Human Emotion Classification using Facial Expressions and CNN Models

Alishana Thorat
Shivajirao S. Jondhale College of Engineering
Mumbai, Maharashtra, India

Kanishka Panpatil
Shivajirao S. Jondhale College of Engineering
Mumbai, Maharashtra, India

Selvavani Mathavan
Shivajirao S. Jondhale College of Engineering
Mumbai, Maharashtra, India

Sneha Kushwaha
Shivajirao S. Jondhale College of Engineering
Mumbai, Maharashtra, India

Savita Sangam, PhD
Shivajirao S. Jondhale College of Engineering.
Mumbai, Maharashtra, India

ABSTRACT
This project aims to teach machines how to recognize human emotions by analysing facial expressions. Using deep learning and the pre-trained VGG16 model, our system identifies six key emotions: happiness, sadness, anger, fear, surprise, and disgust. This system applies transfer learning, data augmentation, and class balancing to improve accuracy and performance. The result is a reliable emotion detection model that can support real-world applications like mental health monitoring, smart assistants, and interactive learning tools.

General Terms
Emotion detection, classification, Machine Learning, CNN , VGG16 Model , data augmentation

Keywords
Emotion detection, facial expressions, deep learning, VGG16, transmission learning, convolutional neural network (CNN), data augmentation, class imbalance.

1. INTRODUCTION
Have you ever noticed how important people rely on facial expressions to understand how someone is feeling — even without saying a word? Human faces can express joy, sadness, wrathfulness, fear, surprise, or even nausea in just a second. These subtle emotional cues play a huge part in how the project interact with others every day.

With today's rapid advancements in artificial intelligence, especially deep learning, the project now have the capability to educate machines to recognize and interpret these feelings just by looking at a face. This project explores one such application— recognizing human emotions from facial images using deep learning.

The goal is to develop a system that can directly identify feelings like happiness, sadness, wrathfulness, fear, surprise, and nausea from an image of a person's face. To do this, the project uses a powerful technique called transfer learning, applying the pre-trained VGG16 model — a type of convolutional neural network (CNN) — which has already learned how to identify general patterns in images.

Rather than training a new model from scratch, the project fine-tune this pre-trained network to concentrate on features specific to human facial expressions. This not only saves time and computing power but also improves accuracy, especially when working with limited data.

The project also applies techniques like data augmentation to increase diversity in training images and use class balancing to make sure all emotions are learned equally.

The main purpose of this project is not just to make a model that works well on paper — it's to move a step closer to making technology more emotionally intelligent. Imagine virtual assistants, mental health tools, or learning platforms that can understand how users feel in real time and respond accordingly. That is the kind of future this project is aiming for.

2. LITERATURE SURVEY
The field of facial emotion recognition (FER) has gained significant momentum in recent years, particularly with the rise of deep learning and transfer learning techniques. A wide range of models has been proposed by researchers aiming to enhance both the accuracy and efficiency of emotion detection systems.

F. Ghaffar [1] introduced a CNN-based approach that showed strong performance on benchmark datasets, showcasing the strength of convolutional neural networks in extracting spatial facial features. Building on this foundation, A. V. Savchenko

[2] Developed a lightweight multi-task learning model capable of identifying facial expressions alongside facial attributes, leveraging shared feature learning to improve overall outcomes. Similarly, A. Saroop et al. [3] implemented a multi- task deep learning framework, proving that integrating multiple facial analysis tasks can strengthen generalization and resilience of the system.

Exploring alternatives to traditional CNNs, A. Chaudhari et al.

[4] proposed ViTFER, a Vision Transformer-based architecture. Their work highlighted the capability of transformers to capture long-range facial dependencies, an essential factor for precise emotion classification.

Focusing on compact models for real-time use, S. Kaur and N. Kulkarni [5] created FERFM using the MobileNetV2 backbone — a resource-efficient solution tailored for applications with limited processing power. Meanwhile, Zhang [6] enhanced the well-known VGG16 model for emotion detection, showing that older architectures can still perform competitively when properly fine-tuned.

Hybrid techniques have also been explored. A. R. Angeline and A. N. Alice [7] proposed a combination of DenseNet-161 and a feature stabilization method, building a robust system that handles multimodal facial data effectively. S. Nathani [8] conducted an in-depth comparison of several transfer learning strategies, focusing on customized CNN and VGG16 variants, and emphasized the importance of model tuning for specific datasets.

Expanding on these innovations, Y. El Boudouris and A. Bohi

[9] Introduced EmoNeXt, a refined version of the ConvNeXt architecture, demonstrating new possibilities in deep learning- based emotion recognition. Broader overviews in the field have also contributed valuable insights. J. Doe and J. Smith

[10] Presented a comprehensive review of deep learning techniques used in FER, identifying key limitations and areas in need of further research. E. Davis and M. Johnson [11] concentrated on lightweight, real-time FER systems optimized for low-power devices. Finally, H. Tanaka and L. Wei [12] tackled the issue of cross-cultural emotion detection, suggesting strategies to help models generalize across diverse populations

3. SYSTEM ARCHITECTURE

In the figure 1 shows flow of system architecture diagram of proposed system and implementation steps.

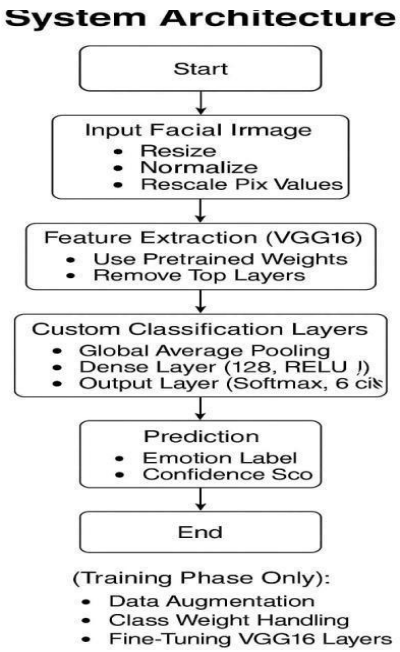


Figure 1: System Architecture

Start- The system commences when an image input is obtained from a predetermined source. This marks the initiation of the emotion recognition pipeline.

Input Facial Image-The facial image is obtained or loaded and must clearly display the subject's face. To get it ready for processing, image is resized to 224×224 pixels, normalized, and the pixel values are adjusted. After these steps, the project ensures that the input data remains consistent, minimizing the impact of lighting differences or variations in image quality.

Image Preprocessing- Data augmentation generates new variations of the input images to enhance the dataset's diversity and enhance the model's ability to generalize. The model's learning process is modified to handle imbalanced emotion categories by adjusting the class weight. The process of fine-tuning involves unfreezing a few of vgg16's higher layers, enabling the model to adapt more efficiently to the specific task of emotion recognition. Normalizing input images is essential for model precision. Preprocessing involves resizing the image to match the input shape of the vgg16 model, normalizing pixel values to be between 0 and 1, and applying techniques such as rescaling. These steps aid the model in comprehending facial features more accurately.

Feature Extraction using VGG16- The pre-processed image is fed into the vgg16 model, which has been trained on a vast dataset (Imagenet). The outermost layers are discarded, and only the convolutional base remains. These base extracts provide comprehensive information about various features, including edges, patterns, and facial structures, which are crucial for emotion recognition.

Custom Classification Layers- The features uprooted by VGG16 are fed into new bracket layers. A Global Average Pooling subcaste reduces the point chart into a manageable vector. This is followed by a thick subcaste with 128 neurons and ReLU activation to learn emotion-specific patterns. Eventually, a SoftMax affair subcaste with six neurons (one for each emotion) generates the probability for each emotional class.

Prediction- The model selects the emotion class with the highest probability as the prediction. Alongside this predicted label, a confidence score is generated to indicate how certain the model is about its prediction. These results are either displayed to the user or used by another system component.

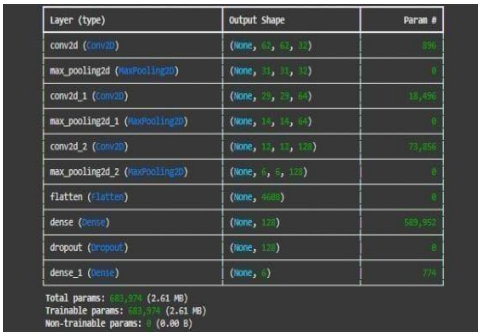
End- Once the prediction and confidence score are produced, the process ends. The results can be saved or passed on for further processing, such as behavior analysis or user feedback systems.

4. METHODOLOGY

This project entails developing a powerful facial emotion recognition system using deep learning techniques, with a particular emphasis on transfer learning and convolutional neural networks. The methodology is divided into crucial stages, such as data preparation, model development, training, evaluation, and performance assessment.

4.1 CNN model-convolutional neural networks-

CNNs are extensively employed in image processing tasks because they possess the capability to automatically identify and extract significant spatial patterns. In this project, CNN played a vital role in recognizing important elements in facial images, such as contours, textures, and shapes — all of which are essential for comprehending emotional expressions.



Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 62, 62, 32)	896
max_pooling2d (MaxPooling2D)	(None, 31, 31, 32)	0
conv2d_1 (Conv2D)	(None, 29, 29, 64)	18,496
max_pooling2d_1 (MaxPooling2D)	(None, 14, 14, 64)	0
conv2d_2 (Conv2D)	(None, 12, 12, 128)	73,856
max_pooling2d_2 (MaxPooling2D)	(None, 6, 6, 128)	0
flatten (Flatten)	(None, 4608)	0
dense (Dense)	(None, 128)	589,952
dropout (Dropout)	(None, 128)	0
dense_1 (Dense)	(None, 6)	774

Total params: 683,974 (2.61 MB)
Trainable params: 683,974 (2.61 MB)
Non-trainable params: 0 (0.00 B)

Figure 2: CNN Architecture

4.2 Transfer Learning using VGG16 - Rather than starting from scrape, the project has employed there-trained vgg16 model that had been trained on the ImageNet dataset. This project stripped down the external layers of the model and employed the convolutional base for rooting features. By espousing this strategy, they were suitable to save time, minimize computational charges, and achieve remarkable results as there-trained layers were formerly complete at relating pivotal visual factors applicable to emotion recognition.

4.3 Constructing the convolutional neural network- Constructed a technical classifier on top of the vgg16 model. This included - a global normal pooling subcaste to reduce point confines - a SoftMax affair subcaste with six units to prognosticate the probability for each of the six feelings

4.4 Data Augmentation- In order to improve model performance and prevent overfitting, the project has utilized data augmentation techniques. This required generating modified versions of images by applying random transformations such as flipping horizontally, rotating, zooming and shearing. These artificial alterations mimic real-life variations in facial expressions and enhance the model's ability to adapt to different situations.

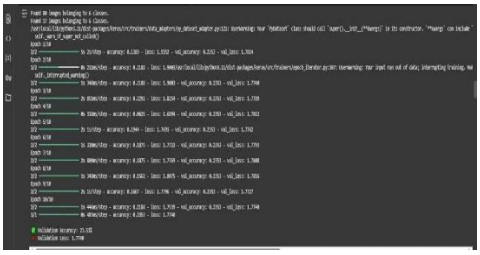


Figure 3: Data Augmentation

4.5 Training the model- The model was trained using the Adam optimizer and the categorical cross-entropy loss function. Imbalanced sentiment classes were addressed using class weighting. In addition to early stopping, enforcing a system to help overfitting during the training phase.

4.6 Evaluation and Prediction- After the training phase, the model was assessed on data that had not been used during the training process. It prognosticated the most probable emotion for each image and also handed a confidence score for its vaticination. This allowed us to assess the model's capability To rightly identify mortal feelings only one address is needed, center all address text. For two addresses, use two centered tabs, and so on. For three authors, you may have to improvise.

Figure 4: Evaluation and Prediction

4.7 Evaluating and visualizing fine-tuned model performance

In order to comprehend the model's learning process more comprehensively, generated two performance graphs: one illustrating accuracy and another depicting
the loss over time.

4.8 Graph 1: accuracy over epochs (fine-tuned).

4.8.1 The X-axis (epochs): In the CNN model represents the number of times the entire training dataset has gone through the model. The range of values in the graph spans from 0 to 9, signifying 10 distinct epochs of training.

4.8.2 The Y-axis (accuracy): Indicates the model's ability to accurately predict the correct emotion. Higher values indicate better performance.

4.8.3 The blue line - Train accuracy (fine-tuned): This line measures the model's performance on the training data. The process begins around 0.6 and experiences fluctuations, ultimately reaching a peak above 0.7, indicating progress over time.

4.8.4 The orange line - validation accuracy (fine-tuned): This line demonstrates the model's ability to accurately predict outcomes on data it has not seen before.
The model's performance remains constantly high at around 0.35 – 0.4, indicating an implicit issue of overfitting — the model excels in training data but struggles with confirmation data.

4.8.5 Conclusion: Although the model shows substantial enhancement on the training data, the confirmation delicacy remains unchanged. This is a strong suggestion of overfitting, suggesting that the model performs exceptionally well on known data but struggles to apply its knowledge to new and strange cases.

4.9 Graph 2: loss over epochs (fine-tuned).

4.9.1 The X-axis (epochs): Same as before, it represents the number of training cycles (epochs).

4.9.2 The Y-axis (loss): Loss is a measure of error how much the model's predictions differ from the actual labels. Reduced loss indicates improved functionality.

4.9.3 The blue line: The train loss, which measures the model's performance on training data, shows a consistent decrease with minor fluctuations, suggesting that the model is effectively learning from the data.

4.9.4 The orange line: The validation loss for the orange line remains consistently high and stable, indicating overfitting — the model struggles to perform well on validation data.

4.9.5 Conclusion: Although there was a decrease in training loss, the consistent and high validation loss suggests that the

model has not learned features that can be generalized to various scenarios. It implies that by incorporating supplementary data or refining regularization techniques, the model's capacity to generalize to unseen data can be improved.

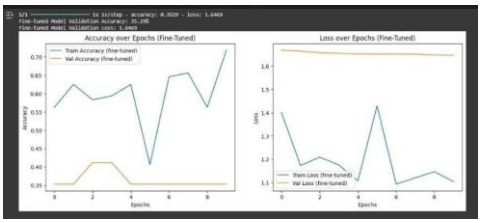


Figure 5: Evaluating and Visualizing Fine-Tuned Model Performance

5. WORKING OF PROJECT

Step 1: Data medication the original step involves arranging the dataset in a manner suitable for training the model. This dataset comprises facial images, each labelled with one of six feelings wrathfulness, fear, happiness, sadness, surprise, or nausea. All images are resized to 224 × 224 pixels to guarantee comity with the vgg16 model. The dataset is latterly divided into three subsets training, confirmation, and testing. To enhance the model's robustness and alleviate overfitting due to limited data, employed data addition ways like gyration, drone, shear, vertical flips, and shifts likewise, all pixel values are acclimated to fall within the range of 0 to 1.

Step 2: Import vgg16 architecture next, we bring in the vgg16 model, a deep convolutional neural network that has been pre- trained on the vast ImageNet dataset. Because the original layers of vgg16 were specifically created for ImageNet's 1000- class type, we count those layers. We keep the convolutional base, which is superb at landing broad visual features analogous as edges and textures firstly, the base model's layers are concrete to retain the learned features and expedite the training process.

Step 3: Add substantiated layers to make the model suitable for emotion type, we mound fresh layers on top of the vgg16 base. These include,

• A global normal pooling caste to flatten the point maps.

• A densely connected neural network with 128 neurons and the ReLU activation function to prize high- position emotional features.

• A hustler caste to erratically kill neurons during training and reduce overfitting.

• And a SoftMax affair caste with six bumps each representing an emotion — to affair class chances

Step 4: Prepare the model former to training, we collect the model by setting the demanded parameters. Given that this is a multi- class type task, we employ categorical cross entropy as the loss function. The Adam optimizer is chosen for its capability to adapt and learn during the optimization process. Also, we assign class weights to regard for any disagreement in the number of images per emotion class. This ensures that passions that are constantly overlooked or underrepresented are given the attention they earn during the training process. The pivotal metric for assessing the model's performance is perfection.

Step 5: Train the Model The training process involves feeding the set dataset into the model. While training, the model's performance on the confirmation set is covered using early stopping which halts training if performance stops perfecting, helping help overfitting. After the original training phase, we dissolve some layers of the VGG16 base to fine-tune them. This allows the model to acclimatize pre-trained features more specifically to the emotion dataset. A lower literacy rate is used during fine- tuning to make precise adaptations to the model's weights.

Step 6: Emotion vaticination in the final step, the trained model is used to classify feelings from new images. Each input image is resized and pre-processed just like the training data. The model also processes the image and labours probability scores for each emotion class. The emotion with the loftiest score is named as the vaticination, along with a confidence value. This system is suitable for both real- time operations (e.g., live webcam analysis) and static offline emotion discovery.

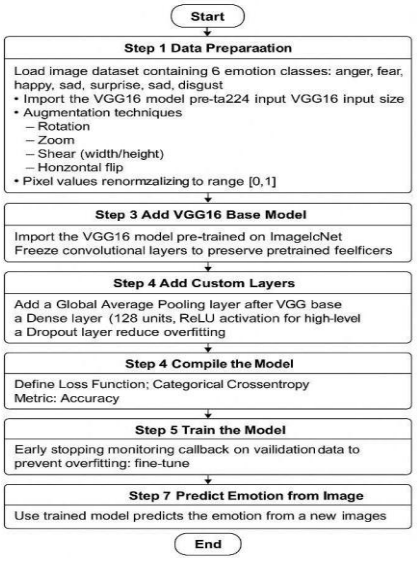


Figure 6: Working of Project

6. FUTURE SCOPE

The future development of this facial emotion recognition system using CNNs and deep learning offers promising possibilities across various real-world applications. Enhancing the model for real-time emotion detection can make it suitable for integration into virtual assistants, online learning environments, and mental health monitoring tools. Future improvements could focus on increasing the model's robustness against variations in lighting, facial angles, and partial occlusions. Additionally, combining facial data with other input types like voice and text could lead to more accurate and context-aware emotion analysis. The adoption of advanced architectures, such as vision transformers or attention-based models, may further boost performance. Expanding the training data to include diverse cultural and demographic representations would also allow the model to perform reliably across different user groups, making it more inclusive and adaptable for global deployment.

7. CONCLUSION

In this project, we successfully designed and implemented a facial emotion recognition system using deep learning techniques, specifically leveraging the VGG16 architecture through transfer learning. This approach allowed us to reduce training time while achieving effective emotion classification results. The model was trained on a dataset containing six

fundamental emotions: anger, fear, happiness, sadness, surprise, and disgust. To enhance its performance and generalization, we applied various data augmentation techniques and addressed class imbalance by assigning appropriate class weight.

9. ACKNOWLEDGEMENT

The authors would like to extend their sincere gratitude to Dr. Savita Sangam, Head of the Department, for her unwavering guidance, continuous support, and insightful encouragement throughout the course of this project titled "Human Image Classification for Sentiment Analysis."

They also express their heartfelt thanks to *Dr. P.R. Rodge*, Principal of the institute, for his visionary leadership and for providing the necessary resources and a supportive environment that greatly contributed to the successful completion of this work.

The consistent support and invaluable contributions of both Dr. Sangam and Dr. Rodge played a pivotal role in shaping the project and ensuring its timely and effective execution.

9. REFERENCES

[1] F Ghaffar, 'Facial Emotions Recognition using Convolutional Neural Net' arXiv preprint arXiv:2001.01456, Jan. 2020. [Online].

[2] A. V Savchenko, 'Facial expression and attributes recognition based on multi-task learning of lightweight neural networks' arXiv preprint arXiv: 2103.17107, Mar.2021. [Online]

[3] A Saroop, S. K. S Gupta, S. K. S Gupta, 'Facial Emotion Recognition: A multi-task approach using deep learning' arXiv preprint arXiv: 2110.15028, Oct 2021. [Online].

[4] A. Chaudhari, A. Agrawal and S Joshi 'ViTFER: Facial Emotion Recognition with Vision Transformers', Journal of Imaging 5 (4) (2022) 80. [Online]

[5] S Kaur and N Kulkarni, 'FERFM: An Enhanced Facial Emotion Recognition System Using Fine Tuned MobileNetV2 Architecture', IETE Journal of Research 10 (2023) 415-418. [Online]

[6] R Zhang, 'Facial emotion detection based on improved VGG-16', Applied and Computational Engineering 2 (3) (2023) 45-50. [Online].

[7] A R Angeline, A N Alice, 'Multimodal Human Facial Emotion Recognition Using DenseNet-161 and Image Feature Stabilization Algorithm' Traitement du Signal 39 (6) (2022)

[8] S Nathani, 'A Comparative Study of Transfer Learning for Emotion Recognition using CNN and Modified VGG16 Models' arXiv preprint

[9] El Boudouris, Y., & Bohi, A. (2025, January). EmoNeXt: An adapted ConvNeXt for facial emotion recognition. arXiv preprint arXiv:2501.08199. [Online]. Available: arXiv: 2501.08199.

[10] Doe, J., & Smith, J. (2023, March). Deep emotion recognition: A comprehensive review of current approaches and future directions. Journal

[11] Davis, E., & Johnson, M. (2024, June). Real-time facial emotion recognition using lightweight CNNs. Proceedings of the IEEE.