iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** II **Month of publication:** February 2023

DOI: https://doi.org/10.22214/ijraset.2023.49221

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue II Feb 2023- Available at www.ijraset.com

A VGG16 Based Hybrid Deep Convolutional Neural Network Based Real-Time Video Frame Emotion Detection System for Affective Human Computer Interaction

Irfan Ahmad Ganie[1], Ankur Gupta[2], Dr. Ashish Oberoi[3]
[1]M. Tech Scholar, [2]Assistant Professor, [3]HOD and Professor, Department of Computer Sciences and Engineering, RIMT University, Gobindgarh Punjab India

Abstract: ***Facial emotion recognition is field of computer vision that deals with recognizing the emotions of the user by facial expression analysis. The development of facial emotion recognition technology has the potential to enhance human-computer interaction by making machines more responsive to human emotions. For recognizing facial emotions, a variety of machine learning methods have been used, such as deep neural networks, support vector machines, decision trees, and among others, Convolutional neural networks (CNNs), recurrent neural networks (RNNs), and their combinations have all been suggested as designs, which have proven especially effective in this field. However, despite substantial advancements in the area, there are still difficulties in attaining accurate and consistent facial expression identification in real-time everyday situations. The effectiveness and usability of face expression detection systems for diverse applications must therefore be improved and efficient systems be developed to expand the focus on further applications of emotions in achieving affective Human computer interaction. My main objective in this paper will be to focus on being able to accurately identify user's emotion in a real world scenario. I will be a designing a hybrid deep learning CNN architectures for the real world implementation of the real-time emotion recognition system for improving Human Computer Interaction***

Keywords: ***Real-time facial emotion Recognition, Affective Computing, Human Computer Interaction.***

I. INTRODUCTION

Emotion is crucial in communication since it is made up of feelings at a certain moment in time, voice intonation, and the transmission of information. Human communication is defined not only by what people say, but also by how they express it. Furthermore, as part of nonverbal communication, facial expressions account for around 55% of message perception, voice intonation accounts for approximately 38%, and actual words account for approximately 7% [14]. "Human beings are creatures born within a linguistic community," according to [2]. Language profoundly separates humans from the rest of life. Our human essence is therefore not a 'natural nature,' but rather a humanly produced nature formed by linguistic rules. It positions us profoundly and ultimately over a differentiating line from everything else that is not 'human'. Simply writing "go home" without any accompanying voice intonation or facial expression could be interpreted as an imperative statement of both extreme positive and negative polarity: for example, if a child is sick and the teacher sends him home; or if a child is misbehaving in class and the teacher sends him home. Humans communicate intentions via facial expressions and intonations, which may be connected to judgments and subjective perception. A micro expression is a facial expression that can easily observe and distinguish it as a communication method in social psychology [3]. The human emotion is one of the primary qualities of a certain emotional state. Emotions impact judgment by influencing how one feels about the item that influences judgment [4]. Facial expression is the result of a facial gesture or expression that shows the position of the muscles on the human face as a form of non-verbal communication as well as an essential way of expressing one's emotions as a form of feelings, intentions, goals, and opinions to others. The logical model of emotion theory describes how people form emotions, and its foundation is the complicated response that incorporates both mind and body [5]. The emotions and surroundings influence the purpose. Emotion determination may be used to develop interfaces that react to a man's various emotive states. Different affective states correspond to different expressions of motor functions; once that state is determined, machines can be made to respond to that state by analyzing the subject's expression, we can record the intensive state, and we can design the machines to respond to a specific state, thereby harmonizing the man machine interactions.

©IJRASET: All Rights are Reserved

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue II Feb 2023- Available at www.ijraset.com

Machines might assist humans in making accurate judgments by recognizing the correct purpose, particularly in illogical circumstances when decisions must be made quicker than a reasonable functioning mind. It is sometimes beneficial to intentionally modify intentional states in order to improve individual performance in stressful vocations and prevent mental diseases from occurring [6]. Recent studies have revealed that multi-modal emotion identification is possible even in real-time under specific conditions [7]. As a result, it is possible to identify the affective state of a specific user in real time.

II. AFFECTIVE COMPUTING AND HUMAN COMPUTER INTERACTION

Affective computing [1] can be broadly defined as a framework by which machine tries to determine the likely affective state of the user by analyzing the expressions of emotions of the user. It aims to create computers with understanding capacities much beyond those of current computer systems. Computing that is related to, derives from, or purposefully manipulates emotion is known as affective computing, giving computers emotional intelligence skills, such as the capacity to perceive and respond intelligently to emotion, the capacity to express emotion (or not express emotion) correctly, and the capacity to manage emotions, is another aspect of affective computing. The latter skill entails managing one's own emotions as well as the emotions of others. The use of technology in social interactions is crucial now more than ever. The majority of people who use computers are not engineers, and they do not have the time to learn and keep up with the specialized skills needed to benefit from a computer's aid. The goal of granting computers emotional capabilities is to assist solve the issue of engaging with complicated systems, resulting in more natural interaction between the two. The conceptual framework for the affective computing can be represented by Fig 1.

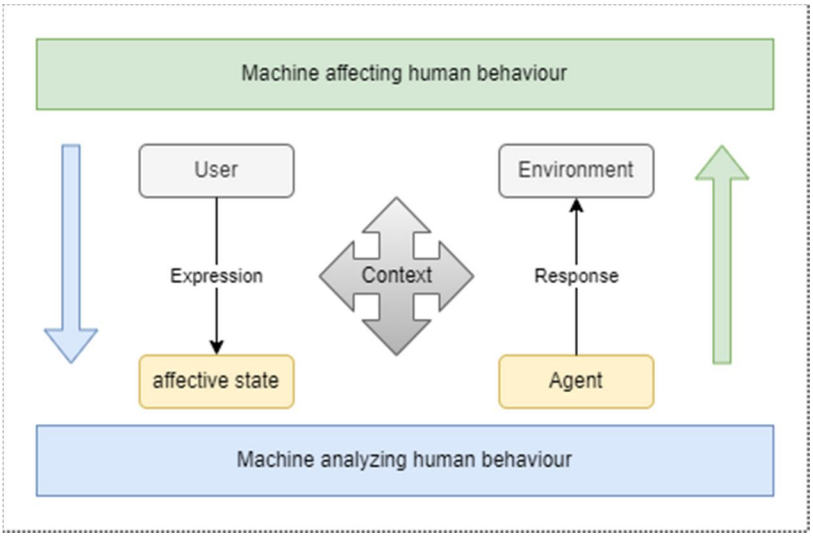


Fig 1 Conceptual framework for affective computing

Human emotions are made up of a network of biasing and regulating processes that operate throughout the body and brain and influence almost everything a person does. The way you move, speak, type, gesture, construct a sentence, or otherwise communicate can all be influenced by your emotions? As a result, there are many cues you can perceive and attempt to correlate with an underlying affective state in order to infer someone's emotion. One can search for various patterns of emotion's influence depending on the available sensors (auditory, visual, textual, physiological, biochemical, etc.). Automated facial expressions determination, vocal inflection recognition, and interpretation about emotion when provided text input regarding goals and activities have been the active areas for machine emotion recognition [9].Thereafter, the signals are analyzed using methods for pattern recognition like hidden Markov models (HMM's), hidden decision trees, auto-regressive HMM's, Support Vector Machines (SVMs) and neural networks. Affective human-computer interaction (AHCI) is a branch that focuses on interactions study of systems that can recognize, interpret, and respond to human emotions in a more effective and personalized manner. AHCI seeks to develop more emotionally intelligent systems capable of understanding and responding to users' emotions in a natural and human-like manner. This is an interdisciplinary field that draws on psychology, computer science, and engineering.

AHCI is accomplished through a variety of techniques for detecting human emotions. Facial expression recognition, is one of the methods that employs computer vision to detect changes in facial expressions that can indicate various emotions such as happiness, sadness, anger.

©IJRASET: All Rights are Reserved

Fear, Disgust,Neutral. Another technique that can detect changes in tone, pitch, and other vocal cues that can indicate different emotions is speech analysis.[10] Changes in a person's emotional state can also be detected using physiological signals such as heart rate and skin conductance. That type of a system can be referred to as a multi-modal system.

III. DETECTING HUMAN EMOTIONS

Human emotion detection is an important part of human-computer interaction. Emotions influence how people connect with one another and with technology. Understanding and recognizing emotions may lead to more individualized and productive interactions in customer service, education, mental health, and marketing, among other fields.

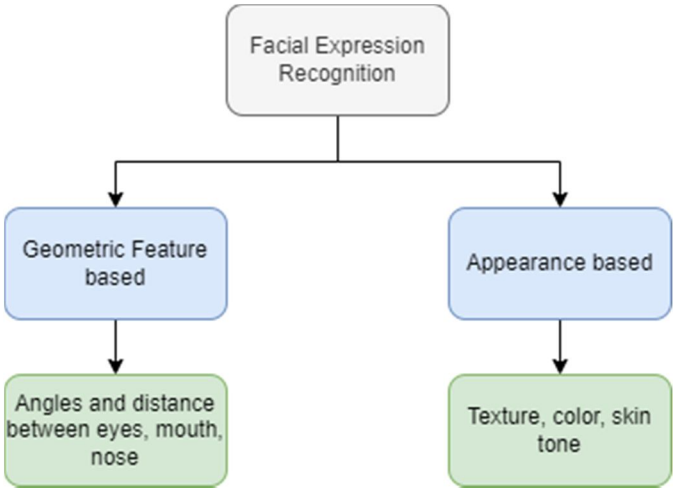
Facial expression recognition is one of the most used ways for identifying human emotions. Facial expression recognition uses computer vision to identify changes in facial expressions that might indicate various emotions such as pleasure, sorrow, or rage [11]. It entails examining the motions of numerous facial muscles, such as the lips, brows, and eyes, to detect an individual's emotional state. With the advancement of artificial intelligence and machine learning, facial expression detection has become increasingly accurate and dependable over time. This technology has gotten a lot of attention in recent years because of its potential to transform areas including healthcare, security, and entertainment. We will go through facial expression recognition in depth in this paper, covering how it works, its uses, and its possible influence on society.

Facial expression recognition is mainly achieved by two methods –

1) Geometric feature based method

2) Appearance based approach

The types of facial expression recognition can be explained using Fig 2.



The geometric feature-based method and the appearance-based approach are the two basic approaches to facial emotion recognition [13]. The geometric feature-based technique entails detecting certain spots on the face, such as the corners of the eyes, nose, and mouth, then calculating the distances and angles between these points to infer the person's emotions. The appearance-based technique, on the other hand, entails examining the texture and color of a person's face to identify their feelings.

IV. AI AND ML IN FACIAL EXPRESSION RECOGNITION

Machine learning (ML) and artificial intelligence (AI) are crucial elements of face expression identification technologies. Models that can identify various emotions and facial expressions may be trained using ML techniques. Because these algorithms are built to learn from data and become better over time, they can recognize facial expressions more precisely and consistently. In order to recognize facial expressions, AI and ML are employed in a variety of methods. For instance using supervised learning, where the system is trained on a large dataset of labelled pictures that correlate to various facial expressions, is one method. This enables the system to discover the patterns and traits connected to each emotion. Once taught, the system can identify facial expressions in fresh pictures and videos and categorize them.

Deep learning algorithms such as convolutional neural networks (CNNs) have been successfully used to the identification of facial expressions. CNNs can find patterns and correlations in massive datasets and are meant to automatically learn features from photos. They are thus excellent at detecting tiny changes in emotions and facial expressions.

Convolutional Neural Network (CNN) is an algorithm included in the deep neural network or Deep Learning family due to the high network depth and is significantly superior when implemented on [12] image data.

The capacity of CNNs to acquire characteristics at many levels of abstraction is one of its primary advantages when it comes to face emotion identification [12]. The network can pick up on low-level details like edges, lines, and corners as well as higher-level details like the location and orientation of face landmarks when it comes to recognizing facial expressions. This enables the network to recognize emotions and facial expressions based on more intricate elements that are difficult to recognize using conventional machine learning methods. The capacity of CNNs to carry out feature extraction and classification in one step is another benefit. End-to-end learning does away with the necessity for manual feature extraction, which may be labor-intensive and prone to mistakes. CNNs may increase the accuracy of facial expression detection while lowering the system's complexity by concurrently learning the features and classification rules.

V. RELATED WORKS

In this section I will cover the background of various studies and publications made in the field of facial emotion determination, ffective computing and human computer interaction. Mehrabian (1981) [14] in his book entitled "Silent Messages: Implicit Communication of Emotions and Attitudes.", captures the 7%-38%-55% rule of non-verbal communication, asserts that body language accounts for 55% of the message in human communication, with words communicating the remaining 7%. This is especially important when we take into account the predominantly non-verbal nature of how emotional content or affect is expressed. A good affect detecting system essentially requires the employment of subtle non-verbal communication techniques to function. Oatley and Duncan (1992) [15], in their research entitled "Incidents of emotion in daily life" and his book entitled "Best Laid Schemes: The Psychology of the Emotions" highlighted that emotion and motivation are related in that how one feels or perceives a situation will heavily influence the course of action that is performed in response. I. A. Ahmad and A. R. Abu-Bakar (2017) [16] "Facial Expression Recognition: A Brief Review of the Literature". This paper provides a comprehensive review of the literature on facial expression recognition. It covers different techniques and approaches that have been used for facial expression recognition, including machine learning, deep learning, and hybrid approaches. It also discusses the challenges and limitations of facial expression recognition and potential future research directions. L. A. Jeni, J. F. Cohn, and T. Kanade (2013) [17] "Facial Expression Recognition: A Survey and New Techniques" This survey paper provides a comprehensive review of the literature on facial expression recognition. It covers different approaches, including appearance-based methods, geometric-based methods, and hybrid methods. The paper also discusses the challenges and open research questions in the field. S. Mollahosseini, B. Hasani, and M. H. Mahoor (2019) [18] "Facial Expression Recognition Using Deep Learning: A Survey" This survey paper provides an overview of the recent advances in facial expression recognition using deep learning. It covers different deep learning architectures that have been used for facial expression recognition, such as convolutional neural networks (CNNs), recurrent neural networks (RNNs), and attention-based models. L. Biswas, S. Saha, and S. K. Das (2020) [19]", *Recent Advances in Facial Expression Recognition: A Brief Review*", this paper provides a brief review of recent advances in facial expression recognition. It covers different approaches, including machine learning, deep learning, and hybrid approaches, as well as the different datasets and evaluation metrics that have been used in the field. The paper also discusses the challenges and future research directions in facial expression recognition. Martin et al (2020) [20], in their research titled "Towards emotion recognition from contextual information using machine learning" highlighted that the cognitive processes that underpin human behaviour are influenced by emotions. Positive emotions may enhance creativity and encourage cooperative conduct, but negative emotions might result in the development of psychiatric problems. Hao-Chiang Koong Lin (2022) [21], in his book entitled "May Computers have Emotions? Establishing an AI Device based on Affective Computing Technologies", in their book they outline the process for creating an emotion system that can be installed on cell phones to allow for emotional expression. The impact of emotional systems on user behaviour is also examined in this research, along with users' views and feelings about emotional information given by computers. The goal of this research is to investigate the elements that designers of emotional machines need to emphasize.

VI. A HYBRID VGG16 DEEP CNN BASED EMOTION RECOGNITION SYSTEM

The basic aim of building a hybrid VGG16 deep CNN based video frame emotion detection system that can effectively detect emotions in real-time and challenging scenarios.

The proposed system works on the analysing the images from FER-2013 dataset, using the VGG16 as the base classification model and adding the dropout and sequential dense model layers over the base model to improve the efficiency of the system to be able to handle the real world scenario.

The resulting model works on extracting the features from the dataset and use the features for model learning process. Once the learning process is completed and the model achieves significant accuracy. The model is saved and deployed to be used in the predictions. The input is then taken from the user's camera or a video module and the prediction is made by the model based on the learning process. The model is implemented in keras, which is a library in python that is used in the deep learning process.
The general architecture of the real time deep learning facial expression recognition proposed system can be illustrated using Fig 3.

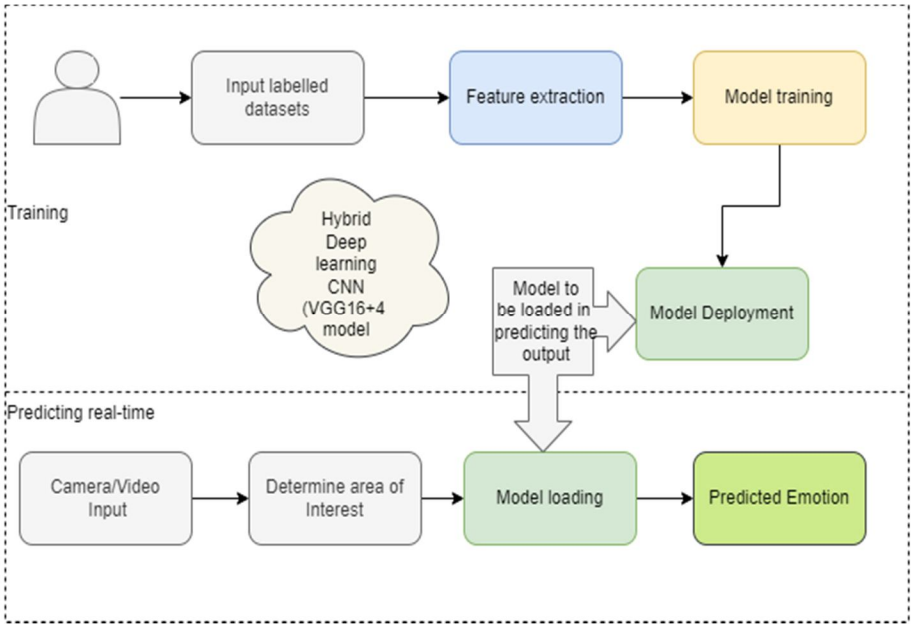


Fig 3 General Architecture of deep learning facial emotion detection system

To be able to identify the emotion more precisely we will adopt a mixed-methods approach, combining VGG16 Convolutional Neural Network classification model and a layer for optimal performance handling the real world scenarios. The following steps were followed in designing the proposed system.

A. Pre-Processing The Dataset

The dataset for this paper is the FER dataset downloaded from Kaggle. FER 2013 is a publicly available dataset for facial expression recognition (FER) research. The dataset was created by Pierre-Luc Carrier and Aaron Courville and released in 2013. The dataset consists of 35,887 grayscale images, each of the size of 48x48 pixels labelled into 7 classes of emotions, including anger, disgust, fear, happiness, sadness, surprise, and neutral. The dataset is annotated with categorical labels, where each image is labeled with the emotion it represents. The dataset is split into training and validation sets, with 28,709 images in the training set, 7178 images in the validation set. Fig 4 shows the sample images available in the dataset



Fig 4 Sample images of FER-2013 dataset

B. Designing the Model

After the input paths and directories are set to generate the image data from directory, we design the model. As discussed, we will be using the VGG16 as a base model and we will add the dropout and a 4 layered sequential model with the base to handle the inputs specific to real world scenarios.

C. Compiling And Training The Model

After the model is defined, the model is stored and compiled. The pattern of losses and accuracy in training and validation is kept for epochs. If the training and validation parameters, such as training loss, training accuracy, validation loss, and validation accuracy, do not improve as the number of epochs grows, the training process is terminated and the model is preserved. After the model compilation is done, the model is trained against the training dataset for a fixed number of epochs. The training process is iterative and the learning process is improved with each epoch**.**

D. Testing The Model

The performance if the model is tested using the validation dataset. The predicted labels are compared against those available in the validation dataset and the model testing is done against the predicted labels versus the actual labels in validation data.

E. Deploying The Model

The model can be deployed in the production environments, in order to classify the facial expressions of the user. The camera module can be hosted on any compatible environment and the model request can be made by a remote call.

F. Hyper Parameter Tuning

Experimented and tried with different hyper-parameters, such as learning rate, number of layers, and batch size, to optimize the performance of the model. Also used various callback techniques to optimize the training process and the performance of the system.

G. Evaluation Of The System

Evaluated the performance of the system based on various metrics such as accuracy, loss, precision, recall and the confusion matrix was plotted on the new images by comparing its predictions to the ground truth labels.

VII. RESULTS

The system achieved a testing and validation accuracy of 85.72% during training for 10 epochs which is about 25% increase over the conventional CNN accuracy. The system also demonstrated a minimal training loss of about 1.75 and a validation loss of about 1.81 which is considerably lower than the conventional CNN models. The training accuracy and the validation accuracy are expected to further increase with the use of a more versatile and colored datasets and a sufficiently large training system, and increase in the number of epochs.

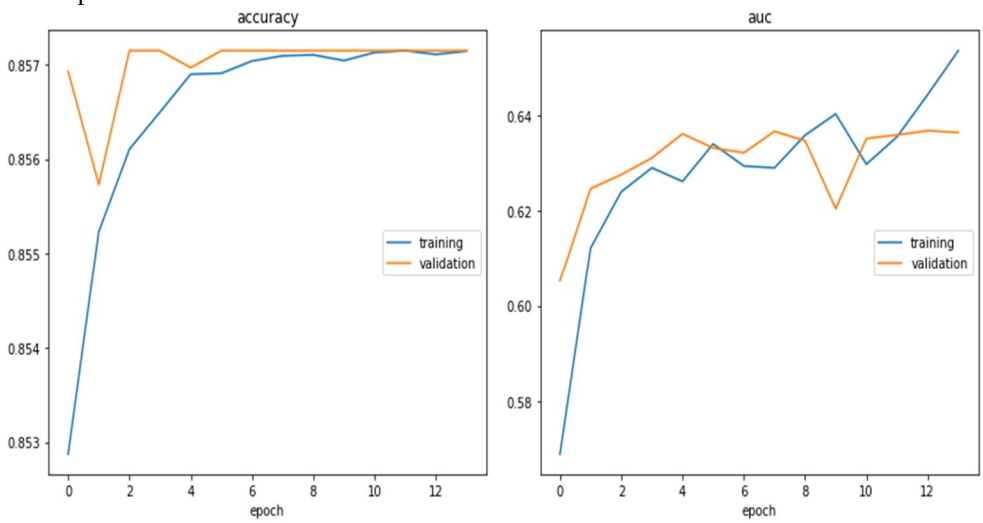


Fig 5 Accuracy and auc after 10 epochs

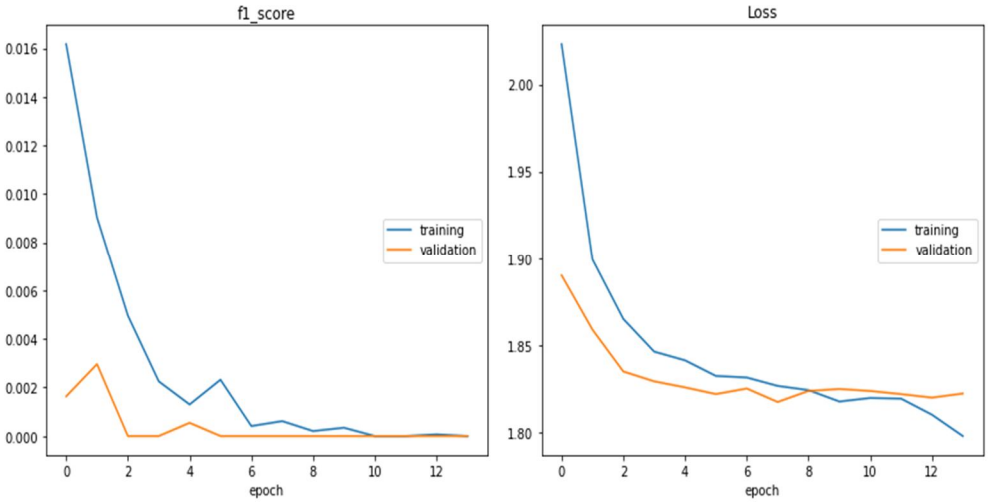


Fig 6 F1_score and loss after 10 epochs

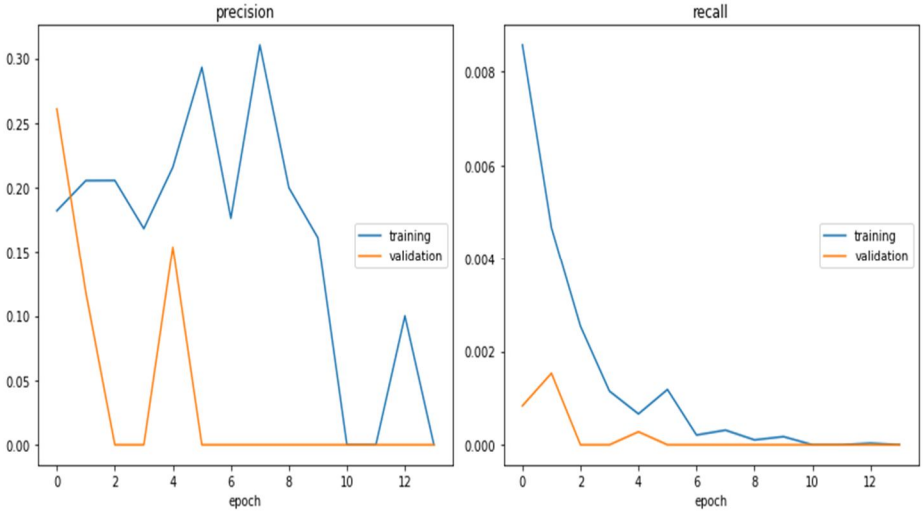


Fig 7 Precision and Recall after 10 epochs

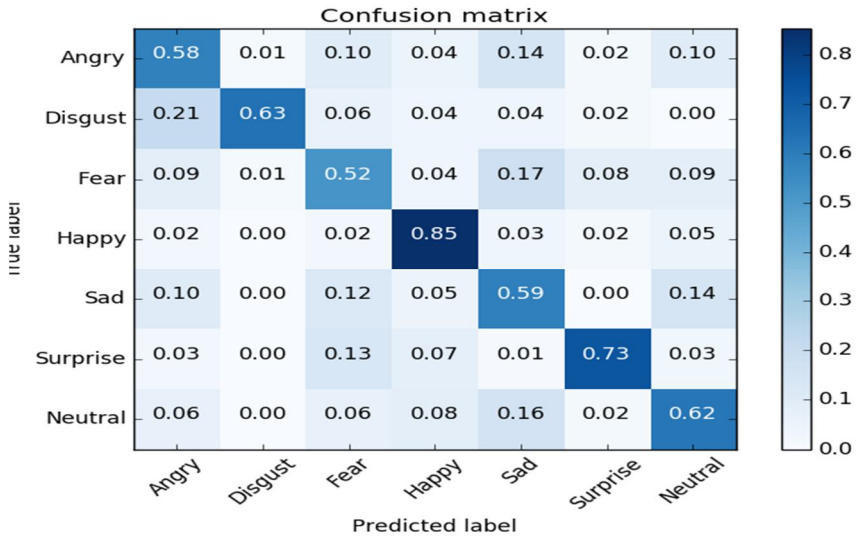


Fig 8 Confusion matrix of the proposed system

TABLE 1 RESULT COMPARISON BETWEEN CONVENTIONAL CNN AND PROPOSED SYSTEM

Regular	Conventional CNN	Hybrid CNN
Accuracy after 2 epochs	46.26%	84.27%
Loss after 2 epochs	1.5%	1.39%
Accuracy after 9 epochs	62.28%	85.71%
Loss after 9 epochs	1.02%	0.81%

VIII. CONCLUSIONS

In conclusion, a new world of opportunities for emotion recognition in computer vision has emerged with the creation of facial emotion detection models. These models are able to precisely identify and categorize human emotions based on facial expressions by utilizing machine learning algorithms and deep neural networks. These models have numerous uses, including in marketing, psychology, and human-computer interface. There are still issues that must be resolved, such as the requirement for extensive and varied datasets and the possibility of bias in the training data. Despite these difficulties, face emotion detection models have made substantial progress, which is very encouraging for the development of emotion identification technology in the future.

Real-time video frame emotion recognition has a wide range of potential applications in the future, as well as several avenues for more study and improvement.

IX. FUTURE WORK

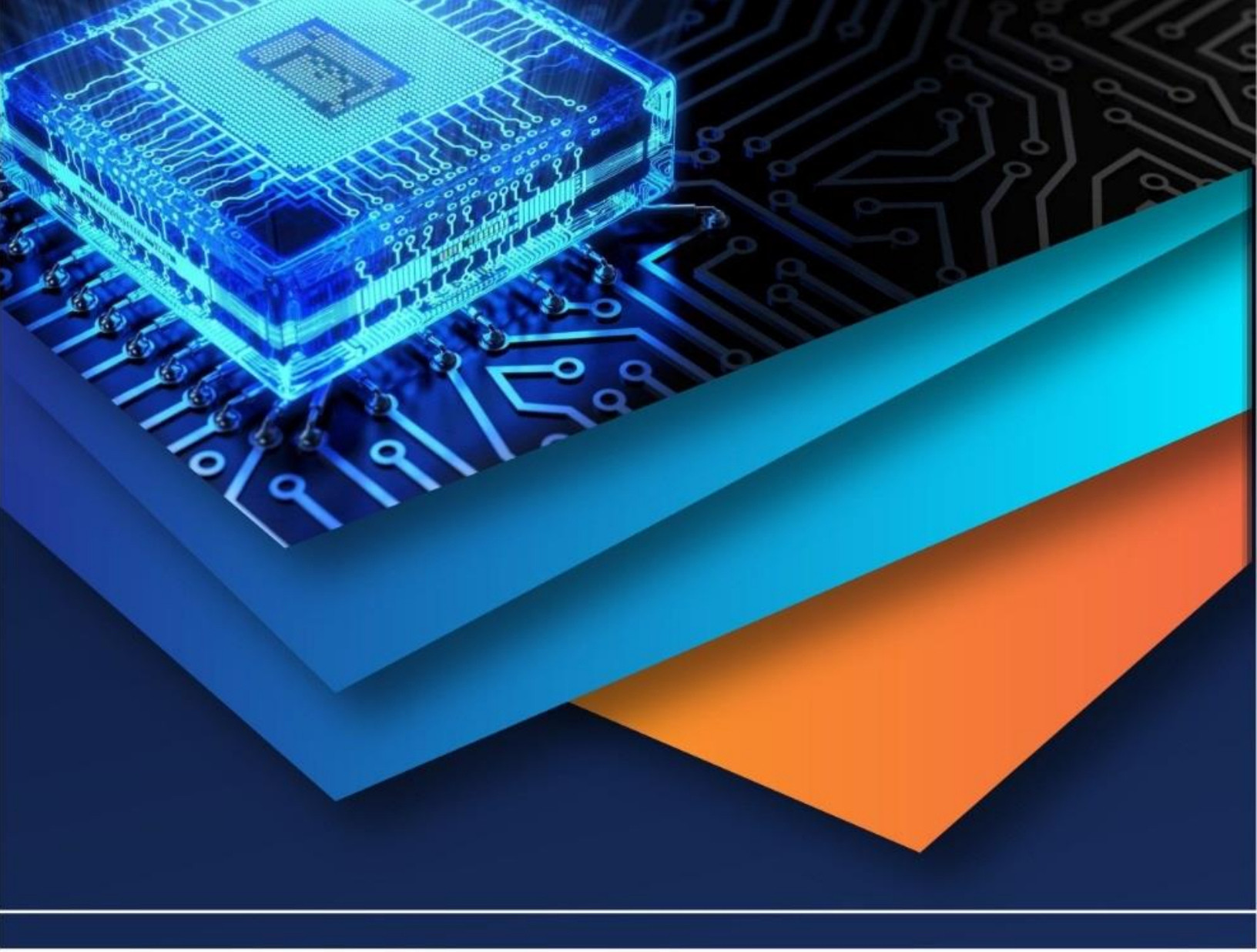
Future research may focus on a variety of proposed system related topics, such as:

1) Cross-modal emotion detection, which involves integrating data from many modalities, such as audio, text, and physiological signals, in order to increase the system's resilience and accuracy.
2) *Emotion Identification In The Wild:* Adapting the system to real-world variability, including a range of linguistic, cultural, and demographic differences.
3) *Real-Time Emotion-Aware Affective Systems:* Developing and testing a system that can recognize emotions in real-time for usage in real-world contexts including tailored recommendations, computer-human interaction, and mental health monitoring.

REFERENCES

[1] R.W. Picard, “Affective Computing," MIT Press, Cambridge, LDV Forum - GLDV Journal for Computational Linguistics and Language Technology, 1997.
[2] http://www.ldvforum.org/2007_Heft1/Bayan_Abu-Shawar_and_Eric_Atwell.pdf
[3] A. Lambert, F. Eadeh, S. Peak, L. Scherer, J. P. Schott, and J. Slochower, “Towards a greater understanding of the emotional dynamics of the mortality salience manipulation: Revisiting the “affect free” claim of terror management theory (in press).” Journal of Personality and Social Psychology, 2014.
[4] Gerald L. Clore, Jeffrey R. Huntsinger, "How emotions inform judgment and regulate thought." Science Direct September 2007, https://doi.org/10.1016/j.tics.2007.08.005
[5] Zhongzhi Shi, in Intelligence Science, "Emotion intelligence." Science Direct 2021.
[6] University of Montreal, “AI can make better clinical decisions than humans: Study”. Science Daily. (2021, September 10) from www.sciencedaily.com/releases/2021/09/210910121712.htm
[7] Kukolja, D., Popović, S., Horvat, M., Kovač, B., Ćosić, K.: “Comparative analysis of emotion estimation methods based on physiological measurements for real-time applications”. Int. J. Hum.-Comput. Stud. 72(10), 717–727 (2014).
[8] R.W. Picard, "Affective Computing," MIT Press, Cambridge, 1997, LDV Forum - GLDV Journal for Computational Linguistics and Language Technology, 1997
[9] Azcarate, Aitor & Hageloh, Felix & Sande, Koen & Valenti, Roberto. Automatic facial emotion recognition, 2005.
[10] Ling Cen, Fei Wu, Zhu Liang Yu, Fengye Hu, "Chapter 2 - A Real-Time Speech Emotion Recognition System and its Application in Online Learning, In Emotions and Technology, Emotions, Technology, Design, and Learning”, Academic Press, 2016, Pages 27-46, ISBN 9780128018569, https://doi.org/10.1016/B978-0-12-8018569.00002-5.
[11] Malinowska, J.K. “What Does It Mean to Empathise with a Robot?” Minds & Machines 31, 361–376 (2021). https://doi.org/10.1007/s11023-021-09558-7
[12] Hassaballah, Mahmoud & Aly, Saleh. (2015). Face Recognition: Challenges, Achievements, and Future Directions. IET Computer Vision. 9. 614-626. 10.1049/iet-cvi.2014.0084.

[13] J. Luo, Z. Xie, F. Zhu and X. Zhu, "Facial Expression Recognition using Machine Learning models in FER2013," 2021 IEEE 3rd International Conference on Frontiers Technology of Information and Computer (ICFTIC), Greenville, SC, USA, 2021, pp. 231-235, Doi: 10.1109/ICFTIC54370.2021.96473
[14] Mehrabian (1981) [27], "Silent Messages: Implicit Communication of Emotions and Attitudes.",University of Virginia, Wadsworth Publishing Company, 1981. ISBN 0534009107, 9780534009106, https://books.google.co.in/books?id=WJgoAAAAYAAJ
[15] Oatley, K., & Duncan, E.. "Incidents of emotion in daily life. In K. T. Strongman (Ed.), International review of studies on emotion", Vol. 2, pp. 249–293). John Wiley & Sons., 1992
[16] Ahmad, S.Z., Abu Bakar, A.R. and Ahmad, N. (2019), "Social media adoption and its impact on firm performance: the case of the UAE", International Journal of Entrepreneurial Behavior & Research, Vol. 25 No. 1, pp. 84-111. https://doi.org/10.1108/IJEBR-08-2017-0299
[17] Jeni, Laszlo A., Jeffrey F. Cohn and Fernando De la Torre. "Facing Imbalanced Data--Recommendations for the Use of Performance Metrics." 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction (2013): 245-251.
[18] Mollahosseini, Ali, Behzad Hasani and Mohammad H. Mahoor. "AffectNet: A Database for Facial Expression, Valence, and Arousal Computing in the Wild." IEEE Transactions on Affective Computing 10 (2017): 18-31.
[19] Biswas, S. Saha, and S. K. Das, "Recent Advances in Facial Expression Recognition: A Brief Review", 2020.
[20] Salido Ortega, M., Rodríguez, LF. & Gutierrez-Garcia, J.O. "Towards emotion recognition from contextual information using machine learning", J Ambient Intell Human Comput 11, 3187–3207 (2020). https://doi.org/10.1007/s12652-019-01485-x
[21] Lin, Hao-Chiang. (2022). "May Computers have Emotions?" Establishing an AI Device based on Affective Computing Technologies. 10.37247/PAELEC.1.22.13.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)