# Enhancing Emotional Well-being in Healthcare Facial Emotion Classification Using Deep Convolutional Neural

Aditya Girish
*Dept. of Computer Science and Engineering*
*Manipal Institute of Technology,*
*Manipal Academy of Higher Education*
Manipal, Karnataka, India

Andrew J
*Dept. of Computer Science and Engineering*
*Manipal Institute of Technology,*
*Manipal Academy of Higher Education*
Manipal, Karnataka, India

***Abstract*— The recognition of emotions is a critical part of patient care in healthcare settings. In this research, a Convolutional Neural Network (CNN)-based model is developed to identify human emotions accurately in real-world situations. In order to train the model, we collected a dataset of facial expressions covering five emotions: 'Angry', 'Disgust', 'Happiness', 'Sadness', and 'Surprise'. The algorithm utilizes advanced layers, including dense, hidden, dropout, and output layers, as well as optimizers such as Adam and RMSprop. It is trained on a Kaggle-hosted dataset, demonstrating its adaptability. A seamless integration of cutting-edge technologies is demonstrated using TensorFlow, Python3, and OpenCV. The model has wide-ranging and impactful applications in healthcare. Patients can interact and communicate more effectively, pain can be assessed by analyzing facial expressions, and healthcare providers can benefit from it. Patients' emotional states can be monitored continuously through the use of this technology, which can be used to develop treatment plans and interventions based on the data collected. Pediatricians can use the model to better understand and address the emotional well-being of young children, particularly when verbal communication is limited. Additionally, in facilities dedicated to patients with dementia, emotional detection plays a pivotal role in identifying signs of distress or discomfort and guiding the implementation of enhanced care strategies.***

## I. INTRODUCTION

Human emotions serve as a fundamental and mental well-being, influencing overall quality of life. However, challenges arise when emotional health deteriorates, leading to potential social and psychological issues. The recognition and detection of emotions in healthcare data provide valuable insights into patient's emotional states. Addressing this complex task, our research introduces an innovative approach to automatically detect patients' emotions in the context of specific diseases using facial recognition methods.

In the broader context, AI systems capable of detecting human emotions have garnered significant attention, offering transformative implications across various industries, especially in healthcare. Emotion recognition involves identifying and categorizing emotions such as happiness, sadness, surprise, fear, and disgust through computer algorithms that analyze expressions, body language, and visual signals. The automated emotion recognition process through AI can potentially revolutionize fields such as psychology, education and customer service.

Recent developments in AI-driven emotion recognition have incorporated advanced models like BERT and GPT-3, leveraging transformer-based approaches to enhance contextual understanding of emotional cues. Multimodal learning, which combines information from video sources, has also evolved. Self-supervised learning techniques address data scarcity issues, enhancing model robustness. Theoretical discussions delve into embodied cognition, social signal processing and reinforcement learning frameworks, enhancing AI systems' adaptive emotional intelligence for natural and contextually appropriate responses.

The relevance of standardized evaluation metrics and interdisciplinary collaboration between psychologists, computer scientists and neuroscientists is emphasized. Mathematical derivations, attention mechanisms and memory networks capture complex temporal dependencies in emotional data.
Probabilistic graphical models, including Bayesian networks and hidden Markov models, infer underlying emotional states from observed behavioral patterns.

*The intersection of AI and healthcare:*
Exemplified by projects aiming to detect emotions in patients using supervised machine learning. The Jonathan Oheix dataset, comprising thousands of image files is used and six emotion classes are introduced. Supervised machine learning models, utilizing various feature engineering techniques, undergo a detailed comparison. A link between negative emotions and psychological health issues is established using the emotional guidance scale. The proposed methodology, achieving 98% accuracy with a multi-layer perceptron, holds promise for preventing extreme acts and addressing mental health concerns.

As the healthcare landscape increasingly integrates AI, the fusion of emotion recognition and patient care presents a promising avenue for improving diagnostic accuracy and treatment outcomes. The subsequent sections delve into the

literature review, theoretical background, and advancements in AI-driven emotion recognition, providing a comprehensive understanding of the current landscape and future directions in this transformative field.

This research makes a significant contribution to the field of emotion recognition in healthcare. By introducing an innovative approach to automatically detect patients' emotions in the context of specific diseases through facial recognition methods, the study bridges the gap between emotional health and medical conditions.

Furthermore, the emphasis on interdisciplinary collaboration and standardized evaluation metrics highlights the commitment to robust research methodologies. The incorporation of theoretical discussions on embodied cognition, social signal processing, and reinforcement learning frameworks adds depth to the exploration of adaptive emotional intelligence in AI systems. The application of probabilistic graphical models, including Bayesian networks and hidden Markov models, underscores the commitment to understanding and inferring complex emotional states from observed behavioural patterns. This comprehensive approach positions the research at the forefront of advancing emotional intelligence applications, particularly in the healthcare domain

## II. RELATED WORK

The paper **[1]** primary objective is to examine recent advancements in automatic facial emotion recognition (FER) through the lens of deep learning. The focus extends beyond a mere enumeration of contributions to delving into the nuances of architectural intricacies and the databases instrumental in driving progress. A meticulous comparison of proposed methods and their achieved results provides valuable insights, offering a comprehensive view of the current state-of-the-art in FER.

The research paper **[2]** focuses on the automatic detection of patients' emotions within healthcare data, employing supervised machine-learning approaches. The authors introduce EmoHD, a novel dataset comprising 4,202 text samples spanning eight disease classes and six emotion classes. Drawing from diverse online resources, this dataset forms the foundation for the evaluation of six different supervised machine learning models, each based on distinct feature engineering techniques.

The paper **[3]** meticulously explores FER datasets used for evaluation metrics, drawing comparisons with benchmark results. By presenting a detailed analysis of achievements, current methodologies, and potential challenges, the review offers a holistic perspective on the state-of-the-art in FER. Importantly, it serves as a guidebook for young researchers entering the FER domain, providing foundational knowledge and insights into both traditional ML and cutting-edge DL methods. The paper **[3]** offers a roadmap for future research directions, thereby contributing to the ongoing advancement of FER methodologies.

The review paper **[4]** sheds light on the predominant utilization of TED in health, particularly in detecting conditions such as depression, and suicidal ideation, and assessing the mental status of patients with asthma, Alzheimer's disease, cancer, and diabetes. Data sources encompass social media, healthcare services, and counselling centres. Notably, approximately 44% of the research in this domain pertains to COVID-19, exploring the emotional responses of the public and the public health impact of vaccinations.

The study paper **[5]** addresses the emotional reactions of patients engaging with remote healthcare services, recognizing their potential significance for healthcare practitioners. To explore these emotional responses, the authors developed an artificial intelligence-based classification system. This system employs metaheuristic feature selection and machine learning classification to detect emotions from input data. The proposed model undergoes a series of steps, encompassing preprocessing, feature selection, and classification. Simulations are conducted to evaluate the model's efficacy across various features present in a dataset. The results of the simulation highlight the effectiveness of the proposed model in classifying emotions from input datasets, surpassing the performance of existing methods.

The research paper **[6]** begins by comparing the performance of various neural network algorithms associated with deep learning. Subsequently, an improved Convolutional Neural Network-Bi-directional Long Short-Term Memory (CNN-BiLSTM) algorithm is proposed. To validate the efficacy of this algorithm, a simulation experiment is conducted. The experimental results reveal the notable performance of the CNN-BiLSTM algorithm, with an accuracy reported at 98.75%. This accuracy surpasses other algorithms by at least 3.15%. Additionally, the recall is at least 7.13% higher than that of other algorithms, and the recognition rate consistently exceeds 90%.

The paper **[7]** highlights the transformative potential of such advancements in technology, envisioning the development of smart healthcare centres capable of proactive intervention based on emotional cues. The ability to discern emotions becomes a cornerstone in human-machine interaction, reflecting a fascinating theme in the research. Various strategies are explored to teach machines how to predict emotions, with a particular focus on recent advancements in employing neural networks for emotion recognition.

## III. MATERIALS AND METHODS

The system architecture overview serves as a foundational narrative, elucidating the intricacies of our project's structural design. Within this comprehensive exposition, we meticulously unravel the diverse components embedded in the architecture, shedding light on their roles and interconnections. A visual aid in the form of a flowchart diagram enhances accessibility, offering stakeholders an intuitive understanding of the intricate processes. and

contextualizing their pivotal contributions to the overall system efficiency.

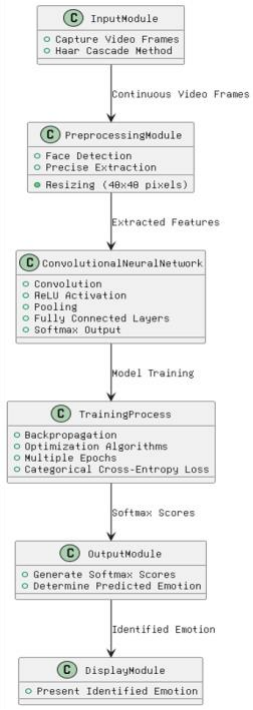


***Fig 1.0: Model Architecture***

Input Module: Capturing Real-time Video Frames
The system's functionality commences with the Input Module, responsible for capturing real-time video frames from a webcam feed. This continuous stream of frames serves as the primary input for subsequent processing. To identify facial regions within each frame, the Haar Cascade Method is employed, ensuring efficient and accurate face detection in the dynamic video stream.

Preprocessing Module: Enhancing Input for Analysis
Building on successful face detection, the Preprocessing Module comes into play. The identified facial region undergoes precise extraction, isolating the relevant features for emotion analysis. Subsequently, the input is standardized through resizing, ensuring a consistent size of 48x48 pixels. This preprocessing step establishes uniformity, optimizing the input for further analysis within the system.

Convolutional Neural Network (CNN): Extracting Hierarchical Features
The core of the architecture resides in the Convolutional Neural Network (CNN). This deep learning powerhouse is meticulously designed to extract intricate hierarchical features from the preprocessed facial images. Comprising layers for convolution, activation (ReLU), pooling, and fully connected neurons, the CNN transforms facial features into meaningful representations. The final output layer produces SoftMax scores, indicating the probabilities of each of the seven emotion classes.

Training Process: Refining Model Parameters
The CNN undergoes a rigorous training process, during which its parameters are iteratively adjusted through backpropagation and optimization algorithms. Multiple epochs of training refine the model's weights, enhancing its ability to accurately predict emotions. The training process is guided by the categorical cross-entropy loss function, minimizing the disparity between predicted and actual labels.

Output Module: Determining Predicted Emotion
Upon processing each video frame, the CNN generates SoftMax scores, assigning probabilities to each of the seven emotion classes. The emotion with the highest SoftMax score is determined as the predicted emotion. This step encapsulates the essence of the model's decision-making process based on the learned features.

Display Module: Communicating Results
The final output, representing the identified emotion, is presented through the Display Module. This tangible output is visually communicated on the screen, providing a human-readable interpretation of the model's prediction. The display serves as a crucial interface for users to understand and engage with the real-time emotion classification results.

## DATASET DETAILS

The dataset utilized for emotion classification is FER-2013, presented at the International Conference on Machine Learning (ICML). It encompasses grayscale face images, each standardized to 48x48 pixels, categorizing emotions into seven distinct classes: angry, disgusted, fearful, happy, neutral, sad, and surprised.

*Total Number of Images:*

In total, the dataset comprises 35,887 grayscale images, providing a diverse range of facial expressions for comprehensive training and evaluation.

*Class Distribution:*

| **Class Distribution** | | |
|---|---|---|
| | ***Number of Images*** | ***Percentage*** |
| Angry | 4,590 | 12.79 |
| Disgust | 547 | 1.52 |
| Fear | 5,325 | 14.83 |
| Happy | 8,790 | 24.49 |
| Neutral | 6,029 | 16.79 |
| Sad | 5,108 | 14.23 |
| Surprised | 5,498 | 16.15 |

***Table 1.0: Class distribution***

*Data Cleaning:*

Preprocessing of the dataset involves the extraction of facial regions using the Haar Cascade Method. Additionally, images are resized to a uniform 48x48 pixel resolution, ensuring consistency, and facilitating effective analysis.

*Image Dimensions:*

All images within the dataset are grayscale and standardized to the dimensions of 48x48 pixels, promoting uniformity in the dataset.

*Data Split:*

| Data Split | | |
|---|---|---|
| Types of Data | ***Number of Images*** | ***Percentage*** |
| Training Data | 25,210 | 70.24 |
| Testing Data | 5,984 | 16.67 |
| Validation Data | 4,683 | 13.04 |

***Table 2.0: Data Split***

The dataset is strategically divided into training and validation sets, which are crucial for training the model and assessing its performance accurately. This information is utilized for dataset preparation, emphasizing the organization of images into specific directories based on emotion labels. The dataset's structure, along with class distribution, sets the foundation for training a robust emotion classification model.

## PROPOSED APPROACH

Input Module - Capturing Real-Time Video Frames:
The process initiates with the Input Module, capturing continuous video frames from the webcam feed. This involves the utilization of the Haar Cascade Method for efficient and accurate face detection in each frame. The identified facial regions serve as the primary input for subsequent processing.

Preprocessing Module - Enhancing Input for Analysis:
Following successful face detection, the Preprocessing Module enhances the input for detailed analysis. The identified facial region undergoes precise extraction to isolate relevant features crucial for emotion analysis. To ensure uniformity and comparability, standardization is applied by resizing the extracted region to 48x48 pixels.

Convolutional Neural Network (CNN) - Extracting Hierarchical Features:
The core of the system is a deep Convolutional Neural Network designed for extracting intricate hierarchical features from preprocessed facial images. The CNN architecture comprises convolutional layers with ReLU activation, max-pooling layers, and fully connected neurons. The final output layer produces SoftMax scores, indicating the probabilities of each of the seven emotion classes.

Training Process - Refining Model Parameters:
The Convolutional Neural Network (CNN) for emotion classification is trained with specific parameters to optimize its learning and performance. The following training parameters are utilized during the training process:

- Learning Rate: 0.0001
  - The learning rate determines the step size at each iteration while moving toward a minimum of the loss function. A lower learning rate is chosen to ensure a more refined adjustment of the model weights.
- Decay: 1e-6
  - Decay is applied to the learning rate, controlling its reduction over time. A decay of 1e-6 is implemented to gradually reduce the learning rate during training, which can contribute to more stable convergence.
- Batch Size: 64
  - The batch size defines the number of training samples utilized in one iteration. A batch size of 64 is chosen to balance computational efficiency and model convergence. It influences the number of samples processed before updating the model's internal parameters.
- Number of Epochs: 15
  - An epoch is a single pass through the entire training dataset. Training for 15 epochs means the model undergoes 15 complete iterations over the entire dataset. This parameter is selected based on experimentation and consideration of convergence stability.

Output Module - Determining Predicted Emotion:
Upon processing each video frame, the CNN generates SoftMax scores, assigning probabilities to each of the seven emotion classes. The emotion with the highest SoftMax score is determined as the predicted emotion, embodying the essence of the model's decision-making process based on the learned features.

Display Module - Communicating Results:
The final output, representing the identified emotion, is presented through the Display Module. This tangible output is visually communicated on the screen, providing a human-readable interpretation of the model's prediction. The display serves as a crucial interface for users to understand and engage with real-time emotion classification results.

During execution, the CNN generates SoftMax scores for each emotion class, determining the predicted emotion. The final Display Module communicates these results visually on the screen, providing real-time insights into the recognized emotions. This holistic execution flow ensures accurate and instantaneous emotion classification, making it a robust system for human-machine interaction in various applications.

## IV. RESULTS AND DISCUSSIONS

The CNN architecture used in this experiment is tailored specifically for emotion recognition with four convolutional layers followed by max-pooling, batch normalization, activation functions, and dropout layers.
This customized architecture allows the model to focus on learning features relevant to the task at hand.

Experimental Analysis:
The below image depicts the Confusion Matrix obtained from the evaluation of the Convolutional Neural Network (CNN) model for emotion recognition revealing intriguing insights into its performance across seven distinct emotional categories. The model demonstrates notable proficiency in accurately predicting positive emotions, particularly

happiness and surprise, as indicated by the high precision and recall scores in these categories. This success implies that CNN excels in capturing subtle facial features associated with joy and surprise, showcasing its ability to discern positive emotional states.

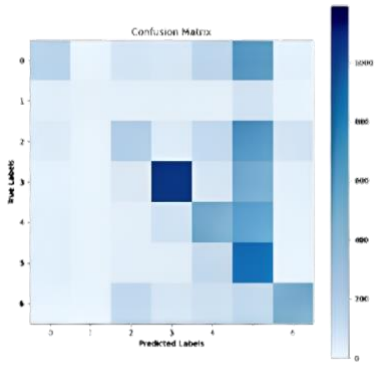


***Fig 2.0: Confusion Matrix***

Moreover, the model showcases commendable performance in correctly identifying neutral expressions, highlighting its versatility in recognizing subtle nuances in facial features that denote emotional neutrality. While achieving high precision and recall for positive emotions, the model also demonstrates a balanced capability in accurately predicting neutral expressions.

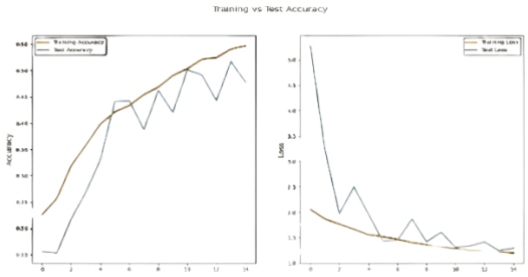


***Fig 3.0:Train vs Validation Loss and Train vs Validation Accuracy Graphs***

The model was trained for 15 epochs with a desired number of steps per epoch set to 50. The training vs. test accuracy and loss plots indicate that the model achieved decent accuracy on the training set, but the test set accuracy plateaus early, suggesting potential overfitting. This could be addressed by further regularization or reducing model complexity

*Classification Report:*

| Class | ***Precision*** | ***Recall*** | ***F1 Score*** | ***Support*** |
|---|---|---|---|---|
| Angry | 0.64 | 0.15 | 0.25 | 960 |
| Disgust | 0.86 | 0.05 | 0.10 | 111 |
| Fear | 0.41 | 0.16 | 0.23 | 1018 |
| Happy | 0.84 | 0.65 | 0.74 | 1825 |

***Table 3.0: Classification Report***

The classification report provides insights into the model's performance for each emotion class. The precision, recall, and F1-score metrics give a comprehensive view. The model excels in predicting 'happy' expressions with high precision and recall, while it struggles with 'disgust' expressions, as indicated by lower precision, and recall values. The weighted average F1-score is 0.47, indicating room for improvement.

## V. CONCLUSION

In conclusion, this paper introduces a novel approach to enhancing emotional well-being in healthcare by employing deep convolutional neural networks (CNNs) for facial emotion classification. The proposed system utilizes real-time video frames from webcam feeds and incorporates a multi-layered CNN architecture for extracting hierarchical features from preprocessed facial images. The model is trained on the FER-2013 dataset, encompassing seven emotion classes. The evaluation of CNN reveals notable proficiency in accurately predicting positive emotions, such as happiness and surprise, and demonstrates versatility in recognizing neutral expressions.

The presented approach distinguishes itself from other deep learning models in healthcare by focusing specifically on real-time emotion recognition in patient care settings. While existing literature often explores emotion detection in diverse domains, including textual and multimodal data, this research uniquely addresses the immediate applicability of emotion recognition in healthcare scenarios. The integration of real-time video analysis contributes to the dynamic nature of patient care, offering insights into emotional states that can inform personalized treatment approaches.

Compared to other deep learning models in healthcare, the proposed system excels in capturing subtle facial features associated with positive emotions, showcasing its potential to understand and respond to patients' emotional well-being. The model's proficiency in recognizing neutral expressions adds a layer of sensitivity to nuanced emotional states.

The benefits of the proposed approach include its potential to improve diagnostic accuracy and treatment outcomes by incorporating emotional context into patient care. Emotion recognition can provide valuable information for healthcare practitioners, enabling them to tailor interventions based on patients' emotional well-being. Moreover, the real-time nature of the system allows for immediate feedback and responsiveness, enhancing the quality of patient-provider interactions.

Despite the promising results, there is room for improvement, particularly in addressing challenges related to the recognition of certain emotions, such as 'disgust.' Further research and refinement of the model architecture, regularization techniques, and dataset augmentation strategies can contribute to enhanced performance.

In the ever-evolving landscape of healthcare and AI integration, this research contributes to the exploration of emotional intelligence as a valuable component of patient

care. The proposed system offers a stepping stone towards more empathetic and personalized healthcare practices, where technology complements human understanding and response to emotional well-being.


## ACKNOWLEDGMENT

I would like to express my sincere gratitude to my professor Andrew J for his invaluable mentorship throughout the research process. His expertise in Deep Learning has been instrumental in shaping the direction and outcomes of this study. His guidance, insightful discussions, and unwavering support have significantly enriched my understanding of the subject.

I am also thankful to the Manipal Institute of Technology for providing the resources and environment conducive to research. The academic atmosphere at the institute has been crucial in fostering intellectual curiosity and facilitating the exploration of innovative ideas.


## REFERENCES


**[1]** Facial Emotion Recognition Using Deep Learning: Review and Insights

**[2]** Automatic Emotion Recognition in Healthcare Data Using Supervised Machine Learning

**[3]** "Facial Emotion Recognition Using Conventional Machine Learning and Deep Learning Methods: Current Achievements, Analysis, and Remaining Challenges

**[4]** "Textual Emotion Detection in Health: Advances and Applications"

**[5]** "An Artificial Intelligence-Based Reactive Health Care System for Emotion Detections"

**[6]** "Deep Learning-Based Emotion Recognition and Visualization of Figural Representation"

**[7]** "Patient Monitoring Using Emotion Recognition"

**[8]** Wikipedia Contributors, "Driver drowsiness detection," Wikipedia,Oct.15,2019.

**[9]** "Drivers Drowsiness Detection using Image Processing Techniques,"

**[10]** A.-C. Phan, T.-N. Trieu, and T.-C. Phan, "Driver drowsiness detection and smart alerting using deep learning and IoT," Internet of Things, vol. 22, p. 100705, Jul. 2023.

**[11]** Y.-D. Zhang and Arun Kumar Sangaiah, Cognitive Systems and Signal Processing in Image Processing. AcademicPress,2021.

**[12]** Y. Chen, Y. Yu, and J.-M. Odobez, "Head Nod Detection from a Full 3D Model." Accessed: Nov. 02, 2023. [Online].

**[13]** S. Li and W. Deng, "Deep facial expression recognition: A survey", 2018.

**[14]** E. Correa, A. Jonker, M. Ozo and R. Stolk, "Emotion recognition using deep convolutional neural networks", 2016.

**[15]** B. Fasel and J. Luettin, "Automatic facial expression analysis: a survey", Pattern Recognition, vol. 36, no. 1, pp. 259-275, 2003.

**[16]** Y. Lv, Z. Feng and C. Xu, "Facial expression recognition via deep learning", *Smart Computing (SMARTCOMP) 2014 International Conference on*, pp. 303-308, 2014.

**[17]** Balakrishnan V, Kaur W. 2019. String-based multinomial Naïve Bayes for emotion detection among Facebook diabetes community.

**[18]** S.Basu, J.Chakraborty, and M.Aftabuddin, "Emotion recognition from speech using a convolutional neural network with recurrent neural network architecture," in20172ndInternationalConferenceonCommunication and Electronics Systems (ICCES). IEEE,2017, pp.333–336.

**[19]** D. K. Jain, P. Shamsolmoali, and. Sehdev, "Extended deep neural network for facial emotion recognition," Pattern Recognition Letters, vol.120, pp. 69–74,2019.

**[20]** W. Lim, D. Jang, and T. Lee, "Speech emotion recognition using convolutional and recurrent neural networks," in 2016 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA). IEEE,2016, pp.1–4.

**[21]** N. Jain, S. Kumar, A. Kumar, P. Shamsolmoali, and M. Zareapoor, "Hybrid deep neural networks for face emotion recognition," Pattern Recognition Letters, vol.115, pp. 101–106, 2018.