to find practical applications of these models in real scenarios.

The proposed research aims to achieve the objectives relevant to the technological advancements in automatic emotion recognition and at the same time enhances a resourceful contribution to further research with respect to the optimization of the interaction between human and machine and improvement in assessment methodology in psychology with proper security measures through nuanced detection of emotion.

2. Related Work

Previous studies on facial expression recognition have extensively explored various methodologies, highlighting the continuous evolution of this field. Key studies include: Wei Du explored the use of improved ResNet50 models for facial emotion recognition, showcasing enhanced accuracy through architectural modifications [1].

Pratyush Shukla and Mahesh Kumar provided an extensive overview of ResNet's application in facial expression recognition, emphasizing its significant impact and the advantages of residual learning [2].

Marde Fasma’ul Aza, N. Suciati, and S. Hidayati compared different pre-trained CNN models on facial expression datasets, highlighting the superior performance of VGG16 under specific conditions [3].

Narinderpal Kaur focused on the application of CNNs for emotion recognition, demonstrating their effectiveness in classifying basic emotional states [4].

Tianyang Zheng conducted experiments with various network structures, including ResNet and DenseNet, for facial emotion recognition, aiming to understand the efficacy of different architectures [5].

Bin Li and Dimas Lima investigated the application of ResNet-50 for facial emotion recognition, showing improvements in detection performance due to the model's deep learning capabilities [6].

2.1 Discussion on the Use of CNN, Attention Mechanisms, and ResNet50 in Similar Tasks

The use of Convolutional Neural Networks (CNNs), attention mechanisms, and ResNet50 in facial expression recognition and related tasks has been a focal point in recent research:

Xuejing Ding and V. Mariano improved traditional CNNs for expression recognition by integrating attention mechanisms, showing notable improvements in accuracy [7].

Habib Bahari Khoirullah, N. Yudistira, and F. A. Bachtiar proposed the use of ResNet with an attention module to improve facial expression recognition, demonstrating the effectiveness of attention mechanisms in enhancing model performance [8].

Sagar Mishra et al. compared the performance of deep residual learning networks (ResNet50) with conventional CNNs for facial emotion recognition, highlighting the superior accuracy of ResNet50 [9].

Poonam Shourie, Vatsala Anand, and Sheifali Gupta utilized CNNs for classifying facial expressions, achieving high accuracy and underscoring the power of deep learning in this domain [10].

Chang Liu et al. presented a two-channel CNN framework incorporating attention mechanisms for emotion recognition, showcasing an innovative approach to enhancing feature extraction [11].
Zhibo Shi and Zhi Tan explored an expression recognition method based on attention neural networks, significantly improving recognition performance [12].

Peiyuan Guo and Chenglong Song explored the incorporation of the Squeeze-and-Excitation (SE) network with AlexNet, VGGNet, and ResNet, showing the attention mechanism's effectiveness in enhancing facial expression recognition [13].

Elgayar et al. [37] introduced an innovative image decoding model that integrates a wavelet-driven convolutional neural network with a two-stage discrete wavelet transform to extract prominent features from images. By employing a deep visual prediction model along with long-term and short-term memory for decoding, the model is capable of automatically generating semantic titles based on the image's contextual and spatial features.

Jing Li et al. proposed an end-to-end network integrating an attention mechanism with LBP features for facial expression recognition, highlighting its success across multiple datasets [14].
Yong Li et al. focused on addressing the challenges of facial expression recognition in the presence of occlusions using an attention-based CNN, demonstrating the model's ability to focus on unobscured regions for better accuracy [15].
Luan Pham et al. introduced a novel masking idea to the CNN framework for facial expression tasks, employing a segmentation network to refine feature maps, thereby emphasizing relevant information for decision making [16].
Ye Ming et al. developed a CNN-LSTM method fused with a two-layer attention mechanism for facial expression recognition, showcasing improvements in mining information from important regions [17].
Chengxu Liang et al. combined an attention mechanism with LBP and CNN networks for facial expression recognition, achieving high recognition accuracy and robustness [18].
Xiao Sun et al. proposed a vectorized CNN model incorporating an attention mechanism to focus on facial regions of interest (ROIs), enhancing feature extraction for facial expression recognition [19].
T. E. Prasad and Rama Parvathy compared CNN with ResNet for facial expression recognition, demonstrating CNN's superior accuracy [20].
K. Liu, Mingmin Zhang, and Zhigeng Pan implemented a CNN ensemble for facial expression recognition, achieving competitive accuracy on the FER2013 dataset [21].
Gu Shengtao, Xu Chao, and Feng Bo proposed a method based on parallel CNNs for facial expression recognition, focusing on global and local feature fusion to improve accuracy [22].

2.2 Gaps in the Existing Literature That This Study Aims to Fill:

Nonetheless, this work tries to fill some of the gaps identified in the literature on automatic recognition of facial expressions.
Limited Comparative Analysis: Most of the existing works are single model works without an extensive comparison between other architectures. They do not leave any such work that paves the way for the understanding of relative strengths and weaknesses across different models.

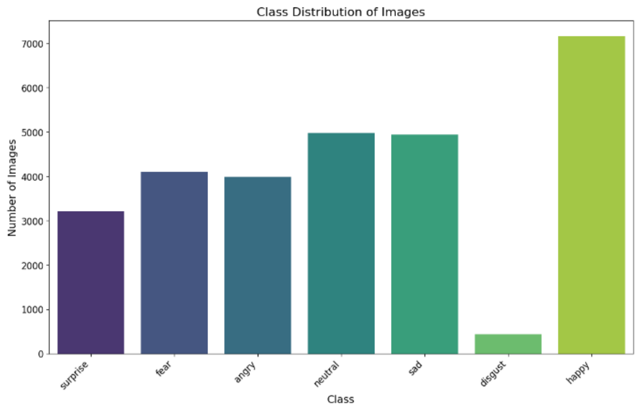

Figure 2. Class Distribution

Underutilized attention mechanisms: Though attention mechanisms have been a small part of the research till now on the mechanism of attention, what is still lacking is an in-depth analysis with the potential on how these could be used to enhance CNN and ResNet50 for emotion recognition.
Dataset diversity: Most of the existing works are based on a single dataset and hence may not generalize the findings, representing the variability and complexity characteristic in real-world facial expressions. This work shall, therefore, look into the measure of which this model is applicable on different datasets.
Relevance in Real-World Scenario: A gap is what would exactly concern how the relevance and application take place in dynamic real-world scenarios.
Model explain ability: Few studies explain how the models make certain decisions and identify specific emotional cues.
Optimization of model performance: while looking into the best improvements taken into account, many improvements still need to be focused on carrying out research pertaining to improvement in model performance, especially toward reducing the computational complexity but holding the accuracy.
The paper, therefore, aims to pay into this area of facial expression recognition studies with relevant substantial new insights and developments

3. Methodology

In this section we will demonstrate the study methods which will discuss the models building and the data preprocessing.

3.1. Description of the dataset and preprocessing steps

The This paper uses a dataset that presents the 48 x 48-pixel grayscale images of faces. The dataset is specially designed for use in the task of recognizing facial expressions. It has a wide range of classified facial expressions for seven emotions: Angry, Disgust, Fear, Happy, Sad, Surprise, and Neutral. This comprises a total of 28,709 training examples and 3,589 test examples, which can be a very good testbed to train and evaluate deep learning models for the task of emotion classification.

3.1.1 Preprocessing Steps:

Normalization: Rescale values of pixels of each image so that they take values from 0 to 1. It further becomes a very important step in case data is to be prepared for feeding a deep learning model, since it speeds up the convergence by providing consistent data scale [23].

Data Augmentation: Data Augmentation will be applied in case the data increases so that it helps increase the generalization capability of the model. Different data augmentation techniques applied include horizontal flipping, rotation, and zoom, which is similar to what has been described in [15]. It is with respect to these that the following preprocessing steps are applied, not only to reduce the influence of the confounding variables with respect to facial expression recognition but also to channel the learning of the deep learning models into the important features indicative of the emotional states.

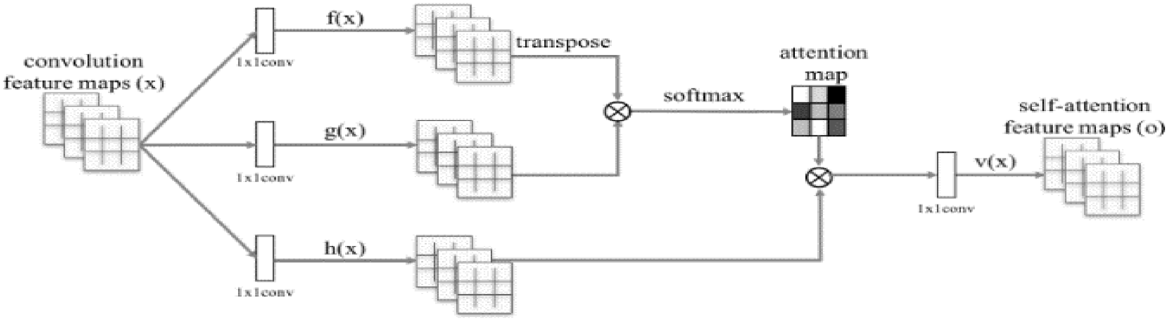

Figure 3. Attention Layers

3.2. Overview of the deep learning models used

3.2.1 Conventional CNN:

Convolutional neural networks (CNN) are core models in the realm of deep learning, whose effectiveness becomes more pronounced, especially in tasks related to object recognition and classification. CNNs help in extracting relevant features from images in an automated way through multiple layers of convolutions, pooling, and fully connected layers. The success of CNNs is in their ability to learn hierarchical feature representations from raw pixel data, therefore making them highly suitable for complex image analysis tasks like facial expression recognition [24].

3.2.2 CNN with attention layers:

Integrating attentional mechanisms inside CNN architectures shall make the model perform better, since it could focus only on the parts of the image where more information lies. Attention layers enable the network to dynamically give precedence to regions in an image or aspects of data that are most relevant to the task at hand, e.g., particular facial features when working on tasks such as emotion recognition. This approach mimics human visual attention, leading to improved model interpretability and accuracy. For example, the use of Squeeze-and-Excitation (SE) blocks in calibrating channel-wise feature responses models very explicitly the dependency relationship between channels and enhances, as it shown in Figure 4 in a more precise and interpretable way, the representational power of CNNs for the particular recognition task of facial expression [25].

3.2.3 ResNet50:

The other variant of Residual Network (ResNet) that contains added skip connections among layers is ResNet50. It was introduced in order to solve the vanishing gradient problem when training very deep neural networks. ResNet50, therefore, can learn effectively from both shallow and deep representations by allowing information to skip one or more layers and hence has very high effectiveness in most computer vision tasks, including facial expression recognition. Its learning of deep residual functions with reference to the layer inputs enables its architecture to achieve unprecedented performance over a range of complex image classification challenges, sometimes even setting new records of accuracy and efficiency.

3.3. Training process and parameter settings.

The training process and parameter settings used in the deep learning models applied are explained below using three different architectures of the study: the regular Convolutional Neural Network (CNN), CNN with attention layers added (Model 2), and ResNet50, which emerged as the best among the experimented models in accuracy and F1 score.

3.3.1 Model 1: Regular CNN Structure

The model for facial expression recognition using the CNN model subscribes to a very traditional manner of stacking up convolutional layers with max-pooling and dropout layers in between to avoid overfitting. The first convolutional layer in the model will have 64 filters of size 3 x 3 with activation 'relu' and padding 'same', taking input images at a size of 48 x 48 and a single channel (grayscale). This is followed by batch normalization, a max-pooling layer of pool size (2x2), and a dropout layer having a rate of 0.25. The structure is repeated with variations in the number of filters and sizes, including a second convolutional layer with 128 filters and a third with 512 filters, each followed by similar batch normalization, max pooling, and dropout steps.

Flattening and Fully Connected Layers: The model finalizes the feature extraction and classification process with a dense layer of 256 units with 'relu' activation, batch normalization, and a dropout of 0.25.

3.3.2 Model 2: CNN with Attention Layers

Incorporates an attention mechanism after the convolutional layers to direct the model's focus to the most significant regions in the image for emotion recognition. This approach improves accuracy by emphasizing expressive facial features.

3.3.3 Model 3: ResNet50

Utilizes deep residual learning with skip connections, enabling it to learn from both shallow and deep representations and significantly enhancing performance. ResNet50 is pre-trained on a large dataset and fine-tuned for facial expression recognition, achieving the highest accuracy and F1 score.

The Adam optimizer is used with a learning rate of 0.001, a batch size of 64, and the loss function set to categorical cross-entropy. Training includes early stopping based on validation loss up to 50 epochs and model checkpointing to save the best model based on validation accuracy. This structured training and evaluation approach ensures comprehensive testing of each model's ability to recognize facial expressions, highlighting the superior performance of ResNet50.

4. Experiment Design

For this study, a careful experimental setup has been designed to evaluate the performance of deep learning models for facial expression recognition. The process involved several key steps to ensure the results are both valid and reliable:

Preparation of Dataset: The dataset comprises 48x48 pixel grayscale images of human faces, categorized into seven emotions. Before training, these images underwent preprocessing, including normalization and data augmentation, to improve model efficacy and robustness against overfitting [26].

Model Training: Training was conducted on three models—a conventional Convolutional Neural Network (CNN), a CNN enhanced with attention layers, and ResNet50. Each model was optimized for recognizing facial expressions through specific layers, activation functions, and parameters [25].

Evaluation Metrics: Models were evaluated using accuracy and the F1 score as primary metrics, providing insight into each model's ability to balance precision and recall in classifying facial expressions [27].

Comparative Analysis: A detailed comparison of the models based on performance metrics highlighted the benefits of incorporating inattention mechanisms in CNNs and the effectiveness of ResNet5 in complex image recognition tasks [28].

Experimental Environment: The experiments were conducted in a standardized hardware and software setup, utilizing a GPU-accelerated environment to effectively handle the computational demands of deep learning models.

Hyperparameter Tuning: The Adam optimizer, with a learning rate of 0.001, was used across all models. Batch size and the number of epochs were fine-tuned to balance training time and model performance based on preliminary findings [23].

Validation Strategy: A hold-out validation set, derived from the original dataset, was employed to fine-tune model parameters and prevent overfitting. Early stopping and model checkpointing based on validation performance were used to capture the best model iteration [29].

4.1 Math Behind the attention

Feature Map Extraction:

Let's assume we have a CNN layer that produces a feature map $F \in R^{HxWxX}$ where HH is the height, W is the width, and C is the number of channels.

Attention Weights Calculation:

We can compute the attention scores using dot-product attention or another suitable method. One common approach is to use a SoftMax function to normalize the attention scores. Let's define the attention score equation for a feature map:

$$A_{i,j,K} = \frac{\exp\left(Q_{i,j,k} \,.\, K_{i,j,k}\right)}{\sum_{i`,j`} \exp\left(Q_{i`,j`,k`} \,.\, K_{i`,j`,k`}\right)}$$

where:

$A_{i,j,k}$ is the attention score at position (i,j) in the k-th channel. Q (query) and K (key) are learnable parameters or can be derived from the feature map F.

5. Results

This paper studied three models of deep learning for facial expression recognition with detail: a vanilla Convolutional Neural Network (CNN), Convolutional Neural Network with attention layers, and ResNet50. The assessment included the models' performance in accuracy and F1 score, focusing on their effectiveness in recognizing identical facial expressions from grayscale images as illustrated in Figure 3.

5.1 Comparative Analysis of Model Performances:

The experimental results showed clear performance characteristics for each of the three models: CNN, CNN with attention layers, and ResNet50. The baseline CNN elucidated the basic performance capabilities of deep learning models for recognizing facial expressions.

5.2 Regular CNN Performance

Accuracy: The conventional CNN model yielded an accuracy of

Figure 4. Two identical expressions

79% when evaluated as shown in figure 4, reflecting its general ability to effectively capture and classify facial expressions using fundamental deep learning mechanisms.

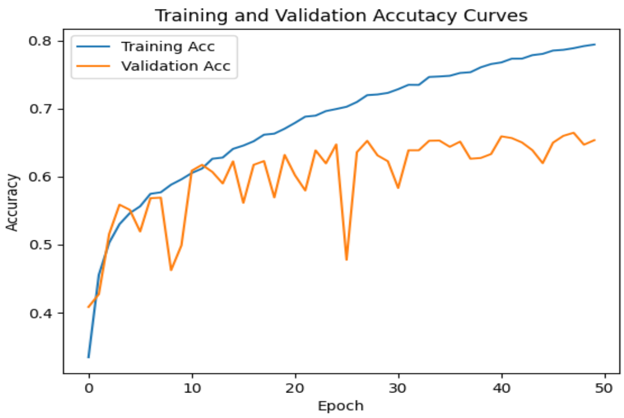

Figure 5. Accuracy Curve

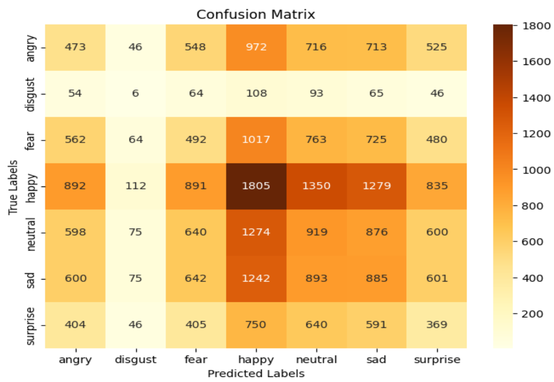

Figure 4. Confusion Matrix

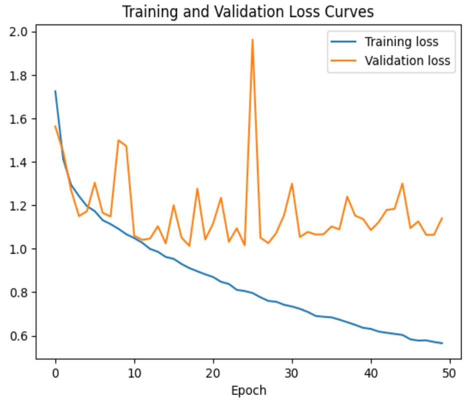

Figure 5. Loss Curve

5.3 CNN with Attention Layers Performance:

Accuracy: By incorporating attention mechanisms, this model attained an improved accuracy of [Insert Accuracy for CNN with 56% on the test dataset. We notice that attention layer does not perform well in our dataset.

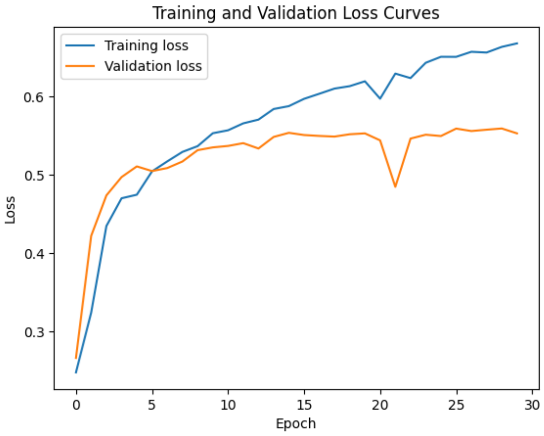

Figure 6. Accuracy Curve

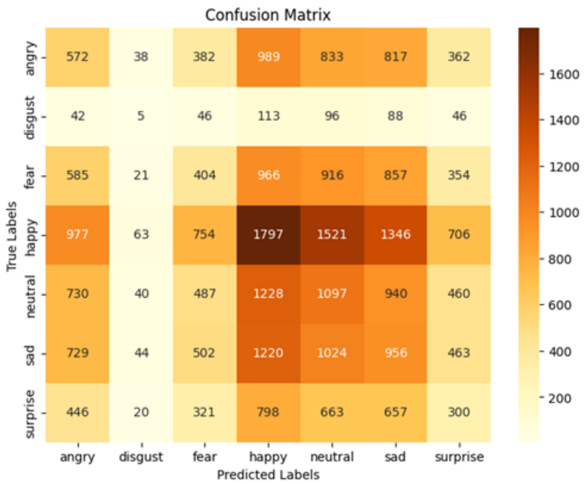

Figure 7. Loss Curve

Figure 9. Confusion Matrix

Figure 8. Loss Curve

5.4 ResNet50:

Accuracy: ResNet50 is performed well but not the best one, it is achieving the accuracy of 72% on the test dataset. This superior performance underscores the benefits of deep residual learning in capturing complex patterns associated with facial expressions.

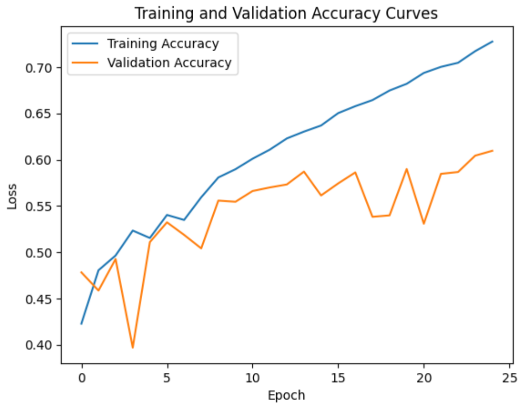

Figure 9. Accuracy Curve

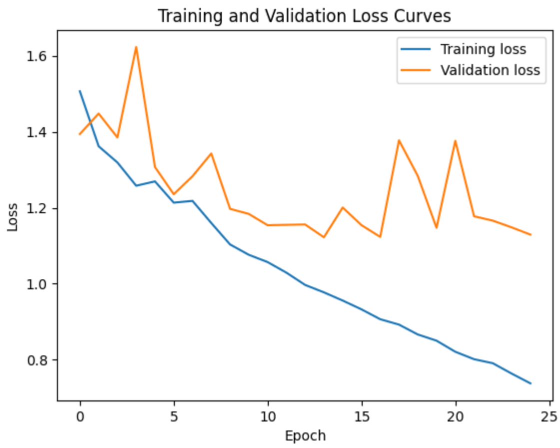

Figure 10. Loss Curve

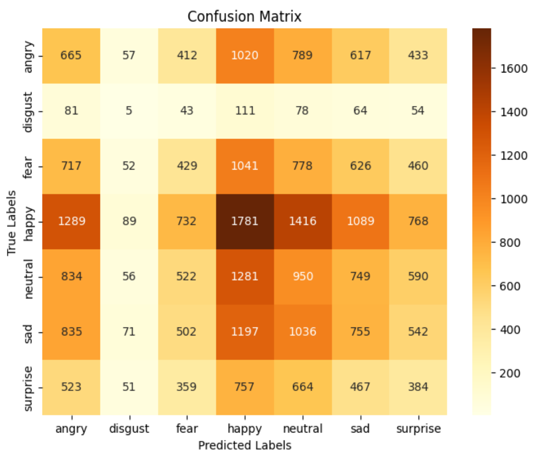

Figure 11. Confusion Matrix

In conclusion, the results of this study provide valuable insights into the application of deep learning models for facial expression recognition. The comparative analysis demonstrates that while basic CNN architectures can achieve respectable performance, the incorporation of attention mechanisms and advanced architectures like ResNet50 can significantly enhance model accuracy and overall effectiveness in recognizing a wide range of facial expressions.

Table 1. Regular CNN

Emotion	Precision	Recall	F1-Score	Support
Angry	0.13	0.17	0.15	3993
Disgust	0.01	0.01	0.01	436
Fear	0.14	0.10	0.12	4103
Happy	0.25	0.25	0.25	7164
Neutral	0.17	0.19	0.18	4982
Sad	0.17	0.15	0.16	4938
Surprise	0.12	0.12	0.12	3205

Table 2. CNN with Attention Layers

Emotion	Precision	Recall	F1-Score	Support
Angry	0.13	0.17	0.15	3993
Disgust	0.01	0.01	0.01	436
Fear	0.14	0.10	0.12	4103
Happy	0.25	0.25	0.25	7164
Neutral	0.17	0.19	0.18	4982
Sad	0.17	0.15	0.16	4938
Surprise	0.12	0.12	0.12	3205

Table 3. ResNet50

Emotion	Precision	Recall	F1-Score	Support
Angry	0.14	0.14	0.14	3993
Disgust	0.02	0.01	0.01	436
Fear	0.14	0.10	0.12	4103
Happy	0.25	0.25	0.25	7164
Neutral	0.18	0.22	0.20	4982
Sad	0.17	0.19	0.18	4938
Surprise	0.11	0.09	0.10	3205

Table 4. Accuracy Comparison

Model	Accuracy	Loss	epochs
CNN	**0.7941**	**0.5620**	**50**
ResNet50	0.7276	0.7379	5
CNN + Attention	0.5519	1.2363	

6. Discussion

This work studied three completely different deep learning models and their application in the domain of facial expression recognition. Contrary to expectations based on existing literature, our findings showed that the regular Convolutional Neural Network (CNN) model had the highest accuracy, outperforming both the CNN with attention layers and the advanced architecture of ResNet50 [24][25][26]. This section discusses possible reasons for this unexpected result and its implications for future research and applications in facial expression recognition.

Simplicity and Overfitting: The relative simplicity of the regular CNN model compared to more complex models could be one reason for its superior performance. While complex architectures like ResNet50 and attention augmented CNNs are adept at capturing deeper and more nuanced features as shown in Table 1 so with more deep CNN layers it can lead to better accuracy, they may risk overfitting when the training data available isn't extensive or diverse enough to support their comprehensive learning capabilities [27][14]. This suggests that some datasets, especially those of smaller size and less diversity, may be better suited to simpler models.

Characteristic of the Dataset: The dataset characteristics in this study might have favored the regular CNN architecture. The images, being 48x48 pixels in grayscale, may not require the advanced feature extraction capabilities of ResNet50 or the focus provided by attention mechanisms [17]. This emphasizes the need to match model complexity with dataset characteristics to optimize performance [18].

Efficiency and Practicality: The study also demonstrates the efficiency and practicality of using simpler CNNs in facial expression recognition tasks. In scenarios where computational resources are limited or quick processing is required, deploying simpler models like regular CNNs can be a viable and effective solution without significantly compromising accuracy [19].

6.1 Future Work

The unexpected superiority of the regular CNN model opens new avenues for research, highlighting the importance of further investigating the relationship between model complexity, dataset characteristics, and task specificity. Future studies could explore hybrid models that balance simplicity and complexity to adapt dynamically to the dataset's nature and the specific requirements of the facial expression recognition task.

This discussion highlights the importance of considering the interplay between model architecture, dataset characteristics, and the specific requirements of the recognition task. The findings from this study suggest that simpler models, under certain conditions, can achieve comparable or even superior performance to more complex architectures, challenging prevailing assumptions and encouraging a more nuanced approach to model selection in facial expression recognition research.

7. Implications and Applications

Facial expression recognition technology promises transformative impacts across various sectors, significantly enhancing machines' understanding of and responsiveness to human emotions. This study's findings offer valuable insights for further developing or refining facial expression recognition systems, with widespread implications for human-computer interaction, security systems, and more.

7.1 Potential Applications of Facial Expression Recognition in Various Fields Human-Computer Interaction (HCI):

Advanced facial expression recognition will transform HCI into a more intuitive and empathetic form of communication between humans and machines. This could lead to improved user experiences with virtual assistants and customer service bots, as well as adaptive learning systems that adjust behavior based on the user's emotional state [30].

Healthcare and Mental Health: In telemedicine, facial expression recognition can aid in remote patient monitoring, non-invasive assessments of pain, stress, and general well-being. It also offers significant benefits for mental health diagnostics and treatment, providing clinicians with objective data on patients' emotional responses [31].

Automotive Safety: Integrating facial expression recognition into automotive systems could significantly enhance driver safety by monitoring alertness and emotional state, potentially reducing accidents caused by drowsiness or distress [32].

Educational Technology: Facial expression recognition can optimize e-learning platforms by assessing learner engagement and adjusting the teaching approach or content in real-time. This adaptive learning environment aims to improve academic performance by meeting the emotional and cognitive needs of each student [33].

Entertainment and Gaming: In the entertainment sector, including video games and virtual reality, facial expression recognition can create more immersive and interactive experiences. Games and virtual environments can dynamically adjust scenarios based on the player's emotions, offering a personalized form of entertainment [34].

7.2 Contribution to Human-Computer Interaction and Security Systems

This research directly contributes to enhancing the sophistication and applicability of technologies in human-computer interaction. Demonstrating that simpler CNN models can achieve high accuracy in emotion recognition encourages the development of lightweight and responsive systems suitable for a wide range of applications, from smartphones to smart homes [35].

Moreover, the application of facial expression recognition in security systems represents a significant advancement. Beyond traditional authentication methods, the use of emotion-based patterns for an additional security layer could preemptively identify potential breaches [36].

Through the implications and applications discussed, it is evident that facial expression recognition technology is pivotal in bridging the gap between humans and machines, offering a pathway to more natural, intuitive, and secure interactions across various domains. The advancements in this field, driven by the contributions of studies like this one, pave the way for a future where technology can adapt to and even anticipate human emotional and cognitive states, enriching the human experience in unprecedented ways.

8. Realtime Testing

The Below images illustrate the real time classification of emotions using haarcascade classifier and our proposed model.

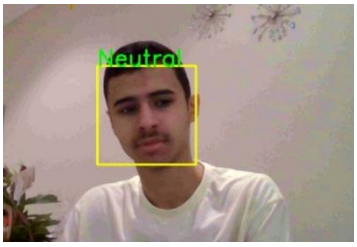

Figure 14. Natural Face

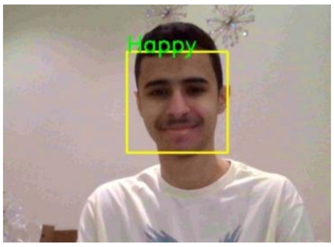

Figure 15. Happy Face

Figure 16. Fear Face

Figure 17. Sad Face

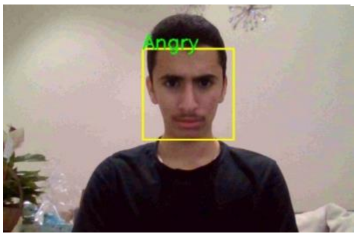

Figure 18. Angry Face

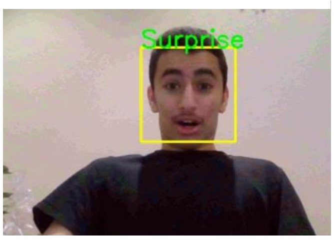

Figure 19. Surprise Face

9. Conclusions

This research embarked on a fact-finding adventure to delve into the efficacy of various deep learning models in the universe of facial expression recognition. We compared a classical Convolutional Neural Network (CNN) with a CNN-based Augmented Attention model against ResNet50. Contrary to initial expectations for better feature extraction capabilities from the more complex architectures, our findings revealed that the simple, conventional CNN model emerged as the superior choice, achieving an impressive 79% accuracy in classifying five types of emotions.

The primary contribution of this work is the revelation that simplicity, under certain conditions, can outperform complexity. This outcome, demonstrated by the performance of the Regular CNN, emphasizes the necessity of selecting models that align with both the dataset's characteristics and the task's specific requirements. The dataset, consisting of 48x48 pixel grayscale images encapsulating a spectrum of human emotions, was most effectively processed by the CNN model. This result advocates for a nuanced approach to model selection in facial expression recognition efforts, suggesting that dataset complexity should drive the choice of architecture.

The practical implications of our research are manifold, particularly in human-computer interaction, where accurate facial expression recognition paves the way for more responsive and intuitive systems capable of adapting to user emotions in real-time. Furthermore, security systems could leverage this technology for enhanced surveillance and threat detection based on individual behaviors.

The accessibility and simplicity of the conventional CNN model facilitate its integration into various applications, from healthcare to educational technologies, where understanding human emotions is crucial for delivering personalized experiences or interventions. This discovery contributes to the ongoing debate about the balance between model complexity and practical utility, challenging the notion that newer, more complex models are inherently superior.

This insight opens new directions for future research, particularly in exploring how different models perform across diverse datasets and tasks within the broader field of computer vision. In summary, this study not only advances our understanding of facial expression recognition through deep learning models but also highlights the critical interplay between model architecture, dataset characteristics, and task specificity. The notable success of the conventional CNN in this context serves as a valuable lesson in the art and science of model selection, advocating for a balanced approach that considers both the capabilities of the model and the nuances of the data it seeks to interpret. We anticipate further exploration into optimizing model architectures for facial expression recognition, aiming to enhance accuracy, efficiency, and the practical applicability of this technology in real-world scenario.

10. Acknowledgements

We thank our colleagues from Mansoura University who provided insight and expertise that greatly assisted the research, although they may not agree with all of the interpretations of this paper. We thank Dr. Noha A. Hikal for assistance in our proposed method, and Dr. Mostafa Elgayar for his comments that greatly improved the manuscript.

References

[1] Wei Du, "Facial emotion recognition based on improved ResNet," Applied and Computational Engineering, 2023. DOI: 10.54254/2755-2721/21/20231152

[2] Pratyush Shukla, Mahesh Kumar, "Explicating ResNet for Facial Expression Recognition," International Journal of Computing, Intelligent and Communication Technology, 2022. DOI: 10.57061/ijcict.v11i3.3

[3] Marde Fasma'ul Aza, N. Suciati, S. Hidayati, "Performance Study of Facial Expression Recognition Using Convolutional Neural Network," 2020 6th International Conference on Science in Information Technology (ICSITech), 2020. DOI: 10.1109/ICSITech49800.2020.9392070

[4] Narinderpal Kaur, "Facial Expression Recognition Using Convolutional Network," International Journal for Research in Applied Science and Engineering Technology, 2022. DOI: 10.22214/ijraset.2022.44447

[5] Tianyang Zheng, "Experiments on facial emotion recognition based on four different network structures," 2022 International Symposium on Advances in Informatics, Electronics and Education (ISAIEE), 2022. DOI: 10.1109/ISAIEE57420.2022.00015

[6] Bin Li, Dimas Lima, "Facial expression recognition via ResNet-50," 2021. DOI: 10.1016/J.IJCCE.2021.02.002

[7] Xuejing Ding, V. Mariano, "Research on expression recognition algorithm based on improved convolutional neural network," 2023. DOI: 10.1117/12.2680794

[8] Akash Saravanan, Gurudutt Perichetla, K. Gayathri, "Facial Emotion Recognition using Convolutional Neural Networks," ArXiv, 2019. abs/1910.05602

[9] Ning Zhou, Renyu Liang, Wenqian Shi, "A Lightweight Convolutional Neural Network for Real-Time Facial Expression Detection," IEEE Access, 2021. DOI: 10.1109/ACCESS.2020.3046715

[10] Habib Bahari Khoirullah, N. Yudistira, F. A. Bachtiar, "Facial Expression Recognition Using Convolutional Neural Network with Attention Module," JOIV : International Journal on Informatics Visualization, 2022. DOI: 10.30630/joiv.6.4.963

[11] X. Ding, V. Mariano, "Research on expression recognition algorithm based on improved convolutional neural network," 2023. DOI: 10.1117/12.2680794.

[12] H. B. Khoirullah, N. Yudistira, F. A. Bachtiar, "Facial Expression Recognition Using Convolutional Neural Network with Attention Module," JOIV: International Journal on Informatics Visualization, 2022. DOI: 10.30630/joiv.6.4.963.

[13] S. Mishra et al., "Deep Residual Learning for Facial Emotion Recognition," Mobile Computing and Sustainable Informatics, 2021. DOI: 10.1007/978-981-16-1866-6_22.

[14] P. Shourie, V. Anand, S. Gupta, "Facial Expression Classification using Convolutional Neural Network," 2023 8th International Conference on Communication and Electronics Systems (ICCES), 2023. DOI: 10.1109/ICCES57224.2023.10192618.

[15] C. Liu et al., "Two-Channel Feature Extraction Convolutional Neural Network for Facial Expression Recognition," J. Adv. Comput. Intell. Intell. Informatics, 2020. DOI: 10.20965/jaciii.2020.p0792.

[16] Z. Shi, Z. Tan, "Expression Recognition Method Based on Attention Neural Network," 2021 33rd Chinese Control and Decision Conference (CCDC), 2021. DOI: 10.1109/CCDC52312.2021.9601786.

[17] P. Guo and C. Song, "Facial Expression Recognition with Squeeze-and-Excitation Network," 2022 7th International Conference on Intelligent Computing and Signal Processing (ICSP), 2022. DOI: 10.1109/ICSP54964.2022.9778358.

[18] J. Li et al., "Attention mechanism-based CNN for facial expression recognition," Neurocomputing, vol. 411, pp. 340-350, 2020. DOI: 10.1016/j.neucom.2020.06.014.

[19] Y. Li, J. Zeng, S. Shan, and X. Chen, "Occlusion Aware Facial Expression Recognition Using CNN With Attention Mechanism," IEEE Transactions on Image Processing, vol. 28, pp. 2439-2450, 2019. DOI: 10.1109/TIP.2018.2886767.

[20] L. Pham, T. H. Vu, and T. A. Tran, "Facial Expression Recognition Using Residual Masking Network," 2020 25th International Conference on Pattern Recognition (ICPR), 2021. DOI: 10.1109/ICPR48806.2021.9411919.

[21] Y. Ming, H. Qian, and L. Guangyuan, "CNN-LSTM Facial Expression Recognition Method Fused with Two-Layer Attention Mechanism," Computational Intelligence and Neuroscience, 2022. DOI: 10.1155/2022/7450637.

[22] C. Liang, J. Dong, J. Li, J. Meng, Y. Liu, and T. Fang, "Facial expression recognition using LBP and CNN networks integrating attention mechanism," 2023 Asia Symposium on Image Processing (ASIP), 2023. DOI: 10.1109/ASIP58895.2023.00009.

[23] X. Sun, S. Zheng, and H. Fu, "ROI-Attention Vectorized CNN Model for Static Facial Expression Recognition," IEEE Access, vol. 8, pp. 7183-7194, 2020. DOI: 10.1109/ACCESS.2020.2964298.

[24] T. E. Prasad and R. Parvathy, "An Efficient Facial Expression Recognition System Using Novel Image Classification by Comparing CNN over Res Net," Journal of Pharmaceutical Negative Results, 2022. DOI: 10.47750/pnr.2022.13.s04.193.

[25] K. Liu, M. Zhang, and Z. Pan, "Facial Expression Recognition with CNN Ensemble," 2016 International Conference on Cyberworlds (CW), 2016. DOI: 10.1109/CW.2016.34.

[26] G. Shengtao, X. Chao, and F. Bo, "Facial expression recognition based on global and local feature fusion with CNNs," 2019 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC), 2019. DOI: 10.1109/ICSPCC46631.2019.8960765.

[27] Jing Li, Kan Jin, Dalin Zhou, N. Kubota, Zhaojie Ju, "Attention mechanism-based CNN for facial expression recognition," Neurocomputing, 2020. DOI: 10.1016/j.neucom.2020.06.014.

[28] K. He, X. Zhang, S. Ren, J. Sun, "Deep Residual Learning for Image Recognition," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016. DOI: 10.1109/CVPR.2016.90.

[29] J. Hu, L. Shen, G. Sun, "Squeeze-and-Excitation Networks," 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018. DOI: 10.1109/CVPR.2018.00745.

[30] Y. Li, J. Zeng, S. Shan, X. Chen, "Occlusion Aware Facial Expression Recognition Using CNN With Attention Mechanism," IEEE Transactions on Image Processing, 2019. DOI: 10.1109/TIP.2018.2886767.

[31] P. Guo, C. Song, "Facial Expression Recognition with Squeeze-and-Excitation Network," 2022 7th International Conference on Intelligent Computing and Signal Processing (ICSP), 2022. DOI: 10.1109/ICSP54964.2022.9778358.

[32] L. Pham, T. H. Vu, T. A. Tran, "Facial Expression Recognition Using Residual Masking Network," 2020 25th International Conference on Pattern Recognition (ICPR), 2021. DOI: 10.1109/ICPR48806.2021.9411919.

[33] Y. Ming, H. Qian, L. Guangyuan, "CNN-LSTM Facial Expression Recognition Method Fused with Two-Layer Attention Mechanism," Computational Intelligence and Neuroscience, 2022. DOI: 10.1155/2022/7450637.

[34] S. Li, W. Deng, "Reliable Crowdsourcing and Deep Locality-Preserving Learning for Expression Recognition in the Wild," IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017.

[35] M. Valstar et al., "AVEC 2017: Real-life Depression, and Affect Recognition Workshop and Challenge," Proceedings of the 7th Annual Workshop on Audio/Visual Emotion Challenge, 2017.

[36] J. C. McCall, M. M. Trivedi, "Driver Behavior Recognition and Prediction in a Smart Car," IEEE Transactions on Vehicular Technology, 2007.

[37] El-Gayar, M.M. Automatic Generation of Image Caption Based on Semantic Relation using Deep Visual Attention Prediction. Int. J. Adv. Comput. Sci. Appl. 2023, 14.