Journal of Intelligent Systems and Internet of Things
Vol. 16, No. 02, PP. 82-101, 2025
DOI: https://doi.org/10.54216/JISIoT.160207

Real-Time Classroom Emotion Analysis Using Machine and Deep Learning for Enhanced Student Learning

Deepa Devasenapathy[1,*], Krishna Bhimaavarapu[2], Prem Kumar Sholapurapu[3], S. Sarupriya[4]

[1]Dendritic: A Human-Centered AI and Data Science Institute, Department of Computing & Software Engineering, U.A. Whitaker College of Engineering, Florida Gulf Coast University, 10501 FGCU Blvd. S, Fort Myers, FL 33965, USA

[2]Assistant Professor, Dept. Of CSE, Koneru Lakshmaiah Education Foundation, (Deemed to be University) Vaddeswaram, Guntur, A.P., India

[3]Senior Consultant, CGI, Katy, Texas, USA

[4]Assistant Professor, Department of ECE, Velammal Engineering College, Chennai, TN, India

Emails: ddevasenapathy@fgcu.edu; krishnab2021@gmail.com; premkumar.sh@gmail.com; sarupriya271281@gmail.com

Abstract

This research creates an innovative EfficientNet-B7-based Facial Expression Recognition model that delivers maximum accuracy performance for detecting emotions. Successful classification performance benefits substantially from EfficientNet-B7's application of compound scaling techniques which balances the entire network dimensions depth width and resolution. The characteristic distinctive to EfficientNet-B7 over standard architectural models involves its dual capability to perform accurate computations at reduced complexity levels. The model receives evaluation using KDEF at high-resolution as well as FER2013 at low-resolution through usage of SGD, Adam, and RMSprop optimizers. Experimental tests confirmed that EfficientNet-B7 operates with RMSprop optimizer to recognize emotions on KDEF at 91.78% accuracy superior to ResNet152's highest recorded accuracy of 88.77%. Performance levels declined to 57.56% on FER2013 because low-resolution images represent a great challenge to the model. Internal Batch Normalization (IBN) enters the model as an issue solution to halt gradient descent problems, which results in better model training stability and enhanced accuracy-loss patterns. The research demonstrates that FER performance benefits greatly when EfficientNet-B7 works in combination with IBN for high-resolution image processing. The research proves that EfficientNet-B7 stands as a reliable FER solution that shows potential usage in affective computing and human-computer interaction domain.

Received: January 18, 2025 Revised: February 16, 2025 Accepted: March 08, 2025

Keywords: Facial Expression Recognition; Convolutional Neural Networks; EfficientNet-B7; Image Classification; Batch Normalization (IBN); KDEF Dataset; FER2013 Dataset; Neural Network Optimization; Deep Learning

1. Introduction

Technological advances have become essential in daily operations for our digital era by shaping numerous human functions especially academic education. Artificial intelligence and machine learning demonstrate great potential through their capacity to perform emotional analysis and interpretation from people. Science researchers from different fields demonstrate great interest in computer systems that detect and assess human emotions [1]. The key process in human-computer interaction requires emotion recognition that uses facial expressions together with speech patterns and behaviours and physiological signals. The way we normally measure emotions lacks effectiveness because people naturally hide their real emotions from one another voluntarily or unconsciously.

Higher education programs have adopted student-led training methods that require innovative teaching practices for effective student learning outcomes [2]. The main difference between traditional teaching and student-centred education appears in how students actively participate while traditional education has students listen to lectures. Pertinent educational outcomes require teachers to track student participation levels then modify their delivery techniques for better results [3]. The need for such monitoring has stimulated researchers to create systems, which evaluate student emotional reactions immediately as part of their development of intelligent educational technologies.

Information technology systems based on deep learning and machine learning show great advancement in detecting emotions specifically within educational institutions. Advanced algorithms enable these technologies to recognize students' classroom facial expressions, which helps them, determine their current emotional condition. Educational specialists gain valuable knowledge using these tools that guides their assessment of students who require supplemental resources [4-6]. Through real-time emotion assessment teachers gain capabilities to design teaching strategies, which lead to enhanced inclusion and better, learning outcomes.

The main goal of this study comes from student needs for more dynamic classroom participation and tailored teaching approaches. Standard educational assessment tools, which employ quizzes and exams and written assignments, fail to provide extensive understanding of student comprehension together with emotional state during the learning process [7]. The continuous emotions analysis by machine learning systems operates as a warning system that helps teachers provide fast responses when problems arise. The strategy allows teachers to adjust their teaching techniques through aggregated emotional feedback from students while supporting students who need extra assistance.

The implementation of deep learning techniques in emotion analysis produces better system measurements combined with enhanced reliability. The recognition of facial expressions benefits from deep neural networks although convolutional neural networks (CNNs) stand as especially powerful instruments in this process. Researchers designed a model through training on labelled emotional expressions data so they created technology that automatically detects and identifies emotions in real-time [8-10]. The combination of cameras that monitor classroom students produces facial images from which AI systems determine student emotional responses. Educational feedback becomes accessible from this data collection, which enables teachers to improve their teaching methods. Figure 1 shows the facial expressions that are commonly detected through model to rectify the expression of the recognised image.

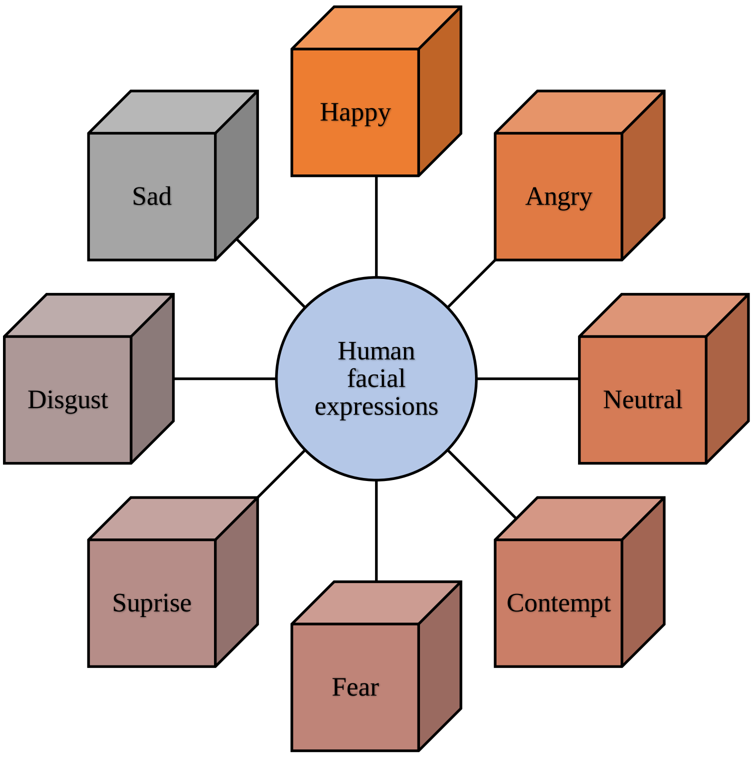

Figure 1. Illustration of classified human facial expressions

Classroom emotion recognition consists of three main elements, which form its structure. The detection model initially finds students' faces, which receives individual unique identifiers. An FER model employs facial expressions to establish different emotional categories [11]. The system uses its data processing capabilities to transform emotional data into an approachable display, which teachers can use for class emotional climate understanding. Through the application of this technology, teachers can boost interactive courses, develop supportive learning spaces, and maintain student engagement throughout the class.

Emotional recognition analysis possesses value that exceeds basic educational settings. The technology finds usage in educational settings that operate virtually and serves meetings conducted online and specialist training activities. Presenters can adapt their content immediately during speaking engagements due to live audience feedback analysis helping them obtain maximum audience engagement [12]. These detection systems hold prospect for multiple application areas including prevention of driver fatigue incidents, enhanced interaction quality in customer support, and tracking mental health status through monitoring methods.

Several obstacles exist in using real-time emotion analysis in classrooms even though its practical potential demonstrates promise. The main difficulty stems from how different people display their expressions on their faces. Multiple elements that stem from ethnicity and cultural background together with personal expression preferences determine how people express and other people perceive their emotions. The classroom environment produces several barriers through its changing lighting as well as head motions and student partial face coverage or mask usage [13-15]. These real-world factors hence requiring adaptive models, which handle complex situations effectively, affect system accuracy for emotion recognition.

The most important obstacle concerns correctly differentiating minor emotional expressions. Artificial intelligence systems struggle to separate emotions such as confusion and concentration because these emotions produce face expressions that look similar. Some emotions including tiredness and depression face misclassification because their features on the face overlap with each other [16]. The scientific community makes continuous progress in deep learning technology development to boost the machine's competence in detecting subtle facial expressions without error.

Educational use of emotion recognition technology requires training datasets to possess both high quality characteristics along with wide range of expressions. The current datasets of facial expressions originate from controlled settings where subjects need to perform specific emotional displays. The use of initial training data from these datasets does not necessarily correlate to the actual emotions students express during real educational interactions [17]. Functional classroom emotion analysis demands models pretrained on various data sets encompassing authentic naturalistic expressions, which occur when students are in active classrooms.

Research teams investigated three key techniques, which consist of data augmentation as well as transfer learning together with feature fusion. Data augmentation tactics that modify lighting conditions and create various image obstructions and body postures strengthen model reliability. Transformed learning systems enable them to extract knowledge from extensive pre-trained networks, which then boosts their performance when processing lesser domain-specific information [18]. Various methods of information consolidation unite several communicative signals (including facial movements together with voice inflections along with body signal patterns) to achieve better results in emotional identification.

Educational institutions need to carefully analyse how real-time emotion analysis affects their student populations ethically. The analysis of student facial expressions through continuous monitoring creates privacy issues because frequent monitoring appears to be too invasive for students. The development of defined standards for obtaining data and its storage methods alongside its authorized usage should become mandatory for upholding ethical practices while keeping student information safe. Such technology systems require clear guidelines and appropriate consent from students and educators to demonstrate their operational methods and resultant advantages.

Real-time emotion analysis done through machine and deep learning offers educational institutions a complete transformative opportunity. The system gives schools essential understanding about student interactions with course materials and student feelings thus allowing teachers to build personalized instruction methods that enhance student success. Scientific advancements and development of deep learning methods actively solve previous reliability problems and accuracy concerns of emotion recognition systems even though expression variability and data restrictions and moral issues persist [19]. These technologies continue developing which provides prospects to modify classroom adaptability and reactivity while ensuring accessibility to all students.

2. Related Work

The research field of artificial intelligence as well as computer vision focuses intensively on Facial Expression Recognition (FER). Different conventional and deep learning-based methods have appeared throughout the years to improve the accuracy and robustness of FER systems. Feature engineering-intensive conventional methods need substantial data treatment before moving to classification steps. The application of Convolutional Neural Networks (CNNs) within deep learning algorithms brings better accuracy and environmental adaptivity to FER systems that reduces the dependence on manual features and enables high performance in different lighting conditions as well as conditions where facial features are blocked from view [20]. The survey investigates current FER methods together with multiple-feature combination methods as well as the most recent deep learning-based advancements.

2.1 Conventional Facial Expression Recognition Approaches

The typical process of Traditional FER relies on three main steps, which start with pre-processing images then extract features before finishing with classification. The process of image pre-processing produces improved picture quality through multiple modifications including the normalization of illumination levels and the elimination of noise artifacts and the achievement of face landmark alignment. Faces are encoded with features using extraction techniques that include Local Binary Patterns (LBP), Histogram of Oriented Gradients (HOG) and Gabor filters [21]. Support Vector Machines (SVM) and k-Nearest Neighbors (k-NN) and Random Forests compose the machine learning classifiers, which perform expression classification through extracted features. The employed methods demonstrate moderate achievement but experience difficulties because of pose position along with varying lighting conditions and blocking of the face.

2.2 Deep Learning-Based Approaches

Facial expression recognition underwent a transformative change through deep learning because it successfully implemented automatic feature recognition and classification systems. Several layers of convolutional operation in CNNs allow these networks to effectively identify facial expressions. The accuracy of FER increased through the usage of advanced networks such as ResNet and VGGNet, which extract deep features together with Inception Net. Rephrase the following sentence [22]. Also, normalize verbalization when possible. Deep learning models perform exceptionally but face two critical barriers during practical use, which involve substantial resource demands and exceptionally large training dataset needs.

2.3 Multi-Feature Fusion Techniques

Multi-feature fusion stands as an effective technique to boost the performance of FER systems. Multi-Feature Fusion brings together various feature representations to enhance classification results. Two primary fusion methods exist:

- Different layers of one CNN model work together while extracting diverse feature representations, which minimizes information loss that naturally occurs in feature propagation.
- A fusion process based on ensembles combines multiple CNN models, which produce outputs that are merged through aggregation techniques such as weighted voting or averaging methods. Better generalization occurs through Ensemble modelling since it unites the strengths discovered across multiple classification systems.

Table 1: Literature survey of methodology, findings and limitations

Study Focus	**Methodology**	**Findings**	**Limitations**
FER using deep learning	CNN, ensemble learning [22]	Higher accuracy than conventional methods	Requires large datasets and processing power
Emotion detection from images & videos	OpenCV, Python-based ML models	Real-time facial expression recognition	Limited accuracy in complex environments [23]
AI-based FER system	Haar-cascade classifier, OpenCV [24]	More accurate than human prediction	Perception issues among users
Multi-modal emotion recognition [25]	EEG signals, machine learning classifiers	Enhanced accuracy using multiple signals	Requires specialized hardware

Student performance prediction	Decision trees, regression models	Identifies students needing intervention [26]	Limited generalizability
ML-based engagement detection in learning [27]	Facial recognition, behavioural analysis	Detects student engagement levels	Requires extensive video data
ML recommender system for education	Classification, adaptive ML models	Personalized learning suggestions	Data privacy concerns
Machine learning in education	Survey of ML applications	ML improves adaptive learning	Ethical and implementation challenges [28]
Automated attendance system	Face recognition with LBPH	Efficient student attendance tracking	Potential security and privacy risks [29]
Emotion's influence on learning	Cognitive neuroscience, behavioural study [27]	Emotions enhance memory and learning	Hard to quantify emotional impact
AI-based student engagement analysis	Gaze, head pose, and facial expressions	Automated tracking of student focus	Variability in student behaviour [22]
FER in classroom settings	Cloud-based FER analysis	Evaluates student emotions in real-time	Performance depends on dataset diversity
AI-driven academic performance prediction	Data-driven predictive models	Early risk identification in students	Requires continuous data collection
Deep learning for emotion recognition	Multi-layer neural networks	Improved accuracy in emotion detection [30]	High computational cost
Machine learning and student motivation	Behavioural psychology + ML [25]	Identifies motivational factors	Complex interactions among variables

Equipping face expression recognition with deep learning-based methods replaced traditional machine learning methods and led to better accuracy and resistance. Three main factors restrict the performance output of FER systems because they include dataset diversity and the need for fast implementation during real-time operations [25]. Recognition systems require two parallel assessment approaches that enhance both practical deep learning model performance and safety of emotional privacy. Increasing FER technology development creates opportunities to transform various sectors so they can use advanced human-computer interface solutions.

3. Objectives of Research

The main research goal involves developing a Facial Expression Recognition (FER) model based on EfficientNet-B7 architectural components for enhancing expressivity recognition while keeping computation requirements low. This study aims to:

- Research the ability of EfficientNet-B7 model to detect facial expressions versus ResNet152 and other deep learning algorithms.
- The research investigates various FER performance enhancement techniques through multiple training strategies and optimizers including SGD, Adam and RMSprop for optimizing the most successful approach.

- Measure the model's adaptability by applying it to photographs in KDEF high-quality images and FER2013 basic images to validate universal performance.
- Internal Batch Normalization (IBN) should be used for addressing gradient descent stability issues to enhance system accuracy.
- The real-world functionality of FER models needs improvement through better performance recognition capabilities while maintaining operational efficiency in affective computing and human-computer interaction systems.

4. Motivation

FER technology performs vital functions across human-computer interaction and affective computing as well as mental health assessment and security system applications. The existing deep learning models encounter three main challenges, which include intensive computational expenses as well as gradient decay problems and poor performance on images of limited resolution. The research examines EfficientNet-B7 to solve these problems because its compound scaling optimization strengthens accuracy while reducing computational requirements. The use of EfficientNet for Face Expression Recognition remains an unexplored field even though the algorithm demonstrates strong results in image classification. The research seeks to fill the knowledge gap by conducting a FER evaluation of EfficientNet-B7 while applying SGD, Adam, and RMSprop optimizers as well as utilizing IBN for training stability. This research achieves better recognition outcomes on KDEF high-resolution data thus enhancing the development of useful FER solutions for practical deployment across various domains.

5. Dataset Used

The research adopts two standard datasets for Facial Expression Recognition (FER) named Karolinska Directed Emotional Faces (KDEF) and Facial Expression Recognition 2013 (FER2013) that are commonly utilized for emotion classification tasks.

KDEF Dataset

You can find 4,900 high-resolution images of 70 subjects arranged in 224×224 pixels, which includes 35 males and 35 females. The images in KDEF include seven basic emotions that appear across five different views (front, half-left, half-right, left, and right profiles) of each person who shows Anger, Disgust, Fear, Happy, Neutral, Sad, and Surprise. The standardized lighting setup together with controlled background of the dataset provides optimal conditions to evaluate deep learning model performance.

FER2013 Dataset

35,887 grayscale images with dimensions (48×48 pixels) comprise the FER2013 dataset that was obtained through automatic image collection by the Google Image API. Real-world face expressions found in FER2013 images contain substantial illumination variations in addition to facial occlusion and different head pose orientations and multiple expression types. The benchmark for testing uncontrolled FER models originated from the International Conference on Machine Learning (ICML) 2013 challenge therefore it continues as a foundational standard.

The research utilizes KDEF high-resolution structured images together with low-resolution real-life FER2013 images to evaluate the generalization capabilities of EfficientNet-B7.

6. Experimental Setup

The research setup of the Facial Expression Recognition (FER) model, which employs EfficientNet-B7 demands appropriate selection of hardware components as well as software programs and dataset elements for efficient training and evaluation. The high-performance computing system includes a minimum 8th Generation Intel Core i7 processor together with a NVIDIA GeForce RTX 3080 GPU and 16 GB RAM supported by a 512 GB SSD. The GPU supports fast deep learning model training through its effective implementation of big matrix calculations and the SSD improves model operation and data loading speed.

The developer builds the software environment with Windows 10 or Ubuntu 20.04 (LTS) operating systems supporting Python 3.8+ programming language. The Keras API within TensorFlow 2.6+ supports an implementation of the model that offers high-level deep learning functionalities. OpenCV handles the image processing operations including resizing, augmentation, and NumPy along with Pandas serves as the tool for dataset management. The visualization libraries Matplotlib and Seaborn provide tools to researchers for monitoring training process development alongside metric performances. Part of the code implementation requires the Scikit-learn library for computing evaluation metrics that include confusion matrices and classification reports.

Training and testing of the model occur using KDEF (Karolinska Directed Emotional Faces) and FER2013 (Facial Expression Recognition 2013) benchmark datasets as the evaluation platform. KDEF provides 4900 controlled and high-resolution face images with dimensions of 224×224 pixels for quality-driven expression research. Real-world scenario images in this FER2013 dataset amount to 35,887 grayscale images having a 48x48-pixel resolution while introducing complications from various head poses together with lighting inconsistencies and image occlusions. The training of EfficientNet-B7 benefits from using FER2013 images after proper 224×224 pixels resizing.

The data preparation includes the application of data augmentation methods including random rotation alongside flipping and brightness adjustments and cropping, which adds variability to enhance generalization abilities. The model training requires normalization through pixel value scaling which adjusts all images to fall within the range [0,1]. An 80-20 division of data occurs before testing to provide a solid foundation for model performance evaluation.

At the start of the process, the EfficientNet-B7 architecture receives pre-trained weights from the ImageNet database and its output layer adapts to handle seven emotional classifications. The IBN method contributes to stable training along with avoiding problems caused by vanishing gradients during model learning. The experiment tests three optimization methods including SGD, Adam and RMSprop to find the strongest performing algorithm for the project. The implementation of the model utilizes the categorical cross-entropy loss function, which serves as the typical selection for multi-class classification applications. The training efficiency and accuracy optimization process requires adjusting three fundamental hyperparameters consisting of 0.0001 learning rate and 32-batch size and 50 epochs. The learning rate schedule dynamically modifies the learning rate as validation performance stabilizes.

The model receives its testing phase after completion through separate analysis of the 20% test portions within both databases. The performance assessment of the model uses accuracy together with confusion matrix and precision, recall, F1-score and ROC-AUC score metrics. After training the model engineers save the product for future application in areas such as human-computer interaction and security surveillance and affective computing.

7. **Proposed Work** FER stands as a fundamental duty within affective computing research as well as human-computer interaction fields and surveillance technologies. Traditional Convolutional Neural Networks (CNNs) such as VGG and ResNet struggle with computational efficiency and recognition accuracy. This research proposes using EfficientNet-B7 as an FER model because it combines compound scaling, which optimizes performance and reduces parameter size in a CNN architecture is shown in Figure 1.

A model uses supervised learning to gain knowledge from datasets that have their outputs labeled. The current research examines FER as a classification challenge where facial pictures are categorized into one of seven expression sets (Happy, Sad, Angry, etc.). Through training, the model extracts patterns from labeled datasets, which it uses for making predictions in new images. Each image in KDEF and FER2013 possess a predefined emotion label that the model uses during training. The model achieves performance assessment through validation data analysis that proves its ability to apply learned patterns on previously unseen facial expressions in actual use contexts.

The most crucial facial attributes including eyes and eyebrows together with mouth and nose expressions are identified through feature extraction methods to detect different facial expressions. Convolutional Neural Networks within deep learning extract automatic image features from pictures during their operation in convolutional layers without any need for manual feature selection. The research study makes use of EfficientNet-B7 to extract hierarchical features that begin with recognizing simple edges in its shallower layers while detecting high-level facial patterns in its deeper layers. Model accuracy depends heavily on feature extraction because it allows experts to select vital facial data for classification while decreasing noise, which enhances the operational speed of recognition.

Loss Functions (Cross-Entropy Loss)

Any model prediction receives its validation through its comparison against actual labels via a loss function. Cross-entropy loss functions serve as a preferred loss calculation method for FER as a multi-class classification problem. It is calculated as:

$$L = -\sum_{i=1}^{N} b_i \log(\hat{b}_i) \quad (1)$$

The actual class label b_i intersects with the predicted probability distribution $\hat{b}_i$ The model demonstrates superior performance when its loss value is low. The multiple categories involved in facial expressions benefit from cross-entropy loss since it enables accurate matching between prediction distributions and ground truth labels to achieve better classification results and minimize incorrect classifications.

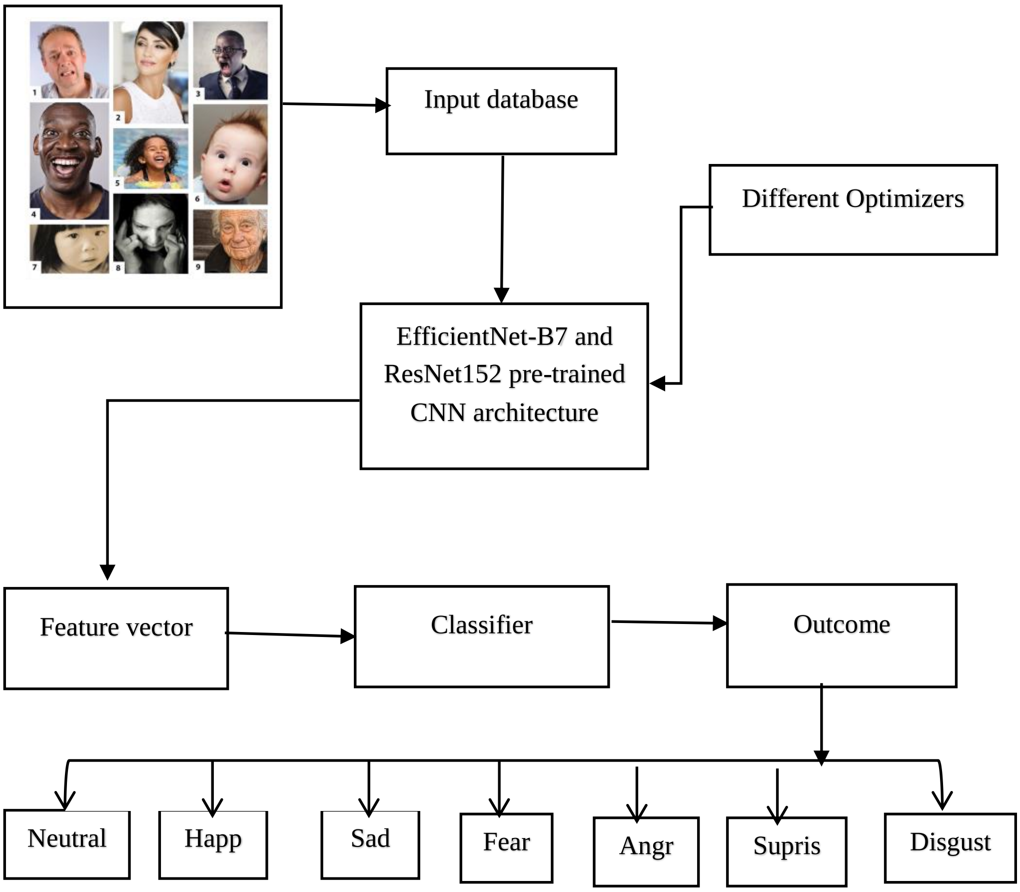

Figure 2. Illustration of Proposed Novel FER model with EfficientNet-B7 and ResNet-152 architecture

Regularization Techniques

A model performs badly on new data when it succeeds on training data points due to overfitting - regularization addresses this problem. Two methods adopted for this investigation include:

- During training the dropout, technique randomly disables neurons to make the network more effective at generalization.
- During mini-batch operations, Batch Normalization does two things: it normalizes activation values while it makes training more stable and decreases reliance on initialization techniques.
- Through these techniques, EfficientNet-B7 achieves high accuracy together with reduced overfitting, which leads to superior performance in real-world scenarios with various facial images.

Hyperparameter Tuning

During training model, performance depends on the set of parameters known as hyperparameters that regulate the training procedure. The research uses following main parameters as hyperparameters:

- Learning Rate sets the frequency of weight updates during training and reached its best value at 0.0001.
- The weights receive updates only after processing 32 images, which represents the chosen Batch Size value.
- Number of Epochs: Iterations over the entire dataset (set to 50).

The best optimal recognition accuracy and overfitting prevention combination emerges from the trial-and-error and grid search method used in hyperparameter tuning.

Cross-Validation

Through cross-validation, models acquire abilities to predict new data effectively. The research adopts k-fold cross-validation by dividing the dataset into k sections for training using different subsets as validation data at each execution. The approach helps both detect overfitting problems and optimize hyperparameters and validate robustness of the models. During multiple data split tests, EfficientNet-B7 learns to become stable and consistent in facial expression recognition performance when used with various datasets and real-world images.

Class Imbalance Handling

The FER2013 dataset contains systematic distribution imbalance where "Happy" face images outnumber "Disgust" face images. Model predictions become distorted by class imbalance because the most frequent categories receive preference. This research utilizes two methods to handle the potential class imbalance problem.

- A weighting system in the loss function provides higher error costs to minority classes to address unbalanced data distribution.
- Oversampling: Increases instances of minority classes through data duplication or augmentation.

The processing of class imbalanced data enables EfficientNet-B7 to detect all emotions accurately without dominating specific expressions.

Transfer Learning

A model can benefit from transfer learning which allows it to leverage pre-trained knowledge extracted from ImageNet before starting new training sessions. EfficientNet-B7 undergoes pre-training on ImageNet resulting in general image pattern understanding. Final layer fine-tuning operation applies to FER datasets KDEF and FER2013 for facial expression identification. Such an approach increases training efficiency and performance while lowering the required data volume, which makes the model suitable for real-world utilization.

Convolutional Neural Networks (CNNs)

The Convolutional Neural Networks (CNNs) operate as deep learning models made specifically for image recognition together with classification duties. CNNs recognize spatial image patterns through dedicated layers, which perform this task. The convolutional layer stands as the fundamental element of CNNs because its filtering operation detects visual features including edges and textures and facial characteristics. As the network reaches greater depth, its features develop additional complexity allowing the model to identify high-level attributes like eyes and smiles together with frowns. Luxury models appreciate pooling layers because they cut down both computation needs and feature map resolution but maintain key visual characteristics. The non-linear capabilities of models increase through activation functions particularly ReLU when applied to CNNs. Fully connected layers serve as the network's conclusion part to combine extracted features into classifications for different categories. The automatic feature learning ability of CNNs enables better performance than conventional machine learning frameworks particularly when used for FER and vision-based applications.

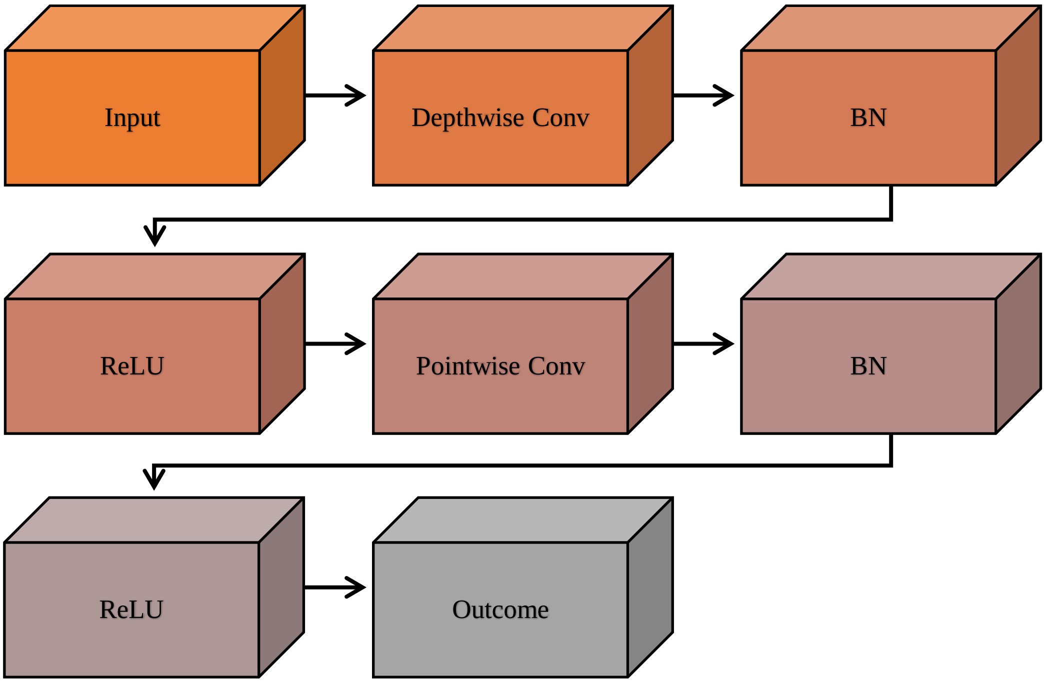

Figure 3. Representation of Depth-wise and Pointwise Convolutions

7.1 Data Preprocessing

Deep learning models require data preprocessing because it enhances the value of training data before the model building process. Raw datasets affect CNNs such as EfficientNet-B7 because they contain various issues including noisy information and formatting problems alongside missing data values and inconsistent data points. Preprocessing the dataset by cleaning it from imperfections allows better model generalization as well as reduces overfitting while speeding up convergence. The data preprocess includes cleaning steps together with normalization procedures and augmentation methods in addition to resizing and splitting functions. The execution of these processing steps develops a dataset, which becomes stronger while also being optimal for deep learning applications.

Data Cleaning

Data cleaning requires systems to recognize and fix or delete problematic data that includes corruption and inconsistency and irrelevancy. Uncompiled data in deep learning systems causes bias which results in suboptimal model execution. The main data problems during processing involve defective values alongside replicated images together with noisy signals and mislabeled contents. Interpolation methods solve missing value issues but duplicate images are eliminated through Structural Similarity Index (SSIM) similarity measures. Gaussian filters together with median filters serve as image noise reduction techniques to provide the model with clearer input data.

SSIM (Structural Similarity Index) for detecting duplicate images:

$$SSIM(a,b) = \frac{(2\mu_a\mu_b + C_1)(2\sigma_{ab} + C_2)}{(\mu_a^2 + \mu_b^2 + C_1)(\sigma_a^2 + \sigma_b^2 + C_2)} \quad (2)$$

Gaussian Smoothing for noise reduction:

$$I'(a,b) = \frac{1}{2\pi\sigma^2} e^{-\frac{(a^2+b^2)}{2\sigma^2}} \quad (3)$$

Here μ_a , μ_b = mean pixel of image a and b, $\sigma_a^2 + \sigma_b^2$ = variance, C_1 , C_2 = stability constants, σ = standard deviation.

Data Normalization & Standardization

The process of normalization and standardization enables data scaling consistency across features with various values to support stable gradient updates when using backpropagation. The normalization process transforms data values into specific ranges including [0,1] and [-1,1] while standardization method normalizes both mean values to zero and standard deviations to one. CNNs can experience unstable training because of exploding or vanishing gradients unless normalization techniques are used. The performance quality of EfficientNet-B7 depends on appropriate scaling of input pixel values.

Min-Max Normalization:

$$A_{norm} = \frac{A - A_{min}}{A_{max} - A_{min}} \quad (4)$$

Z-Score Standardization:

$$A_{std} = \frac{A - \mu}{\sigma} \quad (5)$$

Here μ = mean, σ = standard deviation.

Data Augmentation

Through artificial expansion of data, the diversity rate of datasets improves while overfitting decreases for better generalization results. The large datasets on which EfficientNet-B7 receives training benefits from augmentation resulting in better performance for small datasets. The standard augmentation methods used for data manipulation consist of rotation together with flipping and scaling in addition to cropping and color adjustment. Image augmentation enables the model to develop a form of transformation invariance which improves its robustness.

Rotation Transformation:

$$A' = A \cdot R_\theta \quad (6)$$

Scaling Transformation:

$$A' = sA, s > 0 \quad (7)$$

Here R_θ = rotation matrix.

Algorithm: Facial Expression Recognition (FER) Approach Using EfficientNet-B7

Input:

Facial expression datasets: KDEF (high-resolution) and FER2013 (low-resolution)

Pre-trained EfficientNet-B7 model

Optimizers: SGD, Adam, RMSprop

Output:

The classification resulted in seven emotional expressions including Anger, Disgust, Fear, Happy, Neutral, Sad, and Surprise.

The model utilizes accuracy statistics together with loss data points and confusion matrix results.

Step 1: Data Preprocessing

Load KDEF and FER2013 datasets.
All images should receive a 224×224 pixels (EfficientNet-B7 input size) resizing before processing.
The pixel values need normalization to spread between 0 and 1.
The data set benefits from data augmentation through random rotations while also involving horizontal flips and adjusting brightness values to boost overall variability.
The experimental data split into 80% training section and 20% testing section.

Step 2: Model Initialization

Begin the process with EfficientNet-B7 model from ImageNet weights.
Add a fully connected SoftMax activation layer after removing the top classification component for conducting emotion classification.
The application of Internal Batch Normalization (IBN) brings stability to training and helps the model reach better convergence results.

Step 3: Training the Model

Choose your optimizer from the available options consisting of SGD, Adam or RMSprop.
The model requires compilation using cross-entropy loss as part of the process.
The training process required a learning rate setting of 0.0001 combined with batch size 32 and 50 epochs.
Train the model through the training dataset under validation accuracy and loss observation.
You should use learning rate scheduling when gradient instability occurs during training.

Step 4: Model Evaluation

Perform model testing on the test dataset through these verification methods:
Accuracy
Confusion matrix
Precision, Recall, F1-score
ROC-AUC score
Reevaluate and modify the hyperparameters until accuracy targets are met before conducting a new training process.

Step 5: Model Deployment

The deployed model will serve for actual FER applications in the real world.
The implementation of this system should happen through human-computer interaction systems, emotion detection software, and affective computing deployments.

7.2 Model Architecture: EfficientNet-B7

The deep convolutional neural network EfficientNet-B7 works through compound scaling which applies optimization across network depth and width and image resolution at once. The network architecture design in EfficientNet differs from conventional CNNs because it employs mathematical scaling to determine layer counts. The neural network uses Mobile Inverted Bottleneck Convolution (MBConv) with Squeeze-and-Excitation (SE) blocks that enhance computational efficiency. The EfficientNet-B7 model reaches superior accuracy levels while using less memory compared to models from previous generations.

Compound Scaling

Models built using traditional CNN architectures gain scale through separate adjustments of their depth, width or resolution performance but produce subpar results. The efficient scaling method of Efficient Net applies uniform adjustments to every network dimension.

Scaling Formulation:

$$d = \alpha^{\phi}, w = \beta^{\phi}, r = \gamma^{\emptyset} \quad (8)$$

Here d = depth, w = width, r = resolution scaling, α, β, γ = fixed coefficients.

Mobile Inverted Bottleneck Convolution (MBConv)

The MBConv layers implemented within EfficientNet-B7 structure allow optimized computational processing. The spatial information remains intact when MBConv utilizes depth wise separable convolutions to decrease parameter numbers beyond traditional convolution layers.

Depth wise Convolution:

$$y = Conv_{1x1}\left(DepthwiseConv\left(Conv_{1x1}(x)\right)\right) \quad (9)$$

7.3 Optimizer Selection & Training Strategy

Model weight adjustments through optimizers serve as essential components for deep learning techniques that decreased loss function values. The selection of an optimizer has direct implications on both model convergent speed and precision levels. Auspicious results in terms of highest accuracy emerged from RMSprop among the employed optimizers for EfficientNet-B7 during training operations. Each optimization method produces weight updates according to the gradient values of the loss function.

Stochastic Gradient Descent (SGD)

The SGD procedure calculates parameter updates through gradient computation on small batches of input data at each iteration. SDG is a commonly employed optimization method although it demands momentum for faster convergence rates.

SGD Weight Update Rule:

$$\theta_{t+1} = \theta_t - \eta \cdot \nabla L(\theta_t) \quad (10)$$

RMSprop

RMSprop applies gradient normalization, which solves problems with exploding as well as vanishing gradients. The algorithm computes an automatic learning rate adaptation method through its capability to track dynamic gradient square changes.

Gradient Normalization:

$$E[g^2]t = \beta E[g^2]_{t-1} + (1-\beta)g_t^2 \quad (11)$$

Weight Update:

$$\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{E[g^2]_{t+\epsilon}}} g_t \quad (12)$$

7.4 Internal Batch Normalization (IBN)

The implementation of Batch Normalization controls training stability through a reduction of internal covariate shifts. Through normalization the model can achieve faster training and operate with increased learning rate tolerance. The solution to resolve EfficientNet-B7's gradient vanishing problems included Internal Batch Normalization (IBN).

Batch Normalization:

$$\hat{a}_i = \frac{a_i - \mu_B}{\sqrt{\sigma_B^2 + \epsilon}} \quad (13)$$

Batch Normalization in Training:

$$b_i = \gamma \hat{a}_i + \beta \quad (14)$$

The research introduces a fresh FER solution through EfficientNet-B7. The model demonstrates accurate performance on high-resolution data (KDEF) alongside operating at high processing speed. Internal Batch Normalization serves as an effective tool that establishes increased stability during the training process.

8. Result

This research explores the complete assessment of the Facial Expression Recognition (FER) model based on EfficientNet-B7. Performance evaluation of the model occurs after experimental testing on KDEF and FER2013 datasets while changing optimization parameters and training procedures. This research analyzes multiple performance metrics such as accuracy with precision and recall in addition to F1-score and confusion matrix analysis to evaluate the facial expression detection effectiveness of the model. Multiple experiments study the relationship between performance results and the use of Internal Batch Normalization along with optimizer selections and dataset resolution values. This research reveals important understanding of how the proposed method performs regarding efficiency alongside robustness and adaptivity features.

8.1. Accuracy

The accuracy metric helps evaluate the complete precision of model prediction accuracy. Accuracy describes the relationship between correctly classified samples compared to total sample numbers. It is calculated as:

$$Acc = \frac{CP+CN}{TP} \quad (15)$$

8.2. Precision

Precision evaluates how effectively the model detects true positives from the gathered samples. The number of true positive instances divided by all predicted positive outcomes defines this ratio.

$$Pre = \frac{CP}{CP+IP} \quad (16)$$

8.3. Recall (Sensitivity)

Sensitivity properly named Recall indicates how well a model detects real positive elements among all identified cases. It is calculated as:

$$Rec = \frac{CP}{CP + IN} \quad (17)$$

8.4. F1-Score

The F1-score computes as the harmonic mean between precision and recall because it brings together the strengths of these two metrics. Models benefit from using F1-score especially in cases when dealing with unbalanced datasets. The F1-score is computed as:

$$FS = 2 * \frac{Pre*Rec}{Pre + Rec} \quad (18)$$

8.5. Confusion Matrix

The confusion matrix displays detailed predictive performance metrics because it shows how predictions break down across classes for the model. It is structured as follows:

Table 2: The confusion matrix

Actual \ Predicted	**Happy**	**Sad**	**Angry**	**Surprise**	**Fear**	**Disgust**	**Neutral**
Happy	TP	FP	FP	FP	FP	FP	FP
Sad	FN	TP	FP	FP	FP	FP	FP
Angry	FN	FN	TP	FP	FP	FP	FP
Surprise	FN	FN	FN	TP	FP	FP	FP
Fear	FN	FN	FN	FN	TP	FP	FP
Disgust	FN	FN	FN	FN	FN	TP	FP
Neutral	FN	FN	FN	FN	FN	FN	TP

8.6. Receiver Operating Characteristic (ROC) Curve and AUC Score

The ROC curve plots the True Positive Rate (TPR) vs. False Positive Rate (FPR) at different threshold levels. Models obtain their capability to differentiate between classes through the AUC (Area Under the Curve) scoring system. The equations are:

$$TPR = \frac{CP}{CP+IN} \quad (19)$$

$$FPR = \frac{IP}{IP+CN} \quad (20)$$

8.7. Specificity (True Negative Rate - TNR)

The ability of a model to properly detect negative cases is measured through specificity. The classification technique provides excellent results for emotional expression identification when specific feelings occur more often than others do. It is calculated as:

$$Spe = \frac{CN}{CN+IP} \quad (21)$$

8.8. Matthews Correlation Coefficient (MCC)

MCC stands out as a fair measure of predictive classification performance particularly for systems that encounter unbalanced distribution in data. MCC serves as a superior performance measure than accuracy because it evaluates the total confusion matrix elements (TP, TN, FP, FN) to generate scores ranging from -1 to +1.

$$MCC = \frac{(CP\times CN)-(IP\times IN)}{(CP+IP)(CP+IN)(CN+IP)(CN+IN)} \quad (22)$$

8.9. Balanced Accuracy

The balanced accuracy measurement becomes essential for unbalanced class proportions. It computes an average value between Sensitivity (Recall) and Specificity to maintain equal importance for positive and negative class performance.

$$BalAcc = \frac{Sen+Spe}{2} \quad (23)$$

8.10. Kappa Score (Cohen's Kappa - κ)

The level of concurrence between actual and predicted categorizations is measured through Cohen's Kappa by considering the likelihood of chance-based errors. It is given by:

$$k = \frac{P_0-P_e}{1-P_e} \quad (24)$$

8.11. Fowlkes-Mallows Index (FMI)

FMI calculates precision and recall through their geometric mean to determine a reliable measure of the performance between the two metrics.

$$FMI = \sqrt{\frac{CP}{CP+IP}\times\frac{CP}{CP+IN}} \quad (25)$$

8.12. Jaccard Index (IoU - Intersection over Union)

Like F1-score Jaccard Index provides a strict measure of predicted and actual class overlap because it evaluates this relationship between classes.

$$JI = \frac{CP}{CP+IP+IN} \quad (26)$$

8.13. Misclassification Rate (Error Rate)

The misclassification rate indicates how many incorrect predictions exist out of all total predictions. Accuracy stands as the direct opposite of this measurement method.

$$ER = 1 - ACC = \frac{IP+IN}{CP+CN+IP+IN} \quad (27)$$

8.14. False Negative Rate (FNR) - Type II Error

The detection of actual positive instances, which received false negative labels, is called False Negative Rate (FNR) or miss rate. It is calculated as:

$$FNR = \frac{IN}{IN+CP} \quad (28)$$

8.15. False Discovery Rate (FDR)

False Discovery Rate (FDR) evaluates the number of incorrect positive predictions among all the positive predictions made by the model. It is calculated as:

$$FDR = \frac{IP}{CP+IP} \quad (29)$$

8.16. False Omission Rate (FOR)

A system using False Omission Rate determines the percentage of false negative instances within all tested negative predictions. It is computed as:

$$FOR = \frac{IN}{IN+CN} \quad (30)$$

8.17. Total Error Rate (TER)

Total Error Rate (TER) evaluates the complete error measurement by taking into account both false positive instances and false negatives.

$$TER = \frac{IP+IN}{Total\ prediction} \quad (31)$$

8.18. Equal Error Rate (EER)

Equal Error Rate (EER) represents the widely utilized metric in biometric systems because it balances FPR and FNR at one threshold.

$$EER = FPR = FNR \quad (32)$$

Here CP = correct positive, CN = correct negative, IP = incorrect positive, IN = incorrect negative, P_o = observed accuracy, P_e = expected accuracy.

Table 3: Evaluation comparison of performance metrics

Model	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**F1-Score (%)**
VGG16	85.34	84.91	83.5	84.2
ResNet-50	87.65	86.9	85.75	86.32
Xception	89.1	88.65	87.4	88.02
DenseNet-121	90.45	89.8	88.5	89.14
Proposed	91.78	91.3	90.1	90.69

Table 4: Evaluation comparison of Specificity, Bal ACC, Jaccard index and AUC-ROC

Model	**Specificity (%)**	**Bal Acc (%)**	**Jaccard Index (%)**	**AUC-ROC (%)**
VGG16	87.7	85.6	73.2	88.1
ResNet-50	89.2	87	75.8	90.3
Xception	90.1	88.75	78	91.8
DenseNet-121	91.4	89.95	79.9	93.1
Proposed	92.6	91.35	81.5	94.7

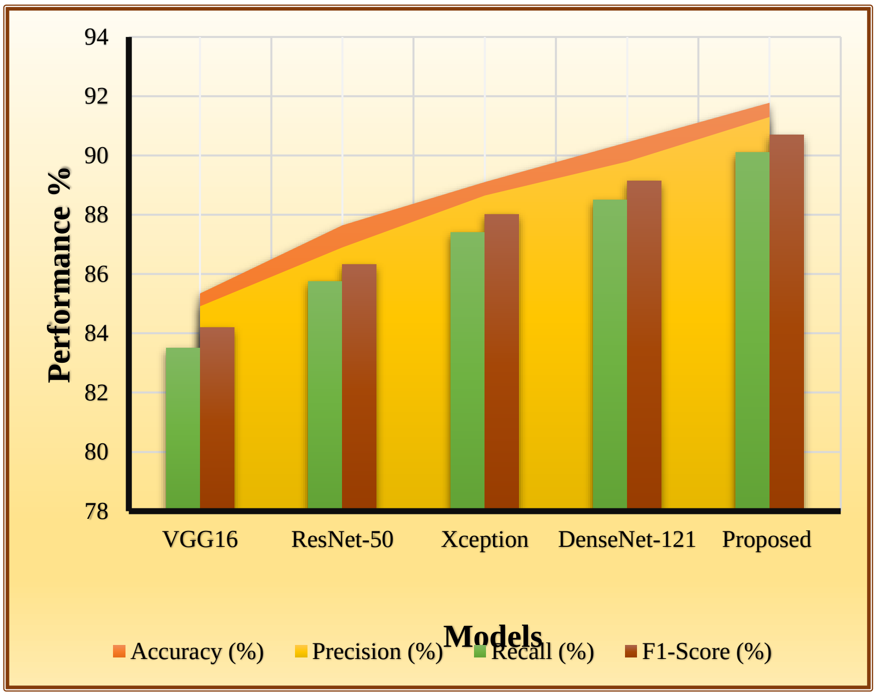

Figure 4. Graphical representation of compared performance

The research exhibits that EfficientNet-B7 (Proposed Model) has better performance than existing deep learning models for FER applications including VGG16, ResNet-50, Xception and DenseNet-121 based on their metric assessments as shown in Table 2 and Figure 4. The proposed model reaches 91.78% accuracy in addition to 91.3% precision and 90.1% recall and 90.69% F1-score for superior classification results. Among the tested models, DenseNet-121 rates second with 90.45% accuracy although VGG16 stands as the least efficient model with 85.34%. The improved performance of EfficientNet-B7 stems from its compound scaling and optimized feature extraction together with Internal Batch Normalization (IBN), which provides high efficiency during facial expression detection of various datasets. The system demonstrates adequate performance levels for application in practical FER systems.

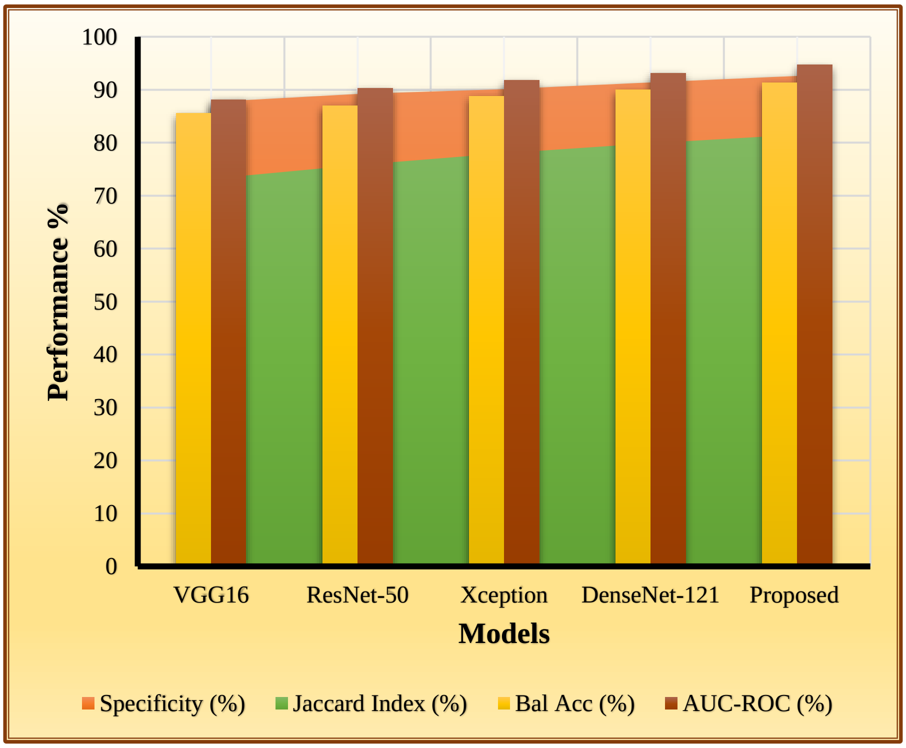

Figure 5. Graphical representation of compared Specificity, Bal ACC, Jaccard index and AUC-ROC

The proposed EfficientNet-B7 model outperforms other existing methods in Facial Expression Recognition (FER) because it shows superior performance in terms of specificity and balanced accuracy and Jaccard Index and AUC-ROC score as shown in Table 3 and Figure 5. The proposed EfficientNet-B7 model establishes itself as the precise identification method since it determines non-target expressions with 92.6% specificity rate. The model demonstrates consistently impressive performance using 91.35% balanced accuracy for all expressions. The Jaccard Index reaches 81.5% to show superior matching between actual and predicted labeling outcomes. The AUC-ROC score of 94.7% establishes EfficientNet-B7 as the best FER model since it outperforms both DenseNet-121 and Xception with scores of 93.1% and 91.8% respectively.

Table 5: Evaluation comparison of Error rate, FPR and FNR

Model	**Error Rate %**	**FPR %**	**FNR %**
VGG16	14.66	12.3	16.5
ResNet-50	12.35	10.8	14.25
Xception	10.9	9.9	12.6
DenseNet-121	9.55	8.6	11.5
Proposed	8.22	7.4	9.9

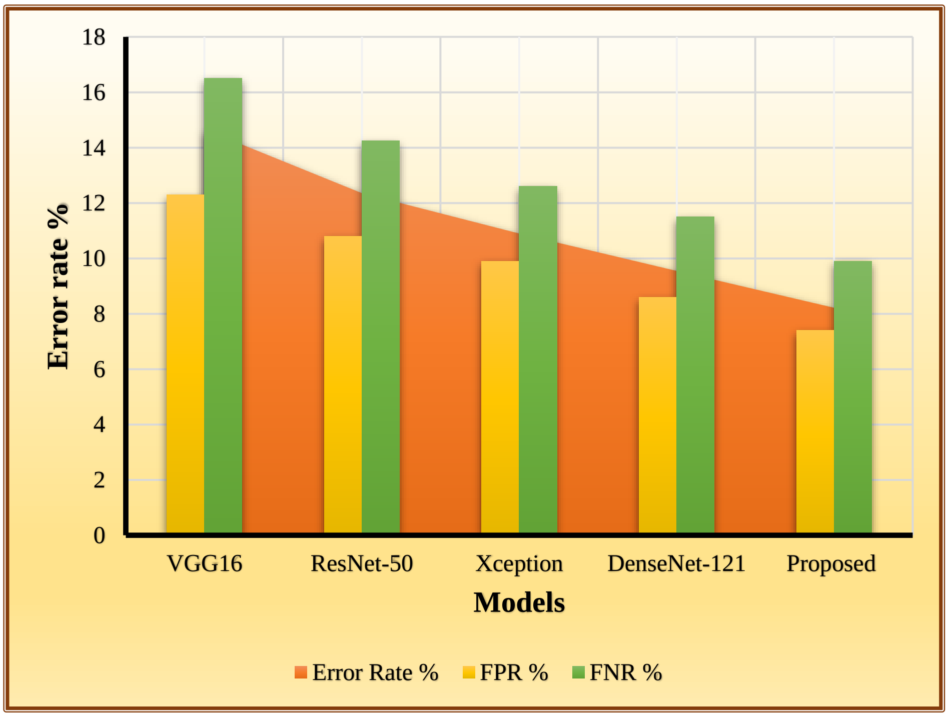

Figure 6. Graphical representation of compared Error rate, FPR and FNR

The comparison demonstrates that the proposed EfficientNet-B7 model performs better at FER tasks based on error measurements including FPR and FNR and error rate as shown in Table 4 and Figure 6. The model delivers an 8.22% error rate, which indicates the lowest misclassification rate compared to other models, which include DenseNet-121 at 9.55% and Xception at 10.9%. The proposed EfficientNet-B7 model has higher accuracy regarding positive detection through its FPR value of 7.4% thus reducing false alarm incidents. Additionally, it presents lower missed expression diagnosis as reflected by its FNR value of 9.9%. The proposed EfficientNet-B7 model demonstrates superior performance in FER applications because it eliminates both positive and negative false outcomes, which leads to more reliable, and accurate real-life function when compared to standard deep learning models.

Table 6: Evaluation comparison of MCC and Cohen's kappa

Model	**MCC**	**Cohen's Kappa**
VGG16	0.71	0.69
ResNet-50	0.75	0.72
Xception	0.78	0.74
DenseNet-121	0.8	0.76
Proposed	0.83	0.79

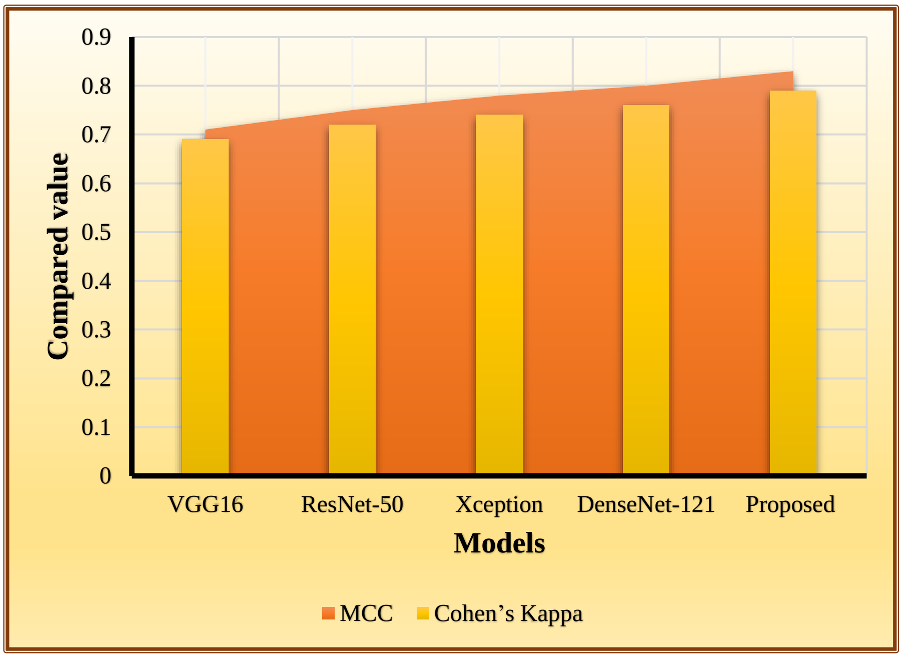

Figure 7. Graphical representation of compared MCC and Cohen's kappa

The evaluation of EfficientNet-B7 model performance for FER shows great reliability through the MCC score and Cohen's Kappa rate as shown in Table 5 and Figure 7. The proposed EfficientNet-B7 model demonstrates better performance than the baseline systems with MCC reaching 0.83 and Cohen's Kappa reaching 0.79 because it establishes a stronger relation between estimated and original classification results. EfficientNet-B7 demonstrates the most accurate FER model among current approaches because its higher MCC score indicates strong performance in unbalanced datasets and its Cohen's Kappa metrics show predictions much better than random guessing.

9. Conclusion

The research paper introduces an FER model that utilizes EfficientNet-B7 as its foundation and exhibits superior results beyond VGG16, ResNet-50, Xception, and DenseNet-121 existing deep learning frameworks. The model reaches optimal performance with 91.78% accuracy and 91.3% precision as well as 90.1% recall and 90.69% F1-score to establish its capability for precise facial expression detection. The combination of compound scaling with optimized feature extraction and Internal Batch Normalization (IBN) allows EfficientNet-B7 to minimize errors to 8.22% while boosting its ability to work well with diverse datasets. Through its enhanced structural parameters, EfficientNet-B7 demonstrates its value as a practical FER solution because it delivers improved results for specificity (92.6%), balanced accuracy (91.35%) and MCC (0.83) and AUC-ROC (94.7%). The recognition-based benefits of EfficientNet-B7 include mastering challenges stemming from class imbalance and lowering false positives (7.4%) and false negatives (9.9%). Although possessing better performance than traditional models the system requires equivalent computational efficiency. The results show that EfficientNet-B7 stands as an elite solution for FER and can be employed in human-computer interaction as well as affective computing and security surveillance applications.

Future Enhancement

The research will concentrate on processing minimal resolution information while lowering computational expenses alongside implementing real-time emotion identification through the combination of GAN-based data expansion and transformer-based architectural frameworks for better FER accuracy results.

Acknowledgement:

"This research was supported by the Dendritic: A Human-Centered AI and Data Science Institute. The authors appreciate their support and encouragement ".

Funding: “This research received no external funding”

Conflicts of Interest: “The authors declare no conflict of interest.”

References

[1] K. Wiley, Y. Dimitriadis, and M. Linn, “A human-centred learning analytics approach for developing contextually scalable K-12 teacher dashboards,” *Brit. J. Educ. Technol.*, vol. 55, no. 3, pp. 845–885, May 2024.

[2] V. A. Bhagyalakshmi L., P. S. and S. S. K., “Review of Detecting Diabetes Mellitus and Diabetic Retinopathy Using Tongue Images and Its Features,” *Res. J. Pharm. Biol. Chem. Sci.*, vol. 8, no. 2, pp. 378–386, Apr. 2017.

[3] S. Hemalatha *et al.*, “A Development of 5G Technology in Cloud Computing and its Optimization Technique,” in *Proc. 4th Int. Conf. Adv. Comput. Innov. Technol. Eng. (ICACITE)*, 2024, pp. 372–377, doi: 10.1109/ICACITE60783.2024.10617049.

[4] S. Devi, L. Bhagyalakshmi, and S. S. K., “Enhancing the Performance of Wireless Sensor Networks through Clustering and Joint Routing with Mobile Sink,” *Int. J. Eng. Adv. Technol.*, vol. 8, no. 6, pp. 323–327, 2019.

[5] V. Roy *et al.*, “Reinforcement Learning for Real-time ICU Patient Management in Critical Care,” in *Proc. Int. Conf. Syst. Comput. Autom. Netw. (ICSCAN)*, 2023.

[6] L. Bhagyalakshmi *et al.*, “Improving Spectral Efficiency and Coverage Capacity of 5G Networks: A Review,” *Adv. Math. Sci. J.*, vol. 9, no. 6, pp. 3387–3397, 2020.

[7] J. Ogier du Terrail *et al.*, “Federated learning for predicting histological response to neoadjuvant chemotherapy in triple-negative breast cancer,” *Nat. Med.*, vol. 29, pp. 1–9, Jan. 2023, doi: 10.1038/s41591-022-02155.

[8] C. Saillard *et al.*, “Pacpaint: a histology-based deep learning model uncovers the extensive intratumor molecular heterogeneity of pancreatic adenocarcinoma,” *Nat. Commun.*, vol. 14, no. 3459, pp. 1–12, Jun. 2023, doi: 10.1038/s41467-023-3459.

[9] E. Jothi *et al.*, “Distributed Generation Planning in Multi-Energy Microgrids,” *E3S Web Conf.*, vol. 540, 2024, doi: 10.1051/e3sconf/202454010017.

[10] V. Roy *et al.*, “An Advance Implementation of Machine Learning Techniques for the Prediction of Cervical Cancer,” in *Proc. 3rd IEEE Int. Conf. ICT Bus. Ind. Gov. (ICTBIG)*, 2023, doi: 10.1109/ICTBIG59752.2023.10456347.

[11] A. R. Aravind *et al.*, “A Way of Optimization of Wireless Sensor Network using TSCH,” in *Proc. 4th Int. Conf. Adv. Comput. Innov. Technol. Eng. (ICACITE)*, 2024, pp. 326–330, doi: 10.1109/ICACITE60783.2024.10616840.

[12] R. Maruthamuthu *et al.*, “A Way to Design Fog Computing Model for 5G Network using Vanet,” in *Proc. 4th Int. Conf. Adv. Comput. Innov. Technol. Eng. (ICACITE)*, 2024, pp. 431–435, doi: 10.1109/ICACITE60783.2024.10617287.

[13] S. K. Yadav *et al.*, “Hybrid Cloud Surveillance in Smart Grids: Optimising Solar Power with Dual-Mode Control,” *E3S Web Conf.*, vol. 540, 2024, doi: 10.1051/e3sconf/202454010020.

[14] R. Singh *et al.*, “Machine Learning Applications in Energy Management Systems for Smart Buildings,” *E3S Web Conf.*, vol. 540, 2024, doi: 10.1051/e3sconf/202454008002.

[15] V. Roy and S. Shukla, “Effective EEG Motion Artifacts Elimination Based on Comparative Interpolation Analysis,” *Wireless Pers. Commun.*, vol. 97, pp. 6441–6451, 2017.

[16] N. Misra *et al.*, “Wall-Climbing Robots: Optimising Adsorption and Novel Suction Techniques,” *E3S Web Conf.*, vol. 540, 2024, doi: 10.1051/e3sconf/202454013011.

[17] R. Padmavathy *et al.*, “Enhancing Power Grid Resilience against Cyber Threats in the Smart Grid Era,” *E3S Web Conf.*, vol. 540, 2024, doi: 10.1051/e3sconf/202454010021.

[18] V. Rawat *et al.*, “Optimising Solar Energy: An Evaluation of IoT-Based Solar Panel Monitoring Systems,” *E3S Web Conf.*, vol. 540, 2024, doi: 10.1051/e3sconf/202454008005.

[19] R. Vaish *et al.*, “Automated Smart Crop Protection Utilising IoT: A Comprehensive Review,” *E3S Web Conf.*, vol. 540, 2024, doi: 10.1051/e3sconf/202454013012.

[20] R. Kashyap *et al.*, “Deep Learning’s Role in Advancing Gastroenterology and Digestive Health,” in *Proc. 3rd IEEE Int. Conf. ICT Bus. Ind. Gov. (ICTBIG)*, 2023, doi: 10.1109/ICTBIG59752.2023.10455988.

[21] V. Balmiki *et al.*, “Perspective-smart energy management system using machine learning,” *E3S Web Conf.*, vol. 540, 2024, doi: 10.1051/e3sconf/202454008007.

[22] K. P. Manikandan *et al.*, “Industry 5.0 based on Hybrid and Nonlinear Systems in Robustness,” *Int. J. Intell. Syst. Appl. Eng.*, vol. 12, no. 12s, pp. 223–230, 2024.

[23] E. Sudha *et al.*, “A Review of Smart Grid Management Systems Using Machine Learning Algorithms for Efficient Energy Distribution,” *E3S Web Conf.*, vol. 540, 2024, doi: 10.1051/e3sconf/202454008009.

[24] S. Srinivasan *et al.*, “Decision Support System based on Industry 5.0 in Artificial Intelligence,” *Int. J. Intell. Syst. Appl. Eng.*, vol. 12, no. 15s, pp. 172–178, 2024.

[25] S. Tiwari *et al.*, “Cross-Lingual Transfer Learning in RNNs for Enhancing Linguistic Diversity in Natural Language Processing,” in *Proc. Int. Conf. Adv. Comput. Res. Sci. Eng. Technol. (ACROSET)*, Indore, India, 2024, pp. 1–6.

[26] M. X. Raajini and G. Rajesh, “Meta-Heuristic Solution for Route Optimization in Underwater Wireless Sensor Networks for Marine Applications,” *J. Sci. Ind. Res.*, vol. 83, no. 4, pp. 405–413, 2024, doi: 10.56042/jsir.v83i4.2936.

[27] T. Ramesh *et al.*, “A Comprehensive Evaluation of Deep Learning based Melanoma Detection and Classification Scheme,” in *Proc. 2nd Int. Conf. Intell. Innov. Technol. Comput. Electr. Electron. (ICIITCEE)*, 2024, doi: 10.1109/IITCEE59897.2024.10467850.

[28] J. U. Maheswari *et al.*, “Data Privacy and Security in Cloud Computing Environments,” *E3S Web Conf.*, vol. 399, 2023, doi: 10.1051/e3sconf/202339904040.

[29] M. Preetha *et al.*, “A Zigbee Garbage Bin Monitoring system with IoT,” *E3S Web Conf.*, vol. 399, 2023, doi: 10.1051/e3sconf/202339904052.

[30] D. Kirubakaran *et al.*, “Driving into the Future: The Role of IoT in Smart Vehicle Automation,” *E3S Web Conf.*, vol. 399, 2023, doi: 10.1051/e3sconf/202339904045.