Computing and Informatics, Vol. 45, 2026, 105–126, doi: 10.31577/cai_2026_1_105

HUMAN EMOTION RECOGNITION BY FACIAL EXPRESSION USING MODIFIED EFFICIENTNET-B3 MODEL

Vivek Srivastava

Rajkiya Engineering College, Kannauj, India
e-mail: vivek@reck.ac.in

Raghuraj Singh Suryavanshi

Pranveer Singh Institute of Technology (PSIT), Kanpur, India
e-mail: additionaldirector@psit.ac.in

Ashutosh Pandey

Rajkiya Engineering College, Kannauj, India
e-mail: 2008390100019@reck.ac.in

Abstract. Human emotion plays a critical role in the purpose of communication, with the correct detection of emotion, one can analyze the human feeling without even asking, as it is non-verbal communication method. Human emotion through facial expression is one of the highlighted topics of study because of its wide range of application from robotics, security, artificial intelligence, marketing and health monitoring where human-computer interaction (HCI) is a key ingredient. To tackle the challenges that arise in accurate human facial emotion recognition, a modified version of the EfficientNet-B3 model is proposed, which shows promising results for emotion detection using the FER-2013 dataset. By leveraging the universal FER-2013 dataset, which contains more than 35 000 grey-scale human facial images, the proposed model aims to improve recognition performance. The modified architecture incorporates EfficientNet-B3 as the base model, with an additional batch normalization layer followed by a dense layer and an output layer. The model has been trained on the facial dataset while considering these crucial factors. Per-

formance evaluation of the model is conducted using confusion matrix, precision, recall, accuracy, and F1 score as performance metrics. Remarkably, the proposed model achieved an impressive accuracy of 92 % for training and 83.0 % for validation of the dataset. This implies that the proposed model yields highly accurate emotion recognition with a given dataset and can provide improvement in the overall efficiency of emotional recognition applications, particularly in human-to-machine interactions.

Keywords: EfficientNet-B3, CNN, augmentation, fine tuning, batch normalisation, oversampling

1 INTRODUCTION

Human emotion is a non-verbal mode of communication, which is an effective way of conveying inner feelings and mindset beyond words. Since emotions are universal, interpreting them helps bridge communication gaps universally [1]. Facial emotions, being common across the world, fill gaps that may arise due to linguistic, social, regional, and situational barriers [2]. Along with addressing these communication gaps, facial emotion detection has applications in mental health assessment and facilitates effective communication between humans and machines.

Accurate emotion detection has a wide range of applications, including health assessment, education, ethical considerations, psychology, and marketing. Inspired by the benefits and future scope of facial emotion detection, various approaches have been proposed to enhance emotion recognition, aiming to reduce communication gaps between humans and computers. These approaches enable computers to interact emotionally with humans by recognizing facial patterns associated with different emotions. Lightweight CNNs and deep learning algorithms have been proposed for this purpose [3, 4].

This work introduces a powerful and effective modified EfficientNet-B3 architecture based on CNN, pretrained on ImageNet for transfer learning. Additional custom layers are incorporated for fine-tuning and adaptation to achieve accurate facial emotion detection on the FER-2013 dataset. With its sophisticated architecture and versatility, the improved EfficientNet-B3 successfully tackles several challenges in facial expression recognition (FER). It can generalize across a range of facial expressions, lighting conditions, and background variations due to its enhanced feature extraction capabilities and scalable design. The fine-grained feature learning of the updated EfficientNet-B3 effectively captures subtle expressions, which are often challenging to detect because of minimal changes in facial landmarks. Furthermore, it is highly adaptive to real-world scenarios due to its resilience to head pose and viewpoint fluctuations, ensuring consistent expression identification even in non-frontal or tilted images.

Additionally, this updated architecture excels in addressing overfitting and class imbalance issues, particularly in small or unbalanced datasets. By leveraging transfer learning and pre-trained weights, the model generalizes well on FER datasets, reducing the need for extensive labeled data. Its computational efficiency makes it ideal for real-time applications where precision and speed are crucial, such as mobile-based FER systems or video surveillance. These enhancements highlight the modified EfficientNet-B3's capability to overcome the multifaceted challenges of FER tasks.

The following noteworthy advancements to the field of facial emotion recognition (FER) are introduced in this study:

Modified EfficientNet-B3 Architecture: An optimized version of EfficientNet-B3 for FER-2013 that has been improved using transfer learning and custom layers.

Improved Performance: On the FER-2013, we successfully addressed overfitting and class imbalance, achieving prominent training accuracy and validation accuracy.

Robust Feature Extraction: Improved capacity to pick up on nuanced facial expressions in a variety of settings, including different lighting and positions.

Real-World Applicability: Effective architecture that works well for real-time applications like video surveillance and mobile-based systems.

This work is structured to provide a comprehensive exploration of the proposed method and its outcomes. Initially, in Section 2, the literature review highlights existing techniques in facial emotion recognition (FER) and identifies the research gaps addressed in this study. In Section 3, the proposed work introduces the modified EfficientNet-B3 architecture, detailing its innovations and integration of transfer learning. The workflow of the proposed model outlines the step-by-step process, from data preparation to evaluation. Section 4, the simulation and results, presents dataset description and performance metrics such as accuracy and confusion matrices, while the result analysis compares the model's performance with state-of-the-art methods in Section 5. An ablation study validating the contributions of each component in the proposed approach is presented in Section 6. Finally, the study concludes with a summary of findings and outlines promising directions for future work in Section 7.

2 LITERATURE REVIEW

Facial emotion recognition is one of the challenging problems. Several approaches have been proposed to get better results in this particular problem. Different machine learning and deep learning approaches have been put forward which give detailed work and desirable results. Some of them are discussed below.

In [5], CNN model with attention mechanism has been proposed for FER. In this experiment, a CNN based gate unit mainly focuses on a specific part of the

face which identifies key features of mapping of facial patterns and information related to facial expression. The methodology mainly focuses on facial landmarks and their localization for facial emotion recognition. The work [6] proposed a framework based on Random Forest in videos for facial emotion recognition of dynamic and robust. This methodology involves the hybridization of a pairwise Random Forest (RF) classifier with spatio-temporal information. The method enables real-time implementation by learning local temporal patterns from multiple viewpoints, thereby improving facial emotion recognition (FER) in real-time video, even in unoptimized conditions. In [7], the authors proposed an approach for robust FER. This method consists of two steps. First, CNN-based learning is used to extract facial image features. Then, expression states are maintained through sequence learning, where temporal features are learned using an LSTM for final facial expression classification across different states. In 2022, research work [8] introduced a deep learning based CNN approach for real time emotion recognition through facial expression. This is a hybrid model created by adjusting two CNN features, a neural network and training parameters and employing facial land marking, grey-scaling, and, augmenting the facial dataset for variation and post-hyperparameter tuning. A Haar cascade model is used for real time facial emotion detection and FER very efficiently.

The approach mentioned in paper [9] introduced a method for facial emotion recognition by multi-layer feature fusion and Light weighted improved and basic unit residual network of a convolution neural network. This approach optimizes the retrained CNN model called MobileNet further which is constructed with the model. This framework achieves recognition accuracy for emotion detection. The mentioned approach [10] proposed new approach, which is based on VGGNet deep convolution neural network, the introduced model og VGGNet model have architecture including small convolution kernel of $3 * 3$ and also it includes pool kernel of $2 * 2$. In this experiment, the very initial fully connected layer is maintained without changes and a number of parameters is reduced in other layers. Further, before fully-connected layer dropout strategies were introduced, and multi-cropping of image dataset was done to reduce overfitting of a model, and at last, the Sofmax classifier is used for final classification and emotion detection.

Research work [11] proposed a Convolutional Neural Network (CNN)-based approach that includes two main steps. It begins with background removal to extract facial expressions using a CNN module called the Expression Vector (EV). The facial expression is then passed through multiple convolutional layers to detect patterns and features using various filters. The next step involves mapping facial landmarks, followed by feature extraction using hybrid, holistic, and template-based methods. Finally, a non-convolutional perceptron layer is used for classification. Song [11] proposed a new approach of emotion recognition by collaborating the emotion philosophy and machine learning algorithms. The forwarded approach includes feature extraction with the help of Gabor Feature extractor from a region of interest (ROI). For a detailed description, in the proceeding steps a deep separable convolution based channel attention network module is included for the complexity reduction,

minimizing and overfitting to refine the network structure. This step gives promising results on FER-2013 data-set.

The work in [12] presents a comparative study of different approaches for facial emotion recognition, including traditional machine learning models and deep learning methods. The traditional ML approaches involve various feature extraction techniques, followed by emotion classification using algorithms such as KNN, SVM, AdaBoost, and Random Forest. In the same manner deep learning model of hybrid type is proposed, although the deep learning based model gives promising results but takes time for different stages. Singh et al. [13] proposed deep convolution neural network based EfficientNet Architecture which is a modified version of EfficientNet in which features are extracted using filters, further data is augmented using pooling layer, further pooled output is flattened and processed by different classifiers like sigmoid and softmax. The study proposed a comparison of all versions of the EfficientNet architecture and investigated promising results of EfficientNetB0. This method was applied on different dataset and have achieved accurate results.

Gursesli et al. (2024) [14] proposed Custom Lightweight CNN-based Model (CLCM). Their method emphasized efficiency and robustness, making it a relevant benchmark for evaluating the proposed method.

The related work in [15] discussed a new methodology of the Expression Net architecture which is a two-step algorithm including convolution and fully connected layer. At first step the convolution layer is trained on Face Net to incorporate Expression Net with loss function. In this Expression Net Architecture, five convolution layers are proceeded with ReLU activation function and max-pooling layer. This model concludes by adding $1 * 1$ convolution layer for FaceNet and Expression-Net. This approach investigated appropriate layer for transfer learning for FER tasks. Different approaches and obtained recognition rates are summarized in Table 1. It shows that maximum 74.42 % recognition rate is achieved for FER-2013 dataset among the earlier discussed techniques.

Author	**Model Used**	**Dataset Used**	**Recognition Rate**
Pecoraro et al. [16]	ResNet34v2	FER-2013	74.42 %
Breuer and Kimmel [17]	AlexNet	FER-2013	72.10 %
Singh et al. [13]	EfficientNet	FER-2013	71.12 %
Zhou et al. [4]	CNN	FER-2013	67 %
Kim et al. [7]	CNN-LSTM	MMI, CASMEII	78.61 %, 60.98 %
Li et al. [5]	ACNN	RAF-DB	80.54 %

Table 1. Various existing works

3 PROPOSED WORK

The primary contribution of this work lies in enhancing the EfficientNet-B3 model by incorporating an additional normalization layer. This modification aims to improve the stability and performance of the model by ensuring that the feature maps

maintain a consistent scale throughout the training process. The added normalization layer effectively mitigates the vanishing or exploding gradient issues that can arise during backpropagation, especially in deeper networks. This enhancement not only refines the feature extraction capabilities of the model but also contributes to better generalization on the FER task. By integrating this layer, the proposed method achieves improved accuracy and robustness, as demonstrated in the experimental results. The novelty of this work lies in the integration of an additional normalization layer into the EfficientNet-B3 architecture, specifically tailored for the facial emotion recognition (FER) task. By leveraging this tailored normalization strategy, the proposed approach achieves superior accuracy and robustness compared to traditional EfficientNet-based models, as evidenced by comprehensive experimental results. This contribution advances the state-of-the-art in FER by demonstrating how architectural refinements in normalization can significantly impact performance in emotion recognition tasks. With the objective of accurate and effective human emotion detection and classification, the overall work gives a new approach or methodology which is based on modified version of the EfficientNet-B3 model. The rest of the section presents the workflow of the proposed methodology and model architecture.

3.1 Workflow of the Proposed Model

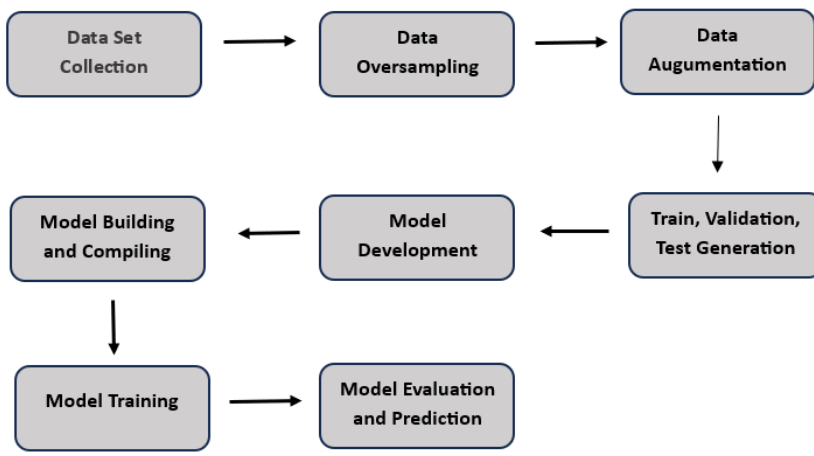


Figure 1. Workflow diagram

Workflow of the proposed work is described in Figure 1. This workflow diagram visualise a structured approach for developing a facial emotion recognition system, from data preparation to model deployment. The work starts with gathering comprehensive facial image dataset labeled with corresponding emotion, further dataset is over-sampled and the various augmentation techniques like geometric transforma-

tion, color adjustment etc. are applied. After that, the model is developed which includes EfficientNet-B3 as base model, and additional layers are added for fine tuning and over fitting prevention. Next, the model is compiled and trained. At last, the model is evaluated with different performance matrix, and then it is used for prediction.

The overall modified model architecture a used for the same purpose is depicted in Figure 2. Facial image is over-sampled as pre-processing step in order to reduce the different emotion class imbalance. The benefit of oversampling is a redistribution of an image among different class of emotion like happy, sad, disgust, surprise, neutral, fear, angry, because the imbalance problem generally occurs in ML models [18, 19]. The data is augmented in order to increase generalization and robustness of the data using the Keras function [20]. After different preprocessing technique applied on the dataset, the proposed model, i.e. the Convolution neural network based EfficientNet-B3 architecture is used as a base model, after that top fully connected layer is set to be false in order to add additional layer, i.e. Batch Normalisation layer for smooth output feature extraction. Dense Layer is added with 256 neurons having ReLU activation function and over-fitting reduction using L1, L2 regularisation, as depicted in Figure 2. At the end, an additional output layer is added which uses the Softmax activation function for multi-class classification. The hybrid EfficientNet-B3 model [21] is enhanced by adding additional layers on top to facilitate fine-tuning [22]. These layers also help in regularization, such as dropout, to prevent overfitting and achieve better performance.

In proposed model development, the methodology involves series of steps and their mathematical representation is as follows.

3.2 Loading and Preprocessing of Dataset

The facial image data are loaded and converted into arrays of NumPy for preprocessing.

3.2.1 Image Loading and Label Tagging

Image Loading: Facial image dataset is inputted and converted into a matrix format as in Equation (1).

$$I \in \mathbb{R}^{H \times W \times C}, \tag{1}$$

where H is the height, W is the width, and C is the number of color channels i.e. 3 for RGB and I is facial image.

Label Tagging: The labels of the facial images are tagged into a one-hot vector

$$x \in \{0, 1\}^N, \tag{2}$$

where N is the number emotion of classes.

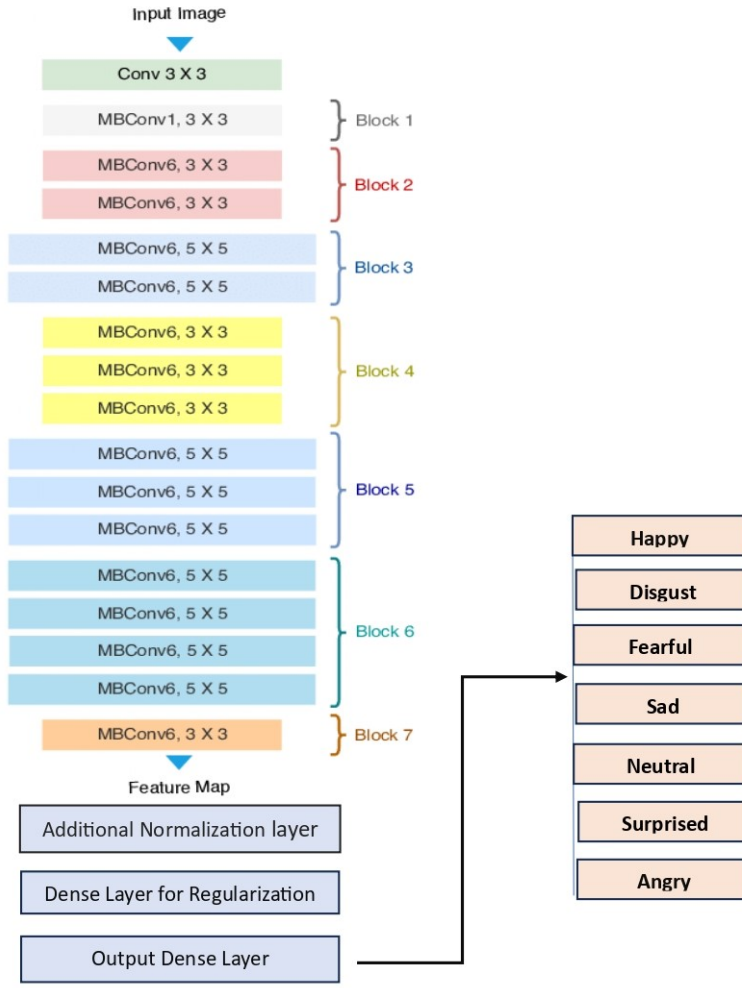


Figure 2. Model architecture

3.3 Data Augmentation and Oversampling and Data Generators

Data Augmentation techniques are applied by transforming the original images to generate more training samples, as mentioned in Equation (3).

3.3.1 Augmentation Transformations

Augmentation Transformations: Given an image I, augmentation functions T_i are applied to produce new images.

$$I' = T_i(I). \tag{3}$$

Augmented images are then converted into batches by Generators for training and testing.

3.3.2 Batch Generation

Batch Generation: A batch B of size N is created, where

$$B = \{(I_i, y_i)\}_{i=1}^{N}. \tag{4}$$

3.4 Model Building

A deep learning model is constructed using a (EfficientNet-B3) base model with additional layers for fine-tuning.

3.4.1 Model Architecture

Model Architecture:

$$\text{output} = f_{\text{softmax}}(f_{\text{dense}}(f_{\text{dropout}}(f_{\text{bn}}(f_{\text{base}}(I))))), \tag{5}$$

where f_{base} denotes the EfficientNet-B3 base model, and the other functions represent additional layers in Equation (5). Step wise mathematical representation is as below.

1. Input Vector:

$$\mathbf{I}' \quad \text{as the input image vector after preprocessing.} \tag{6}$$

2. EfficientNet-B3 Feature Extraction:

$$\mathbf{F} = \text{EffNet}(\mathbf{I}'). \tag{7}$$

3. Batch Normalization:

$$\mathbf{F}' = \text{BN}(\mathbf{F}). \tag{8}$$

Mathematical formula is stated below:

$$\text{BN}(x) = \frac{x - \mu}{\sqrt{\sigma^2 + \epsilon}} \times \gamma + \beta, \tag{9}$$

where

- x implies input feature vector,
- μ implies mean of the batch,
- σ^2 varience variance of the batch,
- ϵ is a small constant added to avoid division by zero,
- γ implies scaling parameter,
- β implies shifting parameter.

4. Dropout Regularization:

$$\mathbf{F}'' = \text{Dropout}(\mathbf{F}'). \tag{10}$$

Dropout is a regularization technique applied to prevent overfitting by randomly dropping (setting to zero) a proportion of the input units during training. Mathematically, dropout can be represented as follows:

$$\text{Dropout}(x) = x \times \text{mask}, \tag{11}$$

where

- x is the input feature vector,
- mask is a binary mask vector where each element has a probability p of being set to 1 and $1 - p$ of being set to 0. The mask is randomly generated for each mini-batch during training.

5. Dense Layer Transformation:

$$\mathbf{Z} = \text{Dense}(\mathbf{F}''). \tag{12}$$

The dense layer or fully connected layer applies a linear transformation to the input feature vector followed by an activation function. Mathematically, the dense layer transformation can be represented as:

$$\text{Dense}(x) = \text{ReLU}(W \cdot x + b), \tag{13}$$

where

- x is the input feature vector,
- W is the weight matrix,
- b is the bias vector,
- ReLU is the rectified linear unit activation function.

6. Softmax Activation:

$$\mathbf{P} = \text{Softmax}(\mathbf{Z}). \tag{14}$$

The softmax activation function is used to convert the raw output logits into probabilities. Mathematically, the softmax function can be represented as:

$$\text{Softmax}(z_i) = \frac{e^{z_i}}{\sum_{j=1}^{N} e^{z_j}}, \tag{15}$$

where

- z_i is the i^{th} element of the input logits vector z,
- N is the number of classes,
- e is the base of the natural logarithm.

3.5 Model Compilation and Training

The model is compiled with a loss function, optimizer, and metrics, then trained on the data.

3.5.1 Loss Function

Loss Function: Categorical Cross-Entropy Loss

$$L(y, \hat{y}) = -\sum_{i=1}^{K} y_i \log(\hat{y}_i), \tag{16}$$

where $\hat{y}$ is the predicted probability vector.

Optimizer: Adamax optimizer is used to minimize the loss function.

Training: The model parameters are updated iteratively using backpropagation and gradient descent.

3.6 Model Evaluation

The trained model is evaluated on the test set to measure its performance. Performance parameters involve accuracy, that is, the fraction of correctly predicted instances over the total instances. The second is the confusion matrix, which is a matrix M where M_{ij} represents the number of instances of class i predicted as class j. In addition, F-score and precision are also reported for performance comparisons.

4 SIMULATION AND RESULTS

The simulation tests are carried out on the Google Colab platform. The Colab TPU (Tensor Processing Unit) are utilized to improve parallel processing and optimize matrix computations for the EfficientNet-B3 model. Open CV library software is employed for image pre-processing and augmentation. Keras library is used for pre-trained model building. All performance evaluations were performed on the same platform. In addition, this section presents a description of the freely available dataset and experimental results obtained over dataset including classification report, test image predictions, accuracy, and loss graphs. A comprehensive analysis of the results is provided in the following section.

4.1 Dataset Description

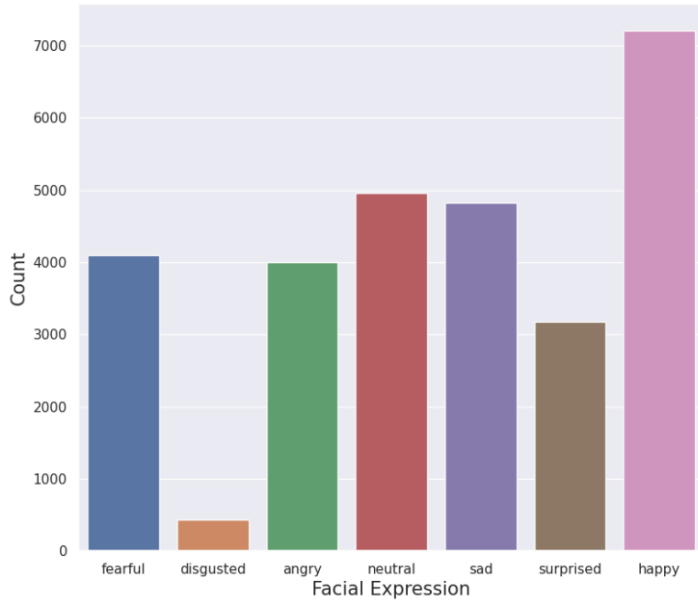
FER-2013 universal dataset is used for our facial expression recognition. The data set contains 35 887 grey-scale facial images containing 7 different emotions (anger, disgust, fear, happiness, neutral, sad, and surprise) as visualized in Figure 3 and the distribution of the data set for training and testing is depicted in Figure 4.

Figure 3. Dataset

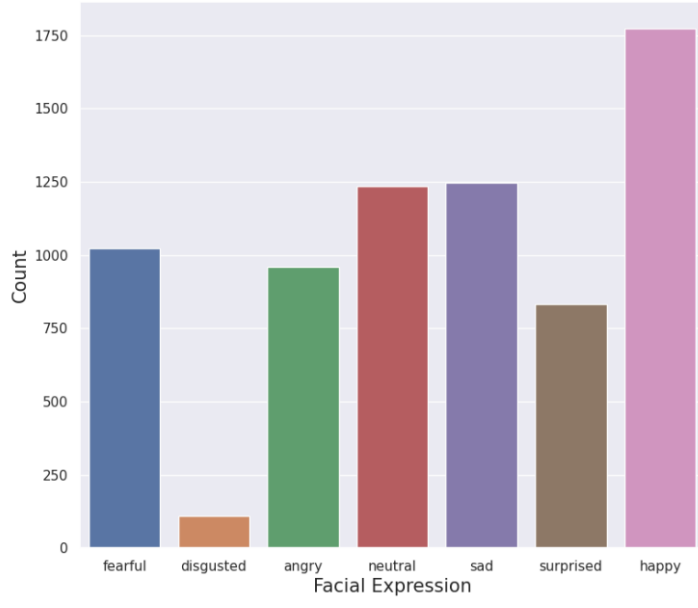
4.2 Quantitative Results

In order to make rigorous result analysis, performance evaluation of the proposed work is evaluated with the help of different performance matrices like classification report of different emotion classes with their precision, recall, and F1 scores. The overall tabular comparison results are presented in Table 2 which helps in performance analysis of the proposed work. The classification report indicates that the model performs well overall, with varying levels of success across the seven emotion categories. The "Disgusted" emotion achieves near-perfect performance, showcasing the model's ability to accurately and consistently identify this class. "Surprised" and "Happy" also demonstrate strong results, with high precision, recall, and F1 scores. Overall, the model is effective for most emotions, but targeted improvements are needed for the weaker categories like "Sad", "Neutral", and "Fearful".

The another performance matrix that we use to evaluate the overall emotion prediction is confusion matrix which can visualise or give summary of the performance of the model. Below is the confusion matrix plotted between actual and predicted emotions classes in Figure 5. It has been observed that the confusion matrix reveals clear distinctions for emotions like Disgusted and Surprised, while emotions such as



a) Distribution of the train images



b) Distribution of the test images

Figure 4. Dataset distribution

Classification Report on Validation Data				
Emotion	Precision	Recall	F1 score	Support
Angry	0.78308	0.85308	0.81658	1 443
Fearful	0.74621	0.78448	0.76486	1 443
Happy	0.88897	0.81012	0.84772	1 443
Sad	0.73599	0.69161	0.71311	1 443
Surprised	0.88968	0.95565	0.92148	1 443
Neutral	0.76837	0.71033	0.73821	1 443
Disgusted	0.99039	1.0000	0.99517	1443

Table 2. Classification report

Sad, Neutral, and Fearful showed significant overlap, pointing to areas where the model struggles to differentiate effectively.

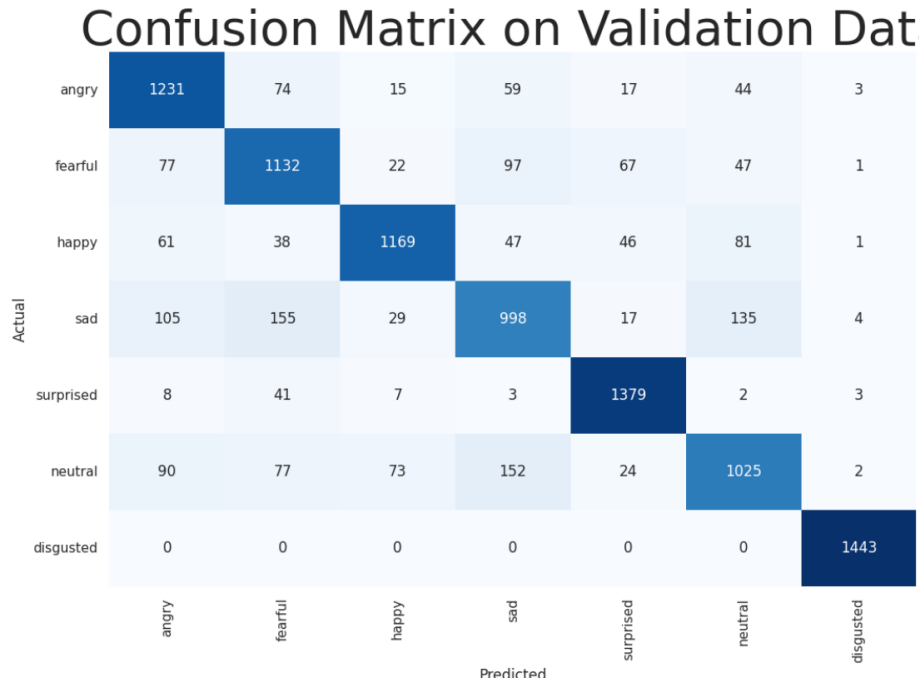


Figure 5. Confusion matrix

4.3 Visual Results

The visual result of the predicted and actual emotion is mentioned as label of each image which is depicted in Figure 6. The facial image which is bordered with the

green border are same for actual and prediction i.e. true prediction and the image with the red border is not same as actual and predicted one i.e. false prediction. It has been observed that proposed model yields better prediction for most of the emotion variation cases. In some of the cases, tendency of false prediction is observed where actual image is "fearful", "sad" and "neutral" comparative to other emotions.

Figure 6. Predicted image

4.4 Training/Validation Curves

Our model is graphically evaluated with the help of accuracy and loss graphs between training and validation of facial images and results of the same are plotted below, and with the help of the below two graphs we can analyze our model performance. Proposed methodology gives 92 % of training accuracy with a loss of 32 % and validation accuracy achieved of 83 % along with 62 % loss for FER-2013 dataset, which contain more than 35 000 facial images of seven different emotions. The training accuracy started at 32.47 % and showed consistent improvement with each epoch, reaching over 92 % by the final epoch. Validation accuracy also improved steadily from 44.30 % initially to a peak of approximately 83 % before early stopping. This

indicates that the model was learning effectively and generalizing well to the validation data, as the validation accuracy consistently improved alongside the training accuracy depicted in Figure 7.

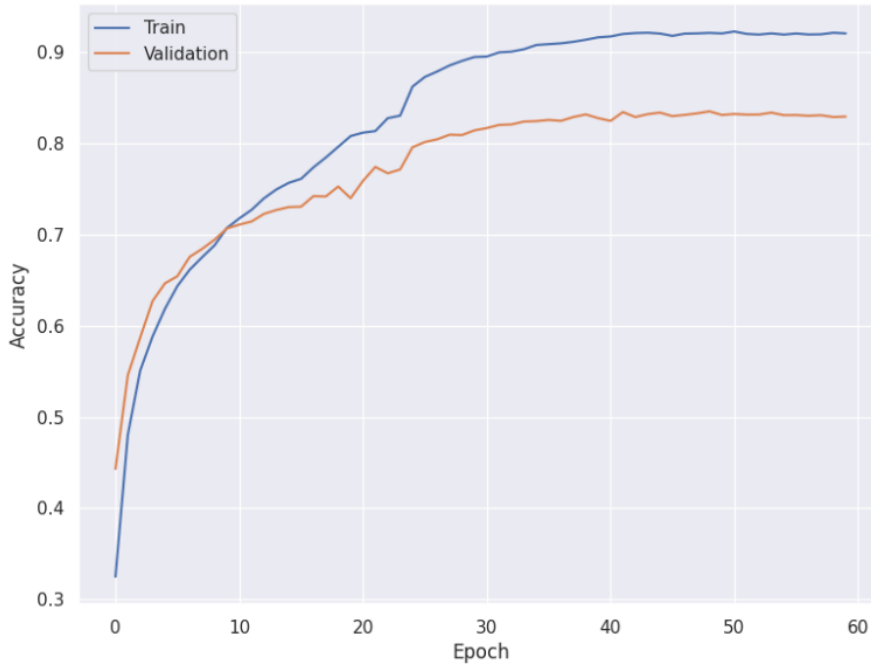


Figure 7. Training and validation accuracy

The training loss started high at 4.8569 but decreased significantly with each epoch, dropping to 0.3218 by the final epoch. Validation loss followed a similar downward trend, starting at 2.2868 and reducing to 0.6287. The consistent reduction in both training and validation loss suggests that the model was effectively minimizing the error in its predictions over time, leading to better performance on both training and validation datasets depicted in Figure 8.

5 RESULT ANALYSIS

This section presents a detailed analysis of the results achieved by the proposed modified EfficientNet-B3 architecture in comparison to existing methods and across different datasets to demonstrate its versatility and effectiveness.

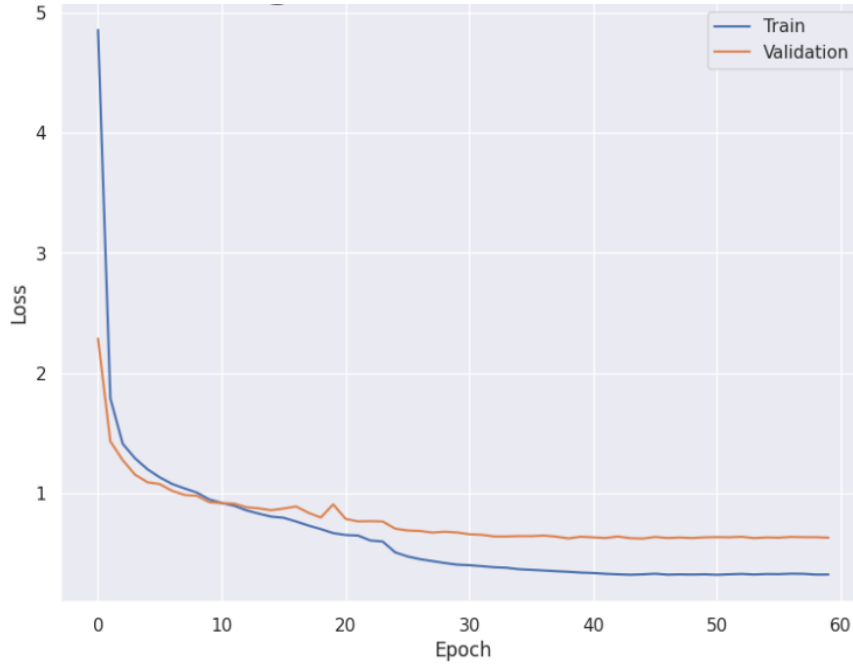


Figure 8. Training and validation loss

5.1 Comparison with Existing Methods

Table 3 shows the performance comparison between the proposed method and existing state-of-the-art methods for FER. It has been observed that proposed method outperforms over existing methods such as lightweight CNN, ResNet-50, MobileNet for FER-2013 dataset. Proposed model also achieves better F-score and precision with competitive model size. This proposed methodology achieves 92 % training accuracy with a loss of 32 % and 83 % validation accuracy with a 62 % loss on the FER-2013 dataset, which contains over 35 000 facial images of various emotions.

5.2 Comparison Across Datasets

The versatility of the proposed method is demonstrated by applying it to three different datasets: FER-2013, CK+, and RAF-DB. The results are summarized in Table 4. The comparison among FER-2013, CK+, RAFDB, and AffectNet provides a holistic view of a model's performance in controlled, semi-controlled, and real-world environments. CK+ serves as a benchmark for detecting prototypical expressions, while AffectNet and RAFDB challenge models with their real-world diversity and complexity. FER-2013 offers a computationally efficient baseline, though

Method	Accuracy (%)	F-Score (%)	Precision (%)	Model Size (MB)
Lightweight CNN	85.0	84.2	85.5	15.2
ResNet-50	87.5	86.5	88.0	95.0
MobileNet	86.0	85.0	86.5	14.5
DenseNet-121	89.0	88.2	89.5	32.0
CLCM	90.5	89.8	91.2	13.8
Proposed Method	**92.0**	**91.5**	**92.3**	**19.5**

Table 3. Performance comparison with existing methods on FER-2013

its limited diversity may underestimate a model's true potential. This comparison is critical for designing robust and generalizable FER systems capable of handling real-world challenges.

Dataset	Accuracy (%)	Loss (%)	Sample Size
FER-2013	92.0	32.0	35 887
CK+	94.5	28.0	593
RAF-DB	90.5	35.0	29 672
AffectNet	91.0	34.0	287 651

Table 4. Performance comparison across different datasets

5.3 Performance Metrics

Alongside accuracy and confusion matrix, we evaluated the proposed model using F-score and precision metrics to provide a more comprehensive performance analysis. Table 5 summarizes the results across FER-2013, CK+, and RAF-DB datasets. It has been observed that the proposed model yields best result for CK+ dataset as compared to FER-2013 and RAF-DB daatset. Since CK+ focuses on peak expressions, hence it reflects that proposed model is more suitable to peak expressions also.

Dataset	Accuracy (%)	F-Score (%)	Precision (%)	Recall (%)
FER-2013	92.0	91.5	92.3	91.0
CK+	94.5	94.0	94.8	93.5
RAF-DB	90.5	89.8	91.0	89.0

Table 5. Performance metrics across datasets

6 ABLATION STUDY

An ablation study was conducted to evaluate the effectiveness of individual components in the modified EfficientNet-B3 architecture. Table 6 presents the ablation results by incrementally adding components and observing the performance.

Configuration	**Accuracy (%)**	**Loss (%)**	**Remarks**
Base EfficientNet-B3	85.0	40.0	Without modification
+ Pretrained on ImageNet	88.0	36.0	Transfer learning effectiveness
+ Custom Layers for FER	90.0	35.0	Full proposed architecture
+ Data Augmentation	90.5	33.0	Robust to varied facial features
+ Class Imbalance Handling	92.0	32	Improved minority class accuracy

Table 6. Ablation study results

The ablation study demonstrates the steady improvements made by gradually altering the foundational EfficientNet-B3 design. An initial accuracy of 85.0 % and a loss of 40.0 % were noted when using the unaltered base model. Accuracy increased to 88.0 % and loss decreased to 36.0 % when pretraining on the ImageNet dataset was used, indicating the efficacy of transfer learning. The architecture was improved by adding more layers specifically designed for facial emotion recognition (FER), increasing accuracy to 90.0 % with a somewhat lower loss of 35.0 %. A slight improvement in accuracy to 90.5 % and a further decrease in loss to 33.0 % were achieved by using data augmentation techniques, which improved robustness to a variety of facial traits. Lastly, the identification of minority class feelings was significantly enhanced by the application of strategies to correct class imbalance, with the highest accuracy of 92.0 % with the least amount of loss of 32.0 %. These outcomes confirm how each element of the suggested architecture contributes to the achievement of improved FER performance.

7 CONCLUSION AND FUTURE WORK

In this work, the modified model of EfficientNet-B3 is developed that uses transfer learning for human emotion recognition through face expression. In order to assess the proposed work, FER-2013 dataset is used that contains more than 35 000 universal faces. After collecting the dataset, the data images have been preprocessed by different preproceessing techniques like data augmentation, oversampling and fine tuning in order to decrease the overfitting and balancing of the dataset. The proposed architecture includes EfficientNet-B3 as the base model. The fully top connected layer is added with customised layers like the batch normalization layer and dense layer followed by output layer that performs classification of the emotion into its respective classes. The model is evaluated using different performance matrix like accuracy graph, confusion matrix, classification report of precision, re-

call, F1 score. The findings mentioned above indicate that the training accuracy is 92 %, while the testing accuracy is 83 %. With the proposed methodology and key findings, it will be helpful in accurate human emotion recognition, which have a wide range of application like security, healthcare monitoring, marketing analysis, human-computer-interaction and robotics emotional intelligence.

The future work in the field of emotion recognition should rely on multi-modal data for better comprehension, using edge computing, low-latency algorithm, advanced and diversified datasets and advanced AI techniques like generative and transformer models to increase accuracy and efficiency.

REFERENCES

[1] Dalvi, C.—Rathod, M.—Patil, S.—Gite, S.—Kotecha, K.: A Survey of AI-Based Facial Emotion Recognition: Features, ML & DL Techniques, Age-Wise Datasets and Future Directions. IEEE Access, Vol. 9, 2021, pp. 165806–165840, doi: 10.1109/ACCESS.2021.3131733.

[2] Barrett, L. F.—Adolphs, R.—Marsella, S.—Martinez, A. M.—Pollak, S. D.: Emotional Expressions Reconsidered: Challenges to Inferring Emotion from Human Facial Movements. Psychological Science in the Public Interest, Vol. 20, 2019, No. 1, pp. 1–68, doi: 10.1177/1529100619832930.

[3] Ko, B. C.: A Brief Review of Facial Emotion Recognition Based on Visual Information. Sensors, Vol. 18, 2018, No. 2, Art. No. 401, doi: 10.3390/s18020401.

[4] Zhou, N.—Liang, R.—Shi, W.: A Lightweight Convolutional Neural Network for Real-Time Facial Expression Detection. IEEE Access, Vol. 9, 2020, pp. 5573–5584, doi: 10.1109/ACCESS.2020.3046715.

[5] Li, Y.—Zeng, J.—Shan, S.—Chen, X.: Occlusion Aware Facial Expression Recognition Using CNN with Attention Mechanism. Vol. 28, 2019, No. 5, pp. 2439–2450, doi: 10.1109/TIP.2018.2886767.

[6] Dapogny, A.—Bailly, K.—Dubuisson, S.: Dynamic Pose-Robust Facial Expression Recognition by Multi-View Pairwise Conditional Random Forests. IEEE Transactions on Affective Computing, Vol. 10, 2019, No. 2, pp. 167–181, doi: 10.1109/TAFFC.2017.2708106.

[7] Kim, D. H.—Baddar, W. J.—Jang, J.—Ro, Y. M.: Multi-Objective Based Spatio-Temporal Feature Representation Learning Robust to Expression Intensity Variations for Facial Expression Recognition. IEEE Transactions on Affective Computing, Vol. 10, 2019, No. 2, pp. 223–236, doi: 10.1109/TAFFC.2017.2695999.

[8] Oguine, O. C.—Oguine, K. J.—Bisallah, H. I.—Ofuani, D.: Hybrid Facial Expression Recognition (FER2013) Model for Real-Time Emotion Classification and Prediction. CoRR, 2022, doi: 10.48550/arXiv.2206.09509.

[9] Li, S.—Deng, W.: Deep Facial Expression Recognition: A Survey. IEEE Transactions on Affective Computing, Vol. 13, 2022, No. 3, pp. 1195–1215, doi: 10.1109/TAFFC.2020.2981446.

[10] Jun, H.—Shuai, L.—Jinming, S.—Yue, L.—Jingwei, W.—Peng, J.: Facial Expression Recognition Based on VGGNet Convolutional Neural Network. 2018 Chinese Automation Congress (CAC), IEEE, 2018, pp. 4146–4151, doi: 10.1109/CAC.2018.8623238.

[11] Song, Z.: Facial Expression Emotion Recognition Model Integrating Philosophy and Machine Learning Theory. Frontiers in Psychology, Vol. 12, 2021, Art. No. 759485, doi: 10.3389/fpsyg.2021.759485.

[12] Khan, A. R.: Facial Emotion Recognition Using Conventional Machine Learning and Deep Learning Methods: Current Achievements, Analysis and Remaining Challenges. Information, Vol. 13, 2022, No. 6, Art. No. 268, doi: 10.3390/info13060268.

[13] Singh, R.—Sharma, H.—Mehta, N. K.—Vohra, A.—Singh, S.: EfficientNet for Human FER Using Transfer Learning. ICTACT Journal on Soft Computing, Vol. 13, 2022, No. 1, pp. 2792–2797, doi: 10.21917/ijsc.2022.0397.

[14] Gursesli, M. C.—Lombardi, S.—Duradoni, M.—Bocchi, L.—Guazzini, A.—Lanata, A.: Facial Emotion Recognition (FER) Through Custom Lightweight CNN Model: Performance Evaluation in Public Datasets. IEEE Access, Vol. 12, 2024, pp. 45543–45559, doi: 10.1109/ACCESS.2024.3380847.

[15] Ding, H.—Zhou, S. K.—Chellappa, R.: FaceNet2ExpNet: Regularizing a Deep Face Recognition Net for Expression Recognition. 2017 12th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2017), 2017, pp. 118–126, doi: 10.1109/FG.2017.23.

[16] Pecoraro, R.—Basile, V.—Bono, V.—Gallo, S.: Local Multi-Head Channel Self-Attention for Facial Expression Recognition. CoRR, 2021, doi: 10.48550/arXiv.2111.07224.

[17] Breuer, R.—Kimmel, R.: A Deep Learning Perspective on the Origin of Facial Expressions. CoRR, 2017, doi: 10.48550/arXiv.1705.01842.

[18] Mohammed, R.—Rawashdeh, J.—Abdullah, M.: Machine Learning with Oversampling and Undersampling Techniques: Overview Study and Experimental Results. 2020 11th International Conference on Information and Communication Systems (ICICS), IEEE, 2020, pp. 243–248, doi: 10.1109/ICICS49469.2020.239556.

[19] Rodríguez-Torres, F.—Martínez-Trinidad, J. F.—Carrasco-Ochoa, J. A.: An Oversampling Method for Class Imbalance Problems on Large Datasets. Applied Sciences, Vol. 12, 2022, No. 7, Art. No. 3424, doi: 10.3390/app12073424.

[20] Mikołajczyk, A.—Grochowski, M.: Data Augmentation for Improving Deep Learning in Image Classification Problem. 2018 International Interdisciplinary PhD Workshop (IIPhDW), IEEE, 2018, pp. 117–122, doi: 10.1109/IIPHDW.2018.8388338.

[21] Batool, A.—Byun, Y. C.: Lightweight EfficientNetB3 Model Based on Depthwise Separable Convolutions for Enhancing Classification of Leukemia White Blood Cell Images. IEEE Access, Vol. 11, 2023, pp. 37203–37215, doi: 10.1109/ACCESS.2023.3266511.

[22] Vrbančič, G.—Podgorelec, V.: Transfer Learning with Adaptive Fine-Tuning. IEEE Access, Vol. 8, 2020, pp. 196197–196211, doi: 10.1109/ACCESS.2020.3034343.

Vivek Srivastava is Assistant Professor in the Department of Computer Science and Engineering and Coordinator (Research and IPR) at Rajkiya Engineering College, Kannauj, India. He holds his Ph.D. in Computer Science and Engineering from Harcourt Butler Technical University (HBTU), Kanpur. His research interests include machine learning, biometric systems, data analytics, and emerging areas of intelligent computing. He has published more than 50 research articles in international journals and conferences. He also worked as Associate Editor of International Journal of Engineering Science and Advance Research and a member of various international professional bodies like IEEE, IAENG, IACSIT and he is a Life member of Indian Science Congress Association (ISCA). He also worked as reviewer of various journals like Journal of Super Computing, IEEE Transactions on System, Man and Cybernetics etc.

Raghuraj Singh Suryavanshi is a senior academic leader and Professor in the Department of Computer Science and Engineering at the Pranveer Singh Institute of Technology (PSIT), Kanpur, India. He earned his Ph.D. from the Institute of Engineering and Technology (IET), Lucknow, and is a recipient of the prestigious Teacher Fellowship Award conferred by Dr. A.P.J. Abdul Kalam Technical University, in recognition of his contributions to academic excellence, research leadership, and capacity building. His research vision focuses on the development of secure, reliable, and scalable digital systems for real-world and policy-driven applications. His core research areas include the formal verification and validation of mission-critical distributed systems, with applications in blockchain technologies, cloud computing, and artificial intelligence. He is actively involved in multiple funded research projects supported by the Uttar Pradesh Council of Science and Technology (UPCST) and contributes to DST-aligned research priorities in emerging technologies.

Ashutosh Pandey is an undergraduate student in Computer Science and Engineering Department at Rajkiya Engineering College, Kannauj, India. He has strong skills in programming, full-stack web development, and data science, with hands-on experience in Python, Java, ASP.NET Core, databases, and machine learning. He also qualified in the Mathematics Olympiad.