Applied Computing Journal https://doi.org/10.52098/acj.20255466
ISSN: 2788-9688 Volume 5, Issue 4, pp 439-451, October 2025
Received: 10/25/2024 Revised: 23/08/2025 Accepted: 24/08/2025

Deep Transfer Learning and Feature Fusion for Improving Facial Expression Recognition on JAFFE Dataset

Merdin Shamal Salih[1], Nechirvan Asaad Zebari[2], Reving Masoud Abdulhakeem[3,*], and Dilovan Asaad Zebari[3,4]

[1] Department of Computer Science, Cihan University-Duhok, Duhok, Kurdistan Region, Iraq.
[2] Department of Information Technology, Lebanese French University, Erbil, Kurdistan Region, Iraq.
[3] Department of Computer Science, College of Science, Nawroz University, Duhok, Kurdistan Region, Iraq.
[4]Department of Information Technology, Technical College of Informatics, Akre University for Applied Sciences, Akre, Kurdistan Region, Iraq.

* Corresponding author: revinkmasoud@nawroz.edu.krd.com

Abstract

Facial Expression Recognition (FER) is extensively used in human-computer interaction where machines can recognize people's emotions through facial expressions. In this study, a hybrid FER framework was proposed using MobileNetV2 deep learning features combined with traditional handcrafted descriptors LBP and HOG to enhance classification performance on small datasets. Additionally, we evaluate our approach using the Japanese Female Facial Expression (JAFFE) dataset, which consists of 213 grayscale images showing seven basic emotions. Data augmentation and transfer learning were applied to increase model generalization. The feature fusion scheme leveraged deep semantic features and local texture descriptors. The feature fusion scheme was used to leverage the benefits of deep semantic features and local texture descriptors, integrated via a dimensionality reduction and classification module with CNN-based architecture. The hybrid method achieved 96.49% accuracy, outperforming MobileNetV2 alone (94.73%) and handcrafted features (95.17%). This demonstrates both the utility of feature fusion to enhance FER accuracy in constrained datasets and indicates the possibility for more reliable emotion recognition systems in live applications.

***Keywords**:* Facial expression recognition, LBP, HOG, deep learning features, fusion.

1. Introduction

Facial expressions (FEs) are the movement of facial muscles which indicate emotional states in a person. Facial expressions are non-verbal communication and so they erect a sort of universal language as well, which surpasses the borders of culture or even language itself. Instead, they help convey emotions like joy, frustration, sorrow, fear, shock and yuck most of the time regardless of whether they speak or not (Liao et al., 2023; Ibrahim et al., 2021). Facial Expression Recognition (FER) is the process of automatically detecting and classifying the facial expression of an individual from digital content, images or videos via computer vision algorithms. FER systems want to mimic the capacity of individuals to understand facial prompts supporting machines that emerge UORBS in real-time to identify emotional states (Li et al., 2020; Liang et al., 2023).

Conventional FER systems mainly work with handcrafted feature descriptors such as Local Binary Patterns (LBP), Histogram of Oriented Gradients (HOG) and Gabor filters (Al-Hatmi & Yousif, 2022). Even though the above methods are computationally efficient and interpretable, they tend to generalize poorly in complex or real-world scenarios. Handcrafted features also often do not encode high-level semantic information that could be important for achieving robust classification. In contrast, deep learning algorithms, especially Convolutional Neural Networks (CNNs) have achieved competitive results on visual recognition by learning hierarchical features from data (Al Kishri et al., 2025). FER studies, particularly on large and diverse FER datasets, CNNs have shown better performance than other types of architectures. However, when used on small datasets like JAFFE CNNs are susceptible to overfitting and may have poor performance (Ni et al., 2022; Zebari et al., 2021; Rusia & Singh, 2021).

To remedy either the obstacles linked with pure heuristic-based handcrafted models and deep learning methods, there are some works from recent research which empower mix-mode devices which leverage each individual. Feature fusion methods integrate low-level texture-based features with high-level deep features generated by pretrained CNNs. Such integration potentially increases the discriminative power of the model and still allows them to maintain computational efficiency. Although small, the Japanese Female Facial Expression (JAFFE) dataset still holds as staple a cornerstone within the FER scientific community. For applying and comparing various recognition techniques, its controllable setting and labeled faces are also ideal, especially after using some of the latest hybrid methods that would work fine as soon as you have one sample per class.

To that purpose in this study, the main goal is to train and test a hybrid facial expression recognition model integrating hand-crafted feature descriptors with deep learning to enhance the classification accuracy on JAFFE dataset. The objectives of this study are to (i) Apply transfer learning using state-of-the-art CNN architectures pretrained on large datasets, (ii) Extract handcrafted features using LBP and HOG descriptors, (iii) Design feature fusion strategy for integrating deep and handcrafted features, (iv) Develop machine leaning classifiers performance evaluation-based classification in fused feature space, and (v) Compare the proposed method with traditional methods including deep learning. Summary The contributions of this paper are summarized as follows:

- A hybrid facial expression recognition scenario based on handcrafted features (LBP and HOG) and deep CNN feature fusion.
- Design feature fusion mechanism for self-expression capability and robustness, especially on small datasets such as JAFFE.
- Benchmarking against standard metrics baselines, including handcrafted and deep models on a comprehensive evaluation for the proposed approach.

- An assessment of experimental outcomes, showing the improved results in both classification accuracy and generalization capabilities compared to pure feature based and deep learning methodologies using hybrid technique.

2. Related Work

Facial expression recognition (FER) is one of the general research topics in recent years, and many studies aimed for higher recognition accuracy by using diverse ways. The traditional approach of handcrafting features like Local Binary Patterns (LBP) and Histogram of Oriented Gradients (HOG). However, the methods can perform poorly with complex variations in facial expressions. To solve these problems, recent research has shifted to deep learning methods that can learn hierarchical features automatically through data.

Here are recent works where they show that combining handcrafted features with deep learning models can improve FER performance: For example, research by Obaid and Alrammahi (2023) introduced a hybrid emotional feature extraction network model that combines Convolutional Neural Networks (CNNs) with handcrafted descriptors to enhance the discriminative power of emotional features for multimedia purposes. It divided the overall facial features into global and local representation, which made it capture both long-ranged dependencies and short-range textures more efficiently to further improve recognition accuracy. Similarly, Chouhayebi et al. introduced a dynamic fusion method based on HOG-TOP and deep learning features that have been tested in the REACT 2023. It employed CNN to extract facial features and fused with dynamic texture feature by using the HOG-TOP descriptor. They fused and learned different types of features with a compact Bilinear model and experienced better results on the INTERFACE05 dataset which was due to performance gains from their Multi-modal Compact Bilinear (MCB) algorithm.

Feature fusion strategies are essential for combining data from various sources efficiently. A recent paper by Li et al. Feature Aggregation: Chang et al. (2025) presented a full-size feature aggregation method that aggregates multiple groups of features, resulting in high quality aggregated feature maps from the detected objects on an image. It gave a good fusion technique for integrating the global and local facial features deeply, to improve the performance of emotion recognition models. This paper by (Chouhayebi et al., 2024) aims to introduce a new method for recognizing facial expressions using deep learning in conjunction with dynamic texture analysis. The proposed model is a combination of encoded features by VGG19, and spatio-temporal data using LSTM. Features were extracted from dynamic facial changes with HOG-HOF descriptors. Skill, Environment and Action features are mixed via Multimodal Compact Bilinear (MCB) pooling and classify with a Support Vector Machine (SVM) to guarantee interpretability. Our method was able to outperform state-of-the-art methods with a margin of over 1% on the eNTERFACE05 dataset, which has been shown to increase accuracy and robustness in emotion recognition. To detect emotions in video data, in Manalu and Rifai (2024) FER using hybrid CNN-RNN architecture. Using the Emotional Wearable Dataset 2020 that includes additional emotions like amusement, enthusiasm, awe, and liking the study proposes three models: MobileNetV2-RNN, InceptionV3-RNN, and a custom CNN-RNN. InceptionV3-RNN had the best performance, having an accuracy of 66% of them. The models achieve good performance for fine-grained emotional states, which not only advances the state-of-the-art of FER but also provides significant potential use in cognitive science and interactive communication systems.

3. Proposed Method

In this work a hybrid framework has been introduced for FER using deep learning and handcrafted features extraction methods as shown in Figure 1. In this work we propose this design to circle around the problem of using only hand-crafted descriptors on small dataset and enhance deep learning model with local texture patterns. Global Pipeline: The four-steps pipeline comprises data preprocessing, feature extraction MobileNetV2 features and handcrafted descriptors, fusion of features and classification by deep or non-deep classifiers.

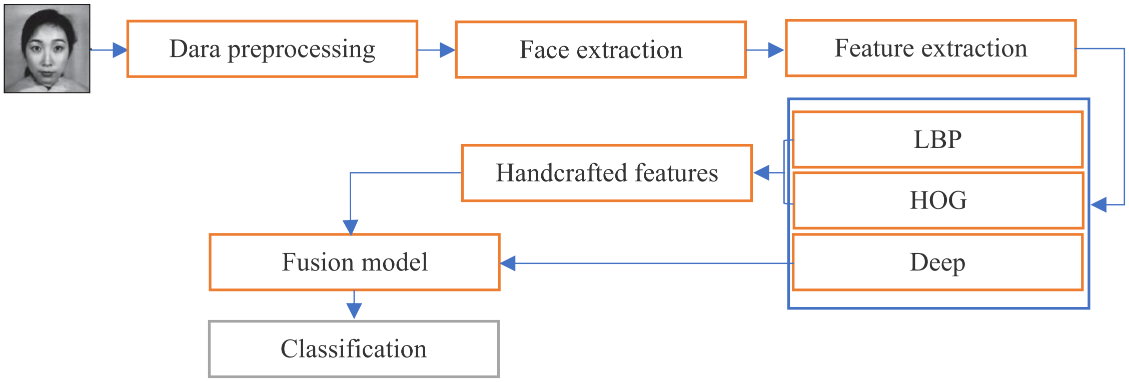


Figure 1: Proposed Model

3.1. Preparing Data

In this work, we use the JAFFE dataset which contains 213 greyscale facial images of seven expressions i. e., anger, disgust, fear, happiness, sadness, surprise and neutral. The resizing was performed to support a constant size of 224 × 224 pixels on the input required by MobileNetV2, which facilitates the extraction of handcrafted descriptors from them. In order to help with generalization, we used different image augmentation approaches like horizontal flips, ±15° rotations, brightness variations and zoom in/out operations. The dataset was split into training ensuring liability distinctions (133 images), validation (23 images) and testing (57 images). In the model training, we preprocessed images to [0, 1] by normalization of pixel intensities.

3.2. Feature Extraction

One of the stages in facial expression recognition is feature extraction, which converts raw image data into an informative format with a focus on relevant attributes of faces. This helps to segment the more discriminatory features, those which may include edges and textures in addition to specific patterns that are needed for robust classification of different emotions. To date, many traditional as well as deep learning-based and handcrafted features are employed for facial expressions spatial and structural information capturing. Feature extraction plays a significant role in the accuracy and robustness of emotion classification systems.

3.3.1 Deep Features

For deep spatial feature extraction, we selected MobileNetV2, a light-weight architecture already established as efficient in vision tasks. These modeling choices of depth wise separable convolutions and inverted residual connections are light weight which is appropriate for training with small data sizes but in constrained computational resources (Ma et al., 2021; Srinivasuet al., 2021). In the other network, a pretrained MobileNetV2 model was used (with its final classification layers removed). We then obtained features from the output of the global average pooling (GAP) layer, which yielded 1280-dimensional feature vector for each image as shown in Figure 2. Furthermore, the JAFFE dataset was fine-tuned to adapt the model for FER tasks, improving the performance of deep features in expression-specific patterns.

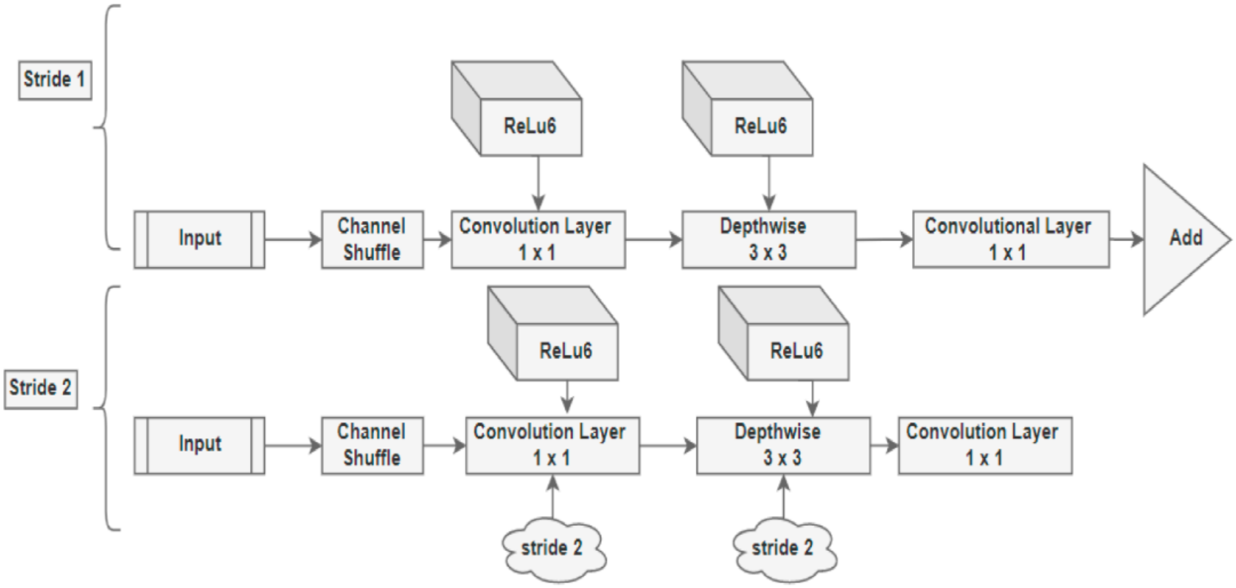


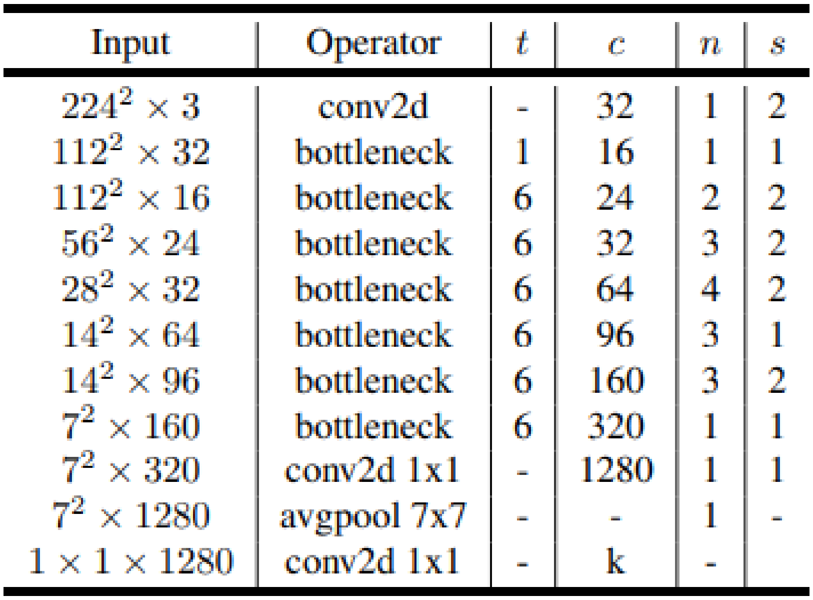
Figure 2: The Architecture of MobileNetV2 Model

Each line describes a sequence of 1 or more identical (modulo stride) layers, repeated *n* times. All layers in the same sequence have the same number *c* of output channels. The first layer of each sequence has a stride *s* and all others use stride 1. All spatial convolutions use 3 × 3 kernels. The expansion factor *t* is always applied to the input size as described in Table 1 (Sandler et al., 2018).

3.3.2 Handcrafted Features

After work on deep features, handcrafted descriptors were extracted at multiple resolutions to preserve small texture and edge patterns (Zebari et al., 2020; Perera and Patel, 2019). The edge and shape related information were encoded using HOG, we computed HOG features with the block size set to 2×2, cell size of 8×8, and used an orientation bin setting at 9. To ensure the micro-textures were well-modelled, LBP was used by discretizing the pixel neighborhood. We used the uniform LBP method with parameters P=8 and R=1 because of its stability in lighting changes and rotation. In order to make them compatible, both HOG and LBP feature vectors were normalized using z-score normalization, reducing variance across feature scales.

Table 1: The basic implementation structure of MobileNetV2 Architecture.



Input	Operator	t	c	n	s
$224^2 \times 3$	conv2d	-	32	1	2
$112^2 \times 32$	bottleneck	1	16	1	1
$112^2 \times 16$	bottleneck	6	24	2	2
$56^2 \times 24$	bottleneck	6	32	3	2
$28^2 \times 32$	bottleneck	6	64	4	2
$14^2 \times 64$	bottleneck	6	96	3	1
$14^2 \times 96$	bottleneck	6	160	3	2
$7^2 \times 160$	bottleneck	6	320	1	1
$7^2 \times 320$	conv2d 1x1	-	1280	1	1
$7^2 \times 1280$	avgpool 7x7	-	-	1	-
$1 \times 1 \times 1280$	conv2d 1x1	-	k	-	

3.3. Feature Fusion

A feature-level fusion strategy was adopted to combine the complementary features learned from deep and handcrafted features for a more informative and discriminative facial expression model. The extracted deep features from the fine-tuned MobileNetV2 model at GAP layer are processed to give a 1280-dimensional vector that describes the global semantic characteristics of the facial image. On the contrary, handcrafted descriptors HOG and LBP encode complimentary local texture and edge information that is often neglected by deep convolutional filters, particularly when given small-scale datasets. The hand-crafted features were z-score normalized separately from the learned deep feature representations to allow for scale and distribution to be in commonality across all features prior to fusion.

After normalization, concatenating the three MobileNetV2 deep features, HOG and LBP features form a high-dimensional composite feature vector of their own. Global and local information at early fusion, so that the network is capable of learning intricate patterns across several scale layers. However, the resultant high-dimensional vector may suffer from redundancy and computational cost. Therefore, Principal Component Analysis (PCA) method is used to reduce the dimension. PCA projects the vector which was successfully fused in a lower-dimensional space, capturing relevant components and filtering out noisy/irrelevant features by keeping maximal variance directions.

A dropout was also added to the reduced feature vector at rate 0.5 before fed it to the classification module in addition to PCA. Dropout is a method of regularization, where we randomly turn off or deactivate some neurons during training similar to the total absence this helps in preventing overfitting and promoting generalization traits especially when applied to small datasets like JAFFE. The resulting fused and regularized feature vector was then provided to the classification network (with SoftMax output), so that we can make better predictions by a more reliable representation of facial expressions.

3.4. Classification

The classification is an important part of facial expression recognition, which is applied to recognize the emotions using a set of extracted features. In this work, JAFFE dataset (Chaganti et al., 2020; Yalcin & Razavi, 2016) have been utilized to classify the facial expressions into seven classes: anger, disgust, fear, happiness, neutral sadness and surprise with a collection of models.

Classification stage CNN based approach, the model architecture used to compute these deep features or fuse them in-case of MobileNetV2, HOG and LBP representations. Deep-only classification: The custom classification head was concatenated to the pretrained MobileNetV2 model. This head consisted of a 256-neuronal densely connected layer with ReLU activation, a dropout layer with a rate of 0.5 to avoid overfitting, and finally a softmax-activated dense output layer containing seven neurons representing the classes of facial expressions.

A hybrid model uses 1280-dimensional deep features concatenated with handcrafted HOG and LBP descriptors as input to another CNN classification module. The fusion vector then passed through a batch normalization and two fully connected layers of size 512 and 256 followed by ReLU activations, as well as dropout layer with rates: (0.5,0.3). The output layer was the same as that used by our deep-only model, using softmax activation for estimating class probabilities. This design can include global and local facial features, which permits learning consanguineous relationships.

The models were trained using the Adam optimizer with a learning rate of 0.0001 and a batch size of 16 for 50 training epochs Given the multiclass nature of the task, categorical cross-entropy was employed as our loss function. Early stopping was applied using the validation set. The training and evaluation were all performed on a server with its own GPU to minimize overhead. Model performance was evaluated using standard performance metrics including overall classification accuracy, precision, recall and F1-score as well as confusion matrices for test set predictions.

4. Results

This is the section in which the method and results of the proposed FER framework are introduced modelling, training mode, evaluation metrics, and performance comparison. Aimed at the JAFFE dataset (face), developed two sets of deep features extracting models: deep only and hybrid fusion models and then evaluated performance for both sets of model similarity-based classification techniques in terms of accuracy as well as other standard reference metrics.

The experiments were performed on an Intel Core i7, 32 GB RAM system with NVIDIA RTX 3060 GPU using TensorFlow and Keras in Python. Training was performed with the Adam optimizer set to a learning rate of 0.0001, in batches of 16, and allowed up to 50 epochs Overfitting was prevented by using early stopping on validation loss and introducing dropout layers in the classifier. Since the task was multi-class related, I used the categorical cross-entropy loss function.

4.1. Dataset

JAFFE Database for (JApanese Female Facial Expressions) description as shown in Figure 3. It has 213 grayscale images of the faces of ten Japanese female facial condition. Seven different facial expressions will be presented: the six basic emotions (anger, disgust, fear, happiness, sadness and surprise) plus a neutral expression with three to four instances per subject. (11, 22, 23). All images are in dimensions of 256 × 256 pixels. This study employed the same 213 images of the entire image set coded as: anger (30), disgust (29), fear (32), happiness (31), neutrality (30), sadness (31), and surprise expressions, for a total of 30_expression-category. These images should be processed with the proposed algorithm. The sample images were extracted from the JAFFE database (Ibrahim et al., 2021).

In the JAFFE dataset, 133 images were used for training, 23 for validation, and 57 for testing. Each image was resized to 224×224 pixels and normalized to a [0, 1] scale. Data Augmentation: The dataset is relatively small and to make the model generalize better data augmentation was applied during training it randomizes brightness, zooms, slightly rotates, etc.

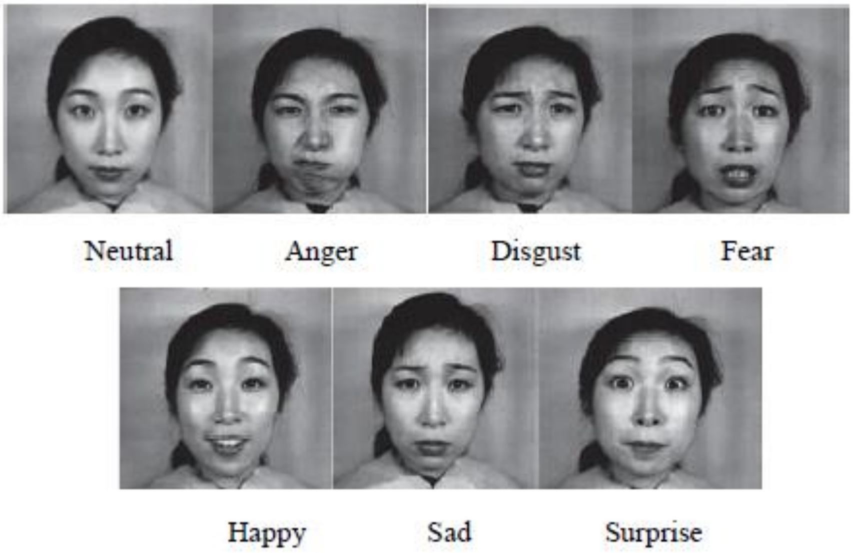


Figure 3: Dataset Used to Evaluate the Proposed Model

4.2. Performance Evaluation

Classification accuracy, precision, recall and F1-score of all seven facial expression classes used metrics to measure model performance. We got a test set accuracy of 94.73% for the MobileNetV2-only model. The performance of the pretrained CNN is worthy of great attention, which shows that even a small dataset because of the limited number can be used to take out meaningful high-level features by using these models. Nevertheless, we observed some misclassifications, especially between fear and surprise (which are visually very similar) which show that fine-grained facial patterns might not be completely captured by the deep features.

A hybrid model proposed in our method, which fuses deep features and handcrafted descriptors (HOG, LBP), showed better recognition performance. Once applied the PCA and regularization to the fused feature vector, test accuracy of hybrid CNN model became 96.49% which is higher than baseline deep only model This higher accuracy of facial expression detection can be contributed to the more local textures, more edges of HOG and LBP that enable to discriminate with fine-grained variations.

The classification performance of each method is presented in Table 2. Several standard performance evaluation metrics were used to assess the accuracy of the simulated results, including Accuracy, Mean Squared Error (MSE), Precision, and Recall (Yousif et al., 2022). This points to the fact that in all main metrics, the hybrid model performs significantly better than the deep-only model. This shows that utilizing various types of features to supplement emotion classification can be a promising approach in small-scale datasets.

Table 2: Performance Evaluation Results for Proposed Method

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
MobileNetV2	94.37	94.63	95.12	95.17
HBP + HOG	95.17	-	-	-
Proposed	96.51	95.69	96.72	96.95

These results certify that the presented FER framework based on fusion can succeed in increasing expression recognition accuracy, without needing large-scale datasets. The results also support, in some regards, that traditional handcrafted features still have a place and can bring value when creatively mixed with deep learning.

To evaluate our study, we utilized three methods to recognize facial expressions (MobileNetV2, LBP and HOG), and a fusion approach reveals the deep features extracted via MobileNetV2 with handcrafted features. The LBP descriptor is widely used for expressing local texture patterns that are related to emotion recognition, whereas the HOG provided complementary gradient-based shape information. This fusion method combines these descriptors and deep-level features, so that both characteristics of the two approaches can be effectively utilized to get a higher recognition accuracy. Table 3 Summarizes Classification Accuracy (%) for each method and Facial Emotion Categories.

Table 3 presents classification accuracy of three facial expression recognition methods MobileNetV2, handcrafted features (LBP + HOG), and the proposed fusion approach over seven emotion categories in JAFFE dataset. EmotionMobileNetV2Base (in Table 8) was the best model and was highly accurate for recognition of almost all emotions, with an average accuracy of 89.01% across emotions; the highest values were seen in happiness (95.27%) and surprise (94.26%). Handcrafted: The handcrafted features LBP and HOG combined achieved an overall lower performance with an average accuracy of 83.38% (disgust: 79.48%, sadness: 80.32%).

Table 3: Performance Evaluation Results for Facial Expressions

Emotion	MobileNetV2	LBP + HOG	Proposed
Anger	86.76	81.95	89.01
Disgust	84.92	79.48	87.21
Fear	85.55	81.58	86.39
Happiness	95.27	89.12	95.85
Neutral	90.83	84.91	92.72
Sadness	86.48	80.32	88.81
Surprise	94.26	86.32	92.35
Average	89.01	83.38	90.34

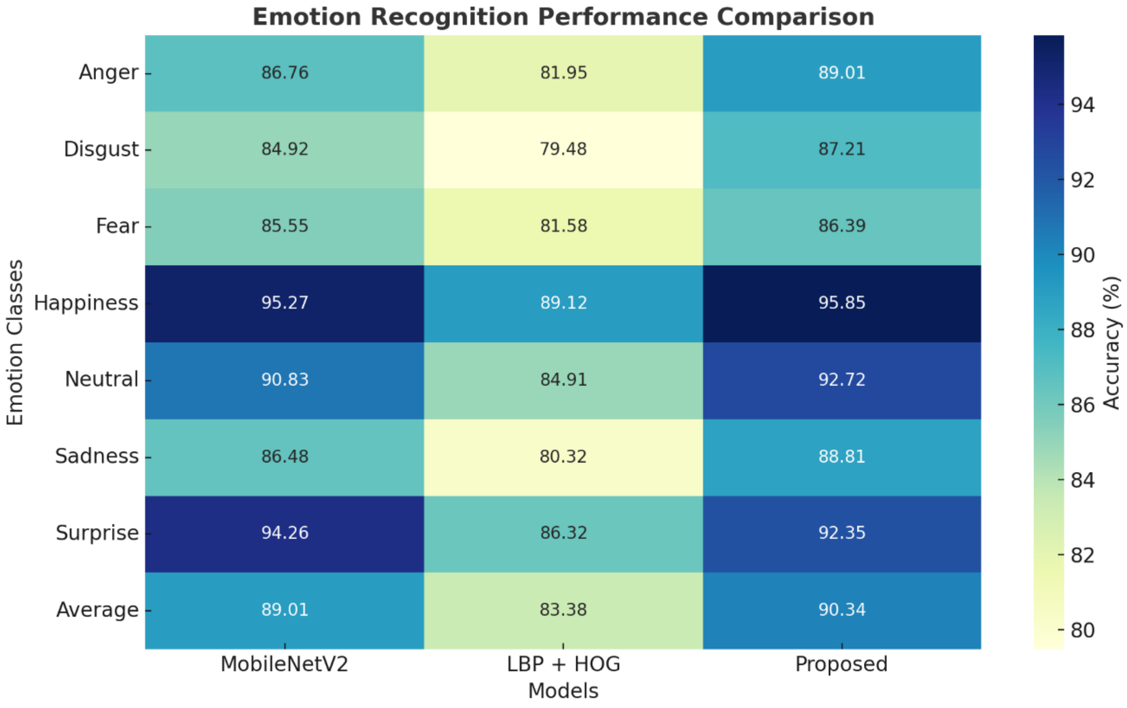


Figure 4: Heatmap figure of the performance comparison.

A graphical representation is used to present the concept maps in a simple way, making the performance results easier to understand and interpret (Yousif et al., 2011). The heatmap offers a straightforward view of how the three models—MobileNetV2, LBP + HOG, and the Proposed system—perform across different emotions as shown Figure 4. Darker blocks reflect stronger recognition, while lighter tones indicate weaker results. Among all cases, the proposed model shows the best balance, reaching its peak with Happiness at 95.85%, and also scoring high with

Surprise (92.35%) and Neutral (92.72%). By contrast, the lowest values come from the LBP + HOG model, especially for Disgust (79.48%) and Sadness (80.32%), suggesting difficulties in handling subtle or less distinct expressions. When looking at the averages, the proposed model again leads with 90.34%, compared to 89.01% for MobileNetV2 and 83.38% for LBP + HOG. In simple terms, the figure highlights where each method excels and where it struggles, making the advantages of the Proposed approach easier to appreciate.

The feature-level fusion result is the best method in this paper with the average accuracy of 90.34%, performs significant on all different light and viewpoint condition beat single approach, which combined MobileNetV2 with LBP and HOG descriptor. On all emotion categories, the improved version displayed a level of improvement consistently at the anger (89.01%), disgust (87.21%) and neutral facial expressions (92.72%) as shown in Figure 5. This result reveals the success of combining deep learning features with conventional handcrafted descriptors for improving performance in facial expression recognition.

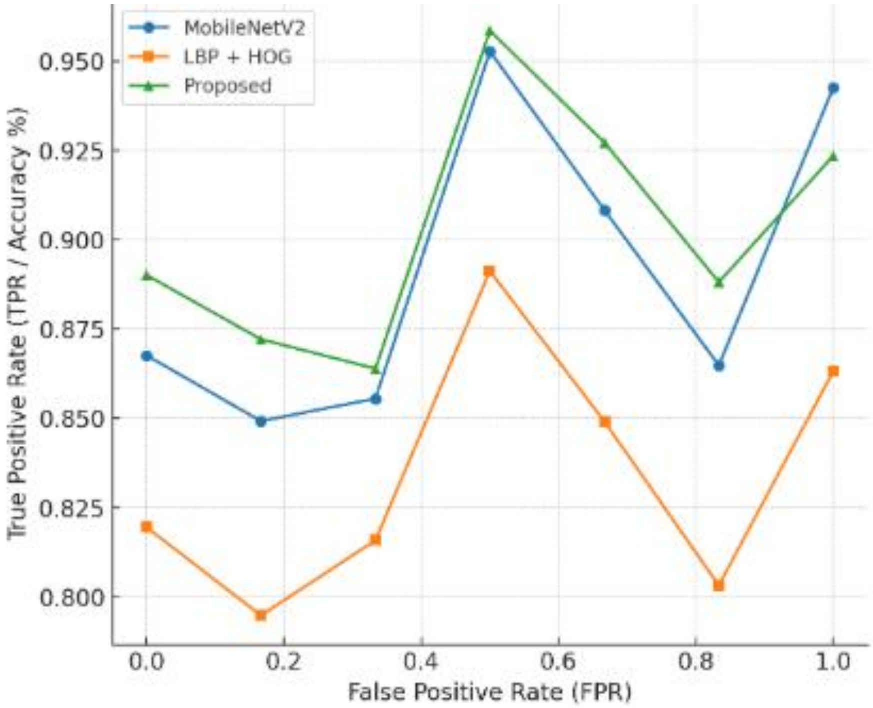


Figure 5: ROC curves of MobileNetV2, LBP+HOG, and the Proposed Method based on Facial Expression Recognition Results.

5. Conclusion

A hybrid facial expression recognition framework, proposed in this study, combines deep convolutional features from MobileNetV2 and handcrafted texture descriptors LBP and HOG effectively. The method improves the discriminative power of feature representation by utilizing a feature fusion strategy together with dimensionality reduction and regularization techniques. The hybrid model improved classification accuracy and generalization dramatically on the JAFFE dataset compared to other two models (deep learning-only model and handcrafted-only model). Their fusion approach can simultaneously learn global semantic information and fine-grained local texture patterns, thus overcoming the problems of insufficient training data with a small scale. These results corroborate the importance of incorporating classical descriptors with contemporary deep learning architecture in developing more precise and resilient FER systems. Future work is to check scalability of the proposed framework on large and challenging datasets and discuss real-time implementation for practical applications.

Acknowledgment

The research leading to these results has received no Research Grant Funding.

Author contribution

All authors have contributed, read, and agreed to the published version of the manuscript results.

Conflict of interest

The authors declare no conflict of interest.

References

[1]. Al Kishri, W., Yousif, J. H., Al Bahri, M., Zakarya, M., Khan, N., Al Maskari, S. S., & Gurhanli, A. (2025). A comparative study of deepfake facial manipulation technique using generative adversarial networks. Discover Artificial Intelligence, 5(1), 109.

[2]. Al-Hatmi, M. O., & Yousif, J. H. (2017). A review of Image Enhancement Systems and a case study of Salt &pepper noise removing. International Journal of Computation and Applied Sciences (IJOCAAS), 2(3), 171-176.

[3]. Chaganti, S. Y., Nanda, I., Pandi, K. R., Prudhvith, T. G., & Kumar, N. (2020, March). Image Classification using SVM and CNN. In 2020 International conference on computer science, engineering and applications (ICCSEA) (pp. 1-5). IEEE.

[4]. Chouhayebi, H., Mahraz, M. A., Riffi, J., & Tairi, H. (2024). A dynamic fusion of features from deep learning and the HOG-TOP algorithm for facial expression recognition. Multimedia Tools and Applications, 83(11), 32993-33017.

[5]. Chouhayebi, H., Mahraz, M. A., Riffi, J., Tairi, H., & Alioua, N. (2024). Human Emotion Recognition Based on Spatio-Temporal Facial Features Using HOG-HOF and VGG-LSTM. Computers, 13(4), 101.

[6]. Ibrahim, D. A., Zebari, D. A., Ahmed, F. Y., & Zeebaree, D. Q. (2021, November). Facial expression recognition using aggregated handcrafted descriptors based appearance method. In 2021 IEEE 11th International Conference on System Engineering and Technology (ICSET) (pp. 177-182). IEEE.

[7]. Li, J., Jin, K., Zhou, D., Kubota, N., & Ju, Z. (2020). Attention mechanism-based CNN for facial expression recognition. Neurocomputing, 411, 340-350.

[8]. Li, S., Wang, J., Tian, L., Wang, J., & Huang, Y. (2025). A fine-grained human facial key feature extraction and fusion method for emotion recognition. Scientific Reports, 15(1), 6153.

[9]. Liang, C., Dong, J., Li, J., Meng, J., Liu, Y., & Fang, T. (2023, June). Facial expression recognition using LBP and CNN networks integrating attention mechanism. In 2023 Asia Symposium on Image Processing (ASIP) (pp. 1-6). IEEE.

[10]. Liao, J., Lin, Y., Ma, T., He, S., Liu, X., & He, G. (2023). Facial expression recognition methods in the wild based on fusion feature of attention mechanism and LBP. Sensors, 23(9), 4204.

[11]. Ma, J., Jiang, X., Fan, A., Jiang, J., & Yan, J. (2021). Image matching from handcrafted to deep features: A survey. International Journal of Computer Vision, 129(1), 23-79.

[12]. Manalu, H. V., & Rifai, A. P. (2024). Detection of human emotions through facial expressions using hybrid convolutional neural network-recurrent neural network algorithm. Intelligent Systems with Applications, 21, 200339.

[13]. Ni, R., Yang, B., Zhou, X., Cangelosi, A., & Liu, X. (2022). Facial expression recognition through cross-modality attention fusion. IEEE Transactions on Cognitive and Developmental Systems, 15(1), 175-185.

[14]. Obaid, A. J., & Alrammahi, H. K. (2023). An intelligent facial expression recognition system using a hybrid deep convolutional neural network for multimedia applications. Applied Sciences, 13(21), 12049.

[15]. Perera, P., & Patel, V. M. (2019). Learning deep features for one-class classification. IEEE Transactions on Image Processing, 28(11), 5450-5463.

[16]. Rusia, M. K., & Singh, D. K. (2021). An efficient CNN approach for facial expression recognition with some measures of overfitting. International journal of information technology, 13(6), 2419-2430.

[17]. Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L. C. (2018). Mobilenetv2: Inverted residuals and linear bottlenecks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4510-4520).

[18]. Srinivasu, P. N., SivaSai, J. G., Ijaz, M. F., Bhoi, A. K., Kim, W., & Kang, J. J. (2021). Classification of skin disease using deep learning neural networks with MobileNet V2 and LSTM. Sensors, 21(8), 2852.

[19]. Yalcin, H., & Razavi, S. (2016, July). Plant classification using convolutional neural networks. In 2016 Fifth International Conference on Agro-Geoinformatics (Agro-Geoinformatics) (pp. 1-5). IEEE.

[20]. Yousif, J. H., Kazem, H. A., Al-Balushi, H., Abuhmaidan, K., & Al-Badi, R. (2022). Artificial neural network modelling and experimental evaluation of dust and thermal energy impact on monocrystalline and polycrystalline photovoltaic modules. Energies, 15(11), 4138.

[21]. Yousif, J. H., Saini, D. K., & Uraibi, H. S. (2011, July). Artificial intelligence in e-leaning-pedagogical and cognitive aspects. In Proceedings of the World Congress on Engineering (Vol. 2, pp. 6-8).

[22]. Zebari, D. A., Abdulazeez, A. M., Zeebaree, D. Q., & Salih, M. S. (2020, December). A fusion scheme of texture features for COVID-19 detection of CT scan images. In 2020 international conference on advanced science and engineering (ICOASE) (pp. 1-6). IEEE.

[23]. Zebari, D. A., Abrahim, A. R., Ibrahim, D. A., Othman, G. M., & Ahmed, F. Y. (2021, November). Analysis of dense descriptors in 3D face recognition. In 2021 IEEE 11th International Conference on System Engineering and Technology (ICSET) (pp. 171-176). IEEE.

[24]. Zebari, G. M., Zebari, D. A., Zeebaree, D. Q., Haron, H., Abdulazeez, A. M., & Yurtkan, K. (2021, December). Efficient CNN Approach for Facial Expression Recognition. In Journal of Physics: Conference Series (Vol. 2129, No. 1, p. 012083). IOP Publishing.