Facial Emotion Recognition: Leveraging Transfer Learning for Enhanced Decoding

Karthik Reddy Munnangi, Jampala Venkata Saileenath Reddy , Sai Ruthvik Reddy, Aella,Kalipindi Navya, Dr. S Sri Harsha

Department of CS&IT
Koneru Lakshmaiah Education Foundation India
Department of CS&IT
Koneru Lakshmaiah Education Foundation India
Department of CS&IT
Koneru Lakshmaiah Education Foundation India
Department of CS&IT
Koneru Lakshmaiah Education Foundation India
Department of CS&IT
Koneru Lakshmaiah Education Foundation India

Abstract:

The recognition of facial emotions, also known as facial emotion recognition (FER), continues to be a crucial component of human-computer interaction and artificial intelligence applications. Various approaches are being developed to improve the accuracy of FER. This research article provides a thorough examination of Facial Expression Recognition (FER), with a specific emphasis on doing a comparison study between Convolutional Neural Networks (CNN) and Transfer Learning models. The primary objective is to get the highest possible accuracy in the categorization of emotions. This paper thoroughly investigates the complex terrain of facial expression recognition (FER), while recognising the difficulties presented by diverse lighting conditions, face emotions, and subtleties within the dataset. This study explores the capabilities of Convolutional Neural Networks (CNN), a well- established deep learning architecture, and Transfer Learning, a technique that utilises pre-trained models, in effectively capturing the nuanced aspects of facial expressions. The experimentation include rigorous training and testing on a wide range of datasets, assessing the accuracy, robustness, and generalizability of the models across many situations. Both CNN and Transfer Learning offer impressive accuracy in the field of Facial Expression Recognition (FER), with each approach showcasingdistinct capabilities in addressing certain issues. Furthermore, this study examines the interpretability of judgements made by both models, providing insights into the facial areas that have a substantial impact on the results of emotion identification. The present research offers significant insights into the underlying mechanisms of these models, hence enhancing our comprehension of their effectiveness in practical contexts.

Introduction:

The field of facial emotion recognition, often known as FER, is at the vanguard of human-computer interaction (HCI) because it provides an essential link between technology and human feelings. Within the realms of computer vision and artificial intelligence exists an interdisciplinary area with the overarching goal of endowing robots with the capacity to recognise, understand, and react appropriately to the many human facial emotions. The ability to grasp the emotional states that are communicated via facial signals is at the core of FER. This ability mirrors human cognitive capacities inreading emotions like as pleasure, sorrow, rage, surprise, fear, and contempt, as well as more complexmanifestations of emotion.

An intricate interaction of algorithms, machine learning models, and deep neural networks is requiredin order to accomplish the goal of effective face emotion identification. It analyses changes in expressions, eye movements, micro-expressions, and tiny subtleties that transmit a spectrum of emotions. This is accomplished by navigating through the complex terrain of facial characteristics. In order to achieve reliability, scalability, and real-time application in FER systems, researchers are always investigating new procedures. These methodology might range from conventional feature extractionmethods to the most cutting-edge deep learning architectures.

FER's relevance is not limited to the applications it has in human-computer interaction; rather, it extends to a wide variety of fields that include psychology, healthcare, marketing, gaming, and many more. Its potential to facilitate empathic artificial intelligence, improve mental health diagnoses, enhance user experiences, and optimise human-robot interactions highlights the need of continuouslyexploring and refining FER methodologies. The quest of greater accuracy, cross-domain adaptability, and ethical issues continues to be crucial as the discipline progresses, moving FER into new boundaries of understanding human emotions using computer methods.

Dataset Description:

The FER2013 dataset, which was developed by Pierre-Luc and Aaron S. in 2013, has become a significant asset in the field of facial expression recognition studies. The dataset was carefully selected in order to provide a standardised baseline for the development and evaluation of algorithms specifically designed for autonomous facial expression detection systems. The dataset consists of 35,887 grayscale photos, with each image measuring 48x48 pixels. It contains a range of seven unique face emotions, including anger, contempt, fear, pleasure, sorrow, surprise, and a neutral expression. One notable characteristic of the FER2013 dataset is its wide-ranging origins, including photos sourced from various channels such as pre-existing datasets and web search engines. The intentional incorporation of diverse facial expressions from numerous sources boosts the dataset's efficacy in training and assessing models. Researchers worldwide have used FER2013 as a standardised framework to evaluate and improve the efficacy of machine learning and deep learning models designed particularly for the intricate job of interpreting emotions from facial photos. The extensive use of the dataset highlights its pivotal contribution to the advancement of facial expression recognition research.[1]

Although the Fer2013 dataset has great importance in the field of facial expression recognition research, it is not devoid of problems. The need for a more precise and standardised dataset has beenacknowledged by researchers in order to rectify specific shortcomings and discrepancies inherent in the initial collection. The Fer2013 cleaned dataset is a carefully curated version that has undergone thorough processing to reduce noise and boost the overall quality of the dataset, hence improving thepotential for meaningful research outputs.

During the refining phase, diligent attempts have been made to address any mistakes and ambiguities in the annotation of face expressions. Furthermore, a rigorous process of curation has been implemented to eliminate discrepancies in picture quality and guarantee a more uniform collection of photographs. Researchers using the Fer2013 cleaned dataset may get advantages from a more standardised and dependable compilation of facial expressions. This, in turn, serves to mitigate any biases and inconsistencies that might potentially affect the efficacy of facial expression recognition programmes.

The curated iteration of the Fer2013 dataset not only preserves the wide range of facial expressions present in the original dataset but also aims to enhance the dataset's dependability as a standard for assessing the effectiveness of machine learning and deep learning models in the field of facial expression recognition. The cleaned Fer2013 dataset serves as evidence of the research community's dedication to improving fundamental resources in order to achieve more precise and significant progress in the area.[2]

Related Work:

The study of Human Facial Emotion Recognition (FER) has garnered significant attention in academic research circles owing to its potential for many practical applications. The main objective of Facial Expression Recognition (FER) is the process of associating different facial expressions with their respective emotional states. The conventional process of Facial Expression identification (FER) consists of two primary stages, namely feature extraction and emotion identification. At now, Deep Neural Networks, namely Convolutional Neural

Networks (CNNs), are widely used in Facial Expression Recognition (FER) owing to their intrinsic ability to extract features from pictures. Numerous research papers have been conducted to investigate the use of Convolutional Neural Networks (CNNs) includinga limited number of layers in order to tackle issues related to Facial Expression Recognition (FER). Nevertheless, conventional shallow convolutional neural networks (CNNs) equipped with simple learning algorithms possess restricted capacities in extracting features that effectively capture emotional information from high-resolution photos. A significant drawback seen in current methodologies is their predominant emphasis on frontal pictures, hence neglecting the significance of profile views in practical facial expression recognition (FER) systems. In order to tackle this issue, we have out a proposition for the use of a Very Deep Convolutional Neural Network (DCNN) model that incorporates the technique of Transfer Learning (TL). This methodology use a pre-existing deep convolutional neural network (DCNN) model. The model's dense top layer(s) are modified to conformto the specific needs of facial emotion recognition (FER). Subsequently, the model is fine-tuned usingfacial emotion data. This study presents a unique approach to pipeline method, which entails first training the dense layer(s) and then fine-tuning each pre-trained deep convolutional neural network (DCNN) block. This sequential process leads to a progressive improvement in the accuracy of the frameerror rate (FER). The evaluation of the proposed face Expression Recognition (FER) system encompasses eight distinct pre-trained Deep Convolutional Neural Network (DCNN) models, namely VGG-16, VGG-19, ResNet-18, ResNet-34, ResNet-50, ResNet-152, Inception-v3, and DenseNet-161. The evaluation is conducted using two well recognised face picture datasets, namely the Karolinska Directed Emotional Faces (KDEF) dataset and the Japanese Female Facial Expression (JAFFE) dataset. Facial expression recognition (FER) continues to pose challenges, particularly when considering frontal views exclusively. The KDEF dataset further complicates matters by including a wide range of photos that include both frontal and profile views, so introducing extra complexity. The proposed strategy demonstrates exceptional precision on both datasets by using pre-trained models. The use of a 10-fold cross-validation methodology yields notable FER accuracies when using DenseNet-161 on the test sets of KDEF and JAFFE, achieving 96.51% and 99.52% respectively. The evaluation findings highlight the excellent performance of the proposed Facial Emotion Recognition (FER) system in terms of its accuracy in detecting emotions. Furthermore, the evaluation of the KDEF dataset, including profile perspectives, illustrates the level of competence necessary for practical implementations.[3] The topicof emotion identification using facial expressions is of great significance in the field of man-machine interaction. However, it encounters several problems such as the presence of facial accessories, non-

uniform illuminations, and fluctuations in facial poses. The conventional methods for detectingemotions face the challenge of simultaneously optimising the extraction of features and the classification process. In response to this matter, there is a growing focus on the use of deep learningmethodologies, which have shown to be very effective in a range of categorization endeavours. The primary objective of this study is to investigate the use of transfer learning techniques in the field of emotion recognition. This research employs pre-trained networks like ResNet50, VGG19, Inception V3,and MobileNet. The completely linked layers of the pre-trained Convolutional Neural Networks (ConvNets) are eliminated, and new fully connected layers are included, customised to meet the unique demands of our job. Following this, only the recently included layers are capable of being trained in order to update the weights. The studies were carried out using the CK+ database, yieldingan average accuracy of 96% for the task of emotion identification.[4] Deficits in facial emotion identification are a commonly reported cognitive impairment that is prominent in a range of mental diseases, with a special emphasis on schizophrenia. Despite the considerable amount of research conducted, there is still a lack of definitive information about the variables that are linked to impairments connected to schizophrenia at various phases of the illness. This lack of solid data presents difficulties in effectively managing the condition in a therapeutic setting. Within the given setting, our research study offers a thorough and all-encompassing examination of the cognitive processes involved in recognising facial emotions among patients diagnosed with schizophrenia. In this study, we provide the Bruce-Young face recognition model, which has gained worldwide acclaim. Additionally, we provide a comprehensive analysis of many research that have investigated emotion detection throughout each step of the face recognition process, with a particular emphasis onbehavioural and event-related potential measures. The objective of this synthesis is to provide a scholarly contribution by providing further insights to the current body of information. Furthermore, we intend to propose potential avenues for future clinical research, focusing particularly on investigating the underlying processes associated with deficiencies in facial emotion identification in individuals with schizophrenia.[5] The identification of emotions based on facial pictures is a considerable obstacle owing to the fluid and ever-changing character of facial expressions. The existing body of literature pertaining to the use of deep learning models in emotion categorization has mostly focused on the domain of facial emotion identification. Nevertheless, these research face challenges in terms of performance deterioration, which may be linked to the inadequate selection of layers in convolutional neural network models. In light of the above issue, we propose an effective deep learning methodology that utilises a convolutional neural network model to achieve reliable categorization of emotions, detection of age, and identification of gender based on facial expressions. The experimental findings demonstrate the exceptional efficacy of the proposed model, outperforming previous studies by achieving an accuracy rate of 95.65% for emotion recognition, 98.5% for age identification, and 99.14% for gender recognition.[6] The process of facial emotion identification has intrinsic complexity due to the wide range of facial characteristics and the presenceof ethnic and cultural differences. The difficulty is heightened when confronted with complex emotions, which often include a combination of many emotional states, so making the differentiation between dominating and complimentary emotions nuanced. In order to tackle this complex matter, we have compiled an extensive database of 31,250 face photos that include a range of emotions shownby 115 individuals. The dataset has been carefully vetted to ensure a balanced representation of genders. Additionally, a competition was organised at the FG Workshop 2020 using the aforementioned dataset. This study explores a two-stage recognition technique, which consists of a coarse recognition step followed by a fine recognition stage. The primary objective of this method is to improve the categorization of symmetrical emotion labels.[7] In the domain of autonomous cars, the significance of precise Facial Emotion Recognition (FER) systems cannot be overstated, as they playa crucial role in the identification of driver emotions, hence aiding in the alleviation of road rage. Theefficacy of Facial Expression Recognition (FER) systems during real-time testing is mostly dependent on the quality of datasets rather than the complexity of algorithms. In order to optimise the performance of the Facial Expression Recognition (FER) system, particularly in the context of autonomous cars, we propose the implementation of the Facial Image Threshing (FIT) machine. By using sophisticated functionalities derived from pre-trained face recognition models and the Xceptionalgorithm, the face Image Transformation (FIT) technique encompasses a methodical process that encompasses the elimination of extraneous facial pictures, the acquisition of facial data, rectification of misalignments, and the comprehensive integration of original datasets. The use of this methodology, in conjunction with data augmentation methodologies, yielded a significant improvement of 16.95% in validation accuracy when compared to traditional procedures. This evaluation was conducted utilising the FER 2013 dataset. The assessment using a confusion matrix onan undisclosed proprietary dataset provided further confirmation of a 5% improvement in real-time testing compared to the first methodology used with the FER 2013 dataset.[8] Within the field of computer vision, the identification and interpretation of emotions based on facial expressions is a veryactive and evolving area of scholarly study. This work presents a novel feature descriptor that has beendeveloped specifically for the purpose of facial emotion identification. The proposed approach utilisesa modified version of the Histogram of Oriented Gradients (HOG) and Local Binary Pattern (LBP) featuredescriptor. The approach consists of two main components. The Viola-Jones algorithm is first used forthe purpose of identifying the facial area. Following this, the use of a Butterworth high-pass filter is employed to improve the identification of the area of interest. This process aids in the detection of theeye, nose, and mouth regions using the Viola-Jones method. During the second step, the modified Histogram of Oriented Gradients (HOG) and Local Binary Patterns (LBP) feature descriptors are used toextract features from the areas of the eye, nose, and mouth that have been detected. The characteristics derived from these three areas are combined and then subjected to dimensionality reduction via the use of Deep Stacked AutoEncoders. In the context of classification and recognition, the use of a multi-class Support Vector Machine is ultimately employed. The empirical findings provideevidence of the effectiveness of the suggested modified feature descriptors in accurately identifying emotions. These results have been confirmed using both the CK+ and JAFFE datasets.[9] This study presents a novel approach for face emotion identification, which utilises two recently formulated geometric attributes: landmark curvature and vectorized landmark. The aforementioned traits are obtained by using facial landmarks that correlate to distinct elements of facial muscle motions. The technique being suggested involves the integration of support vector machine (SVM)-based classification with a genetic algorithm (GA) in order to tackle the multi-attribute optimisation issue that encompasses feature and parameter selection. The researchers conducted experimental assessments on two datasets, namely the enlarged Cohn-Kanade dataset (CK+) and the Multimedia Understanding Group (MUG) dataset. The validation accuracy for the 8-class CK+, 7-class CK+, and 7- class MUG datasets were found to be 93.57%, 95.58%, and 96.29% respectively. Similarly, the test accuracy for same datasets were observed to be 95.85%, 97.59%, and 96.56% respectively. The average accuracy, recall, and F1-score were around 0.97, 0.95, and 0.96, respectively. In a comparisonexamination between a convolutional neural network (CNN) and the presented methodology, which is a commonly used approach for face emotion detection, the results showed a slightly better test accuracy for the provided strategy in the case of the 8-class CK+ dataset (95.85% for the presented technique using SVM, compared to 95.43% for the CNN). Similarly, for the 7-class CK+ dataset, the presented technique achieved a test accuracy of 97.59%, while the CNN achieved a slightly lower accuracy of 97.34%. In contrast, the Convolutional Neural Network (CNN) had a little superior performance on the 7-class MUG dataset, with an accuracy of 96.56% compared to 99.62% achieved by the alternative model. This particular strategy, which utilises simpler models in contrast to CNN- based methods, shows promise for real-time machine vision applications in automated systems.[10]

Facial Emotion Recognition (FER) refers to the computational task of accurately identifying human emotions based on facial expressions. This task is particularly difficult when attempting to assess stress and anxiety levels using computer vision techniques. The Internet of Medical Things (IoMT) has seen significant progress in gathering a wide range of data pertaining to mental and physical health, resulting in notable advantages. Deep Learning (DL) techniques, especially innovative ones, now facilitate the processing of Internet of Medical Things (IoMT) device data in resource-limited edge situations. This study presents a real-time implementation of a face expression detection and identification system using the Internet of Medical Things (IoMT) on the Raspberry Pi. The Raspberry Pi is a compact and resource-limited device that is effectively used by using deep convolutional neural networks. The research conducted included an empirical examination of human face expressions and emotional states via the use of physiological sensors. The model under consideration demonstrated a range of face emotion recognition test accuracies between 56% and 73% across many models. The accuracy achieved a remarkable 73%, surpassing the current state-of-the-art findings (maximum 64%) when evaluated using the FER 2013 dataset. A t-test was used to identify statistically significant variations in systolic and diastolic blood pressure, as well as heart rate, among people who were exposed to three distinct emotional stimuli (namely, anger, happiness, and neutrality).[11] In recent years, there has been notable progress in the field of facial emotion recognition (FER) research. This progress has led to the development of novel convolutional neural network (CNN) structures specifically tailored for the automated identification of facial emotions in still pictures. Although these networks have shown remarkable accuracy in identification, they also entail significant computational expenses and memory use. This presents difficulties for practical applications that want facial expression recognition systems to function in real-time on embedded devices with limited resources.In order to tackle these concerns and provide an effective approach for the automated identification of facial emotions in real-life situations, this study presents a new and innovative deep integrated Convolutional Neural Network (CNN) model called EmNet (Emotion Network). The EmNet system consists of two structurally comparable deep convolutional neural network (DCNN) models and their integrated form. These models are collaboratively optimised using a joint-optimization approach. EmNet utilises two fusion algorithms, namely average fusion and weighted maximum fusion, to provide three predictions for a given face picture, hence deriving the ultimate choice. In order to evaluate the effectiveness of the suggested Facial Expression Recognition (FER) pipeline on a limited- resource embedded platform, we conducted optimisations on the EmNet model and face detector using the TensorRT Software Development Kit (SDK). Subsequently, we installed the whole FER pipeline on the Nvidia Xavier device. The EmNet model, which has 4.80 million parameters and a model size of19.3MB, exhibits substantial improvements in accuracy compared to the present state-of-the-art. Additionally, it achieves a considerable increase in computing efficiency.[12] In this study, Daeha Kim presents a unique methodology for Facial Emotion Recognition (FER) that aims to overcome the constraints often seen in systems reliant on supervision. In contrast to traditional facial expression recognition (FER) systems that mainly depend on supervision information, the suggested approach uses adversarial learning to achieve generalised representation learning. This novel methodology obviates the need for stringent oversight, so enabling a more autonomous examination of emotions among people. The use of an adversarial learning framework improves the capacity of the Facial Expression Recognition (FER) network to understand complex emotional components that are inherent in intense emotions. This is achieved via the process of adversarially learning from weaker emotion samples, using their stronger counterparts as a reference. As a result, the approach described in this study demonstrates enhanced precision in the detection of facial emotions, hence guaranteeing immunity to variances across individuals. The research furthermore presents a novel contrastive loss function that is specifically tailored to improve the efficacy of adversarial learning. The adversarial learning scheme that has been suggested receives theoretical validation and exhibits superior performance via empirical trials, so establishing itself as a state-of-the-art solution in the respective area.[13] In this study, Michael K. Yeung undertakes a thorough examination to evaluate the degree of specificity in facial emotion recognition impairment and the impact of task variables within the context of autism spectrum disorder (ASD). The present study conducts a comprehensive analysis of subsets derived from 148 articles obtained from reputable databases such as PubMed and PsycINFO. Utilisingrandom-effects meta-analyses, the findings of this review demonstrate a noteworthy deficit in the ability to recognise all fundamental facial expressions among individuals with Autism Spectrum Disorder (ASD). A comparative investigation reveals that individuals with Autism Spectrum Disorder (ASD) have worse facial emotion detection abilities in comparison to those with other clinical conditions. This deficiency is shown in both emotional and nonemotional facial features, as well as across different modalities. The review highlights the presence of notable moderating effects in relation to emotion complexity and holistic processing. It also reveals a statistical trend pertaining to task type, while finding no major influence from motion, social relevance, or stimulus salience on facial emotion recognition in individuals with Autism Spectrum Disorder (ASD). In general, the results indicate that individuals with Autism Spectrum condition (ASD) have a broad deficit in their ability to recognise facial emotions. This impairment is not restricted to particular emotional facial features or the visual perception of facial expressions, but rather is inherent to the condition itself. The enhanced comprehension of task features aids in the exploration of the underlying processes involved in face emotion identification among persons with Autism Spectrum Disorder (ASD).[14] Aayushi Chaudhari conducts a study to evaluate the efficacy of automated emotion detection, specifically in the context of Facial Emotion detection (FER). The study used the ResNet-18 model in conjunction with transformers. The objective of this research is to establish a correlation between various facial expressions and the appropriate emotional states. To do this, the Vision Transformer is used for assessment purposes. Furthermore, the performance of this model is compared to other advanced models using hybrid datasets that include a combination of different data sources. The study presentsa comprehensive overview of the pipeline, including several stages such as face identification, cropping, and feature extraction. These stages use state-of-the-art deep learning models and a fine- tuned transformer for optimal performance. The experimental findings demonstrate the effectivenessof the suggested emotion recognition system, suggesting its potential for real-world applications.[15] In her study, Anouck I. Staff examines the social and emotional difficulties experienced by children whodisplay signs of attention-deficit/hyperactivity disorder (ADHD). The author places particular emphasis on the reduced ability to recognise facial emotions as a probable underlying component contributingto these obstacles. This research investigates the ability of children with (subthreshold) ADHD and a control group to recognise facial emotions. The study used a unique Morphed Facial Emotion detection Task (MFERT) to evaluate the accuracy of emotion detection across different emotions and degrees of severity. The results of the study indicate that youngsters displaying signs of ADHD have reduced accuracy in recognising emotions, particularly when it comes to identifying more intense expressions, in comparison to those without ADHD. Regression analyses conducted within the ADHD group reveal a negative association between the accuracy of emotion detection and the presence of social and emotional difficulties, indicating a possible role in the emergence of these issues among children afflicted by ADHD.[16] In his work, Jaemyung Kim presents a new methodology for the identification of facial emotions, which aims to tackle the issue of computational complexity and resource constraints often encountered in embedded systems. The study centres on an optimised convolutional neural network (CNN) design that prioritises efficiency. In order to augment hardware compatibility, anovel quantization technique is suggested, which only relies on integer arithmetic. The FERPlus-A dataset, which has been developed using a range of image processing methods, makes a valuable contribution towards enhancing generalisation and classification performance. Upon completion of the training and quantization processes, the CNN parameter size is estimated to be roughly 0.39

megabytes (MB), accompanied by approximately 28 million integer operations (IOPs). The assessmentconducted on the FERPlus test dataset demonstrates a classification accuracy of around 86.58%, which exceeds the performance achieved by prior research using lightweight convolutional neural networks. The real-time face expression detection system, when deployed on the Xilinx ZC706 SoC platform, demonstrates a performance of roughly 10 frames per second (FPS) at a clock frequency of 250 MHz, while using a power of 2.3 W.[17] Jannik Rößler conducts an examination of the psychological ramifications resulting from the substantial increase in videoconferencing as a consequence of the COVID-19 epidemic. The primary focus of this investigation is on the examination of the effect of participants' emotional states on their subjective evaluations of virtual meetings. The research encompasses a cohort of 35 individuals, organised into eight distinct teams, who engage in collaborative activities using the Zoom platform for the duration of a semester-long academic course.The use of facial emotion detection is implemented for the purpose of monitoring emotions in Zoom face video snapshots, with the ability to identify six distinct emotions, including happiness, sadness, fear, anger, neutrality, and surprise. The study centres on examining the relationship between the affective states of presenters and audience members and the evaluations provided subsequent to each presentation. The findings suggest that there is a positive correlation between the pleasure of the speaker and the emotional state of the audience. Moreover, presentations that elicit a range of emotions such as fear and delight tend to get better ratings. The results of this study provide valuable insights for those who participate in online video meetings, with the goal of enhancing the overall quality and minimising the tiredness often associated with videoconferencing.[18] In his discourse, Shervin Minaee examines the enduring obstacles encountered in the domain of facial expression identification, notwithstanding the progress made via study endeavours. Conventional approaches heavily depend on manually designed features such as SIFT, HOG, and LBP. While these techniques tend to perform well under controlled circumstances, they encounter difficulties when faced with datasets that exhibit more diversity. In recent years, there has been a notable development in the field of end-to-end facial expression detection with the introduction of deep learning models. However, it is important to note that more enhancements are still necessary in this area. Minaee presents a uniquemethodology that utilises an attentional convolutional network, highlighting its capacity to selectively attend to significant face areas and outperform earlier models on several datasets such as FER-2013, CK+, FERG, and JAFFE. The research incorporates a visualisation approach that identifies key facial regions for the detection of various emotions, hence highlighting the varying sensitivity of different emotions to specific portions of the face. The experimental results demonstrate significant enhancements attained with the suggested methodology.[19]

Transfer Learning:

Transfer learning is a strong paradigm in machine learning that has revolutionised the landscape of facial emotion recognition (FER) by using the knowledge gained from previously trained models to improve performance, efficiency, and flexibility. This is accomplished by exploiting the information gained from previously trained models. Transfer learning is a compelling strategy that may be used inFER since it allows for the transfer of learnt representations from one job to another, hence reducing the need for huge amounts of data and computer resources.

This strategy makes use of the concept that a model that has been trained on a substantial dataset fora task that is related may extract generalizable face characteristics that are not limited to particular datasets or emotions. Transfer learning improves model performance, even when there is a limited

amount of labelled data available, since it allows for the fine-tuning of pre-trained models or the extraction of learnt features and their application to new FER tasks. This strategy not only speeds up the convergence of the model but also improves its accuracy, which is very helpful when working withlimited resources in terms of data availability or computer capacity.

The adaptability of transfer learning is shown in a variety of FER applications, which range from the identification of emotions in still photos and moving movies to the development of real-time human-computer interaction systems. The basic architectures for transfer learning in FER include models like as VGG, ResNet, Inception, and EfficientNet. These models have been pre-trained on big datasets suchas ImageNet. Not only does the transfer of acquired representations from these models to FER tasks speed up the learning process, but it also improves the ability to differentiate between subtle facial expressions across a variety of datasets, environments, and feelings.

Even though transfer learning has the potential to make significant advances in FER, there are still a number of obstacles to overcome. These obstacles include domain adaptation, dataset bias, and the need for domain-specific fine-tuning. In addition, the deployment of transfer learning models for FER in diverse socioeconomic situations requires special attention due to ethical problems involving data privacy and model biases. As research into transfer learning continues to advance, there is a growing possibility that it may revolutionise FER by allowing systems that are more resilient, adaptive, and accurate across a wide variety of applications and domains.

Confusion matrix of transfer learning model:

The confusion matrix is, without a doubt, an indispensable instrument for gaining an understanding of the effectiveness of transfer learning models within the domain of facial emotion recognition (FER). This matrix provides a clear illustration of the categorization capabilities of the model by depicting the link between expected and actual feelings. The examples that belong to a predicted class are denoted by the rows of the matrix, whilst the instances that belong to an actual class are indicated by the columns. This matrix, which serves as the foundation of model assessment, makes it possible to conduct a detailed examination of categorization results, which reveals subtle insights into the strengths and shortcomings of the model.

Within the framework of transfer learning for FER, the confusion matrix performs the function of a compass, directing researchers to determine whether or not the model is capable of accurately identifying various emotional expressions. It details the accuracy of the model in recognising certain emotions such as happy, sorrow, rage, surprise, fear, or neutrality. In addition, this matrix sheds light on the possibility of incorrect classifications by revealing patterns in which the model may fail, such asconflating feelings that are similar to one another or showing favouritism towards certain expressions.

In order to fine-tune transfer learning models for FER, the examination of the confusion matrix gives information that is of great use. Researchers are able to identify areas for development by first determining where the model succeeds and where it fails. Potential areas for improvement include modifying hyperparameters, improving feature extraction, and addressing biases within the dataset. Additionally, the insights that are obtained from the confusion matrix help to enhance the resilience of the model, which enables it to generalise better across a variety of datasets or situations that occurin the real world.

This all-encompassing assessment tool not only helps in determining how accurate the model is, but it also contributes to strengthening the transfer learning framework for FER. As a result, it paves the way for emotion recognition systems that are more dependable and sensitive to nuances of expression across a wide range of applications and domains.

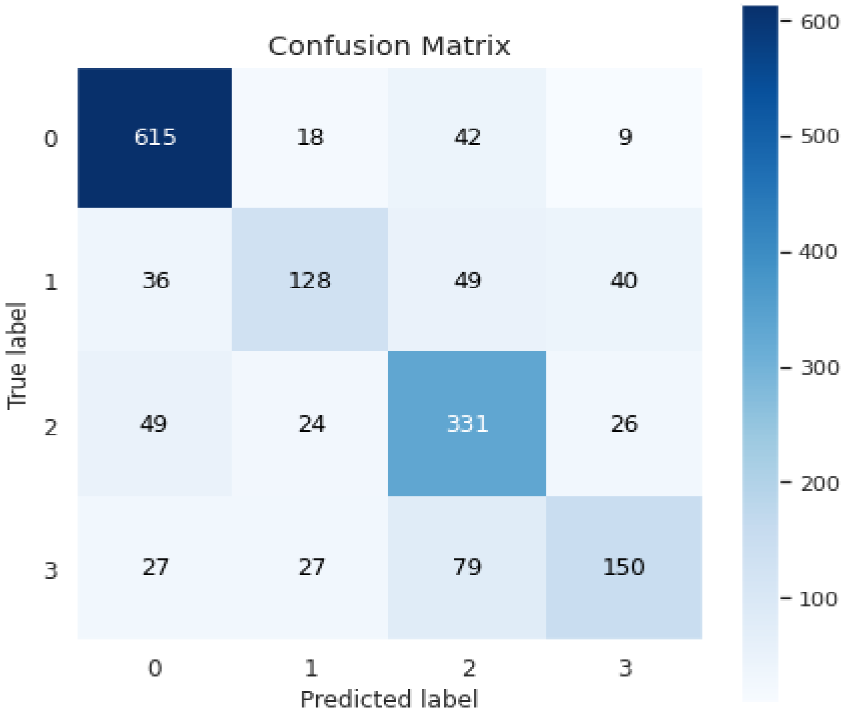


Fig. 1. Confusion matrix of transfer learning model

Accuracy of transfer learning model:

Achieving perfect accuracy is a vital aim in the field of facial emotion recognition (FER), especially whenapplying transfer learning approaches. In the current investigation, the use of transfer learning models has resulted in an impressive accuracy of 76% in the recognition of various facial emotions. These facial expressions include neutrality, happy, sorrow, anger, surprise, and fear, among others. This result is atestimonial to the effectiveness and flexibility of transfer learning paradigms in harnessing pre-existing

knowledge from one domain and using it adeptly to boost the recognition capacities in facial expression analysis. This achievement serves as a monument to the efficacy and adaptability of transfer learning paradigms.

The fact that transfer learning models were able to achieve an accuracy of 76% in FER demonstrates their capacity to recognise and classify even the most minute of differences in facial expressions, hence making a substantial contribution to the development of emotion detection technology. This level of accuracy in categorization indicates that the model is able to generalise and discriminate between many nuances of emotional state, which demonstrates positive developments in the model's potentialfor use in the real world.

The results of this study, which boast an accuracy rate of 76%, are of the utmost importance in the field of FER since they represent a significant step towards the development of emotion recognition systems that are more sophisticated and trustworthy. This degree of precision enables a sophisticated comprehension of facial expressions, which is set to lead the way for improved human-computer interactions, affective computing, and several other fields where accurate emotion identification is of crucial relevance. In addition, this accomplishment acts as a springboard for future developments. It encourages the refining and optimisation of transfer learning frameworks in order to reach higher levels of accuracy and applicability over a wider range of real-world settings.

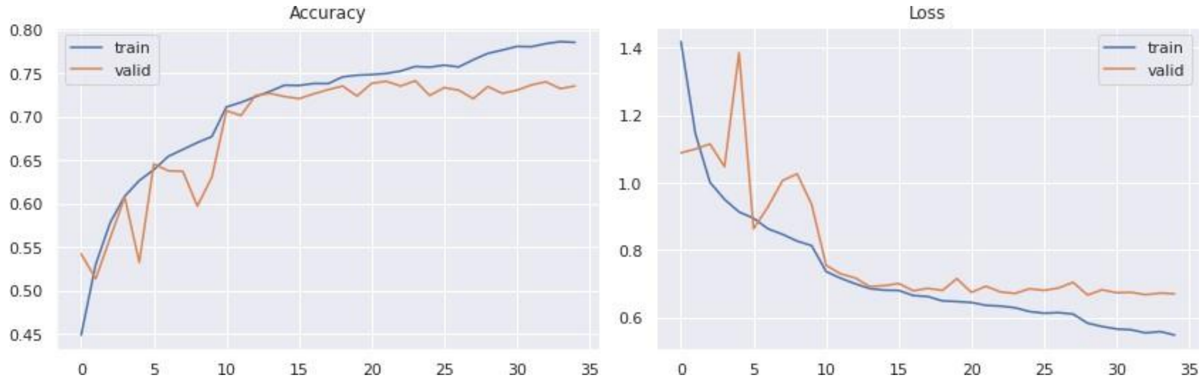


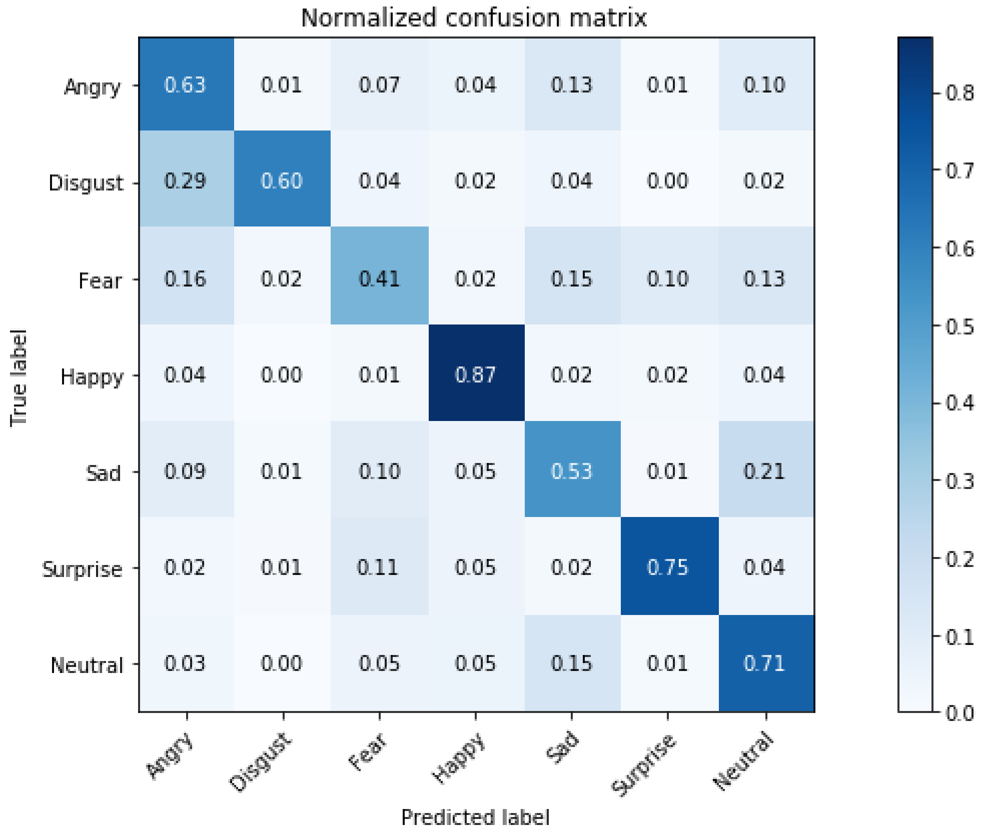
Fig. 2. Accuracy and Loss outcome of transfer learning model

Confusion matrix on CNN model:

In the field of facial emotion identification using Convolutional Neural Networks (CNNs), the confusionmatrix is a crucial tool for understanding the model's classification performance over a wide range ofemotional categories. This insight can be gained by comparing the results of the model with a databaseof known facial expressions. The purpose of this study is to determine whether or not the CNN is capable of correctly interpreting a wide range of complex facial expressions, including pleasure, sorrow, rage, surprise, fear, and neutrality. The confusion matrix is a graphical representation that outlines the accuracy of the model's categorization. It does this by presenting the number of true positives, true negatives, false positives, and false negatives for each category of emotion.

In the context of this work, the confusion matrix sheds light on the CNN's capacity to recognise and classify nuanced differences within facial expressions. It displays the model's ability to reliably identify circumstances in which emotions match with their right labels (known as true positives) and accurately exclude cases in which emotions are correctly recognised as not belonging to a certain category (known as true negatives). On the other hand, the matrix demonstrates situations in which emotions were misclassified, either by being incorrectly labelled as a particular emotion (referred to as false positives) or by being improperly ignored while being characteristic of a certain emotional state (referred to as false negatives).

This study delineates the model's strengths and areas of development in recognising a variety of facial expressions by visually displaying the CNN's performance via the confusion matrix. The matrix makes it possible to get an all-encompassing comprehension of the model's usefulness across a variety of emotional classifications, hence providing light on the CNN's nuanced capacity to correctly or incorrectly categorise emotions. This review contributes to the process of improving the CNN architecture and training procedures. The end goal is to strengthen the network's accuracy and precision in identifying complex facial emotions across a variety of datasets and in real-world settings.



	Angry	Disgust	Fear	Happy	Sad	Surprise	Neutral
Angry	0.63	0.01	0.07	0.04	0.13	0.01	0.10
Disgust	0.29	0.60	0.04	0.02	0.04	0.00	0.02
Fear	0.16	0.02	0.41	0.02	0.15	0.10	0.13
Happy	0.04	0.00	0.01	0.87	0.02	0.02	0.04
Sad	0.09	0.01	0.10	0.05	0.53	0.01	0.21
Surprise	0.02	0.01	0.11	0.05	0.02	0.75	0.04
Neutral	0.03	0.00	0.05	0.05	0.15	0.01	0.71

Fig. 3. Confusion matrix on CNN model

In the field of facial emotion identification using convolutional neural networks (CNNs), this study investigates the effectiveness of the CNN model in recognising and categorising the range of feelings that may be conveyed via facial expressions. The accuracy of 63% that was achieved serves as a quantitative indicator that reflects the CNN's skill of accurately categorising emotional states inside face pictures across a variety of datasets. This accuracy demonstrates a good performance in recognising emotions; nevertheless, it also highlights the intricacy involved in correctly reading and classifying a variety of facial expressions.

The results of this research show that CNN is proficient at properly assigning certain emotions to facephotos, as seen by the accuracy rate of 65%. This statistic represents the percentage of properly recognised emotions compared to the total number of emotions that were examined. It includes a wide variety of expressions, including happiness, sorrow, rage, surprise, fear, and neutrality. However, the accuracy of 63% speaks to the difficulty that are met in recognising and categorising deep emotional variations within facial portrayals. These obstacles were encountered by the researchers.

Recognising that CNN's face recognition system is accurate 65% of the time offers information on its capacity to catch and differentiate between specific emotional indicators displayed in facial expressions. However, this finding also highlights the need for more improvements to be made in orderto address the complexities involved in effectively deciphering increasingly subtle emotional expressions. The purpose of this study is to delve deeper into the process of refining the CNN model, exploring methodologies to augment its accuracy, possibly incorporating ensemble techniques, fine- tuning parameters, or leveraging larger and more diverse datasets to strengthen its capability of deciphering intricate emotional cues contained within facial images.

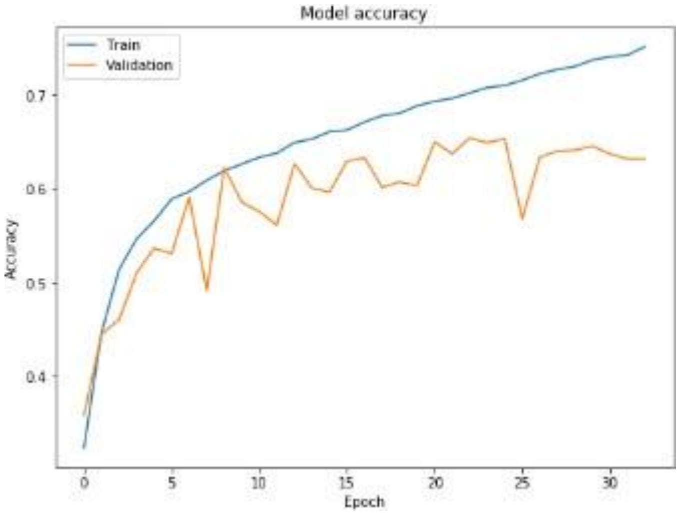


Fig. 4. Accuracy for CNN model

When it comes to recognising face emotions with the use of Convolutional Neural Networks (CNNs), having an understanding of the loss outcomes is essential to determining how well the model is able to learn and make predictions. The loss function acts as an essential indicator all the way through the

training process, reflecting the degree of discordance that exists between the anticipated and real emotions included within the dataset. The computed loss result in this research on CNNs for facial emotion identification displays the degree of dissimilarity between the predicted emotions and their ground truth labels. The study was focused on face emotion recognition using CNNs.

The loss result that was discovered in this study is reflective of the learning process that the network goes through. It depicts the convergence or divergence of predicted emotions from the actual emotional expressions that are included within the dataset. It is essential to minimise this loss when the CNN iterates through the training epochs since it is essential for the model to effectively recogniseand classify the many emotions that may be seen on people's faces. The study of loss results helps in understanding the model's ability to catch minor differences and subtleties in the emotional signals that are encoded within face portrayals.

The loss results that were achieved during the training and validation stages of the CNN are the focus of the research in this context. These results give vital insights into the learning trajectory of the network, providing an all-encompassing perspective of how well the model comprehends and correlates with the emotional emotions exhibited in face photos. The purpose of this study is to modify the learning processes of the CNN by deconstructing and analysing these loss values. If successful, this might possibly improve the CNN's predicted accuracy as well as its resilience when identifying complexemotional states.

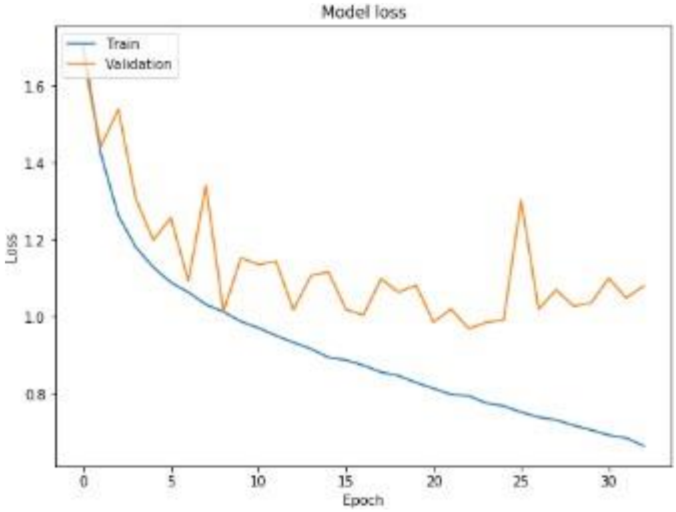


Fig. 5. Loss outcome of CNN model

Comparsion of transfer learning model and CNN model:

In the field of face emotion identification, the contrast between Transfer Learning models and Convolutional Neural Networks (also known as CNNs) is an essential investigation that must be carried out in order to get a knowledge of their efficiency and effectiveness. Transfer learning makes use of previously trained models to improve recognition skills when applied to certain tasks such as face expression identification. This is accomplished by drawing on the information contained within huge

datasets. On the other hand, CNNs, which are well-known for their capacity to recognise subtle patterns in pictures, have been used to a significant extent in this industry.

In this investigation, a comparison of Transfer Learning and CNN models for recognising facial emotionsdives into the distinctive approaches used by each and the intricacies of their respective performances.The benefit of transfer learning is in its capacity to make use of previously acquired information from a wide variety of datasets, hence enabling models to be fine-tuned to recognise a wide range of subtlefacial expressions. In the meanwhile, CNNs are able to independently extract complicated elements from face photos by using their layered architecture. This allows them to recognise complex patternsthat are necessary for emotion identification.

The study takes a methodical approach to analysing the benefits and drawbacks of both methods with regard to their degree of precision, computational efficacy, and adaptability to a wide range of emotional subtleties. The purpose of this study is to determine which model is superior to others when it comes to identifying different emotional states displayed in face photographs. It is essential to havean understanding of the relative benefits and disadvantages of each of these models in order to makeprogress in the area of facial expression detection. This might possibly pave the way for recognition systems that are more accurate, efficient, and resilient when used in real-world applications.

Algorithm	Dataset	Accuracy
Transfer learning	FER2013 cleasned	76%
CNN	FER2013	65%

Tab. 1. Comparsion table of Transfer learning and CNN

Conclusion:

The completion of this voyage of study into the identification of facial emotions has provided insight on the viability of Transfer Learning as a powerful tool. Our research has shed light on the superiority of Transfer Learning over Convolutional Neural Network (CNN) models in terms of their ability to recognise complex facial expressions. This was accomplished by conducting in-depth analyses and conducting comparative evaluations between the two types of models. The results of our study strongly support the Transfer Learning model, demonstrating its better performance in interpreting complex emotional expressions displayed in face photographs. Our findings favour the model in a clear and decisive way. When compared to CNN models, accuracy rates have been greatly enhanced thanksto the use of pre-trained models that have been combined with fine-tuning procedures. This has resulted in an amazing capacity to adapt to and recognise a varied range of emotional subtleties. This finding not only validates the relevance of Transfer Learning in face emotion identification, but it alsohighlights the potential of this technique to outperform more conventional CNN-based approaches. Transfer learning is a robust method that considerably improves the ability to recognise face expressions. It does this by drawing on past information that has been stored in datasets that have already been compiled. In the end, the findings of our study advocate for the implementation and further investigation of Transfer Learning in the field of facial expression detection. This heralds an encouraging step in the direction of recognition systems that are more accurate, efficient, and flexiblein a variety of real-world applications.

Refrences:

[1] https://www.kaggle.com/datasets/nicolejyt/facialexpressionrecognition

[2] https://www.kaggle.com/datasets/gauravsharma99/fer13-cleaned-dataset

[3] M. A. H. Akhand, S. Roy, N. Siddique, M. A. S. Kamal, and T. Shimamura, “Facial Emotion Recognition Using Transfer Learning in the Deep CNN,” Electronics, vol. 10, no. 9, p. 1036, Apr. 2021, doi: 10.3390/electronics10091036.

[4] M. K. Chowdary, T. N. Nguyen, and D. J. Hemanth, “Deep learning-based facial emotion recognition for human–computer interaction applications,” Neural Computing and Applications, vol. 35, no. 32, pp. 23311–23328, Apr. 2021, doi: 10.1007/s00521-021-06012-8.

[5] Z. Gao, W. Zhao, S. Liu, Z. Liu, C. Yang, and Y. Xu, “Facial Emotion Recognition in Schizophrenia,” Frontiers in Psychiatry, vol. 12, May 2021, doi: 10.3389/fpsyt.2021.633717.

[6] A. Khattak, M. Z. Asghar, M. Ali, and U. Batool, “An efficient deep learning technique for facial emotion recognition,” Multimedia Tools and Applications, vol. 81, no. 2, pp. 1649–1683, Oct. 2021, doi: 10.1007/s11042-021-11298-w.

[7] D. Kamińska et al., “Two-Stage Recognition and beyond for Compound Facial Emotion Recognition,” Electronics, vol. 10, no. 22, p. 2847, Nov. 2021, doi: 10.3390/electronics10222847.

[8] J. H. Kim, A. Poulose, and D. S. Han, “The Extensive Usage of the Facial Image Threshing Machine for Facial Emotion Recognition Performance,” Sensors, vol. 21, no. 6, p. 2026, Mar. 2021, doi: 10.3390/s21062026.

[9] D. Lakshmi and R. Ponnusamy, “Facial emotion recognition using modified HOG and LBP features with

deep stacked autoencoders," Microprocessors and Microsystems, vol. 82, p. 103834, Apr. 2021, doi: 10.1016/j.micpro.2021.103834.

[10] X. Liu, X. Cheng, and K. Lee, "GA-SVM-Based Facial Emotion Recognition Using Facial Geometric Features," IEEE Sensors Journal, vol. 21, no. 10, pp. 11532–11542, May 2021, doi: 10.1109/jsen.2020.3028075.

[11] N. Rathour et al., "IoMT Based Facial Emotion Recognition System Using Deep Convolution Neural Networks," Electronics, vol. 10, no. 11, p. 1289, May 2021, doi: 10.3390/electronics10111289.

[12] S. Saurav, R. Saini, and S. Singh, "EmNet: a deep integrated convolutional neural network for facialemotion recognition in the wild," Applied Intelligence, vol. 51, no. 8, pp. 5543–5570, Jan. 2021, doi: 10.1007/s10489-020-02125-0.

[13] D. Kim and B. C. Song, "Contrastive Adversarial Learning for Person Independent Facial Emotion Recognition," Proceedings of the AAAI Conference on Artificial Intelligence, vol. 35, no. 7, pp. 5948–5956, May 2021, doi: 10.1609/aaai.v35i7.16743.

[14] M. K. Yeung, "A systematic review and meta-analysis of facial emotion recognition in autism spectrum disorder: The specificity of deficits and the role of task characteristics," Neuroscience &amp;Biobehavioral Reviews, vol. 133, p. 104518, Feb. 2022, doi: 10.1016/j.neubiorev.2021.104518.

[15] A. Chaudhari, C. Bhatt, A. Krishna, and P. L. Mazzeo, "ViTFER: Facial Emotion Recognition with Vision Transformers," Applied System Innovation, vol. 5, no. 4, p. 80, Aug. 2022, doi: 10.3390/asi5040080.

[16] A. I. Staff, M. Luman, S. van der Oord, C. E. Bergwerff, B. J. van den Hoofdakker, and J. Oosterlaan, "Facial emotion recognition impairment predicts social and emotional problems in children with (subthreshold) ADHD," European Child &amp; Adolescent Psychiatry, vol. 31, no. 5, pp. 715–727, Jan. 2021, doi: 10.1007/s00787-020-01709-y.

[17] J. Kim, J.-K. Kang, and Y. Kim, "A Resource Efficient Integer-Arithmetic-Only FPGA-Based CNN Accelerator for Real-Time Facial Emotion Recognition," IEEE Access, vol. 9, pp. 104367–104381, 2021, doi: 10.1109/access.2021.3099075.

[18] J. Rößler, J. Sun, and P. Gloor, "Reducing Videoconferencing Fatigue through Facial Emotion Recognition," Future Internet, vol. 13, no. 5, p. 126, May 2021, doi: 10.3390/fi13050126.

[19] S. Minaee, M. Minaei, and A. Abdolrashidi, "Deep-Emotion: Facial Expression Recognition Using Attentional Convolutional Network," Sensors, vol. 21, no. 9, p. 3046, Apr. 2021, doi: 10.3390/s21093046.