Journal of Engineering Technology and Applied Physics

Mind Care Solution Through Human Facial Expression

Ali Asghar[1], Muhammad Awais[1,*], Syed Hassan Raza Naqvi[1], Muhammad Umar Mehboob[1], Reqad Ali[1] and Jawaid Iqbal[2]
[1]*Software Engineering Department, Capital University of Science and Technology, Islamabad, Pakistan*.
[2]*Cyber Security Department, Riphah International University, Islamabad, Pakistan.*
[*]*Corresponding author*: mawaiskhan1808@gmail.com, *ORCiD*: 0009-0006-3101-524X
https://doi.org/10.33093/jetap.2024.6.2.2
Manuscript Received: 17 January 2024, Accepted: 16 February 2024, Published: 15 September 2024

***Abstract* — Using proposed system psychologists can use technology to make decisions which can provide ease for both patients and psychologists. Psychologists can check the progress of patients by analysing emotions reports of patient over time. Using historical data and emotion detection technology psychologists can make more accurate decisions. Using proposed system patient and psychologists don't have to go to anywhere they only need a device and internet. Based on the characteristics of patient emotion psychologist only need report generated by system and prescribe medicine in emergency situation. Proposed system improves consultancy method by using machine learning emotion detection algorithm. Proposed system detects facial emotion of patient by using CNN with HAAR cascade classifier. We use FER 2013 dataset to train our model. We use VGG 19 architecture to train our model for optimization function to enhance the accuracy of model. We use RELU. We use DJANGO framework for integration with frontend. Result of our model on dataset 82.3% after find tuning the accuracy goes to 82.3% to 92%. We use recall and F1 method to check the performance of model. We trained model on the testing dataset which have gray scale images and 48*48pixel images to achieve his performance. To achieve our accuracy goal, we split dataset into trainee validation and testing dataset. We use CNN and achieve 93% accuracy in our system which help patient to get feedback only selected question and psychologist. Patients select psychologist to answer questions of psychologist system stores emotions of patient against every question to generate emotion report. Psychologist can analyze emotion report to provide better prescription to patient.**

Keywords—Emotion detection, CNN, Feature extraction, Psychologist, Patient.

I. INTRODUCTION

Despite verbal communication, human emotions being non-verbal communication plays a very important role in our lives. Facial expressions are very crucial to detect one's emotional state. Automatic facial expressions detection in machine learning have gained a lot of popularity. It has various applications in our daily life such as in security systems, lie detection and in healthcare systems. In Machine learning we train a model on labelled dataset. A model is a simple algorithm or mathematical structure. These models help computers to make predictions like human emotion recognition.

Based on the kind of characteristics that are employed, either appearance-based features or geometry- based features, emotion identification systems may be split into two groups. Appearance characteristics, such as wrinkles and furrows, define the texture of the face brought on by expression. The contour of the face and its features, such as the lips or eyebrows, are described by geometrical characteristics [1].

The desire to enhance every element of the interface between people and computers has grown in recent years. It is asserted that for human-computer intelligent interaction to be fully successful, the computer must be able to communicate with the user intuitively, much like human-human communication does. Humans communicate with one another mostly via words, but they may also express their emotions and emphasize key points of their speech through body language.

Visual, verbal, and other physiological ways are used to express emotions. Humans frequently communicate their emotions through their facial expressions. Although there isn't a clear, widely accepted definition of what constitutes an emotion, it is apparent that emotions play a crucial role in human life.

When welcoming, one smiles; when perplexed, one frowns; and when furious, one raises his or her voice. The interaction is substantially enhanced by our understanding of emotions and our ability to respond

Published by MMU PRESS. URL: https://journals.mmupress.com/index.php/jetap/index

to others' facial expressions. A growing body of research is demonstrating that emotional intelligence includes emotional talents [2].

In recent years, psychologists are using technology driven approaches to better serve their patients. Technology helps psychologists to understand the mental health issues of patients. One such technology usage is the use of real time emotion detection systems which can monitor and analyze real-time emotions of psychological patients. With accurate and detailed information about patient's emotional states such systems can enhance the quality of feedback and improve patient outcomes. Psychologists' expertise lies in their abilities to analyze patient's emotional states by diagnosing patient mental health conditions.

Traditionally, psychologists were relying on self-reported emotions and therapy sessions to understanding patient behavior and emotions. But such reports were no reliable because they were not real-time and influenced by patient's current state of mind. Sometime patients try to conceal their emotions so psychologists can't predict true emotion state of patients in such situations. So psychologist need a real time emotion detection system which can detect emotions of patient and provide a report of emotions. Our system aims to enhance the traditional assessment methods by utilizing advance emotion detection technology.

The majority of methods now in use center on categorizing seven fundamental emotions—neutral, pleased, startled, fearful, angry, sad, and disgusted—that have been discovered to be shared by all cultures and subgroups. Following the Facial Action Coding System (FACS), further in-depth methods try to either categorize which Action Units (AU) are engaged or gauge their intensity. The dimensional technique, which treats facial expressions as regression in the Arousal-Valence space, is used in fewer research [3].

In this paper we used CNN to detect human expressions. In the field of computer vision and machine learning, convolutional neural network (CNN) is a very used algorithm. The CNN has three layers. The convolutional layer extracts feature from image, the pooling layers do normalization and finally, the fully connected layer makes prediction. We used CK+ dataset to train our model.

We introduced the Haar cascade classifier which detects face from every frame of video. Haar cascade is a pre-trained classifier to extract features from human face. The classifier finds the patterns in images by scanning them. The extracted features are then used by model to detect emotion. Mainly our system predicts 8 emotions including happy, sad, fear, disgust, contempt, anxiety, anger and surprise. Our system provides a real-time emotion detection for patients by analyzing facial expressions. The real-time information provides accurate and unbiased report of patient emotions which helps psychologists to make more accurate and informed decisions.

II. LITERATURE REVIEW

System identified late-fusion hybrid network that blends 3D convolutional networks (C3D) with recurrent neural networks (RNN). RNN and C3D encode motion and appearance information in many forms. In particular, RNN models both the appearance and the motion of video concurrently, whereas C3D uses appearance information derived by convolutional neural networks (CNN) across individual video frames as input. Without adding any extra emotion-labeled video clips in the training set, our system's identification accuracy was 59.02% when combined with an audio module, compared to 53.8% for the EmotiW winner [4].

It is suggested to use CNNs to analyze pre-processed pictures in order to get rid of distracting elements and concentrate the network's attention on variances resulting from class labels. To do this, transform pictures into LBP codes, which are resilient to variations in light. These binary codes must be uniquely mapped to metric space using an estimate of the EMD in order to be subject to CNNs. Finally, a weighted average of their predictions is used to merge several CNN architectures and representations into an ensemble [5].

A methodical strategy to building many specialized models using deep learning methods, each of which focuses on a single modality. Among them are a relational auto encoder, which addresses the spatio-temporal aspects of videos, a deep belief net focused on representing the audio stream, a KMeans-based "bag-of-mouths" model that extracts visual features around the mouth region, and a convolutional neural network that focuses on capturing visual information in detected faces. Investigate several techniques for combining cues from these modalities into a single common classifier. This delivers predictions with a significantly higher accuracy than the strongest single-modality classifier [6].

Introduce a rule-based approach for reliable face detection and facial expression recognition using a convolutional neural network. The outcome demonstrates accurate grin identification with a recognition rate of 97.6% for 5600 still photos. The suggested system successfully distinguished between talking and smiling using the saliency score derived from voting visual inputs [7].

Comprehensively exploit the previous knowledge that the local facial Action Units (AUs) may be used to breakdown the appearance changes produced by facial expression into a deep architecture called AU-aware Deep Networks (AUDN) for facial expression identification. consists of two layers—a convolution layer and a max-pooling layer—that are intended to create an over-complete representation of all expression-specific appearance variations across all potential locations. An AU-aware receptive field layer is created to search subsets of the over-complete representation, each of which is intended to simulate the combination of AUs as accurately as possible. The final expression recognition process uses concatenated hierarchical features that are learned using restricted Boltzmann machines [8].

The potential benefits of the family of local binary pattern descriptors for FACS Action Unit detection should be examined. For static AU analysis, we contrast local binary pattern and local phase quantization. We expand the purely spatial representation LPQ to a dynamic texture descriptor we call Local Phase Quantization from Three Orthogonal Planes (LPQ-TOP) to record the dynamics of face expressions. We contrast this with the Local Binary Patterns from Three Orthogonal Planes (LBP-TOP). A completely automatic AU identification method assesses the effectiveness of these descriptors using posed and spontaneous expression data gathered from the MMI and SEMAINE databases. Results indicate that LPQ-based systems outperform LBP-based systems in terms of accuracy rate [9].

Using the face recognition SDK and the Microsoft Kinect for Windows sensor V2 to identify eight expressions with the goal of recognizing emotions from facial expressions. Using Visual Studio and Mat lab, we implemented our application for emotion identification [10].

Introduces the emotion recognition system for those who need it most. The incoming voice signal is first read and put via the preprocessing to boost the signal even further. Frequency-based characteristics including spectral flux, spectral centroid, spectral crest, and spectral roll-off are used to extract the features. Deep belief networks (DBN), which are trained using the standard gradient descent method (SGD) and the moth search optimization algorithm (MSA), are then used to classify the emotions. Evaluation metrics including the False Rejection Rate (FRR), Accuracy, and False Acceptance Rate (FAR) are used to assess how well the suggested technique of emotion identification performs. Comparing the performance of the suggested approach with that of the current methods reveals the efficacy of the proposed method [11].

The proposed approach makes full use of MTCNNs, particularly those that have been augmented with the dominant emotion, emotion score function, and top emotion features. It has been demonstrated that preprocessing the training and test datasets increases feature representation efficiency and accuracy detection [12].

FS-CNN is a suggested technique that is used to identify faces in large-scale pictures. Next, facial landmarks are analyzed to forecast expressions for emotion identification. Convolutional neural networks and patch cropping are the two phases that make up the FS-CNN. In the first step, faces are identified in high-resolution photos, and the face is then cropped for further processing. In the second step, a convolutional neural network was utilized to analyses scale invariance on pyramid photos and forecast face emotion based on landmark analytics. The UMD Faces dataset was used to train and test the suggested FS-CNN. Approximately 95% of the time, the mean average accuracy allowed for high performance [13].

Gives a broad review of face expression analysis that is automated in RGB, 3D, thermal, and multimodal dimensions. Describe and categorize the state-of-the-art techniques in accordance with the new taxonomy that will be defined for the topic, covering all stages from face detection to facial emotion identification. display the benchmarking of the most effective techniques along with the significant datasets [14].

Face recognition has made considerable use of CNNs. However, face expression recognition on CNNs has not been properly tested. develop and evaluate a CNN model for recognizing facial expressions. Other pre-trained deep CNN models are evaluated using the CNN model's performance as a baseline. We assess the effectiveness of pre-trained object recognition models Inception and VGG and compare them with pre-trained face recognition model VGG-Face. Every experiment is run on the CK+, JAFFE, and FACES face databases, which are accessible to the general public [15].

Introduced a cutting-edge framework for high-accuracy automated facial expression identification. In particular, because it contains precise and thorough information about emotions, a high-dimensional feature made up of the combination of face geometry and appearance aspects is introduced to facial expression identification. Furthermore, by extracting robust and discriminative features from the data, the deep sparse auto encoders (DSAE) are trained to recognize the facial expressions with high accuracy.

The findings of the experiment show that the proposed framework can identify seven different facial expressions with a high identification accuracy of 95.79% on the expanded Cohn-Kanade (CK+) database. This is significantly better than the performance of the other three state-of-the-art approaches, which are as low as 3.17%, 4.09%, and 7.41% [16].

Presented through the use of transfer learning techniques to identify emotions. Pre-trained Resnet50, vgg19, Inception V3, and Mobile Net networks are utilized in this study. We remove the pre-trained ConvNets' fully connected layers and install our own fully connected layers appropriate for the task's instruction count. Lastly, the only way the recently added layers may be trained is to update the weights. The CK+ database was used in the experiment, which produced an average accuracy of 96% for issues involving emotion recognition [17].

III. METHODOLOGY

A. *System Architecture*

System architecture in Fig. 1 represents the working of a system in which patient chooses psychologist, psychologist adds own questions. Patient answers via camera. Video converted to frames, face detection important in CNN. Dataset images resized for uniformity. Input image passed through 64-filter convolutional layer. Convolutional layer extracts image features using filters. This technique improves feature map quality by emphasizing local contrast. Normalization training speed. CNN contains feature extraction convolution tool. It has m convolutional, pooling, and fully connected layers. First layer extracts various input image features. Pooling layer reduces overfitting by generalizing features. Third layer is dense for classification. Emotions detected are saved to database. CNNs learn mapping from labeled datasets. Trained CNNs make predictions on new data. Psychologist views patient emotion report.

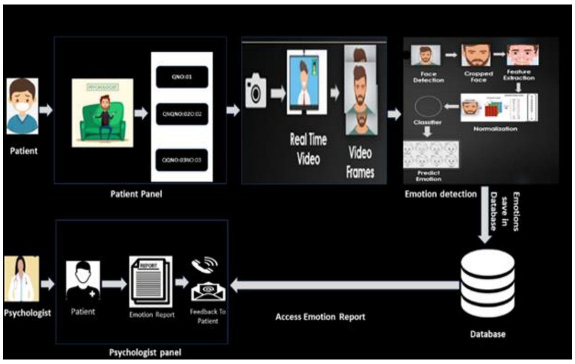

Fig. 1. System architecture.

B. Block Diagram

Block diagram in Fig. 2 explains the patient may consult with the psychologist of their choosing at any time. The system offers a variety of psychologists along with patient ratings. Patient may then choose a psychologist for consultation. The technology displays several psychological inquiries from the psychologist when the patient selects him or her. Now, be patient and click on each to respond. The patient's speech and feelings are recorded by our system. The technology automatically recognizes the patient's emotions. The patient merely needs to click on the questions and respond to them. After that, the technology stores the feelings in a database. They may edit their profile, including passwords and email addresses. Psychologists can get feedback from patients. The patients' list is visible to the psychologist. The system automatically adds the patient to the patient list for the psychologist module when a patient picks a psychologist. The psychologist next views the patient's report from the emotion analysis.

The report includes helpful information, such as emotional context for each question. Following examination of the data, the psychologist contacts the patient by phone or email with recommendations. The psychologist has the ability to edit questions, such as by deleting outdated ones or adding new ones. Administrators have full administrative control over patients and psychologists. Administrators have access to the profile of the psychologist. Following that, he can accept or deny psychologists' requests to sign up for the system. In the event of a rejection, the administrator will email psychologists the rationale for the denial. The system will immediately email the psychologist in the event of acceptance. He can update patients, remove current patients, and add new patients to the patient list. Psychologists can get feedback from patients. Keep track of patient comments for future patients.

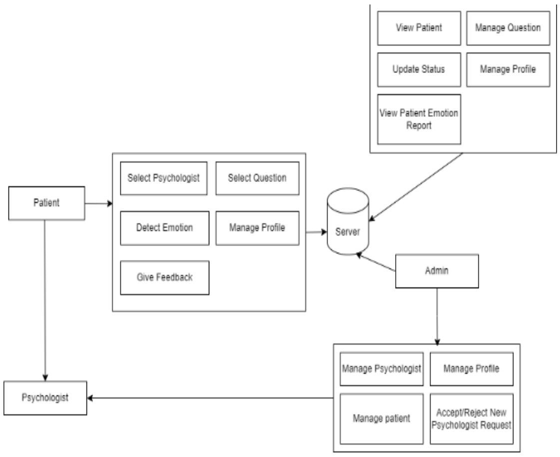

Fig. 2. Block diagram of designed system.

C. Flow Chart

Flow chart shows the login and user management process of the designed system in Fig. 3. Click start, then log in. If user is not existing, create new account option shown in flowchart. New users can register and access the system. After login, user selects user type as admin, psychologist, or patient. If psychologist selected, flowchart shows access to question management tools. They can access patient list and info. Psychologists provide feedback to the patient after receiving their emotional report. Psychologist can view, edit, and manage their profiles.

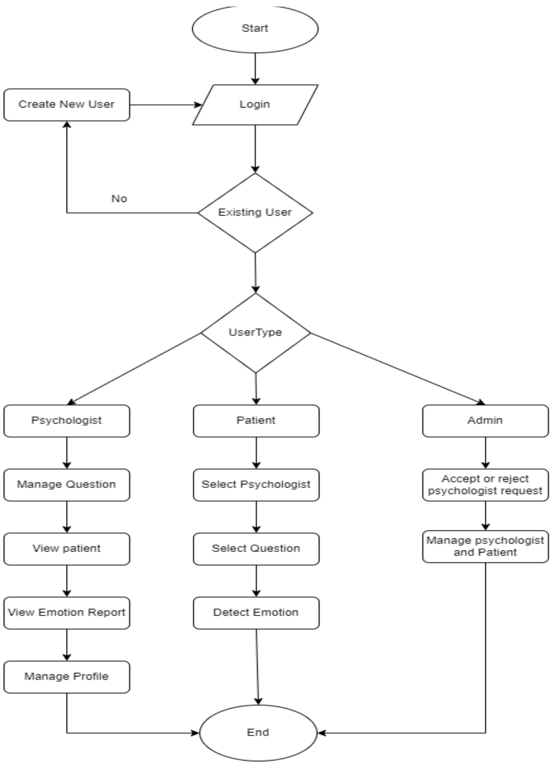

Fig. 3. Flow chart of designed system.

Patients choose their answers. Psychologists have unique sets of patient related questions. The system detects emotions from patient responses to the psychologist's questions. Allows comprehensive patient report generation. The flowchart includes an admin user who can approve or reject psychologists' account requests. The admin manages profiles and ensures system efficiency.

D. Solution

Our proposed system used CNN with Cascade classifier to detect real-time emotions. The training, test and deployment of CNN involves dataset preparation, data preprocessing, model architecture, model training, model evaluation, fine tuning and hyper parameter tuning and model deployment.

We used FER 2013 dataset, which contains facial images labeled with seven different emotions (anger, disgust, fear, happiness, sadness, surprise, and neutral). Split the dataset into training, validation, and testing sets. A common split is 80% for training, 10% for validation, and 10% for testing. We resized all images into 48x48 to ensure uniformity. It is time and power consuming to train model without converting them to RGB and grayscale. So we convert all images into gray scale. Then we normalize the pixel values to a range between 0 and 1.

We used the VGG19 architecture to train our model. In VGG19 architecture which has 9 layers, with 16 convolutional layers and 3 fully connected layers with 5 blocks of convolutional layers. VGG19 architecture use 64 filters. Then we used ReLU optimization function to enhance the accuracy of model. The output layer has eight neurons, corresponding to the eight emotions in the dataset, with a softmax activation function.

First we Initialize the model with random. Then we use loss function categorical cross-entropy to measure the model's performance. We used Adam optimizer to update the model's weights during training. Then we used FER to train the model using the training and validate dataset. After monitoring the model during training process we adjusted hyper parameters for batter accuracy.

After training we evaluate the trained model on the testing dataset to assess its performance. Then we used recall, and F1 to check the performance of the model. The accuracy of our model on testing dataset is 82.3%. In fine-tuning and Hyper Parameter Tuning we experiment with different learning rates, weight decay, dropout rates, and batch normalization. After fine-tuning the accuracy of model increased to 93.5% from 82.3%.

After saving the model we integrate with our application Django. We used haar-cascade classifier to detect the face of person from real-time video. How cascade classifier works, haar-cascade is a popular algorithm used to detect different objects including face from images. First the haar-cascade detects like features from the image. The feature can be of different types like edge feature, and line feature. Integral Image is an image which is used to speed up the haar-like features. Integral Image is images that is computed from the original image. This image is middle representation of the image that can be used for efficient calculation of sum of pixel intensities for rectangular region. Every pixel of integral image stores sum of all the pixel intensities above and to the left of it, including integral image. This can be computed using the following formula:

Integral Image (x, y) = Integral Image (x-1, y) + Integral Image (x, y-1) - Integral Image (x-1, y-1) + Image (x, y)

AdaBoost algorithm is usually used for Haar cascade classifier training. The actual training is done using two types of samples. One sample is positive sample which contains the images of face. While the other sample is called negative sample which contains images of non-face regions. So classifier can easily distinguish between these positive and negative samples by separating these two classes. The trained Haar cascade classifier has multiple stages, including few weak classifiers. In overall process each subsequent stage is more complex than other. to reduce false positives every cascade performs additional checks. Sliding window technique is used by haar cascade classifier to scan the entire image or video frame at different scales and positions. The classifier detects the haar-like features from image. The last step is to predict the emotion when a trained Convolutional Neural Network (CNN) is given an image as input for emotion prediction, it goes through a series of steps to process the image and produce the predicted emotion. Here's an explanation of the process:

In Preprocessing step classifier convert the image into suitable format for CNN. Usually classifier resizes the image to a specific size, converts it to grayscale, and normalizes the pixel values to 0 and 1. After that we pass the image from series of convolutional layers in the CNN. Every convolutional layer has filters to extract features from image. Then an activation function like ReLU is used to add non-linearity into the network. Then we used pooling layer to reduce spatial dimensions of feature maps. We used flattening to convert the output of last convolutional and pooling layers into a 1-dimensional vector. This helps the subsequent fully connected layer to process the extracted features. Then fully connected layers learn high-level representations to make predictions from the extracted layer. The last fully connected layer produce the actual output.

IV. RESULTS AND DISCUSSION

By utilizing an algorithm for machine learning emotion detection. The suggested method uses CNN and HAAR cascade classifier to identify the patient's facial mood. In order to improve the accuracy of the model, RELU, we trained our model using the FER 2013 dataset using the VGG 19 architecture for the optimization function. The DJANGO framework was utilized to integrate with the front end. Our model yielded an accuracy of 82.3% on the dataset; following fine-tuning, the accuracy increased to 92%. Recall and the F1 approach were utilized to assess the model's performance. In order to attain his performance, we trained the model using the testing dataset, which contains 48*48 pixel and grayscale pictures. We divided the dataset into testing and trainee validation datasets in order to meet our accuracy target. The accuracy of our model on testing dataset was 82.3%. In fine-tuning and Hyper Parameter Tuning we experiment with different learning rates, weight decay, dropout rates, and batch normalization. After fine-tuning the accuracy of model increased to 93.5% from 82.3%.

Our system detects the emotional state of the people. Emotion is much important in our life and effect the life of peoples that are connect with you. We use an algorithm and structure to help computers to predict emotions. The interaction between people and computer is increasing day by day. Computer are able to detect emotion more accurately than people. People communicate with one another mostly via words, but express their emotions via speech and body language also. We use this system by technology driven approaches to better serve their patients. This technology helps psychologists to understand the mental health issues of patients. Detailed information about patient's emotional states and systems can enhance the quality of feedback. Our system main aims to enhance the traditional assessment methods by utilizing advance emotion detection technology. This method categorizes in seven fundamental emotions—neutral, pleased, startled, fearful, angry, sad, and disgusted. Facial Action Coding System (FACS) method try to either categorize which Action Units (AU) are engaged. We used CNN to detect human expressions. The CNN has three layers. The convolutional layer extracts feature from image, the pooling layers do normalization and finally, the fully connected layer makes prediction. We used CK+ dataset to train our model. We introduced the Haar cascade classifier which detects face from every frame of video. The classifier finds the patterns in images by scanning them. The extracted features are then used by model to detect emotion. Our system provides a real-time emotion detection for patients by analyzing facial expressions. First we Initialize the model with random model then we use loss function categorical cross-entropy, to measure the model's performance. We used Adam optimizer to update the model's weights during training. Then we used FER to train the model using the training and validate dataset. After monitoring the model during training process we adjusted hyper parameters for better accuracy.

V. CONCLUSION

Using proposed system which uses CNN and haar-cascade classifier to detect emotions of patients in his/her session with psychologist. We trained over model using train and test split for FER dataset and achieved 93% accuracy in our system which help patient to get feedback only selected question and psychologist. Patients select psychologist to answer questions of psychologist system stores emotions of patient against every question to generate emotion report. Psychologist can analyze emotion report to provide better prescription to patient.

ACKNOWLEDGEMENT

Thankfully, we are aware of our parents' affection for us. Last but not least, we would like to thank our family and friends whose prayers made it possible for us to finish this project.

REFERENCES

[1] M. F. Valstar, M. Mehu, B. Jiang, M. Pantic and K. Scherer, "Meta-analysis of The First Facial Expression Recognition Challenge," *IEEE Trans. Sys., Man, and Cybernet., Part B (Cybernetics)*, vol. 42, no. 4, pp. 966-979, 2012.

[2] N. Sebe, M. S. Lew, Y. Sun, I. Cohen, T. Gevers and T. S. Huang, "Authentic Facial Expression Analysis," *Image and Vision Comput.*, vol. 25, no. 12, pp. 1856-1863, 2007.

[3] H. W. Ng, V. D. Nguyen, V. Vonikakis and S. Winkler, "Deep Learning for Emotion Recognition on Small Datasets using Transfer Learning," in *Proc. of the 2015 ACM on Int. Conf. Multimodal Interaction*, pp. 443-449, 2015.

[4] Y. Fan, X. Lu, D. Li and Y. Liu, "Video-based Emotion Recognition using CNN-RNN and C3D Hybrid Networks," in *Proc. of the 18th ACM Int. Conf. Multimodal Interaction*, pp. 445-450, 2016.

[5] G. Levi and T. Hassner, "Emotion Recognition in The Wild via Convolutional Neural Networks and Mapped Binary Patterns," in *Proc. of the 2015 ACM on Int. Conf. Multimodal Interaction*, pp. 503-510, 2015.

[6] S. E. Kahou, X. Bouthillier, P. Lamblin, C. Gulcehre, V. Michalski, K. Konda and Y. Bengio, "Emonets: Multimodal Deep Learning Approaches for Emotion Recognition in Video," *J. Multimodal User Interfaces*, vol. 10, pp. 99-111, 2016.

[7] M. Matsugu, K. Mori, Y. Mitari and Y. Kaneda, "Subject Independent Facial Expression Recognition with Robust Face Detection using A Convolutional Neural Network," *Neural Netw.*, vol. 16, no. 56, pp. 555-559, 2003.

[8] M. Liu, S. Li, S. Shan and X. Chen, "Auaware Deep Networks for Facial Expression Recognition," in *10th IEEE Int. Conf. and Workshops on Automatic Face and Gesture Recogn.*, pp. 1-6, 2013.

[9] B. Jiang, M. F. Valstar and M. Pantic, "Action Unit Detection using Sparse Appearance Descriptors in Space-Time Video Volumes," in *2011 IEEE Int. Conf. Automatic Face & Gesture Recogn.*, pp. 314-321, 2011.

[10] H. A. Alabbasi, P. Moldoveanu and A. Moldoveanu, "Real time Facial Emotion Recognition using Kinect V2 Sensor," *IOSR J. Comput. Eng. Ver. II*, vol. 17, no. 3, pp. 2278-2661, 2015.

[11] A. V. Haridas, R. Marimuthu and B. Chakraborty, "Emotion Recognition System for Specially Needed People with Optimized Deep Learning Algorithm," in *Fourth Int. Conf. Inventive Sys. and Contr.*, pp. 115-121, 2020.

[12] V. D. Rangari, "Dominant Facial Emotion Recognition using MTCNN in Python," *Int. Res. J. Modern. Eng. Technol. and Sci.*, vol. 4, no. 10, pp. 650-653, 2022.

[13] Y. Said and M. Barr, "Human Emotion Recognition Based on Facial Expressions via Deep Learning on High-Resolution Images," *Multimedia Tools and Appl.*, vol. 80, no. 16, pp. 25241-25253, 2021.

[14] C. A. Corneanu, M. O. Simón, J. F. Cohn and S. E. Guerrero, "Survey on RGB, 3D, Thermal, and Multimodal Approaches for Facial Expression Recognition: History, Trends, and Affect-related Applications," *IEEE Trans. Pattern Analy. and Mach. Intellig.*, vol. 38, no. 8, pp. 1548-1568, 2016.

[15] A. Sajjanhar, Z. Wu and Q. Wen, "Deep Learning Models for Facial Expression Recognition," in *2018 Digital Image Comput.: Techniq. and Appl.*, pp. 1-6, 2018.

[16] N. Zeng, H. Zhang, B. Song, W. Liu, Y. Li and A. M. Dobaie, "Facial Expression Recognition via Learning Deep Sparse Autoencoders," *Neurocomputing*, vol. 273, pp. 643-649, 2018.

[17] M. K. Chowdary, T. N. Nguyen and D. J. Hemanth, "Deep

Learning-Based Facial Emotion Recognition for Human–Computer Interaction Applications," *Neural Comput. and Appl.*, vol. 35, pp. 23311-23328, 2021.