JCBI
Journal of Computing & Biomedical Informatics
ISSN: 2710 - 1606
Volume 06 Issue 02
2024

Research Article
https://doi.org/10.56979/602/2024

Convolutional Approaches in Transfer Learning for Facial Emotion Analysis

Ahmed Faraz[1*], Muhammad Fuzail[1], Ali Haider Khan[2], Ahmad Naeem[1], Naeem Aslam[1], and Mueed Ahmed Mirza[3]

[1]Department of Computer Science, NFC Institute of Engineering and Technology Multan, Multan, 59030, Pakistan.
[2]Department of Software Engineering, Faculty of Computer Science, Lahore Garrison University, Lahore, 54000, Pakistan.
[3]Department of Computer Science, University Of Lahore, Lahore, 54590, Pakistan.
[*]Corresponding Author: Ahmed Faraz. Email: farazahmedkhan1997@gmail.com

Received: January 03, 2024 **Accepted:** February 21, 2024 **Published:** March 01, 2024

Abstract: The scientific community has shown significant interest in facial emotion recognition (FER) due to its possible applications. The primary function of Facial Expression Recognition (FER) is to associate various facial expressions with their respective emotional states. Feature extraction and emotion recognition are the primary constituents of conventional FER. The inherent feature extraction capabilities of Deep Neural Networks, particularly Convolutional Neural Networks (CNNs), have resulted in their extensive utilization in Facial Expression Recognition (FER) currently. While previous studies have explored the utilization of multi-layer shallow convolutional neural networks (CNNs) for addressing facial expression recognition (FER) tasks, a significant drawback of these models is their limited capacity to accurately extract features from high-resolution photos. Many of the existing methods also exclude profile views, which are crucial for real-world facial expression recognition (FER) systems, in favor of frontal photographs. This research introduces a highly complex Convolutional Neural Network (CNN) model that incorporates Transfer Learning (TL) to enhance the precision of Facial Expression Recognition (FER). The proposed approach for satisfying the FER criteria involves utilizing a pre-trained DCNN model and fine-tuning it with facial expression data. Subsequently, it substitutes the dense higher layer(s) of the model. To improve the precision of Facial Expression Recognition (FER), a new approach is claimed that consists of iteratively applying the fine-tuning technique on each of the pre-trained DCNN blocks. The validation Facial Expression Recognition FER system of eight DCNN models trained previously, especially VGG-16 and VGG-19 is undertaken. The authentification process is run through two data sets, namely KDEF and JAFFE. But as tricky as the FER is, including the various perspective analysis which is part of the KDEF dataset, the proposed approach is a stand out in the high accuracy that it records. VGG-16 achieved the highest FER accuracies of 93.7% on the KDEF test set and 100% on the JAFFE test set using a 10-fold cross-validation. The assessment emphasizes the benefits of the proposed Facial Emotion Recognition (FER) system, particularly in its ability to reliably detect emotions. It demonstrates promising outcomes on the KDEF dataset, specifically in the context of profile views.

Keywords: Emotion detection; Convolutional neural network (CNN); Transfer learning; Deep Learning.

1. Introduction

Nonverbal communication is a universal language, encompassing facial expressions, voice intonations, and bodily movements. Among these, facial expressions, extensively studied by Ekman and Friesen, are fundamental to understanding human emotions. Happiness, sorrow, anger, fear, surprise, disgust, and neutrality are universal emotional states recognized through facial cues. The scientific community, spanning psychology, psychiatry, healthcare, and human-computer interaction, increasingly focuses on facial expression research, with automated emotion recognition gaining significant interest.

Facial Emotion Recognition (FER) involves discerning and categorizing emotions based on facial expressions. Feature extraction and emotion identification are two key aspects, utilizing techniques like discrete wavelet transform and linear discriminant analysis. Recent years have seen the rise of Convolutional Neural Networks (CNNs) in FER, automating information extraction from images. While complex CNNs like VGG-16, Resnet, Inception, and DenseNet enhance accuracy, challenges like vanishing gradients and limited emotional changes in facial expressions persist.

A reliable FER system must analyze facial expressions from various perspectives, yet current approaches often focus solely on frontal views for simplicity. The goal is to design a system capable of identifying emotions from any viewpoint effectively. The research employs transfer learning and Deep Convolutional Neural Networks (DCNN) to develop an FER system. Learned models such as VGG16 inherently get knowledge from large datasets without the need to start training with random weights and bring in order. Fine-tuning these models with the facial emotion data sets particularly points out side views help to improve identifying emotion which leads to more accuracy. While facial expressions serve to communicate, we are not saying their function ends there and we are even including a use of facial expressions in the field of human-computer interaction. Deep learning and transfer learning present promising solutions despite challenges. Recognition of facial expressions gets deeper than one would think and in practical applications and in communications between man and computer it will assist in their work.

The research is about finding new ways in scientific understanding and see what others have not seen before by unveiling those to the public. There is potential of doing something significant, making a dent in complicated matters, finding insights. These fuel love for science. The challenges and gaps in knowledge within the field fuel the investigation, aiming to offer fresh insights and practical solutions for societal betterment.

Interdisciplinary collaboration is a key inspiration, allowing the integration of insights from various fields for a holistic perspective. Working with experts and mentors enriches the research endeavors. The realization that the work could leave a lasting impact on the field, encouraging further exploration and innovation, serves as a continuous source of inspiration. Overcoming obstacles and persevering in the academic journey is driven by the opportunity to challenge assumptions and contribute valuable research findings, leaving a legacy in the pursuit of knowledge. The motivation behind this study stems from a blend of intellectual curiosity, a dedication to filling knowledge gaps, and a desire to bring about positive change. The overarching research objective focuses on advancing the field through the development of robust facial emotion recognition systems. These systems aim to adeptly identify and comprehend human emotions conveyed through facial expressions, tackling challenges such as emotion categorization, facial expression variability, contextual understanding, and the intricacies of human emotions. The conviction is that rigorous research and dedicated exploration can contribute to both immediate advancements and the inspiration of future scholars and innovators.

Deep learning, a cutting-edge technique, is being employed in the field of facial emotion recognition (FER) within machine learning. Convolutional neural networks, often known as CNNs, have been extensively utilized in numerous research investigations. These papers are available in the digital repository of academic research. Table 1 presents a thorough summary of research conducted using deep learning, highlighting the many architectural frameworks employed in these studies. The table offers a comprehensive summary of several architectures employed to enhance understanding in the relevant field, with each entry encompassing a distinct facet of deep learning approach. The table is purposeful, for it intends to do more than just present the familiar layout of deep learning techniques. This is so that readers can gain a deeper understanding of the complex nature of deep learning research, which is so often evolving.

Table 1. Deep Learning-Based Face Emotion Detection Approaches

Study Authors	Architecture	Features	Classification Method	Notable Aspects
Zhao and Zhang [28]	DBN + NN	Unsupervised Feature Learning	Neural Network	Combination of DBN for feature learning and NN for classification

Pranav and Colleagues [29]	CNN	N/A	N/A	Utilization of a typical CNN architecture for FER
Mollahosseini and Colleagues [30]	Inception + CNN	N/A	N/A	Complex design with multiple inception layers and CNN
Pons and Masip [31]	Network of 72 CNNs	N/A	N/A	Ensemble of CNNs with different filter sizes
Wen and Colleagues [32]	Ensemble of CNNs	N/A	N/A	Training multiple CNNs and selecting a subset for the final model
Ruiz-Garcia [33]	Encoder Weights from Convolutional Auto-encoder	N/A	N/A	Use of encoder weights for CNN weight initialization
Ding [34]	FaceNet2ExpNet	N/A	Transfer Learning	Extension of deep face recognition architecture to FER
Jain [35]	Hybrid Architecture with CNN and RNN	N/A	CNN + RNN	Investigation of hybrid architecture with CNN and RNN
Shaees [36]	Hybrid Architecture with Transfer Learning	Support Vector Machines (SVM)	Transfer Learning + SVM	Utilization of pre-trained AlexNet for feature extraction
Bendjillali [37]	CNN	DWT	N/A	Feature extraction using Discrete Wavelet Transform (DWT)
Liliana [38]	Deep CNN	N/A	N/A	Utilization of a deep CNN architecture with 18 layers
Shi [39]	CNN	N/A	N/A	Exploration of CNN for clustering data in FER
Ngoc [40]	Graph-based CNN	Facial Landmarks	N/A	Utilization of graph-based CNN for FER with focus on landmarks
Jin and Team [41]	CNN-based Strategy with Unlabeled and Labeled Data	N/A	N/A	Consideration of both unlabeled and labeled data in CNN-based strategy
Porcu [42]	Data Augmentation Approaches	Synthetic Images + Other Approaches	N/A	Testing effectiveness of data augmentation with synthetic images

Deep learning has been put into practice in the field of facial emotion recognition, utilizing the CNN approach in particular. Throughout our research we will focus on several main objectives, which include the exploration of both traditional and modern methods of emotion recognition using deep neural networks. In addition to that, we plan to do also the research that helps to improve facial emotion recognition models' precision and robustness due to the transfer learning. The project will attempt to evaluate the potentials and efficiency of deploying pre-trained deep CNN models for the identification of the feelings which are displayed over a person's face. Subsequently, the examination aims to address and to consider the questions attempting to accomplish deep learning and transfer learning by using a facial emotion recognition. Table 2 (1) presents the summary results of the field re-search done by most studies. It showed the different approaches and characteristics used. The information that is presented in the table provides a condensed summary of the substantial research that was carried out by a number of different researchers. By going deeper into the particulars of each entry, readers have the opportunity to learn additional

knowledge regarding the distinct approaches and characteristics that have contributed to the formation of the area of each study. For the purpose of acquiring a full comprehension of the numerous approaches that were utilized in the broader span of the study subject, this compilation is an excellent resource.

Table 2. Machine Learning-Based Face Emotion Detection Approaches

Study Authors	Methodology	Features	Classification Method	Notable Aspects
Xiao-Xu and Wei [16]	WEF + FLD + KNN	Wavelet Energy Feature	K-nearest neighbor (KNN)	Early use of wavelet energy feature
Zhao and Colleagues [17]	KNN	PCA + NMF	K-nearest neighbor (KNN)	Feature extraction using PCA and NMF
Feng and Colleagues [18]	LBP + LP	Local Binary Pattern (LBP) Histogram	Linear Programming (LP)	Multi-section image feature extraction
Zhi and Ruan [19]	2D Discriminant Locality Preserving Projections	Facial Feature Vectors	N/A	Use of 2D discriminant locality preserving projections
Lee and Colleagues [20]	Boosting + CT	Contourlet Transform (CT)	N/A	Extension of wavelet transform to 2D contourlet transform
Chang and Huang [21]	RBF Neural Network + FER with Face Recognition	N/A	Radial Basis Function (RBF)	Integration of FER with face recognition
Shih and Colleagues [22]	SVM	DWT + PCA	Support Vector Machine (SVM)	Comparison of feature extraction methods with SVM
Shan and Colleagues [23]	SVM	Local Statistical Data + LBPs	Support Vector Machine (SVM)	Evaluation of alternative facial representations
Jabid and Colleagues [24]	LDP	Local Directional Pattern (LDP)	N/A	Investigation of local directional pattern (LDP)
Alshami and Colleagues [25]	SVM	Center of Gravity Descriptor + Facial Landmarks Descriptor	Support Vector Machine (SVM)	Use of SVM with different feature descriptors
Liew and Yairi [26]	SVM + Various Classification Techniques	Gabor + Haar + LBP	KNN, LDA, etc.	Comparison analysis of different techniques and features
Joseph and Geetha [27]	Various Classification Approaches + FER	Facial Geometry	Logistic Regression, LDA, KNN, Decision Trees, Naive Bayes, SVM	Evaluation of facial geometry for classification

In terms of contributions, the research offers valuable insights into the practical application of deep learning, with a specific focus on CNNs, in facial emotion recognition. It illuminates the technical aspects and potential benefits of incorporating deep learning techniques into this domain. The study also delves into the advantages and challenges of transfer learning, emphasizing its role in accelerating progress in facial emotion recognition and the importance of leveraging pre-trained models. Notably, the research provides experimental results demonstrating the efficacy of pre-trained deep CNN models in enhancing the accuracy and robustness of facial emotion recognition systems. Additionally, it underscores the

significance of interdisciplinary collaboration between computer scientists, psychologists, and neuroscientists in advancing the field of human-computer interaction and emotional recognition.

In the upcoming section, after conducting an in-depth analysis of the existing literature to provide a foundation for our study. Following the literature review, the Materials and Methods chapter will explain the novel approach employed for facial detection, offering a comprehensive account of the procedures and unique characteristics. Subsequently, the Results and Discussion section will present and analyze the empirical findings within the context of prior research, contributing to the ongoing learned discourse. Chapter six will wrap up the study by enumerating main issues, title areas that were covered, and pointing out to space for any further study. The structure observed here is in the framework of which the research articles are usually subsequentially presented in a systematic and cohesive torrent.

2. Materials and Methods

This work is dedicated to the design of the deep convolutional neural networks (DCNNs) combined with Transfer Learning (TL) and to the implementation of the Facial Emotion Recognition using this approach. Understanding how we get the properties of the CNN layers by the mechanism involved in this process is the key to getting the overall idea of this strategy. The beginning CNN level is for picking out simple visual features, but with higher layers it becomes able to learn complex variations and the lower layers would be the expert at identifying slight textures and shapes composition. Taking up pre-trained complete DCNN models, for instance VGG-16, which are famously proven to perform well on image classification benchmarks like ImageNet means higher accuracy than random initialization.

In this FED (Facial Emotion Detection) system I utilise Transfer Learning to train a pre-trained DCNN model in the recognition of the facial expression, this method reinforces specific identification of facial expressions. Succeeding sections provide an elaborated account of theoretical ideas behind FER approaching an explanation of the suggested optimal deviance detection method with appropriate examples. The architecture of a FED model based on Transfer Learning and DCNNs, which is dedicated to facial emotion detection is presented in the exhibit. This framework has two components. A pre-trained DCNN to extract local feature maps from the original network, complemented with a new set of layers to explicitly realize Facial Expression Recognition (FER).

Revitalizing a pre-trained DCNN involves two crucial steps: switching the predecessor of the model for the new classifier and paring the model. Essentially, a fully-connected dense layers are often added as a classifier. In the realm of transfer learning, we select an existing model, compute its size similarity, and choose from three com-mon methods: not training the entire model, tuning the introverted layer(s) at a time while the others merely update, or rather focusing on the classifier. For tasks, the model's training is most relevant to its original training context, training of the classifier and the first few layers may be sufficient. Also it is imperative to train the whole model, recalibrating its key layers and the extra classifier.

For the framework of facial expression recognition (FER) to make the most of its features, accurate choice of component and training techniques must be seriously considered. Implementation of this research is based on a systematic approach that starts by going through as stepwise procedures to fit the model depending on the tasks. See the FED system flow chart that includes the rather well-developed VGG-16 model for object detection performance, below. Recognizing emotions exactly are being provided by the introduction of emotional data and alteration to several layers and this AI system do this precisely due to its further possibilities.

The model architecture is designed by replacing the final dense layers of an existing model with new ones customized for categorizing facial images into seven emotion classes: it also produces specific human emotions, such as, fear, anger, disgust, sadness, joy, surprise, and neutrality. These dense layers, also referred to as fully connected layers, take input data of a specific dimension and generate an output vector with the required dimension. The output layer, in this case, has seven neurons, each representing a distinct emotion.

The fine-tuning process involves training the modified architecture, which includes both the convolutional basis of the pre-trained model and the integrated dense layer(s). This fine-tuning utilizes a preprocessed emotion dataset that includes operations like resizing and cropping. During testing, a trimmed image is input into the system, and the emotion with the highest likelihood is considered the system's determination.

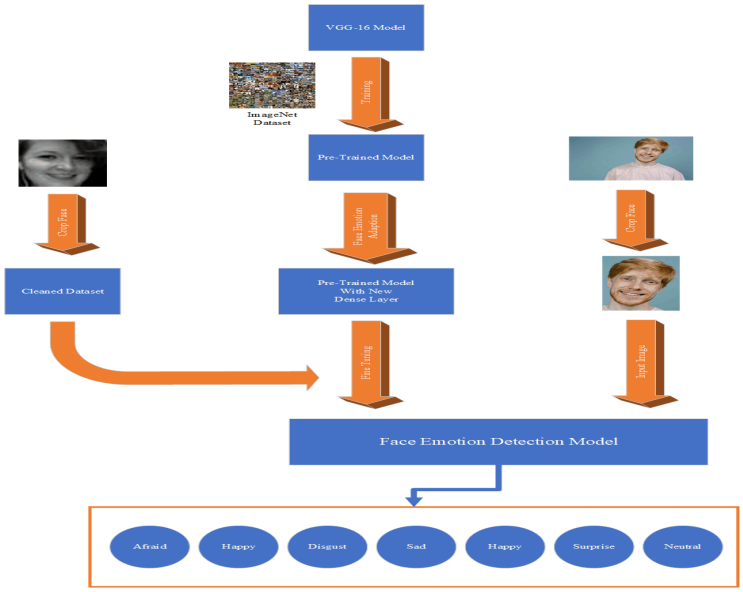

Figure 1. Proposed Face Emotion Detection Architecture

While VGG-16 is used in this illustration, it's noteworthy that alternative pre-trained DCNN models like ResNet, DenseNet, or Inception can be substituted. The suggested model's dimensions depend on the pre-trained model's dimensions and the structure of the added dense layer(s).

The schematic architecture of the proposed model is presented in the figure 2, below, featuring the pre-trained VGG-16 model and additional dense layers for Facial Emotion Recognition (FER). The highlighted green section illustrates the added component, consisting of three sequentially placed fully connected layers.

Optimal face emotion recognition is accomplished by meticulously adjusting the model parameters. The VGG-16 pre-trained base model is utilised when data undergoes resizing, rescaling, or any other form of alteration. The model is being improved by tunings fine-tuned layers and the overall framework of the model to recognise emotion while pre-processing the dataset and correcting the mistakes. Factors such as input dimensions, batch size, and optimization with the modified version of stochastic gradient descent - Adam is pondered over in the parameter settings for effective training and validation. The best facial emotion recognition is through the employment of efficient methodology where perfect model parameters are adjusted. VGG-16 pre-trained base model operates on data any time images undergo resizing, rescaling or any other type of transformation. Using this model, the dense layers are fine-tuned to perform emotion recognition through a dataset of emotions pre-processed. Factors such as input shapes, batch sizes, and Adam optimisation are very important in the process of configuring the parameters of the model. This step is critical to carry out accurate training and validation.

2.1 Dataset

In this reseach work, extensively studied facial expression recognition using two widely acknowledged datasets: For the JAFFE (Japanese Female Facial Expressions), we have [43] and KDEF (Karolinska Directed Emotional Faces) [44]. We have made an exception and chosen our datasets based on their noteworthy aspects and the hitherto level of the emotional analysis of facial expressions.

JAFFE dataset is non discriminatory in the sense that it highlights only cultural representation and strict regulation, whereas the majority of assets are pictures of Japanese females featured with facial expressions. Recognizing capabilities of cultural nuances leading to various facial expressions is a key factor in modeling robust and culture-adherent emotion recognition system. The JAFFE dataset [43] has the benefit of being utilised for training and evaluating the algorithms targeted towards affect recognition task. The entities emotions and neutral expression are six in total as visually depicted in Figure 3.

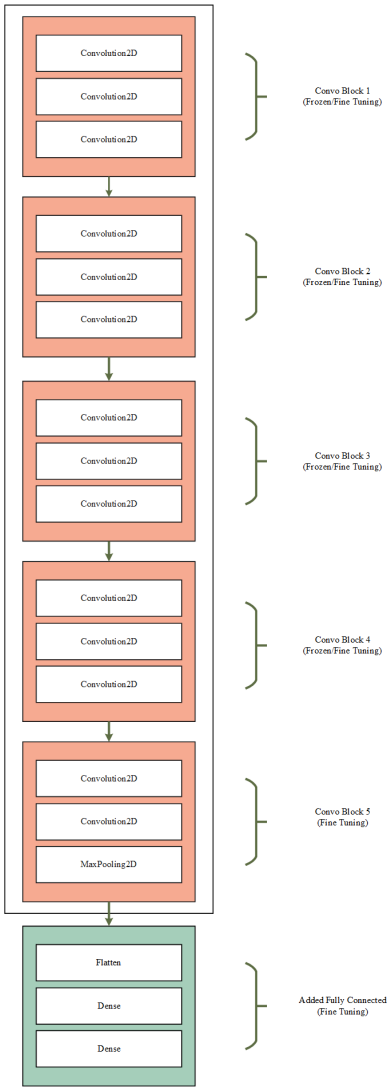

Figure 2. Illustration of VGG-16 Layers and Dense Layer

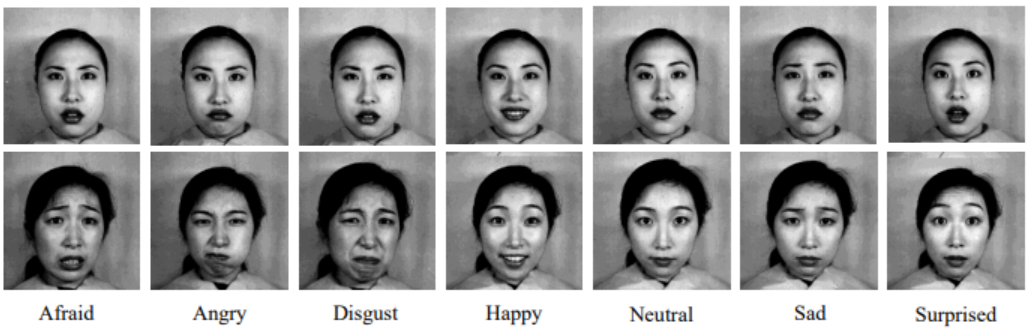

Figure 3. JAFFE Dataset Sample Images

Moreover, the KDEF dataset [46] from the Karolinska Institute in Sweden works pro-vided wider demographic coverage, including different ages and nationalities together with neutral state of mind and seven different moods, such as demonstrated in Fig 4. The introduction of this dataset to the research informs the analysis and preparation of algorithms testing emotional face recognition across a wide sample from various ethnics and emotional states.

Figure 4. KDEF Dataset Sample Images

This study set out to overcome three main limitations in our knowledge of facial expression recognition by combining together sufaces from JAFFE and KDEF datasets. This integration provides a more comprehensive representation of diverse demographics and cultural nuances, offering valuable insights for the development of robust and culturally sensitive emotion recognition algorithms.

2.2 Parameter

In this work, a robust facial emotion detection model was meticulously developed and trained with carefully chosen parameters. The input photos underwent data augmentation, resizing to 224x224 pixels, and normalization. Picture data generators used batch sizes of 32 and 8 for training and validation rounds, respectively. The VGG16 pre-trained base model, with weights initialized from ImageNet, served as an effective foundation. A 512-unit dense layer without activation optimized the model for emotion recognition, followed by a concluding layer with 7 units using softmax activation for predicting distinct emotions. Key factors, including input form and trainable weights, were adjusted for targeted adaptation. The model's structure is outlined in the table. For comprehensive training, the Adam optimizer with a learning rate of 5e-5 was employed, using categorical cross-entropy as the loss function and categorical accuracy as the measure. The systematic approach of this model spans 30 epochs, with 6 phases for training and 3 for validation, as illustrated in table 3.

Table 3. Proposed Method Parameter Value on each Layer

Layer Type	Parameter	Value
Input	Image Dimensions	(224, 224, 3)
Data Augmentation	Rescale Factor	1./255
	Target Size (Generator)	(224, 224)
	Batch Size (Train Gen.)	32
	Batch Size (Validation Gen)	8
Pre-trained Base Model	Model	VGG16 (pre-trained on ImageNet)
	Weights	imagenet
	Include Top	False (Excluding fully connected layers)
	Input Shape	(224, 224, 3)
Fully Connected Layers	Dense Layer 1	Units: 512, Activation: None
	Dense Layer 2 (Output)	Units: 7, Activation: Softmax
Model Summary	Trainable Weights Before	4
	Trainable Weights After	2 (Fully connected layers)
Compilation	Optimizer	Adam, Learning Rate: 5e-5
	Loss Function	Categorical Cross entropy
	Metrics	Categorical Accuracy
Training	Epochs	30
	Steps per Epoch	6
	Validation Steps	3

3. Results

In this section, metrics are defined for evaluating model performance. The training results help identify optimal parameter values. The KDEF and JAFEE datasets use these measures to assess the proposed

model and its limitations. Subsequently, we analyze and make decisions based on the outcomes from these datasets.

3.1 Evaluation Metrics:

3.1.1 Accuracy:

Accuracy is given by

$$accuracy = \frac{no\ of\ correct\ prediction}{total\ no\ \ prediction}$$

3.1.2 Loss:

Categorical cross-entropy is used as the loss function and is given by

$$loss = -\sum_{c=1}^{m}\left(y_{o,c}\log\left(p_{o,c}\right)\right)$$

The projected probability is represented by the symbol p, and there are m classifications, including happy, sad, neutral, fear, angry, disgust, and surprise. The binary indication y can take the values of either 0 or 1.

3.2 Confusion Matrix

The confusion matrix depicts combinations of real and expected values: True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). Utilizing these values, we calculate precision, recall, and F-score. For emotion predictions, scenarios include accurate predictions (TP), incorrect predictions (FP), and correct predictions of incorrect emotions (TN). Consider an image of a cheerful classroom where the red part represents TP for predicting happiness. The blue portion has FP values for predicting emotions like sadness, anger, indifference, or terror. The yellow area shows TN values for accurately predicting the absence of grief, anger, neutrality, or fear. The green area indicates FN values for incorrectly predicting unhappiness despite the positive projection.

3.3 Recall

Recall is given by

$$recall = \frac{TP}{TP + FN}$$

3.4 Precision

Precision is given by

$$precision = \frac{TP}{TP + FP}$$

3.5 F-Score

F-Score is given by

$$FScore = \frac{2 \times recall \times precision}{ecall + precision}$$

3.6 Expermental Setup

This study meticulously designed the experimental setup for optimal facial emotion recognition using deep learning techniques. Google Colab, a cloud-based platform with free GPU and TPU access, served as the chosen computing infrastructure. Resource allocation was efficiently managed, eliminating the need for significant local hardware. Google Drive facilitated centralized dataset access, making image loading and processing seamless within the Colab environment.

Tensor Processing Units (TPUs) were employed to enhance model training efficiency and computational speed. Google's TPUs, specialized hardware accelerators for machine learning, significantly reduced training time, enabling faster iterations.

Critical to the face emotion detection model's effectiveness were the hyperparameters. The learning rate was optimized at 5e-5 during training. Batch sizes of 32 for training and 8 for validation were chosen for efficient gradient updates. Image preprocessing involved standardizing pixel values by dividing them by 255. The widely-used Adam optimizer with a learning rate of 5e-5 was employed.

The VGG16 architecture, pre-trained on ImageNet, was fine-tuned to capture intricate facial expression elements. The training process spanned 30 epochs, striking a balance between model convergence and processing time, as illustrated in the table 4.

Table 4. Experimental Parameters values

Parameter	Value
Learning Rate (LR)	5e-5
Batch Size	32 (for training), 8 (for validation)
Training Ratio	Not specified (default settings)
Image Preprocessing	Rescaling by 1./255
Optimizer	Adam with LR of 5e-5
Training Epochs	30
Architecture	VGG16 with imagenet weights, fine-tuned

In this study investigated the impact of different picture input sizes on model performance using JAFFE and KDEF datasets. Various dimensions, including 48x48, 64x64, 128x128, 224x224, and 360x360 pixels, were examined. Results indicated that the model's performance was influenced by picture input size, evident in accuracy percentages. Notably, the JAFFE and KDEF datasets responded differently to input size variations, emphasizing the importance of optimizing image dimensions for each dataset to achieve optimal model performance, as illustrated in the figure 5 below.

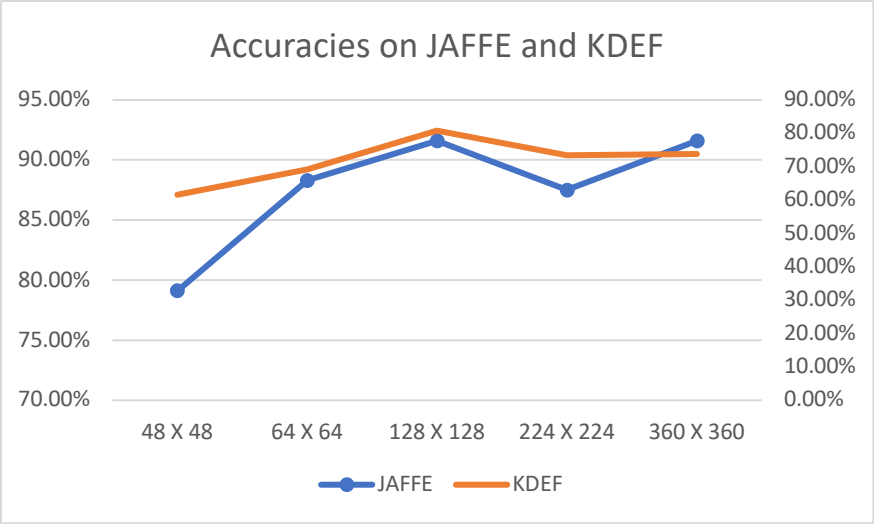

Figure 5. Image Size and their Accuracies on JAFFE and KDEF

3.7 Results

This section assesses the performance of our proposed model on established benchmark datasets. Initial experiments with a standard CNN gauge its performance, followed by examining the impacts of various fine-tuning modes using VGG-16. The model's performance is further evaluated by comparison with several pre-trained Deep Convolutional Neural Network (DCNN) models. Results, showcased in the table, depict accuracies on the test set using 2 × 2 MaxPooling and a 3 × 3 size kernel in a typical CNN. Experiments were conducted on both KDEF and JAFFE datasets, considering input sizes from 360 × 360 to 48 × 48. The reported accuracies represent the highest levels achieved after 50 iterations with specific configurations, with a randomly selected 10% of the total data serving as the test set.

The findings reveal a positive correlation between larger input image sizes and accuracy, up to a certain threshold, evident in both KDEF and JAFFE datasets. For instance, adjusting the input size to 360 × 360 yielded the highest accuracy of 73.87% on the KDEF dataset, while a smaller input size of 48 × 48 resulted in 61.63%. The importance of larger image sizes in enhancing model performance aligns with the intuitive notion that more extensive images inherently carry more information.

Surprisingly, the largest input size of 360 × 360 did not yield the highest accuracy, as indicated in table 5. Optimal results for both datasets were achieved with a picture dimension of 128 × 128, aligning with the model's inherent optimization for processing data of this size. Larger input sizes would demand increased data quantity and a more comprehensive model for improved performance, but in this context, the model is well-optimized for an input size of 128 × 128.

Table 5. Image Input Size and Accuracies on KDEF and JAFFE

Image Input Size	JAFFE	KDEF

48 X 48	79.1%	61.6%
64 X 64	88.3%	69.3%
128 X 128	91.6%	80.8%
224 X 224	87.5%	73.4%
360 X 360	91.6%	73.8%

The proposed approach leverages complex CNN architectures with Transfer Learning (TL) to mitigate overfitting in datasets with limited information. By applying TL techniques to pre-trained models, the strategy aims to enhance the model's generalization across diverse input sizes and datasets, improving reliability in facial emotion identification tasks.

A set of experiments explores the significance of fine-tuning within the transfer learning-based system. Table 6 summarizes test results for the proposed model using VGG-16 in various fine-tuning modes, evaluated on the KDEF and JAFFE datasets. Figure 6 illustrates emotion classification by the KDEF dataset, with accuracy values derived from the 10% test set of randomly selected data.

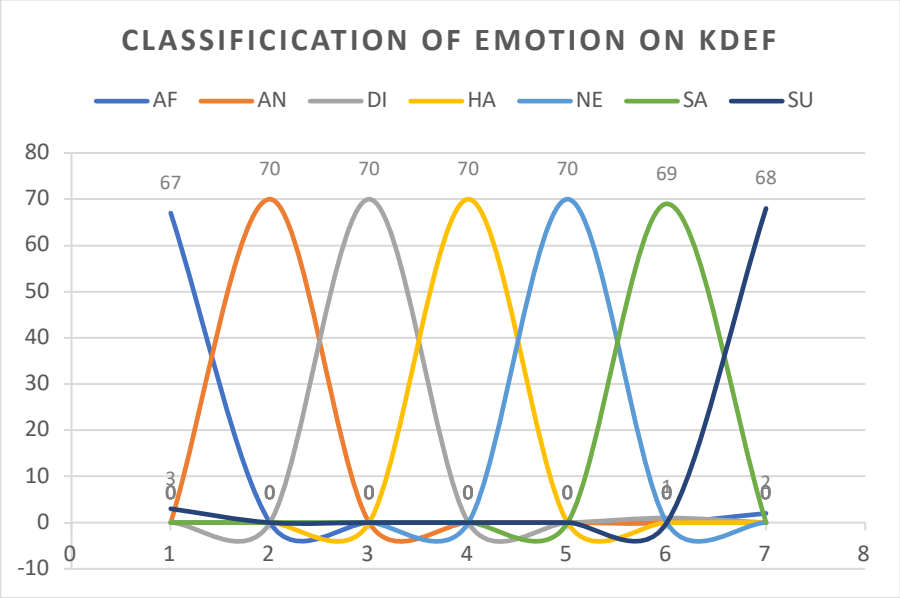

Figure 6. Classification of Emotion on KDEF Dataset

Throughout the tests, 50 training iterations were conducted for each fine-tuning mode. Specifically, during the entire process of fine-tuning the model, the dense layers were trained for the initial 10 iterations, followed by an allocation of the next 40 iterations to incorporate Conv blocks from the pre-trained foundation. Furthermore, the outcomes for the situation when the complete model is trained from the beginning, with random initialization, for a total of 50 iterations are displayed to offer a comprehensive assessment of the effectiveness of the recommended transfer learning strategy.

Table 6. Mode of Training and Accuracies of JAFFE and KDEF

Mode of Training	KDEF	JAFFE
Only Dense Layers	77.55%	91.67%
Dense Layers and VGG16 Block	91.83%	95.83%
Entire Model (Dense Layers and Full VGG16 Base)	93.47%	100.0%

To gain a deeper comprehension of the relative effectiveness of different fine-tuning alternatives, examine the information shown in Table 6. The combination of dense layer and VGG-16 Block 5 fine-tuning produces particularly remarkable results, surpassing those achieved by fine-tuning the dense layer alone. Due to this disparity, it is crucial to modify the last block, specifically Block 5, of the pre-trained VGG-16 model. Using the entire VGG-16 base for fine-tuning, as opposed to just Block 5, leads to improved

outcomes. This realization emphasizes the essentiality of the full VGG-16 architecture in enhancing the model's performance through fine-tuning.

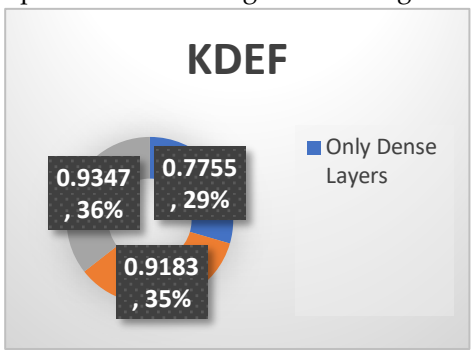

Figure 7. Mode of training on KDEF with Accuracies

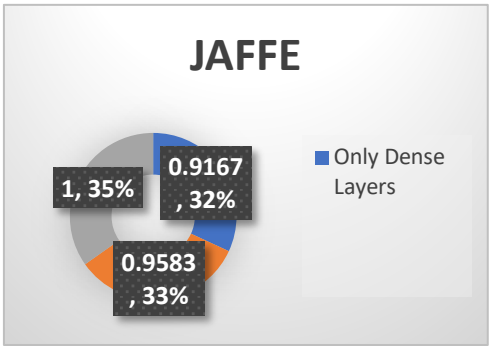

Figure 8. Model of training on JAFFE with Accuracies

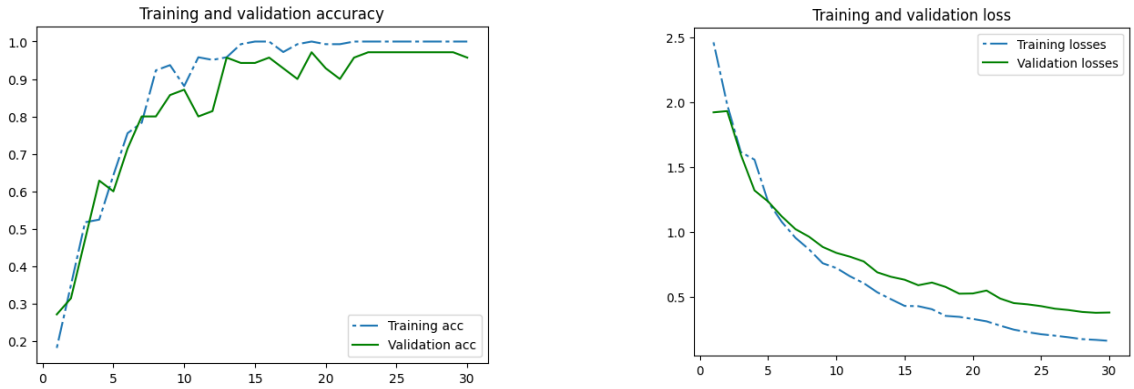

Figure 9. Show Accuracies and loss of the Proposed model

The maximum achievable accuracies in this context are very remarkable. They reach 93.47% for the KDEF dataset and 100% for the JAFFE dataset as shown in the figure 9. The effectiveness of the recommended transfer learning-based system is evident from these results, particularly when employing a tuning method that incorporates the dense layers and the last block of the pre-trained VGG-16 model. This method delivers exceptional precision in facial emotion recognition tests, and the findings also indicate that it is superior to utilize the complete VGG-16 basis for fine-tuning.

4. Discussion

This section evaluates the effectiveness of the proposed facial expression recognition (FER) strategy by comparing its performance with the leading approaches in emotion recognition, using the KDEF and JAFFE datasets. Table 6 presents a comprehensive understanding of the recognition accuracy of the test set, the separation of training and test data, and the distinct characteristics of each strategy. The analysis integrates both conventional and deep learning-based approaches.

The JAFFE dataset is widely utilized by most of the current methodologies. Nevertheless, the dataset is quite small, comprising only 213 samples. Ultimately, only a small number of approaches considered 210 samples. Nevertheless, the KDEF dataset comprises a considerable number of 4900 pictures, encompassing both side and profile views. Previous studies have only utilized a small portion of this dataset, often selecting subsets of it, such as 980 frontal photographs [17,34,35], or even fewer images [45], rather than the entire dataset.

Categorizing images that only display the front is significantly simpler compared to those that depict both the front and the profile. The table enumerates various methodologies employed in prior research to partition samples into training and test cohorts. The comparison table facilitates understanding of the effectiveness of each method by demonstrating the correlation between each method and the techniques used in feature selection and classification.

Table 7. Comparison of Accuracies with Our Method

Author	Year	Method	Accuracies in %	
			KDEF	JAFFE
Liew and Yairi, [17].	2015	Features are extracted using Gabor, Haar, LBP, and other methods, and then classified using SVM, KNN, LDA, and other techniques.	82.40	89.50
Zhao and Zhang [22]	2015	DBN is employed for the purpose of unsupervised feature acquisition, while NN is utilized for classification.		90.95
Ruiz-Garcia et al. [36]	2017	To set the CNN's initial weights, one uses a Stacked Convolutional Auto-Encoder (SCAE).	92.52	
Alshami el al. [35]	2017	Put SVM's Facial Landmarks and Centre of Gravity descriptors to work	90.80	91.90
Jain et al. [45]	2018	A hybrid deep learning architecture combining Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN).		94.91
Joseph and Geetha [46]	2019	Facial geometry-based feature extraction employing many classification techniques, including Support Vector Machines (SVM) and k-Nearest Neighbours (KNN).	31.20	
Bendjillali et al. [24]	2019	Using CNN for image improvement, feature extraction, and classification		98.63
Our Proposed Method with VGG-16		a pre-trained deep convolutional neural network (DCNN) for transfer learning.	**93.47**	**100**

(Table 7) shows that by utilizing the JAFFE and KDEF datasets, our proposed solution surpassed previous methods in identifying face emotions. This achievement was made possible by employing a pre-trained VGG-16 deep convolutional neural network (DCNN) for transfer learning. Liew and Yairi (2015) employed Support Vector Machines (SVM), k-Nearest Neighbors (KNN), and Linear Discriminant Analysis (LDA) as their classification algorithms. The use of Gabor, Haar, and Local Binary Pattern (LBP) characteristics resulted in accuracy rates of 82.40% for KDEF and 89.50% for JAFFE, correspondingly. Zhao and Zhang (2015) achieved a classification accuracy of 90.95% by employing Deep Belief Networks (DBN) to acquire unstructured features and neural networks (NN) for the remaining tasks. Ruiz-Garcia et al. (2017) achieved a 92.52% accuracy rate by employing a Stacked Convolutional Auto-Encoder (SCAE) to determine the initial weights of a Convolutional Neural Network (CNN). The researchers from Alshami et al. (2017) reported high accuracies of face recognition at 90.80% by combining Support Vector Machine classifiers and face landmarks feature to define their positions and with centre of gravity descriptors at 91.90%. The Jain, et al. (2018) succeeded in the achievement a accuracy of 94.91% and they #introduced deep learning architecture that amalgamated Convolutional Neural Networks (CNNs) and Recurrent Neural Neworks (RNNs). Joseph and Geetha (2019) are able to attain successful result by applying a series of classifi-cation tech-niques i.e. SVM and KNN to extract geometric based features from photos. As a result, Bendjillali et al. (2019) have reported an astounding 98.63% accuracy in the image enhancing, feature extracting, and classification processes through the application of a CNN. Our suggested solution outperforms existing approaches in facial emotion identification, as seen by achieving a 93.47% accuracy rate for KDEF and a perfect score of 100% for JAFFE as shown in (Figure 10). This demonstrates a significant advancement in the state-of-the-art in this field.

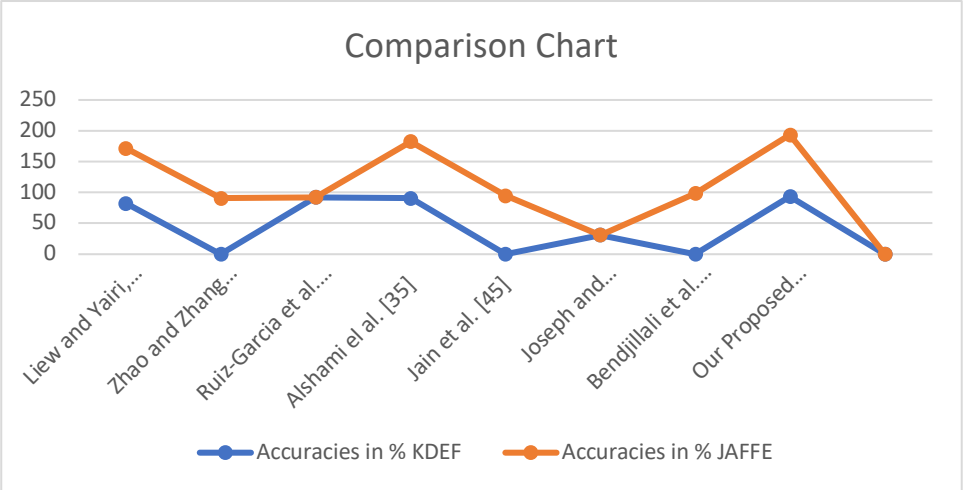

Figure 10. Comparison with Previous Approaches on KDEF and JAFFE

5. Conclusions

The research begins by using benchmark datasets and evaluates the performance of the proposed facial emotion recognition model using both traditional CNNs and the VGG-16 architecture, along with other fine-tuning approaches. The study reveals that larger dimensions are associated with enhanced model accuracy, prompting an investigation into the impact of input picture sizes on accuracy. The importance of a well-optimized model design for certain input dimensions is emphasized by the fact that the highest level of accuracy is achieved when using an image size of 128 × 128.

In order to mitigate the problem of overfitting, especially in datasets with limited data points, the recommended approach involves employing Transfer Learning (TL) and complex Convolutional Neural Network (CNN) structures. The objective of the model is to enhance the accuracy of face emotion identification tasks by utilizing transfer learning on pre-trained models. This approach aims to improve the ability of the model to generalize across different input sizes and datasets.

For future studies, we will do further research on the impact of fine-tuning of the transfer learning-based system. When comparing the results achieved by fine-tuning only the dense layer with the results achieved by combining the dense layer with VGG-16 Block 5 fine-tuning, the latter shows outstanding precision. It should be noted that the final block of the pre-trained VGG-16 model, namely Block 5, needs to be modified. The efficacy of the proposed transfer learning technique is evidenced by the superior results obtained by fine-tuning with the entire VGG-16 base.

The visualization's provided allow you to assess the performance of the model across different fine-tuning configurations. These visualization's include charts depicting accuracy and loss. It is worth noting that the highest achievable accuracies for the KDEF dataset are an impressive 93.47%, and for the JAFFE dataset, an astonishing 100%. Utilizing the dense layers and final block of the pre-trained VGG-16 model for fine-tuning demonstrates the remarkable effectiveness of the transfer learning-based system, as indicated by these findings.

Ultimately, the findings indicate that employing the complete VGG-16 basis for fine-tuning enhances overall performance, and the proposed model demonstrates exceptional accuracy in tasks involving the identification of facial emotions. The primary discovery of this study is the identification of the optimal arrangement of deep learning models for detecting facial emotions. This emphasizes the need of employing transfer learning and fine-tuning strategies to achieve superior outcomes.

Author Contribution: The concept that was presented resulted from the collaboration of Ahmed Faraz, Ahmad Naeem, Muhammad Fuzail, Kamran Abid, Ali Haider Khan, and Mueed Ahmed Mirza, with Ahmed Faraz taking the lead in developing the theory and conducting the computations. It was the responsibility of Muhammad Fuzail and Kamran Abid to verify the analytical methods. Ahmad Naeem and Ali Haider played a crucial role by motivating Ahmed Faraz to explore a specific topic and overseeing the study's outcomes. Throughout the process, each author actively engaged in discussions about the findings, contributing to the final publication.
Conflicts of Interest: The authors declare no conflict of interest.

References

1. L. Zhang, S. Wang, and B. Liu, "Deep learning for sentiment analysis: A survey," Wiley Interdiscip Rev Data Min Knowl Discov, vol. 8, no. 4, p. e1253, Jul. 2018, doi: 10.1002/WIDM.1253.
2. S. Li and W. Deng, "Deep Facial Expression Recognition: A Survey," IEEE Trans Affect Comput, vol. 13, no. 3, pp. 1195–1215, Apr. 2018, doi: 10.1109/TAFFC.2020.2981446.
3. T. Baltrusaitis, A. Zadeh, Y. C. Lim, and L. P. Morency, "OpenFace 2.0: Facial behavior analysis toolkit," in Proceedings - 13th IEEE International Conference on Automatic Face and Gesture Recognition, FG 2018, Institute of Electrical and Electronics Engineers Inc., Jun. 2018, pp. 59–66. doi: 10.1109/FG.2018.00019.
4. J. Booth, A. Roussos, S. Zafeiriou, A. Ponniah, and D. Dunaway, "A 3D Morphable Model learnt from 10,000 faces." [Online]. Available: http://www.ibug.doc.ic.ac.uk/resources/lsfm
5. 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, no. 5, pp. 2439–2450, May 2019, doi: 10.1109/TIP.2018.2886767.
6. Z. Zhu and Q. Ji, "Robust real-time eye detection and tracking under variable lighting conditions and various face orientations," *Computer Vision and Image Understanding*, vol. 98, no. 1, pp. 124–154, Apr. 2005, doi: 10.1016/j.cviu.2004.07.012.
7. 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, no. 5, pp. 2439–2450, May 2019, doi: 10.1109/TIP.2018.2886767.
8. J. Pena-Garijo *et al.*, "Specific facial emotion recognition deficits across the course of psychosis: A comparison of individuals with low-risk, high-risk, first-episode psychosis and multi-episode schizophrenia-spectrum disorders," *Psychiatry Res*, vol. 320, p. 115029, Feb. 2023, doi: 10.1016/J.PSYCHRES.2022.115029.
9. Q. Meng, X. Hu, J. Kang, and Y. Wu, "On the effectiveness of facial expression recognition for evaluation of urban sound perception," *Science of the Total Environment*, vol. 710, Mar. 2020, doi: 10.1016/j.scitotenv.2019.135484.
10. T. Hussain, B. Yang, H. U. Rahman, A. Iqbal, F. Ali, and B. shah, "Improving Source location privacy in social Internet of Things using a hybrid phantom routing technique," *Comput Secur*, vol. 123, Dec. 2022, doi: 10.1016/J.COSE.2022.102917.
11. T. Akter *et al.*, "Improved transfer-learning-based facial recognition framework to detect autistic children at an early stage," *Brain Sci*, vol. 11, no. 6, Jun. 2021, doi: 10.3390/brainsci11060734.
12. M. Rahul, N. Tiwari, R. Shukla, D. Tyagi, and V. Yadav, "A New Hybrid Approach for Efficient Emotion Recognition using Deep Learning," *International Journal of Electrical and Electronics Research*, vol. 10, no. 1, pp. 18–22, 2022, doi: 10.37391/IJEER.100103.
13. P. Kumar and B. Raman, "A BERT based dual-channel explainable text emotion recognition system," *Neural Netw*, vol. 150, pp. 392–407, Jun. 2022, doi: 10.1016/J.NEUNET.2022.03.017.
14. P. Tzirakis, G. Trigeorgis, M. A. Nicolaou, B. W. Schuller, and S. Zafeiriou, "End-to-End Multimodal Emotion Recognition using Deep Neural Networks."
15. S. Lai, H. Xu, X. Hu, Z. Ren, and Z. Liu, "Multimodal Sentiment Analysis: A Survey," May 2023, [Online]. Available: http://arxiv.org/abs/2305.07611
16. X.-X. Qi and W. Jiang, "Application of Wavelet Energy Feature in Facial Expression Recognition."
17. L. Zhao, G. Zhuang, and X. Xu, "Facial expression recognition based on PCA and NMF," *Proceedings of the World Congress on Intelligent Control and Automation (WCICA)*, pp. 6826–6829, 2008, doi: 10.1109/WCICA.2008.4593968.
18. X. Feng, M. Pietikäinen, and A. Hadid, "Facial expression recognition based on local binary patterns," *Pattern Recognition and Image Analysis*, vol. 17, no. 4, pp. 592–598, Dec. 2007, doi: 10.1134/S1054661807040190/METRICS.
19. R. Zhi and Q. Ruan, "Facial expression recognition based on two-dimensional discriminant locality preserving projections," *Neurocomputing*, vol. 71, no. 7–9, pp. 1730–1734, Mar. 2008, doi: 10.1016/J.NEUCOM.2007.12.002.
20. C.-C. Lee, C.-Y. Shih, W.-P. Lai, and P.-C. Lin, "An improved boosting algorithm and its application to facial emotion recognition," *Journal of Ambient Intelligence and Humanized Computing 2011 3:1*, vol. 3, no. 1, pp. 11–17, Oct. 2011, doi: 10.1007/S12652-011-0085-8.
21. C. Y. Chang and Y. C. Huang, "Personalized facial expression recognition in indoor environments," *Proceedings of the International Joint Conference on Neural Networks*, 2010, doi: 10.1109/IJCNN.2010.5596316.
22. F. Y. Shih, C. F. Chuang, and P. S. P. Wang, "PERFORMANCE COMPARISONS OF FACIAL EXPRESSION RECOGNITION IN JAFFE DATABASE," *https://doi.org/10.1142/S0218001408006284*, vol. 22, no. 3, pp. 445–459, Nov. 2011, doi: 10.1142/S0218001408006284.
23. C. Shan, S. Gong, and P. W. McOwan, "Facial expression recognition based on Local Binary Patterns: A comprehensive study," *Image Vis Comput*, vol. 27, no. 6, pp. 803–816, May 2009, doi: 10.1016/J.IMAVIS.2008.08.005.
24. T. Jabid, M. H. Kabir, and O. Chae, "Robust Facial Expression Recognition Based on Local Directional Pattern," *ETRI Journal*, vol. 32, no. 5, pp. 784–794, Oct. 2010, doi: 10.4218/ETRIJ.10.1510.0132.
25. H. Alshamsi, V. Kepuska, and H. Meng, "Real time automated facial expression recognition app development on smart phones," *2017 8th IEEE Annual Information Technology, Electronics and Mobile Communication Conference, IEMCON 2017*, pp. 384–392, Nov. 2017, doi: 10.1109/IEMCON.2017.8117150.
26. C. F. Liew and T. Yairi, "Facial expression recognition and analysis: A comparison study of feature descriptors," *IPSJ Transactions on Computer Vision and Applications*, vol. 7, pp. 104–120, 2015, doi: 10.2197/ipsjtcva.7.104.
27. 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.
28. X. Zhao, X. Shi, and S. Zhang, "Facial Expression Recognition via Deep Learning," *IETE Technical Review*, vol. 32, no. 5, pp. 347–355, 2015, doi: 10.1080/02564602.2015.1017542.

29. E. Pranav, S. Kamal, C. Satheesh Chandran, and M. H. Supriya, "Facial Emotion Recognition Using Deep Convolutional Neural Network," *2020 6th International Conference on Advanced Computing and Communication Systems, ICACCS 2020*, pp. 317–320, Mar. 2020, doi: 10.1109/ICACCS48705.2020.9074302.
30. A. Mollahosseini, D. Chan, and M. H. Mahoor, "Going deeper in facial expression recognition using deep neural networks," *2016 IEEE Winter Conference on Applications of Computer Vision, WACV 2016*, May 2016, doi: 10.1109/WACV.2016.7477450.
31. G. Pons and D. Masip, "Supervised Committee of Convolutional Neural Networks in Automated Facial Expression Analysis," *IEEE Trans Affect Comput*, vol. 9, no. 3, pp. 343–350, Jul. 2018, doi: 10.1109/TAFFC.2017.2753235.
32. G. Wen, Z. Hou, H. Li, D. Li, L. Jiang, and E. Xun, "Ensemble of Deep Neural Networks with Probability-Based Fusion for Facial Expression Recognition," *Cognit Comput*, vol. 9, no. 5, pp. 597–610, Oct. 2017, doi: 10.1007/S12559-017-9472-6/METRICS.
33. A. Ruiz-Garcia, M. Elshaw, A. Altahhan, and V. Palade, "Stacked deep convolutional auto-encoders for emotion recognition from facial expressions," *Proceedings of the International Joint Conference on Neural Networks*, vol. 2017-May, pp. 1586–1593, Jun. 2017, doi: 10.1109/IJCNN.2017.7966040.
34. H. Ding, S. K. Zhou, and R. Chellappa, "FaceNet2ExpNet: Regularizing a Deep Face Recognition Net for Expression Recognition," in *2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017)*, IEEE, May 2017, pp. 118–126. doi: 10.1109/FG.2017.23.
35. N. Jain, S. Kumar, A. Kumar, P. Shamsolmoali, and M. Zareapoor, "Hybrid deep neural networks for face emotion recognition," *Pattern Recognit Lett*, vol. 115, pp. 101–106, Nov. 2018, doi: 10.1016/j.patrec.2018.04.010.
36. S. Shaees, H. Naeem, M. Arslan, M. R. Naeem, S. H. Ali, and H. Aldabbas, "Facial Emotion Recognition Using Transfer Learning," *2020 International Conference on Computing and Information Technology, ICCIT 2020*, Sep. 2020, doi: 10.1109/ICCIT-144147971.2020.9213757.
37. R. I. Bendjillali, M. Beladgham, K. Merit, and A. Taleb-Ahmed, "Improved facial expression recognition based on DWT feature for deep CNN," *Electronics (Switzerland)*, vol. 8, no. 3, Mar. 2019, doi: 10.3390/electronics8030324.
38. D. Y. Liliana, "Emotion recognition from facial expression using deep convolutional neural network," in *Journal of Physics: Conference Series*, Institute of Physics Publishing, Apr. 2019. doi: 10.1088/1742-6596/1193/1/012004.
39. M. Shi, L. Xu, and X. Chen, "A Novel Facial Expression Intelligent Recognition Method Using Improved Convolutional Neural Network," *IEEE Access*, vol. 8, pp. 57606–57614, 2020, doi: 10.1109/ACCESS.2020.2982286.
40. Q. T. Ngoc, S. Lee, and B. C. Song, "Facial landmark-based emotion recognition via directed graph neural network," *Electronics (Switzerland)*, vol. 9, no. 5, May 2020, doi: 10.3390/electronics9050764.
41. X. Jin, W. Sun, and Z. Jin, "A discriminative deep association learning for facial expression recognition," *International Journal of Machine Learning and Cybernetics*, vol. 11, no. 4, pp. 779–793, Apr. 2020, doi: 10.1007/S13042-019-01024-2/METRICS.
42. S. Porcu, A. Floris, and L. Atzori, "Evaluation of data augmentation techniques for facial expression recognition systems," *Electronics (Switzerland)*, vol. 9, no. 11, pp. 1–12, Nov. 2020, doi: 10.3390/electronics9111892.
43. "Japanese Female Facial Expression (JAFFE) Database." Accessed: Jan. 07, 2024. [Online]. Available: https://www.kasrl.org/jaffe_download.html
44. M. G. Calvo and D. Lundqvist, "Facial expressions of emotion (KDEF): Identification under different display-duration conditions," *Behav Res Methods*, vol. 40, no. 1, pp. 109–115, Feb. 2008, doi: 10.3758/BRM.40.1.109/METRICS.
45. N. Jain, S. Kumar, A. Kumar, P. Shamsolmoali, and M. Zareapoor, "Hybrid deep neural networks for face emotion recognition," *Pattern Recognit Lett*, vol. 115, pp. 101–106, Nov. 2018, doi: 10.1016/J.PATREC.2018.04.010.
46. 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/METRICS.