Comprehensive Exploration of Facial Emotion Recognition using Conventional Machine Learning and Transfer learning Models

C Saravanan
Vellore Institute of Technology
M Poonkodi
poonkodi1679@gmail.com

Vellore Institute of Technology
Prem Sankar
Vellore Institute of Technology

Research Article

Keywords: Facial Emotion Recognition, Support vector machines, ResNet, EfficientNet, Accuracy, MobileNet, MUG Facial Dataset

Posted Date: May 3rd, 2024

DOI: https://doi.org/10.21203/rs.3.rs-4304090/v1

Additional Declarations: No competing interests reported.

Comprehensive Exploration of Facial Emotion Recognition using Conventional Machine Learning and Transfer learning Models

Saravanan C[1], Poonkodi M [2]* and Prem Sankar N[3]*

[1] School of Computer Science Engineering, Vellore Institute of Technology, Chennai, India
[2] School of Computer Science Engineering, Vellore Institute of Technology, Chennai, India
[3] School of Computer Science Engineering, Vellore Institute of Technology, Chennai, India
poonkodi1679@gmail.com

Abstract. Facial emotion recognition plays a vital role in enhancing human-computer interaction by allowing machines to perceive and react to human emotions. This paper conducts an in-depth exploration of various methodologies employed for recognizing facial emotions, emphasizing both traditional machine learning techniques and contemporary transfer learning models. We delve into a variety of algorithms such as support vector machines, and sophisticated neural networks like ResNet, EfficientNet, and MobileNet, assessing their efficacy using the standard MUG Facial Expression dataset. These models are tested to discern complex patterns in facial expressions, vital for accurate emotion detection. Our extensive analysis sheds light on the capabilities and constraints of each approach, providing valuable insights that pave the way for further research and practical deployments in this dynamic field. This comprehensive review aims to guide future advancements and enhance the practicality of facial emotion recognition systems.

Keywords: Facial Emotion Recognition, Support vector machines, ResNet, EfficientNet, Accuracy, MobileNet, MUG Facial Dataset

Statements and Declarations

Competing interests: This research was conducted solely by the student with academic supervision from the teacher; no external funding, support, or resources were involved.

1 Introduction

The field of facial emotion recognition (FER) is a cornerstone in advancing human-computer interaction (HCI), with profound implications in areas like security, healthcare, and entertainment [1]. The capability to precisely decode and react to human emotions is crucial for creating intelligent systems that interact with users in a seamless and intuitive manner. Historically, FER has leaned on traditional machine learning techniques, necessitating manual extraction and selection of features.

The advent of transfer learning models, however, has ushered in a new era in the domain, offering enhanced and automated methodologies for the recognition of complex emotional expressions [2]. This paper endeavors to conduct an exhaustive examination of FER, employing both classical machine learning and cutting-edge transfer learning models. We initiate our discussion by addressing the core concepts and challenges linked to FER, followed by a review of the conventional machine learning techniques historically used in this area.

Subsequently, we delve into the recent progress in transfer learning, scrutinizing the application of architectures such as convolutional neural networks (CNNs), ResNet [3], EfficientNet [4], and MobileNet [5] in FER tasks. Moreover, we assess the efficacy of these models on a widely recognized dataset in the FER research community, the MUG Facial Dataset [6]. Through this comparative analysis, we aim to underscore the advantages and limitations of each methodology, offering valuable insights for both academics and practitioners in the field.

Our exploration also touches upon the significance of data preprocessing, feature extraction, and model optimization techniques in enhancing the overall performance of FER systems. Additionally, we consider the ethical considerations and challenges associated with the implementation of FER technologies in real-world settings. Ultimately, our study seeks to contribute to the continuous development of more precise, efficient, and user-centric FER systems, facilitating more empathetic and responsive interactions between humans and machines.

2 Related Work

Al-Atroshi explored the development of expression detection through transfer learning, focusing on preprocessing steps like face detection, landmark localization, and normalization. The study utilized Convolutional Neural Networks (CNNs) for both feature extraction and classification, achieving an accuracy of 87.65%. [16]

Muhammad Wafi compared feature extraction methods (FL, LBP) for facial expression recognition using the ELM neural network. The study found that FL had superior performance, and the proposed aELM method slightly improved the basic ELM performance, reaching accuracies of 88.07% on CK+ and 83.12% on JAFFE datasets. The paper suggests potential improvements by addressing randomly generated input weights in future research. [17]

Jaiswal introduced a transfer learning-based method for facial emotion detection, evaluated on JAFFE and FERC-2013 datasets. The proposed model outperformed existing ones, achieving an accuracy of 98.65% on both datasets. [18]

Ketan Sarvakar described a network comprising six 2D convolution layers, two max pooling stages, and two fully connected layers, processing 48x48 pixel preprocessed face inputs. The network uses four different filter sizes, incorporates maximum pooling, and applies dropout to reduce overfitting, differing from traditional CNNs. [19]

C. Dalvi provided a comprehensive overview of Facial Expression Recognition (FER), highlighting the need for balanced datasets and offering comparisons of CNN models. The paper discusses challenges, proposes solutions, and outlines emerging trends, emphasizing the rich potential in this dynamic research area. [20]

Yahui Nan introduced the A-MobileNet model for FER, incorporating an attention module and dropout technology for improved feature extraction and overfitting prevention. The model outperformed lightweight MobileNet series models, achieving accuracies of 84.49% on RAF-DB and 88.11% on FER Plus, demonstrating its effectiveness in fine-grained expression analysis. [21]

Ali I's study evaluated models on CK+, JAFEE, and RAF-DB datasets. The images were preprocessed using GAN, and facial landmarks were identified using MediaPipe face mesh. The KNN model outperformed others with an accuracy of 97%. [22]

3 Proposed Work

.

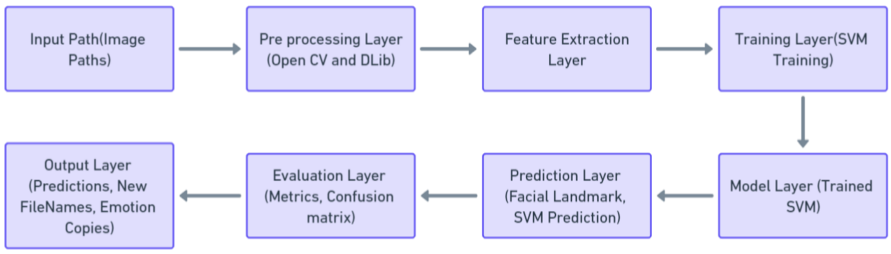

Fig. 1. Workflow of the proposed SVM Facial Annotation Architecture

The depicted workflow in Figure 1 outlines the architecture for recognizing facial expressions utilizing Support Vector Machines (SVM). The initial step involves gathering a dataset of facial images, essential for the SVM model's training, validation, and evaluation phases. Preprocessing these images includes face detection and facial landmark extraction, which are critical features for the model. These landmarks undergo normalization to maintain uniformity across the dataset. The collected data is split into subsets for training and validation. The SVM model undergoes training on the training subset to categorize various facial expressions. Optimization of the model's hyperparameters, such as the kernel type and regularization parameter, is conducted to improve its performance. To assess the model's ability to generalize to new data, cross-validation techniques are applied. Upon successful training, the SVM model is utilized to automatically label new facial images with their respective expressions.

This automated labeling process aids in efficiently categorizing facial expressions within extensive image datasets.

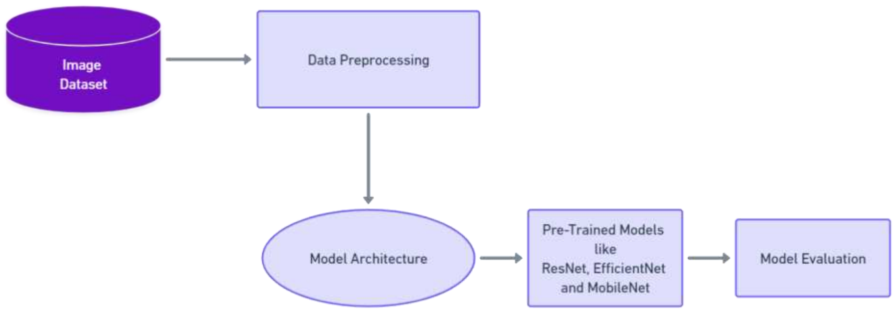

Fig. 2. Workflow of the proposed TL Pipeline

The workflow depicted in Figure 2 describes the pipeline used for developing transfer learning models for facial emotion recognition. Initially, image datasets are collected, which are essential for training, validating, and testing the transfer learning models. Following collection, the datasets undergo preprocessing, which involves resizing images to a standard size for consistency. The images are then converted into tensor format, suitable for processing by neural networks. Additionally, pixel values are normalized to range between 0 and 1 (from the original range of 0 to 255) to facilitate faster training by keeping the magnitudes of neural network weights small. The datasets are divided into three subsets for training, validating, and testing the model. Transfer learning models such as ResNet, MobileNet, and EfficientNet are employed, and after training, the model's performance is evaluated using metrics such as precision, recall, accuracy, and F1 score.

3.1 Experimental Data

In our study, we utilized the MUG Facial Expression Database [6], which contains sequences of images depicting facial expressions performed by 86 individuals. The subjects were positioned in front of a camera against a backdrop of a blue screen. Each image is stored in jpg format with dimensions of 896×896 pixels and a file size ranging between 240 and 340 KB. The database grants access to images of 52 subjects for authorized users, amounting to a total of approximately 38GB of data. The data is organized according to the subject ID, with all available sequences included in the subject's folder. The image sequences can be found in the 'take*' subfolders, with the first part of the sequences categorized into folders named anger, disgust, fear, happiness, neutral, sadness, and surprise.[6]

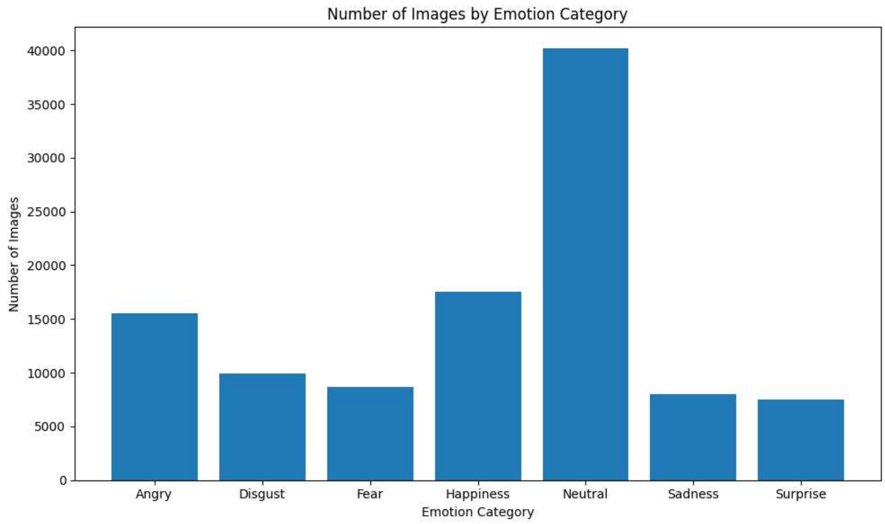

Fig 3. Number of images by Category

3.2 Data Preprocessing

In the SVM Facial Annotations workflow, the preprocessing of facial images involves several mathematical and algorithmic steps aimed at refining the data for effective classification by a Support Vector Machine (SVM) model. Initially, face detection is executed using a Haar Cascade classifier. This involves calculating the Haar features in each image and using these features in a cascade of boosted classifiers.

Face Detection with Haar Cascade where Haar features are computed as differences in pixel intensities across adjacent rectangular regions, rapidly computed via integral images.

$$Face\ Region = CascadeClassifier(Haar\ Features)$$

$$S(x, y) = \sum x' \leq x, y' \leq yI(x', y')$$

Once the region of interest (face) is detected, it is converted to grayscale to simplify the image data, reducing it to a single intensity value per pixel.

Once the region of interest, or the face, is detected, it is converted to grayscale to simplify the image data, reducing it to a single intensity value per pixel using the formula.

$$I_{Gray}(x, y) = 0.299 \times I_R(x, y) + 0.587 \times I_G(x, y) + 0.114 \times I_B(x, y)$$

Where I_R, I_G and I_B the red, green, and blue components of the image, respectively.

Subsequently, facial landmarks are identified using the dlib library's shape predictor, a regression tree-based ensemble learning tool. The shape predictor models the relationship between pixel intensities around candidate landmark regions and the actual landmark positions, trained using a gradient boosting method. These landmarks delineate crucial facial features and their spatial arrangement, effectively transforming the facial features into a vector of coordinates. This vector serves as input for the SVM classifier, ensuring the SVM receives high-quality, relevant features for classification and enhances its ability to accurately recognize facial expressions.

To ensure consistency across different images, the extracted landmarks undergo a normalization process, involving scaling and translation to align them in a standard coordinate frame. This step helps in minimizing variations due to differences in face sizes and orientations. The normalized landmarks are then flattened into a one-dimensional array, converting the coordinate data into a format that can be used as input features for the SVM classifier.

In parallel, the emotion labels associated with each image, which represent the facial expressions, are encoded into numerical values using a Label Encoder. This encoding is necessary because machine learning models, including SVM, require numerical labels for the classification task. These preprocessing steps collectively transform the raw facial images into a structured and standardized form, making them suitable for training and evaluating the SVM model for facial expression recognition. This process ensures that the model can effectively learn from the features and make accurate predictions about the emotions conveyed in new, unseen images.

In our transfer learning framework, the initial step of preprocessing is crucial for the effective training of models. The process begins by standardizing the sizes of images to ensure that every input fed into the neural network is consistent in dimension. This uniformity is vital as it eliminates any bias toward varying image scales, allowing the network to learn from the data more efficiently. Once the images are resized, they are converted into tensors. Tensors, essentially multi-dimensional arrays, are the preferred format for numerical input in transfer learning systems. This conversion enables the images to be easily manipulated and processed by the underlying algorithms.

Following the conversion to tensors, normalization is applied, where the range of pixel values is adjusted to a standard scale, usually between 0 and 1. This is achieved by dividing the original pixel values, which range from 0 to 255, by 255. Normalization is not merely a procedural step; it is a crucial phase that aids in accelerating the convergence of the training process. It does this by regulating the gradients and enhancing the stability and consistency of the changes in the network's weights during training. Additionally, the dataset is split into two distinct sets: a training set and a test set. The training set acts as the instructor for the model, enabling it to learn and make predictions, while the test set serves as an evaluator, assessing the neural network's predictive capability on unseen data.

Beyond preparing the data for input, the preprocessing phase also aims to enhance the quality of the data used for learning. This often involves advanced techniques such as data augmentation, which includes applying various transformations like rotation, scaling, and flipping to artificially expand the original images. This not only increases the volume of training data but also introduces a level of variation that can strengthen models, enabling them to perform well when faced with new and diverse datasets. After preprocessing, the data is ready for the subsequent stage of transfer learning, where the model is trained. The data is now clean, well-structured, and primed for optimal learning.

3.3 Dataset Preparation and Facial Annotation using SVM

In this project, we undertook the detailed task of creating and annotating a dataset for recognizing facial emotions using a Support Vector Machine (SVM) classifier. The process involved several crucial steps to ensure precision and effectiveness. Initially, we used the Haar Cascade classifier from OpenCV [7] to detect faces in each image. Following this, we utilized the dlib library [8] to extract 68 key facial landmarks, such as the corners of the mouth and the tip of the nose. These landmarks were then flattened into a one-dimensional array for use as input features in our SVM model.

$$Landmarks = shape\_predictor(I_{Gray}\ , Detected\ Face)$$

The shape predictor estimates the connection between the intensity of pixels surrounding potential landmark areas and their true positions, trained using a gradient boosting technique. These landmarks outline essential facial features and their layout, essential for facial expression analysis. They convert the facial features into a coordinate vector, serving as input for the SVM classifier.

To prepare our data for classification, we extracted emotion labels from the image filenames, encoded as two-letter abbreviations (e.g., 'ha' for happiness, 'sa' for sadness). These labels were then converted into numerical values using the LabelEncoder. We split our dataset into training and validation sets, using the training set to train an SVM classifier for multi-class classification. The SVM model was then tested on the validation set, where we computed various metrics such as accuracy, precision, recall, and F1 score [9] to assess the model's ability to accurately classify emotions.

For each image in our dataset, the trained SVM model predicted the emotion based on the facial landmarks, and this predicted emotion was used to annotate the image.

The annotated images were then saved in a separate directory, organized into subdirectories based on their predicted emotions, which facilitates further analysis and usage of the annotated dataset. Additionally, to increase the diversity and robustness of our dataset, we included images from a separate set of subjects and performed the same annotation process, exposing our model to a wider range of facial expressions and variations.

The outcome of this thorough process is a richly annotated dataset of facial images with emotion labels, which holds significant potential for various applications in emotion recognition, human-computer interaction, and beyond. This dataset not only serves as a valuable resource for training and testing emotion recognition models but also contributes to the advancement of research in understanding and interpreting human emotions through facial expressions.

WORKFLOW 1: FACIAL ANNOTATION USING SVM		
	IMPORT NECESSARY LIBRARIES	
	LOAD FACE CASCADE CLASSIFIER AND FACIAL LANDMARK PREDICTOR	
	DEF DETECT_FACES_AND_LANDMARKS(IMAGE_PATH, MAX_LANDMARKS):	
		READ THE IMAGE AND CONVERT IT TO GRAYSCALE
		INITIALIZE AN EMPTY LIST FOR LANDMARKS
		FOR EACH FACE DETECTED, PREDICT LANDMARKS AND ADD TO THE LANDMARKS LIST
		RETURN THE FIRST MAX_LANDMARKS LANDMARKS
	ENCODE EMOTION LABELS TO NUMERICAL VALUES USING LABELENCODER	
	BUILD AND TRAIN AN SVM MODEL FOR MULTI-CLASS CLASSIFICATION AND SAVE THE TRAINED SVM MODEL	
	MAKE PREDICTIONS FOR EACH IMAGE USING THE TRAINED SVM MODEL:	

	FOR EACH IMAGE PATH, DETECT LANDMARKS AND PREDICT THE EMOTION, EXTRACT INFORMATION FROM THE IMAGE PATH
	STORE THE PREDICTION IN A LIST AND COPY THE IMAGE TO THE CORRESPONDING EMOTION FOLDER

Table 1. The pseudocode for workflow of Facial Annotation using SVM Model

3.4 Transfer Learning Architectures and Evaluation

The Figure 2 shows the complete workflow of the proposed Transfer learning architecture. The architecture shows detailed steps taken for emotion recognition using transfer learning. We first loaded in the dataset after proper annotation using the SVM model. And pre-processed the dataset using various preprocessing steps like resizing, random cropping, flipping, rotation, and color adjustments, to improve the robustness of the model against variations in input data. All images were converted into tensors and normalized to ensure Consistent input distribution, which is crucial for the stable convergence of transfer learning models.

In our approach, we utilized a combination of sophisticated transfer learning models: EfficientNet [13], MobileNet [14], and ResNet [15]. Each of these architectures was selected for its proven effectiveness in classifying images, offering distinct advantages. The ResNet model, particularly the ResNet-50 variant, is renowned for its ability to overcome difficulties in training deep networks through the use of residual learning. MobileNet stands out for its efficiency, making it ideal for scenarios with limited computational resources. EfficientNet, meanwhile, offers a scalable architecture that delivers superior accuracy while requiring fewer parameters.

The evaluation of the transfer learning models extended beyond simple accuracy. Precision, recall, and F1 score metrics were used to provide a comprehensive assessment of the model's predictive performance. Precision measured the accuracy of positive predictions, recall indicated the model's ability to find all relevant instances within the dataset, and the F1 score provided a balance between precision and recall, particularly useful in the context of class-imbalanced datasets.

Accuracy measures the proportion of correctly classified instances out of the total instances.[12]

$$\text{Accuracy} = \frac{No.of\ Correct\ Predicted\ Samples}{Total\ Number\ of\ Samples}$$

Precision measures the proportion of correctly identified positive instances out of all instances classified as positive.[10]

$$\text{Precision} = \frac{True\ positives}{(True\ positives\ +\ False\ positives)}$$

Recall (Sensitivity) measures the proportion of correctly identified positive instances out of all actual positive instances.[9]

$$\text{Recall} = \frac{True\ positives}{(True\ positives\ +\ False\ Negative)}$$

The F1 score is the harmonic mean of precision and recall, providing a balance between the two metrics.[11]

$$F1 = \frac{2\times Precision\ \times Recall}{Precision+Recall}$$

Additionally, a confusion matrix was employed as a visual tool to understand the classification performance for each emotion category. This matrix detailed the instances of correct and incorrect predictions, offering a granular view of the model's strengths and weaknesses across the emotional spectrum.

4 Results and Discussion

The Table 2. Shows The performance metrics for SVM Facial Annotations reveal varied effectiveness across different emotions. Anger demonstrates moderate precision (62%) and recall (67%), with an F1-score of 64%, indicating the model's reasonable reliability, but highlighting the need for refinement. Disgust is identified with high accuracy, as evidenced by its precision (92%), recall (86%), and F1-score (89%). Fear's balanced performance is reflected in its F1-score of 84%. Both happiness and surprise achieve perfect precision scores of 100% and exhibit high recall, leading to impressive F1-scores of 97% and 92%, respectively. This suggests the model's exceptional proficiency in recognizing these emotions. Neutral emotion is characterized by a notably high recall, indicating its consistent detection, while sadness boasts perfect precision, showcasing its precise identification. Overall, the SVM classifier excels in detecting emotions like happiness, surprise, and disgust, though it shows some limitations in accurately identifying anger.

The performance Indicators for the SVM model reveal a high degree of precision In its outcomes, with an overall accuracy reaching 87%. The model's precision rate is a notable 88.3%, highlighting its capability to generate accurate predictions. Additionally, its recall rate of 87.12% demonstrates the model's effectiveness in capturing all essential instances. An F1 Score of 87.34% is indicative of a well-tuned model that strikes an excellent balance between exactness and completeness in its results. Such metrics denote the model's dependable performance in data classification tasks, establishing its significance in the realm of predictive analytics.

Table 2. Precision, Recall and F1-Score of the SVM model for facial annotation

Emotion	Precision	Recall	F1-score
Angry	62%	67%	64%
Disgust	92%	86%	89%
Fear	80%	89%	84%
Happiness	100%	94%	97%
Neutral	83%	96%	89%
Sadness	100%	80%	89%
Surprise	100%	85%	92%

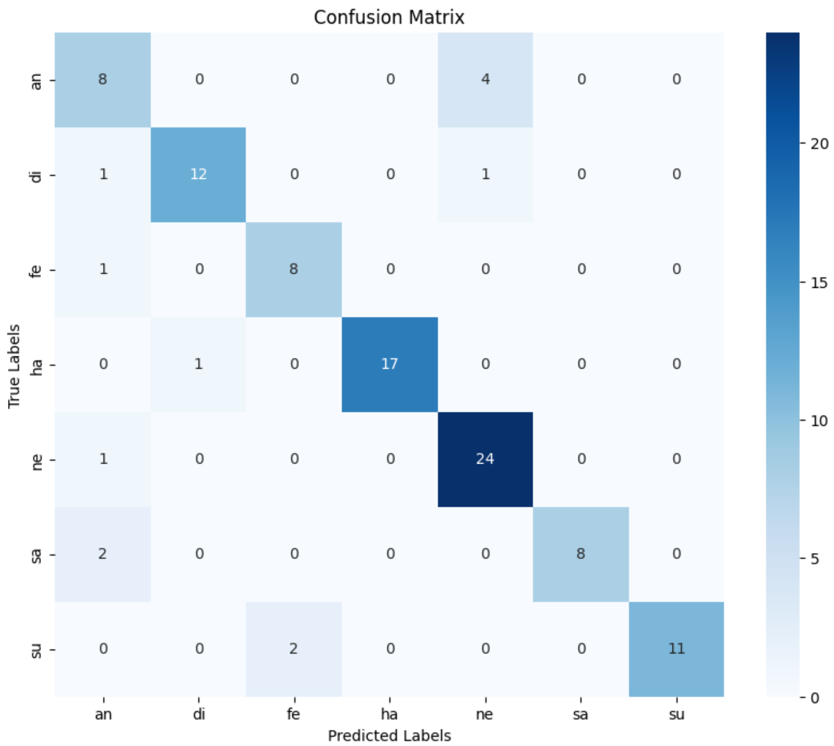

Fig 3. Confusion Matrix of SVM Model for every emotion

Table 3. Transfer learning algorithm's evaluation metrics in Percentage.

Algorithm	Accuracy	Precision	Recall	F1-score
MobileNet	73.66%	74.84%	73.66%	72.17%
ResNet	67.13%	72.20%	67.13%	66.05%
EfficientNetB0	76.69%	77.30%	76.69%	76.56%

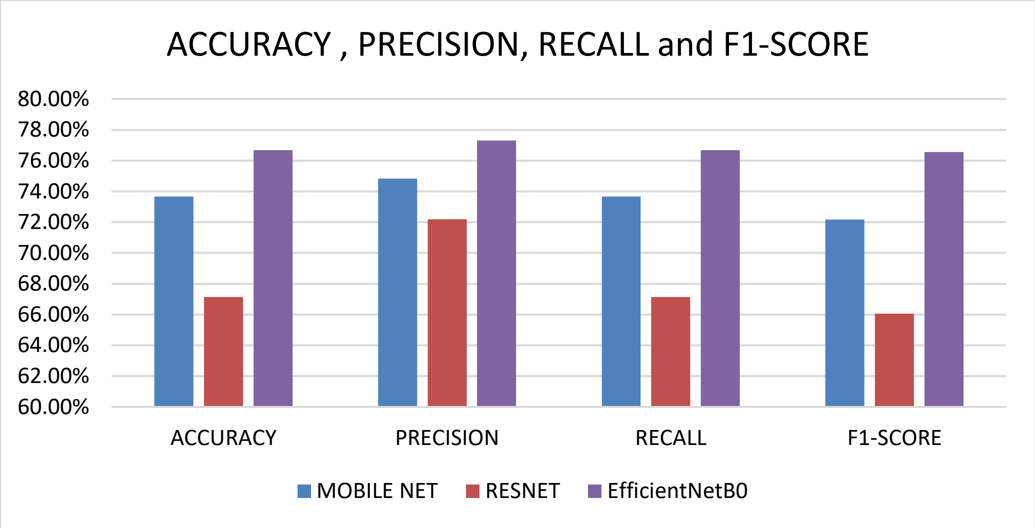

Fig. 4. Evaluation metrics in % of the used transfer learning algorithms

Table 3 shows the comparative analysis of three advanced transfer learning models reveals the following: MobileNet demonstrates a uniform performance with an accuracy, recall, and precision in the mid-70s percentile, and an F1-score slightly below at 72.17%, suggesting reliable outcomes across the board. ResNet, while having the modest numbers among the trio, posts an accuracy and recall of 67.13%, coupled with a slightly higher precision of 72.20%, indicating its effectiveness in making correct predictions when it identifies cases, despite a lower detection rate as reflected in its F1-score of 66.05%. EfficientNetB0 emerges as the frontrunner with superior accuracy and recall rates at 76.69% and the highest precision of 77.30%, culminating in an F1-score of 76.56%, which signals its superior balance in both accurate and consistent prediction capabilities.

5 Conclusion

In summarizing this work, the initial focus was on grasping the significance of recognizing facial emotions. A dataset was curated for this purpose, utilizing transfer learning to detect facial emotions. The SVM Model was employed for annotating facial expressions through landmarks, achieving a commendable overall accuracy of 87%. Subsequent to this, a comparative evaluation of transfer learning models was undertaken, examining ResNet, MobileNet, and EfficientNet. MobileNet demonstrated consistent results, with accuracy, precision, and recall all in the approximate 74% range and an F1-score at 72.17%. ResNet presented lower performance with scores mainly around 67%, yet it had a relatively higher precision of 72.20%, indicating its effectiveness in making correct predictions, despite missing some cases. Efficient-

NetB0 surpassed its peers with the topmost scores—an accuracy of 76.69%, precision of 77.30%, and both recall and an F1-score just slightly lower at 76.56%, positioning it as the most proportionate and precise model among those tested, thus the most trustworthy for regular predictive tasks. Advancements could be made through further refinement of the models' parameters and, with increased computational resources, the models could potentially be executed on the entire dataset instead of a fraction, possibly enhancing performance further.

References

1. Corneanu, C. A., Simón, M. O., Cohn, J. F., & Guerrero, S. E. (2016). Survey on RGB, 3D, thermal, and multimodal approaches for facial expression recognition: History, trends, and affect-related applications. IEEE Transactions on Pattern Analysis and Machine Intelligence, 38(8), 1548-1568.
2. Li, S., Deng, W., & Du, J. (2017). Reliable crowdsourcing and deep locality-preserving learning for expression recognition in the wild. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2852-2861.
3. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 770-778.
4. Tan, M., & Le, Q. V. (2019). EfficientNet: Rethinking model scaling for convolutional neural networks. Proceedings of the 36th International Conference on Machine Learning (ICML), 6105-6114.
5. Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., ... & Adam, H. (2017). MobileNets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861.
6. Aifanti, N., Papachristou, C., & Delopoulos, A. (2010). The MUG facial expression database. 11th International Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS), 1-4.
7. Bradski, G. (2000) The OpenCV Library. Dr. Dobb's Journal of Software Tools, 120; 122-125.
8. King, D.E. (2009). Dlib-ml: A Machine Learning Toolkit. Journal of Machine Learning Research, 10, 1755-1758.
9. Powers, D.M.W. (2011). Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness and Correlation. Journal of Machine Learning Technologies, 2(1), 37-63
10. Davis, J., & Goadrich, M. (2006). The relationship between Precision-Recall and ROC curves. Proceedings of the 23rd International Conference on Machine Learning, 233-240.
11. Chinchor, N. (1992). MUC-4 evaluation metrics. Proceedings of the 4th Conference on Message Understanding, 22-29.
12. Fawcett, T. (2006). An introduction to ROC analysis. Pattern Recognition Letters, 27(8), 861-874.
13. Hoang, V. T., & Jo, K. H. (2021, July). Practical analysis on architecture of EfficientNet. In 2021 14th International Conference on Human System Interaction (HSI) (pp. 1-4). IEEE.

14. Sinha, D., & El-Sharkawy, M. (2019, October). Thin mobilenet: An enhanced mobilenet architecture. In 2019 IEEE 10th annual ubiquitous computing, electronics & mobile communication conference (UEMCON) (pp. 0280-0285). IEEE.
15. Wu, Z., Shen, C., & Van Den Hengel, A. (2019). Wider or deeper: Revisiting the resnet model for visual recognition. Pattern Recognition, 90, 119-133.
16. Facial Expression Recognition Based on Transfer learning: An Overview. (2023). Iraqi Journal of Science, 64(3), 1401-1425.
17. Wafi, Muhammad & Bachtiar, Fitra & Utaminingrum, Fitri. (2023). Feature extraction comparison for facial expression recognition using adaptive extreme learning machine. International Journal of Electrical and Computer Engineering (IJECE).
18. A. Jaiswal, A. Krishnama Raju and S. Deb, "Facial Emotion Detection Using Transfer learning," 2020 International Conference for Emerging Technology (INCET), Belgaum, India, 2020, pp. 1-5
19. Ketan Sarvakar, R. Senkamalavalli, S. Raghavendra, J. Santosh Kumar, R. Manjunath, Sushma Jaiswal,Facial emotion recognition using convolutional neural networks,Materials Today: Proceedings,Volume 80, Part 3,2023,Pages 3560-3564,ISSN 2214-7853.
20. Dalvi, Chirag & Rathod, Manish & Patil, Shruti & Gite, Shilpa & Kotecha, Ketan. (2021). A Survey of AI-based Facial Emotion Recognition: Features, ML & DL Techniques, Age-wise Datasets and Future Directions. IEEE Access.
21. Yahui Nan, Jianguo Ju, Qingyi Hua, Haoming Zhang, Bo Wang,A-MobileNet: An approach of facial expression recognition,Alexandria Engineering Journal,Volume 61, Issue 6,2022,Pages 4435-4444,ISSN 1110-0168,
22. Ali I. Siam, Naglaa F. Soliman, Abeer D. Algarni, Fathi E. Abd El-Samie, Ahmed Sedik, "Deploying Machine Learning Techniques for Human Emotion Detection", Computational Intelligence and Neuroscience, vol. 2022, Article ID 8032673, 16 pages, 2022.