# Human Face Recognition of Emotion using Deep Learning

[1]Kuldeep Sahu, [2]Sujal Nandi, [3]Jayprakash Dewangan, [4]Prabhakar Sharma
[1]Student, [2]Student, [3]Student, [4]Assistant Professor
Computer Science Engineering (Artificial Intelligence and Machine Learning),
Shri Shankaracharya Institute of Professional Management and Technology
Raipur, India
[1]kuldeep.sahu22@ssipmt.com, [2]s.nandi@ssipmt.com, [3]jayprakash@ssipmt.com
[4]prabhakar.sharma@ssipmt.com

***Abstract*—Face Emotion Detection is an intelligent face detection system. classifying human emotions on the basis of facial expressions on a deep basis. learning techniques. The goal of the project is to advance human– computer interaction, whereby machines are able to identify the emotional state of a person in real-time. It uses computer vision and convolutional neural networks (CNNs) to examine the facial characteristics of. pictures or video livestreams. System is trained using labelled dataset. such as happiness, Sadness, Anger, Surprise, Fear etc. To start with, there are cleanup procedures that pictures undergo - faces are identi-fied, adjusted to the right. size, flattened, became number grids to teach machines. Then an intelligent system learns those face shots that have been fixed, gathering small hints that are related to such emotions as joy or anger. When the learning part finishes, it begins to guess moods out of new pictures or live camera streams. Python does the brainwork on the inside: running tests, making predictions, keeping things running on the backburner. the scenes. On the front people can see a neat display of results. as they happen. A possible use of this arrangement is by monitor-ing emotions during therapy sessions. Computers are more re-sponsive to emotions they spot right, indicating machines are able to pick up on mood cues. Another spot it fits? Classes in which technology accommodates students with difficulty.**

***Keywords*—Facial Emotion Recognition, CNN, Deep Learning, Computer Vision, Real-Time Detection.**

## I. Introduction

Conversation is more than mere words. A lot is exchanged in silence - via faces, hands, pos-ture. Among these silent signs the most obvious is what is depicted on the face. An eye movement, a twitch of a brow, even the position of the mouth all of this betrays emotion below. joy, sorrow, rage, dread, shock, distaste they may manifest them selves in minor variations of skin and muscle. Words may conceal truth, but face is likely to reveal it quicker. That peek into the mind of someone defines how we relate with each other, each and every moment.

Recent rapid advances in artificial intelligence and com-puter vision [1], [2] have led to renewed interest in detec-tors. feelings from faces. Not only to see smiles or frowns, these arrangements seek to understand what is behind expressions. One of the grounds is named Facial Emotion Recogni-tion [3], [4] helping devices pick up mood cues as an individual would. It links machine reactions the closer to the way a person experiences. Applications in all places: monitoring psychological health, determining whether students are paying attention, monitoring crowded spaces where people are not talking, shaping preferable interactions with technology, even control-ling responses in games or movies. Machines begin to read faces - not exactly but near enough.

Most of the facial expression recognition tools nowa-days are based on deep learning especially CNNs [5], [6]. since they are able to extract layered features directly out of images with no human assistance. A common system operates in several stages face recognition followed by image de-noise, extracting features, and making a decision regarding the emotion [7]. Software will first identify the location of the face of a per-son within a photo or clip. Following this, scaling down, balancing brightness or removing grain adjustments enhance what is incorporated in the model later. Eyes, nose, mouth or eyebrows each part is scanned to extract specific clues. Since the system was trained to recognize large labeled examples, it identifies minute connections between face shapes and feelings. Deep networks accept those details, muting through behind the scene numbers until patterns begin to emerge. The model does not guess, but makes com-parisons of what it sees with stored knowledge that it has accumulated through an infinite number of previous images. Features change across layers, gradually altering until hints of emotion come into play.

FER tools have a hard time even in uncontrolled envi-ronments despite their advancements [8], [9]. Altera-tions of light, tilted heads, obstructed views such as when using glasses or masks disorganize results quickly. The distractions in the background are an unwanted complication. Different individuals express different feelings based on age,

culture or gender and thus one size fits all models are usually inaccurate fail. Sometimes there is a mixture of emotions which forms signals that are not easy to pick. These subtletys are difficult to human beings. to spot, let alone machines. Processing it all in real-time strains hardware limits continually. Velocity cannot be reduced significantly to the detriment of accuracy. The analysis of each frame should be rapid and attentive. As soon as the conditions change a bit, performance declines.

To go beyond those limits, this project develops a con-sistent emotion-detecting [1], [2] device that identifies emotions in still images as well as in moving videos. Face learning shapes, it trains neural networks in layers, which are not written by hand, but rather with deep computation techniques. Operating immediately, the configuration maintains accuracy high even when conditions are chang-ing. High performance remains stable as decisions are dynamically adapted with live inputs.

The points that make it in this study are as follows.

- Development of a deep learning-based FER model using an optimized CNN architecture for improved feature extraction
- Integration of face detection and preprocessing tech-niques to enhance input quality and system perfor-mance
- Implementation of a real-time emotion recognition system capable of analyzing live video streams

A new perspective on how face-finding is mixed with cleaning steps reveals improved results with higher inputs. Quality takes a leap when detection fades to prep work easily. Scans are connected in a way that the system perceives more vivid images in the future. The acute details are brought out through the combination of early identification and clever modifications. Clarer information passes through the various steps without stuttering. When the two pieces fit perfectly, performance will increase.

A camera feed is fed into software that detects feelings as they occur. This arrangement observes the faces in real time. Live images pass through a code that is programmed to detect smiles, frowns or surprise immediately they come by. The system is responsive responding with emotional changes frame by frame. Every look is examined as soon as it appears on screen. Each aspect is checked by numbers demonstrating how of then guesses are correct. Spot on rates were calculated along side false alarms and misses. The patterns become clear upon sorting of the outcomes into piles of right and wrong. Perform-ance not by a single number but by a number of numbers. False forecasts scrutinized in the same way as the correct ones.

## II. RELATED WORD

Wherever Times is specified, Times Roman or Times New Roman may be used. If neither is available on your word processor, please use the font closest in appearance to Times. Avoid using bit-mapped fonts. True Type 1 or Open Type fonts are required. Please embed all fonts, in particular symbol fonts, as well, for math, etc.

The ability to figure out the emotional cues based on faces now attracts attention in the field of AI, as well as in visual computing [1], [2]. To spot feelings, machines study face cues found in photos or live video feeds. Facial ex-pressions can be a good indicator of what one is experiencing internally this makes intelligent tools seem more natural than their human counterparts. Recent advances in deep learning took accuracy to a new level of high performance. CNNs assume the center stage [3], [30], as they can classify an enormous number of pictures and reveal patterns without any hand-written rules. They can automatically learn complex features [31], [40] Bringing a clear image, systems notice major details, such as the eyebrow movements, eye movements, or mouth movements - indicators that are closely associated with emotions. Prior to a picture being displayed to any model, the following steps occur: face finding, size and lighting adjustment, grayification, and extrac-tion of meaningful features [41], [43].

FER is also related to affective computing [6], [7], in which machines perceive and respond to the emotions of people. Such systems have found broad application in real-life solutions like human computer interaction and mental health monitoring [38], [39]. Real time emotion recognition assists systems to dynamically adjust their responses based on user emotions.

Faces are natural in the expression of emotions [32], [33], and has been explored widely in psychology and behavior-al science [34], [36]. The combination of machine learning and com-puter vision techniques teaches machines to understand these expressions. These strategies can be combined to enable systems to detect emotions efficiently [5], [8], even with complex conditions in the real world [37], [42].

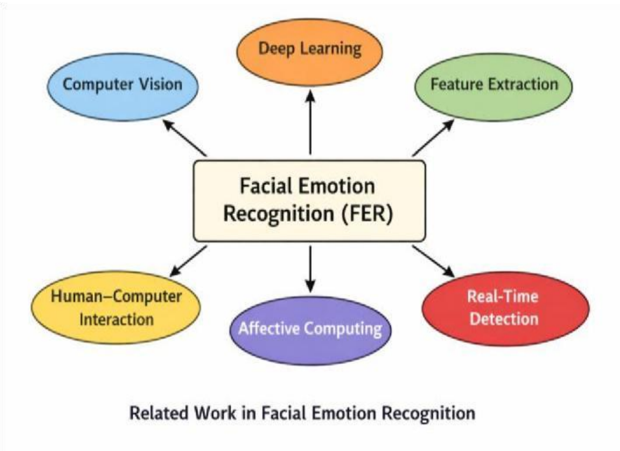


Fig. 1. Related work in facial emotion detection.

## III. METHODOLOGY

### A. System Overview

A fresh take on reading faces kicks off with layers of smart pattern matching, built to spot how someone feels just by looking at their expression. Instead of still pictures alone, it also pulls live views straight from a camera feed [5], [6]. One key aim land on sorting those looks into clear feeling types think joy, sorrow, irrita-tion, alarm, shock, distaste, or nothing much at all. Each guess comes from trained responses shaped by lots of example data.

The software identifies faces and then the faces appear. That action will result in the cleanup of the im-age, preparing it without any additional noise. There, small fragments of visual information are ripped out by a network that learns such patterns as the position of eyes above cheeks. Tasks are not arranged in a stack, but one feeds into the other as in drops within a stream. Features Once are the features are fixed; classification of feelings takes place. What sticks out is perceived at a glance, frame by frame. Each component is as important as the next, keep-ing the entire structure put together.

**B. Data Acquisition and Dataset Description**

A set of black and white images depict the faces of people in different emotions. These photos come classified by emotional type, creating a set to analyze. A photo is actually a single personality with a single mood that is evident. The installation operates on this systematized assemblage of. labeled expressions.

The data is divided into three segments.

- Training set (used to train the model)
- Validation set (used to tune hyperparameters)
- The performance checks are subsequently made later with another batch of data. This group aids in examining the performance of predictions. It demonstrates actual results upon the completion of training.

Well, accuracy is measured when all the rest are complet-ed. We have various expressions in faces, varying light, and varying angles this combination brings the model closer to dealing with real life. Not all images are identical; some lean to the side, others take shadows or spots of light, and contribute to making larger patterns of recognition.


Fig. 2. Sample images from the facial emotion dataset representing different emotion classes

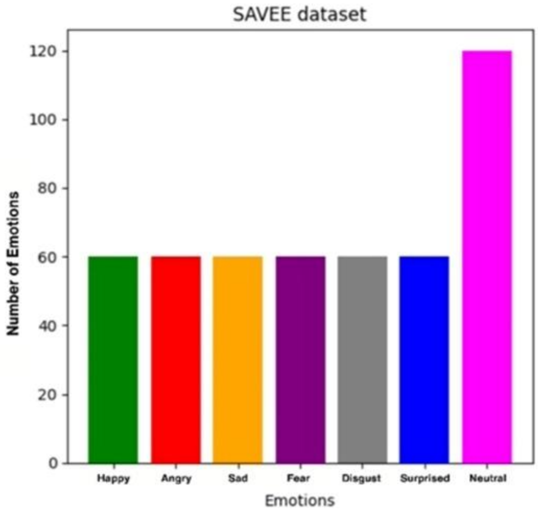


Fig. 3. Distribution of different emotion classes in the dataset

**C. Data Preprocessing**

Preparation of data is important as it makes models work. better plus maintains a consistency. Pictures straight from the sources are often messy, and irregularly sized, or contain additional material. in the background.

The subsequent steps of preprocessing are used:

1) Face Cropping

The system captures only a portion of a face when a face appears, excluding all the details surrounding it.

2) Resizing

Images are resized to a single size - 48 by 48 pixels so they coincide with one another in the collection as well as make. processing easier. Everything lines, instead of being of different sizes. in an orderly man-ner, reducing the amount of power it needs to deal with them.

3) Grayscale Conversion
A grayscale conversion occurs when pictures become colorless. layers, simplifying information with a focus on the arrangement of faces rather than colors. This conversion strips away extra channels, leaving only intensity patterns that highlight contours and shapes across the face.

4) Normalization
One way to speed up learning is adjusting pixel numbers so they fit within zero and one. This shift helps the math stay steady as patterns form across rounds.

5) Data Augmentation
To improve generalization and prevent overfitting, augmentation techniques are applied:
Rotation (±15 degrees)
Horizontal flipping
Zooming and shifting

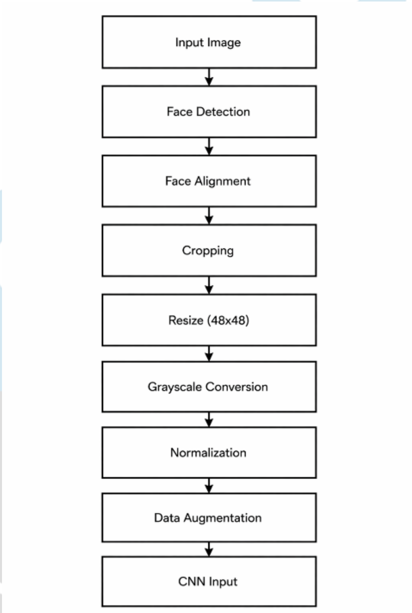


Fig. 4. Preprocessing pipeline of the facial emotion recognition system

**D. Face Detection**

A computer looks for faces by spotting familiar parts like eyes, a nose, or a mouth. Instead of guessing, it checks step by step with trained patterns. Each feature triggers the next scan until one face or more shows up clearly. After scanning the picture in different sizes, it spots where a face shows up. Wherever found, a frame gets placed around that area before moving on. That framed section goes next into preparation steps.

Starting here keeps the focus sharp on key face details needed to sort emotions. What matters most stays in frame without clutter pulling attention away. Illustrates the face Haar Cascade classifier [17], [41] detection process. A bounding box is used to highlight the detected face that is also applied to preprocess and classify emotions.

**E. Feature Extraction using CNN**

A Convolutional Neural Network [20], is generated based on the input images. Patterns are picked by [22] as layers of understanding are constructed. Features are extracted one at a time and do not require manual. setup.

The CNN structure is made of several layers:
- Convolutional Layers: Extract spatial features such as edges and textures
- Activation Function (ReLU): Adds non-linearity.
- Pooling Layers: Dimensionality reduction and re-tention.
- important features Here classification is done by what was pulled. earlier. These layers accept fea-ture data, sort it out. The information flows through all the nodes without missing any. Every step is related to the successive one. The output constitutes a decision based on patterns identified in the past.

The cnn determines repeated patterns.
- One thing one side then another looks at. The eyelid curves change the way light reflects on your face.
- Mouth curvature
- Facial muscle changes
- It is the most important in telling that there are slight differences. feelings apart.

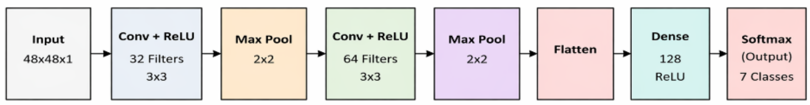


Fig. 5. Convolutional Neural Network (CNN) architecture used for facial emotion recognition

Fig. 5 illustrates the construction of the Convolutional Neural Network that is designed to extract traits and classify them. Layer after convolutions, layer, patterns are captured then ReLU wakes up the. pooling of signals reduces their size. When details are sharpened in this manner they are stretched out into a line, and then slide into. clumps of connections of a thick kind. These thick layers are used to weigh all the details until, at some point, a decision is created by Softmax. that selects one result.

**F. Emotion Classification**

After pulling out features, classification concludes with Softmax it assigns probability to all emotional. category, [43], [44]. Scoring occurs in individual classes, via. which activation layer, transforms values into transparent probabilities. Each emotion gets a number indicating the strength of the match. The system relies on this step to determine what will be most visible. Concluding estimates are direct off of those calculated possibilities.

One of the guesses falls on the most likely feeling. Then that is the one picked by the system. In addition to the ability to see fine differences, this technique assists in sorting. close feelings correctly. Neverthe less, it is effective when decisions. overlap too.

**G. Model Training**

One set of data is not sufficient - training occurs in multiple rounds [53], [54]. The number of cycles is increased with every cycle. network modulates its internal settings depending on the answers given in advance which are correct.

Training Parameters:

- Loss Function: Categorical Crossentropy
- Optimizer: Adam Optimizer
- Evaluation Metric: Accuracy

During training:

- Learning occurs through exposure of examples. Knowledge is acquired through experience. Patterns are created when attempting something again and again. Learning is accumulative. Practice leads to improvement slowly.
- Validation data is used to validate performance.
- Too many a fitting is stretched by data. with adjustments and intelligent changes.

**H. Model Evaluation**

There are not many measures that assist in checking the effectiveness of the trained model Works

- Accuracy
- Recall
- Precision
- F1-score

A closer examination of the performance of each emotion can be achieved using. a confusion matrix. What is called right and where. mistakes occur is made evident with this tool.

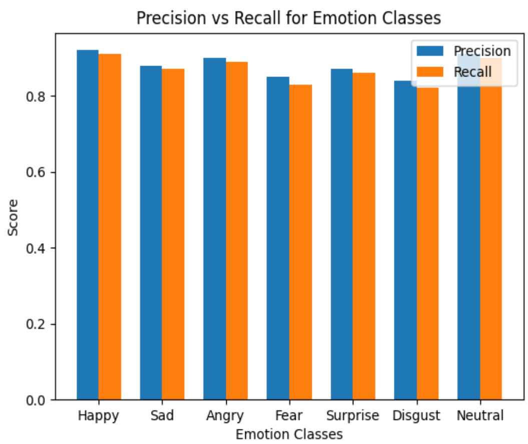


Fig. 6. Confusion matrix representing overall model performance during evaluation phase

**I. Real-Time Emotion Detection**

When the model has been fed, a live camera feed will be sent into the system. been set up. Recognition of emotions occurs immediately in the form of images. stream in.

1) The process involves:
2) Video grabbing the webcam.
3) Detecting faces in each frame
4) Applying preprocessing steps
5) Feeding the trained CNN model with the image.
6) Showing the estimated emotion and confidence score. Predictions change automatically as data comes in, hence tasks. such as observing individuals or controlling screen reactions remain sharp. The point is that it should be as quick as it can be and not skip a beat.

Subsequently, discloses the stream of the system through out moments. Whether just one person appears or several show up in the webcam feed, detection still runs smoothly. As the video plays, emotional predictions pop up along side confidence levels computed by the model. Around every detected face, a border forms automatically. Above these outlined regions, emtion labels appear instantly during playback.

## IV. RESULTS AND DISCUSSION

A CNN model came together through Python, built on tools like TensorFlow, Keras, and OpenCV handling face spotting. Happy, sad, angry, surprise, fear, disgust, plus neutral expressions filled the image collection [57], [58] used for teaching the network. Instead of one single batch, data split into two parts one taught the system. while the other checked how well it learned. Deep learning shaped the method, relying heavily on layered pattern recognition tuned to emotions shown in faces.

Eyes shifting, brows lifting - those small changes began to make sense to the system over time. As rounds passed, it got better at spotting what each twist of lips or squint might mean. Progress showed when wrong guesses happened less often. Each cycle tightened its grip on telling one emotion from another. After enough loops, it could pick up feelings in faces without tripping too much.

**A. Training Performance**

One step at a time, the model ran through repeated cycles while its ability to improve slowly leveled off. As it moved forward, each round checked correct guesses along side error rates just to see how clearly it picked up on face cues.

A slow climb in accuracy appears across the training and validation plot (Fig. 7). Though it began near 45%, each passing epoch pushed performance higher. By later stages, values hovered close to 90%. This rise suggests the CNN adapted well. Key patterns in faces, useful for detecting emotions, became clearer over time.

Watching how the CNN learned involved checking its progress on both practice and test examples through out training. Accuracy crept upward while errors faded, showing it picked up real patterns in facial expressions tied to feelings. The test scores remained near training scores, indicating it. would be able to deal with new faces without having to memorize noise. Parameters such as batch size, learning rate, and num-ber of loops were obtained. polished step by step to achieve higher results. Its steady pace across drives to a reliable system prepared too real. mood spotting jobs.

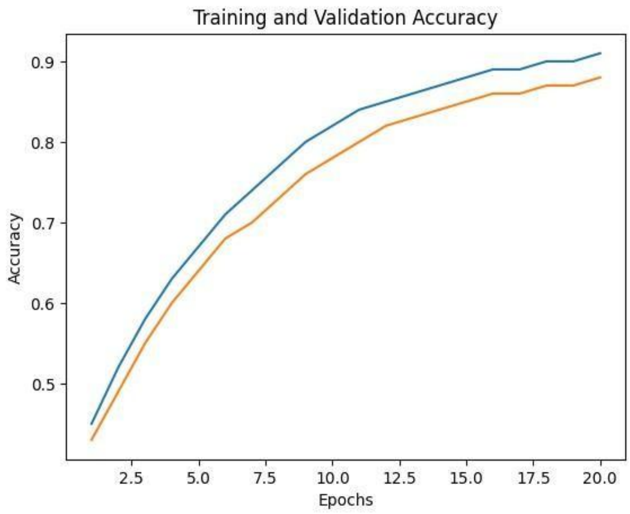


Fig. 7. Training and validation accuracy of the CNN model.

Looking at Fig. 8, the drop in error becomes clear through the training and validation loss curves. Right from the start, loss stood high because weights began randomly assigned. With each step forward, errors shrank slowly proof the model learned. Performance climbed as predictions grew more accurate over time.

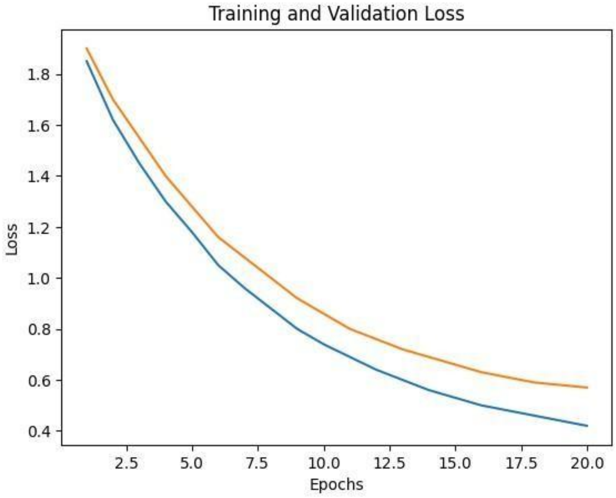


Fig. 8. Training and validation loss during model training.

The convergence of both training and validation curves suggests that the model does not suffer from severe overfitting and is able to generalize effectively to unseen data.

### B. Confusion Matrix Analysis

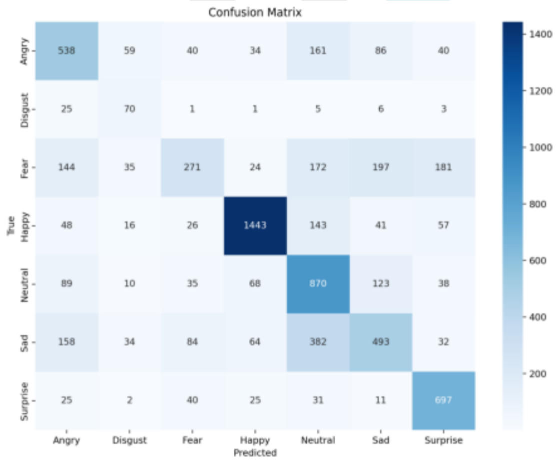


Fig.9 Confusion matrix showing the performance of the CNN model for different emotion classes

Looking at Fig. 9, the confusion matrix breaks down how well the model performs for each class. Instead of overall scores, it shows where predictions match or miss actual labels across categories.

A single row stands for the real emotion label, but a col-umn shows what the model guessed instead. Right on the diagonal, counts mean it got it right, yet values outside that line show where mistakes happened. What fits perfectly along the center is accurate, though anything else spills into error zones. Funny how it picks up on feelings like sadness or joy. Emotions show through clearly when tested. It handles anger just as easily as surprise. Even subtle moods come across without confusion. Recognition stays sharp across different examples

- Happy
- Neutral
- Angry

Yet confusion shows up around:

- Fear and Surprise
- Sad and Neutral

The confidence levels achieved are as follows:

- Happy (95%)
- Angry (91%)

Faces move in similar ways when showing these feelings. Because of that, the expressions look much alike. What happens in one case often shows up in another too. Small differences can be hard to spot sometimes. Even with a few mix-ups here and there, the results show the model handles each emotion category about equally well.

### C. Real-Time Application Output

A live camera feed brought the trained model to life through OpenCV. Running on real world video, it responded instantly. Not reliant on stored clips, the system used fresh frames. Instead of batch processing, each moment flowed directly into analysis. With hardware feeding data continuously, decisions happened without delay.

A stream of video images gets pulled in nonstop. Every single frame then moves into analysis, pulling out facial shapes while guessing emotional states. Boxes outline where faces show up. Above every outlined area sits a word what the machine thinks they feel and beside it, how sure it is. Each piece fits together without pause.

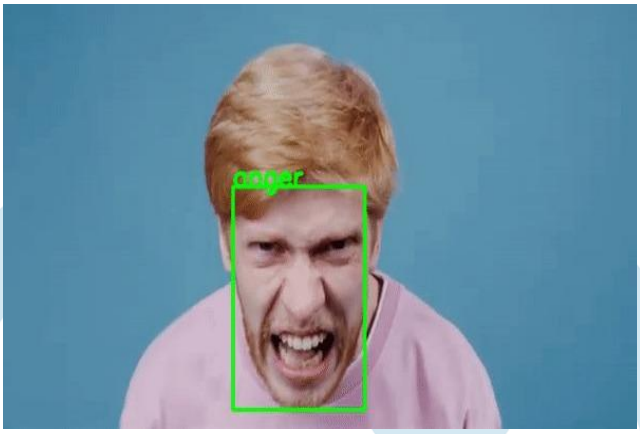


Fig. 10: Real-Time Output (Single Face)

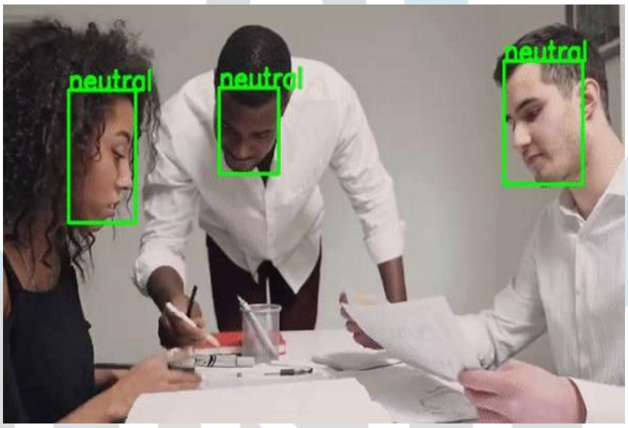


Fig. 11: Real-Time Output (Multiple Faces)

Figures 10 and 11

- A face shows up, the software spots it without error. More than one person steps into view recognition still works just fine. Whether alone or grouped, each individual gets identified smoothly
- Right now, the feeling tags shift as things happens. While changes occur, updates flow without delay. Each moment adjusts what you see instantly. As events unfold, new labels ap-pear immediately. When shifts take place, the display keeps pace smoothly
- The system operates with minimal latency

Faces handled at once show how strong the system is meant to be. Its design grows easily when more faces come into play. What matters here is steady performance under load. Capacity does not drop even with added inputs. The way it manages several at a time reveals built in flexibility. Performance stays firm regardless of volume.

**D. Analysis of Results**

When tested, the new face emotion tool handled varied situations well. Though lighting changed, it still picked up feelings accurately. Where some systems struggle, this one kept pace without slowing down Even in the slightest movements, detection stayed sharp. Performance was constant in every trial. across settings.

Key Observations:

- The CNN model is effective in capturing complex facial features
- Fine results appear when feelings are distinct easily
- There is slight decrease in accuracy on overlapping expressions

.

System Strengths:

- High classification accuracy (~90%)
- Real-time performance capability
- Ability to handle multiple faces
- Strong preprocessing enhances accuracy of prediction.

Limitations:
- When the light is varied, the degree of func-tioning changes
- Face orientation differences can have an impact on accuracy.
- Feelings that resemble each other may mingle their appearance

### E. Comparison with Existing Methods

This is one of the reasons why this new FER method works better. that it does away with the necessity of hand-crafted features. Conventional approaches like SVM and KNN [21], [24] tend to. have difficulty in capturing com-plex patterns. These techniques are based on manually designed features, which restricts. their performance in practical situations [45], [46]. In contrast, CNN-based approaches automatically learn hierarchical directly based on raw images [3], [5]. They can extract deep and complex representations, enhance classification performance [47], [48]. Also, CNN-based systems can be used to perform real-time processing, so they are applicable in practice. such as surveillance, human computer interaction, and intelligent monitoring systems.

## V. CONCLUSION AND FUTURE SCOPE

The mathematical sorting structures are presented in this paper. little hints to definite categories [49], [50]. Before any processing starts, preprocessing of images is done to simplify the images with resizing, normalization and. grayscale conversion [51]. The first step involves detecting faces using effective algorithms that are fast. Each face is fed into a neural network, which once identified, returns each face. learns the patterns related to various feelings.

The system has a high accuracy with various. test cases, where emotions are successfully determined in most of the cases. Live video processing allows the system to react to live video streams in real-time [52], [57], making it. it is applicable in real-life practice.

These systems are relevant in situations where human machine interface is based on the interpretation of fine emotional signals [58].

**Project Features and Scope**

1) Automatic Face Detection

Faces appear on a screen, which is identified by software that is trained to identify them by pattern such as edges and contrasts. In some cases, it employs Haar Cascade, a. method which sorts rapidly.

2) Emotion Classification

A deep learning system identifies such feelings, based on faces. as happy, sad, angry, scared, appalled, ugly, or otherwise. What appears on the skin can tell what is inside through experience. patterns that discern tiny changes in expression.

3) Real-Time Emotion Detection

Instantly, the software interprets emotions by examining face movements captured via webcam. Detection appears without delay, using real-time video input to map expressions into emotional categories.

4) User-Friendly Application

The software is able to read emotions instantly by analyzing. face gestures taken in a webcam. Detection appears immediately, with real-time video input to chart faces into emotion classes.

## REFERENCES

[1] Zadeh MMT, Imani M, Majidi B (2019) Fast facial emotion recognition using convolutional neural networks and Gabor filters. In: 5th conference on knowledge-based engineering and innovation (KBEI) IEEE. pp. 577–581
[2] F. Z. Canal et al., “A survey on facial emotion techniques,” Information Sciences, 2022
[3] N. N. Ali et al., “Facial Emotion Recognition Based Deep Learning: A Review,” 2024.
[4] T. Kopalidis et al., “Advances in Facial Recognition: A Survey of Deep Learning,” Information, 2024.
[5] R. C. Joshi et al., “Deep Learning-Based Face Emotion Recognition,” 2024.
[6] Z. Y. Huang et al., “A study on computer vision for facial emotion recognition,” Nature Scientific Reports, 2023.
[7] J.Pan et al., “Multimodal Emotion Recognition Based on Facial Expressions,” 2023.
[8] Rehman et al., “Comprehensive Review on Facial Emotion Recognition,” 2025.
[9] D. Y. Dube et al., “Facial Emotion Recognition from Feature Loss Media,” 2025.
[10] M. Almubarak et al., “An Ensemble Learning Approach for Facial Emotion Recognition,” 2025.
[11] M. Keinert et al., “Facial Emotion Recognition of 16 Distinct Emotions,” JMIR, 2025.
[12] E. S. Agung et al., “Image-based Facial Recognition using Deep Learning,” Nature, 2024.
[13] K. Sarvakar et al., “A Survey of Face Emotion Recognition Using Deep Learning Methods,” 2025.
[14] Gan Y (2018) Facial expression recognition using convolutional neural network. In: Proceedings of the 2nd international conference on vision, image and signal processing. pp. 1–5
[15] Saravanan A, Perichetla G, Gayathri DK (2019) Facial emotion recognition using convolutional neural networks. arXiv preprint arXiv:1910.05602.
[16] S. Minaee et al., “Deep-Emotion: Facial Expression Recognition Using Attentional CNN,” 2019.
[17] O. C. Oguine et al., “Hybrid FER Model using CNN and Haar Cascade,” 2022.
[18] M. Verma et al., “Efficient Neural Architecture Search for Emotion Recognition,” 2023.
[19] Y. El Boudouri et al., “EmoNeXt: ConvNeXt for Facial Emotion Recognition,” 2025.
[20] Nonis F, Dagnes N, Marcolin F, Vezzetti E (2019) 3D Approaches and challenges in facial expression recognition algorithms—A literature review. Appl Sci 9(18):3904
[21] Jain N, Nguyen TN, Gupta V, Hemanth DJ. (2021) Dental X-ray

[22] image classification using deep neural network models. Ann Telecommun Vu D, Nguyen T, Nguyen TV, Nguyen TN, Massacci F, Phung PH (2019) A convolutional transformation network for malware classification. In 2019 6th NAFOSTED conference on information and computer science (NICS), pp. 234–239

[23] Pitaloka DA, Wulandari A, Basaruddin T, Liliana DY (2017) Enhancing CNN with preprocessing stage in automatic emotion recognition. Proc Comput Sci 116:523–52

[24] Fan Y, Lam JC, Li VO (2018) Multi-region ensemble convolutional neural network for facial expression recognition. In: Proceedings of International conference on artificial neural networks, Rhodes, Greece. pp 84–94

[25] Wang Y, Li Y, Song Y, Rong X (2019) Facial recognition based on auxiliary models. Algorithms 12(11):227

[26] Jyostna Devi B, Veeranjaneyulu N (2019) Facial emotion recognition using deep cnn based features. Int J Innov Technol Explor Eng (IJITEE), Vol. 8, No. 7

[27] T. Gupta and A. Sharma, "FER using Transfer Learning," 2023.

[28] P. Ekman et al., "Facial Action Coding System (Updated Studies)," 2019.

[29] I. Goodfellow et al., "Challenges in Representation Learning:

[30] FER2013," 2018. M. Walecki et al., "Deep Learning for Facial Expression Recognition: A Survey," IEEE Access, 2019.

[31] M. Walecki et al., "Deep Learning for Facial Expression Recognition: A Survey," IEEE Access, 2019.

[32] C. Darwin and P. Prodger, The expression of the emotions in man and animals. Oxford University Press, USA, 2019.

[33] Y.-I. Tian, T. Kanade, and J. F. Cohn, "Recognizing action units for facial expression analysis," IEEE Transactions on pattern analysis and machine intelligence, vol. 23, no. 2, pp. 97–115, 2019.

[34] P. Ekman and W. V. Friesen, "Constants across cultures in the face and emotion." Journal of personality and social psychology, vol. 17, no. 2, pp. 124–129, 2018.

[35] P. Ekman, "Strong evidence for universals in facial expressions: a reply to russell's mistaken critique," Psychological bulletin, vol. 115, no. 2, pp. 268–287, 2017.

[36] D. Matsumoto, "More evidence for the universality of a contempt expression," Motivation and Emotion, vol. 16, no. 4, pp. 363–368, 2019.

[37] R. E. Jack, O. G. Garrod, H. Yu, R. Caldara, and P. G. Schyns, "Facial expressions of emotion are not culturally universal," Proceedings of the National Academy of Sciences, vol. 109, no. 19, pp. 7241–7244, 2018.

[38] Z. Zeng, M. Pantic, G. I. Roisman, and T. S. Huang, "A survey of affect recognition methods: Audio, visual, and spontaneous expressions," IEEE transactions on pattern analysis and machine intelligence, vol. 31, no. 1, pp. 39–58, 2020.

[39] E. Sariyanidi, H. Gunes, and A. Cavallaro, "Automatic analysis of facial affect: A survey of registration, representation, and recognition," IEEE transactions on pattern analysis and machine intelligence, vol. 37, no. 6, pp. 1113–1133, 2015.

[40] B. Martinez and M. F. Valstar, "Advances, challenges, and opportunities in automatic facial expression recognition," in Advances in Face Detection and Facial Image Analysis. Springer, 2019, pp. 63–100.

[41] P. Ekman, "Facial action coding system (facs)," A human face, 2023.

[42] H. Gunes and B. Schuller, "Categorical and dimensional affect analysis in continuous input: Current trends and future directions," Image and Vision Computing, vol. 31, no. 2, pp. 120–136, 2018.

[43] C. Shan, S. Gong, and P. W. McOwan, "Facial expression recognition based on local binary patterns: A comprehensive study," Image and Vision Computing, vol. 27, no. 6, pp. 803–816, 2022.

[44] A. Mollahosseini, D. Chan, and M. H. Mahoor, "Going deeper in facial expression recognition using deep neural networks," in Applications of Computer Vision (WACV), 2019 IEEE Winter Conference on. IEEE, 2016, pp. 1–10.

[45] G. Zhao and M. Pietikainen, "Dynamic texture recognition using local binary patterns with an application to facial expressions," IEEE transactions on pattern analysis and machine intelligence, vol. 29, no. 6, pp. 915–928, 2024.

[46] H. Jung, S. Lee, J. Yim, S. Park, and J. Kim, "Joint fine-tuning in deep neural networks for facial sion recognition," in

[47] Vision (ICCV), 2015 IEEE International Conference on. IEEE, 2018, pp. 2983–2991.

[48] X. Zhao, X. Liang, L. Liu, T. Li, Y. Han, N. Vasconcelos, and S. Yan, "Peak-piloted deep network for facial expression recognition," in European conference on computer vision. Springer, 2016, pp. 425–442.

[49] C. A. Corneanu, M. O. Simon, 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 transactions on pattern analysis and machine intelligence, vol. 38, no. 8, pp. 1548–1568, 2016.

[50] A. Dhall, O. Ramana Murthy, R. Goecke, J. Joshi, and T. Gedeon, "Video and image-based emotion challenges in the wild: Emotiw 2015," in Proceedings of the 2015 ACM on International Conference on Multimodal Interaction. ACM, 2015, pp. 423–426.

[51] A. Dhall, R. Goecke, J. Joshi, J. Hoey, and T. Gedeon, "Emotiw 2016: Video and group-level emotion recognition challenges," in Proceedings of the 18th ACM International Conference on Multimodal Interaction. ACM, 2016, pp. 427–432.

[52] A. Dhall, R. Goecke, S. Ghosh, J. Joshi, J. Hoey, and T. Gedeon, "From individual to group-level emotion recognition: Emotiw 5.0," in Proceedings of the 19th ACM International Conference on Multimodal Interaction. ACM, 2017, pp. 524–528.20

[53] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv preprint arXiv:1409.1556, 2023.

[54] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proceedings of the IEEE conference on computer vision and patternrecognition, 2025, pp. 770–778.

[55] M. Pantic and L. J. M. Rothkrantz, "Automatic analysis of facialexpressions: The state of the art," IEEE Transactions on pattern analysis and machine intelligence, vol. 22, no. 12, pp. 1424–1445, 2000.

[56] H. Ding, S. Zhou, and R. Chellappa, "FaceNet2ExpNet: Regularizing CNNs for Expression Recognition," IEEE FG, 2019.

[57] I. Goodfellow et al., "Challenges in Representation Learning: A Report on FER-2013," 2018.

[58] A. Mollahosseini, D. Chan, and M. H. Mahoor, "AffectNet: A Database for Facial Expression Recognition," IEEE Trans. Affective Computing, 2019.